PLFNets: Interpretable Complex-Valued Parameterized Learnable Filters for Computationally Efficient RF Classification

Sabyasachi Biswas, *Graduate Student Member, IEEE*, Cemre Omer Ayna, *Graduate Student Member, IEEE*, and Ali Cafer Gurbuz, *Senior Member, IEEE*

***Abstract*— Radio frequency (RF) sensing applications such as RF waveform classification and human activity recognition (HAR) demand real-time processing capabilities. Current state-of-the-art techniques often require a two-stage process for classification: first, computing a time-frequency (TF) transform, and then applying machine learning (ML) using the TF domain as the input for classification. This process hinders the opportunities for real-time classification. Consequently, there is a growing interest in direct classification from raw IQ-RF data streams. Applying existing deep learning (DL) techniques directly to the raw IQ radar data has shown limited accuracy for various applications. To address this, this article proposes to learn the parameters of structured functions as filterbanks within complex-valued (CV) neural network architectures. The initial layer of the proposed architecture features CV parameterized learnable filters (PLFs) that directly work on the raw data and generate frequency-related features based on the structured function of the filter. This work presents four different PLFs: Sinc, Gaussian, Gammatone, and Ricker functions, which demonstrate different types of frequency-domain bandpass filtering to show their effectiveness in RF data classification directly from raw IQ radar data. Learning structured filters also enhances interpretability and understanding of the network. The proposed approach was tested on both experimental and synthetic datasets for sign and modulation recognition. The PLF-based models achieved an average of 47% improvement in classification accuracy compared with a 1-D convolutional neural network (CNN) on raw RF data and an average 7% improvement over CNNs with real-valued learnable filters for the experimental dataset. It also matched the accuracy of a 2-D CNN applied to micro-Doppler (μD) spectrograms while reducing computational latency by around 75%. These results demonstrate the potential of the proposed model for a range of RF sensing applications with enhanced accuracy and computational efficiency.**

***Index Terms*— Activity recognition, explainable artificial intelligence (AI), Gammatone, Gaussian, micro-Doppler (μD), radar, radio frequency (RF) sensing, Ricker, sinc.**

Received 31 July 2024; revised 6 October 2024; accepted 16 October 2024. Date of publication 24 October 2024; date of current version 6 November 2024. This work was supported in part by the National Science Foundation (NSF) under Award 2047771, in part by the Air Force Research Laboratory (AFRL) through ML-RCP Program under Award SPC-1000011554-GR131315, and in part by U.S. Engineer Research and Development Center under Grant W912HZ-21-2-0053. *(Corresponding author: Ali C. Gurbuz.)*

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

Digital Object Identifier 10.1109/TRS.2024.3486183

I. Introduction

RECENT developments in affordable solid-state transceivers and computationally efficient graphics processing units (GPUs), along with novel deep learning (DL) approaches, have significantly enhanced the usability of radio frequency (RF) sensors across a wide range of applications, including human activity recognition (HAR) [1], [2] and RF signal classification [3], [4].

This technological convergence has increased the demand for real-time classification systems for RF-based applications. However, current state-of-the-art techniques often require a two-stage process for complex I/Q RF signal classification. First, time–frequency (TF) analysis or other radar signal processing techniques are applied to generate 2-D (or higher 3-D and 4-D) radar data representations. The most common of these data forms are time-varying range-profile, range–Doppler, range–angle maps, and TF spectrograms [5], [6]. While some studies explore joint domain classification [7], [8], [9], micro-Doppler (μD) signature remains the prevalent data form in HAR problems [10], [11]. Calculating the μD representation of RF data requires heavy computation with many windowed short-time Fourier transforms (STFTs) in addition to the need for parameter optimization for window type, window size, FFT length, and overlap size between windows. The second step following the generation of RF data representation is feeding the generated image into a deep neural network (DNN) for activity classification [12], [13], [14], [15]. This two-step RF classification scheme is computationally intensive and exhibits high temporal latency in inference, limiting its real-time applicability. While some efforts have been made to implement TF spectrogram generation with DL-based approaches [16], [17], such methods still tend to be computationally expensive for real-life applications.

The problems mentioned above necessitate a scheme to directly process raw RF data. The solution to using raw RF data is to process complex-valued (CV) numbers by representing the data in different channels. The CV spectrograms, unlike absolute values, contain the full information like phase shift and Fourier components. Yang et al. [18] propose a recurrent neural network (RNN) that decodes 2-D I/Q radar data sequences, where both I and Q channels are supplied

concurrently into the RNN. This model classifies two distinct human activities with 93% average accuracy. The downside of this model is that training requires high computational complexity, with the network training phase spanning over 240 h using four Nvidia GeForce GTX 1080 GPUs. In [19], an adaptive magnitude thresholding approach has been shown to outperform state-of-the-art DL methods with up to a tenfold reduction in memory usage and training time, making it suitable for embedded platform deployment. However, it still involves a two-step process: first generating μD signatures and applying adaptive magnitude thresholding, followed by the classification process. Loukas et al. [20] provide a similar strategy by supplying range profile information along with the raw I/Q data and replacing the old RNN with a long-short term memory (LSTM) network. The studies [21], [22] propose a novel CV CNN architecture for RF-based HAR. The spectrogram values were used in the logarithmic scale. The study [22] devised a model named FourierNet that implements windowed Fourier transform in the 1-D time-series data preprocessing as a part of the network. Finally, [23] looked into the representation of the CV RF data in the forms of range–time, range–Doppler map, or spectrogram. But this study also required a two-stage process for classification which does not solve the problem for real-time applications.

An important issue in RF signal processing with neural networks is dissecting the learned model to gain insight. This is not possible with the black-box nature of the traditional neural networks. The study [24] is one of the pioneers in introducing learnable filters for addressing the interpretability issue by developing learnable sinc filters for the speaker identification task from raw audio recordings. This model is named SincNet and uses a 1-D layer of sinc functions with their low and high cutoff frequencies transformed into learnable parameters. With this design, bandpass filtering by neural networks was made easy for dissection, and the model was enabled to learn high-level, physically interpretable parameters. In recent times, several studies have focused on the use of filter banks for enhancing audio signals. Sivapatham et al. [25] used Gammatone filter banks in a CNN for monaural speech signal enhancement. They first used nonlearnable 64-channel Gammatone filter banks to filter the audio signal input, which was then fed into a simple CNN architecture. Khan and Yener [26] introduced parameterized learnable Ricker wavelets as a wavelet decomposition layer within the CNN model for a phone recognition task. However, these methods were developed for audio signal classification. In [27], 2-D-learnable Gaussian filters were introduced for image classification. All the above-mentioned methods worked on real-valued data, whereas RF signals are CV, requiring CV-CNN layers for effective classification.

The authors' previous study [28] demonstrates the application of a CV neural network operating directly on the CV 1-D slow-time data after pulse compression via the CV-SincNet architecture. CV-SincNet maps the real-valued neural network operations into the complex number domain to directly process the 1-D complex slow-time RF data. CV-SincNet also learns complex sinc functions that translate to one-sided bandpass filters in the Fourier domain. Each CV sinc filter is parameterized by center frequency and bandwidth. This architecture was shown to provide enhanced classification performance compared with a 2-D CNN that uses the RF μD-spectrograms with lower computational load.

This work builds on our previous study [28] by introducing a more generalized framework for parameterized learnable filterbank functions (PLFs). Unlike the previous approach, which was limited to the sinc filter, we now present a generalized parameterized filter learning approach that can be extended to other parameterized filter representations that are suitable for backpropagation. We show the proposed concept on bandpass filters, such as Gaussian, Gammatone, and Ricker wavelet filters, demonstrating that we can learn parameters of structured functions that can directly and effectively operate on time-domain RF signals. Furthermore, we have developed a more versatile neural network architecture, expanding beyond ASL sign recognition to be applicable to other application domains such as HAR and modulation recognition. In addition, we have simplified the overall network structure to reduce computational latency, making it better suited for real-time applications. The analyses in this article highlight the advantages of the generalized PLF models, including higher classification accuracy, faster prediction times, and improved interpretability.

The contributions of this article are summarized as follows.

1) *Generalized PLF Formulation:* Develop a flexible framework for learning parameters of more general PLFs such as sinc, Gaussian, Gammatone, and Ricker.
2) *PLF-Based CNN Architecture:* Integrate PLFs into CNNs to extract meaningful features directly from time-series IQ radar data.
3) *Evaluation on Synthetic and Real-World Datasets:* Assess the proposed approach on both the synthetically generated and real-world datasets.
4) *Comparison With Existing Methods:* Increased performance compared with the traditional real and CV CNNs and real-valued PLF-based CNNs.
5) *Physical Interpretation and Efficiency:* Demonstrate the physical relevance of learned filters and the potential for real-time classification with reduced latency.

The organization of the article is as follows. Section II discusses the formulation of different filters along with implementation details of all the neural network blocks for CV operations. Section III presents the datasets and results for all compared networks. Finally, Section IV concludes this article and discusses future directions.

II. Proposed Method

A. Parameterized Learnable Filter Formulation

Let us assume discretized 1-D slow-time complex IQ radar data $x[n]$ to be processed by a CNN architecture. In this case, the first layer in a standard CNN uses a set of 1-D finite impulse response filters, $h_j[n] j = 1, \ldots, J$, with the input data, which results in the jth convolution output $y_j[n] = x[n] * h_j[n] = \sum_{\ell=0}^{K-1} x[\ell] h_j[n-\ell]$, where K is the total length of the input data. When the neural network is trained over the data, all the elements of the filters $h_j[n]$ are learned.

The convolution operation is mainly filtering in the frequency domain with the frequency response of the filter. However, since standard CNNs can learn an unrestricted filter $h_j[n]$, the learned filter can have a very random frequency-domain response. Here, instead of letting CNNs learn any random filter, we propose to put a structure to the learned filters. Let the filters be defined by a function $f_{\boldsymbol{\theta}}[n]$ that is parameterized by a vector of parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_P]$ where P is the total number of parameters of the function. In this case, instead of learning every element of a filter, CNNs will learn the parameters $\boldsymbol{\theta}$ of the function. If the number of parameters is comparably less than the length of the filter, $P \ll N$, the neural network learns a much smaller number of parameters. For a neural network architecture to be able to learn the parameters of the defined function within the backpropagation algorithm, it is critical that the defined function should be fully differentiable with respect to each parameter $\theta_p p = 1, \ldots, P$. Under such differentiability conditions, parameters $\boldsymbol{\theta}$ can be updated within gradient-based optimization approaches used in neural network training.

Although any differentiable function $f_{\boldsymbol{\theta}}[n]$ can be used within the defined architecture, in terms of radar-based activity or modulation recognition problem, the underlying features of the targeted classes are in the TF domain. Hence, we target to use $f_{\boldsymbol{\theta}}[n]$ that corresponds to bandpass filters. Since we can design our own filter functions, we can use windowing operations such as Hamming windows to have lower sidelobes or better frequency selectivity. Finally, the output of the proposed generalized PLF layer will be the convolution of the raw radar data with the structured filters as $y_j[n] = x[n] * f_{\boldsymbol{\theta}_j}[n]$.

1) Complex Sinc Filter: Our previous work [28] focused on a complex sinc filter rather than this generalized framework, and it can be seen as a subcase of the framework presented here, where the structured filter can be represented as follows:

$$f_{\boldsymbol{\theta}}[n] = 2\theta_1 \text{sinc}(2\pi\theta_1 n) \times e^{j2\pi\theta_2 n} \tag{1}$$

where $\text{sinc}(x) = \sin(x)/x$ is the sinc function and $\boldsymbol{\theta} = [\theta_1, \theta_2]$ has two learnable parameters. Since the frequency response of the sinc function in (1) is a bandpass filter the parameter θ_1 corresponds to the bandwidth, and θ_2 is the center frequency of the bandpass filter. In addition, since the complex filter in (1) is differentiable with respect to its learnable parameters $[\theta_1, \theta_2]$, it is compatible with backpropagation.

While the complex sinc filter is an example of the proposed PLF architecture, in this study we propose several different filters we can use within the generalized structure as detailed in Section II-A2–II-A4.

2) Complex Gaussian Filter: Another filter function could be a Gaussian filter. The impulse response of a Gaussian filter is characterized by a sinusoidal wave modulated by a Gaussian function. The time-domain formulation of a complex Gaussian filter with respect to variance (θ_1) and center frequency (θ_2) is given by

$$f_{(\theta_1,\theta_2)}[n] = \sqrt{\frac{2}{\sqrt{\pi}\theta_1}} e^{-\frac{n^2}{2\theta_1^2}} \cdot e^{j2\pi\theta_2 n} \tag{2}$$

where $n \in [-(N/2), ((N-1)/2)]$ and the 3-dB bandwidth B of the filter is related to the θ_1 parameter as

$$\theta_1 = \frac{\sqrt{\log(2)}}{\pi \times B}. \tag{3}$$

The filter function defined in (2) is differentiable in terms of both the parameters, and hence it is compatible with backpropagation.

3) Complex Gammatone Filter: A Gammatone filter is a type of linear filter commonly used in auditory modeling to simulate the filtering characteristics of the human ear [29]. It is named for its mathematical form, which combines a gamma distribution function with a sinusoidal tone. The time-domain formulation of a complex Gammatone filter with respect to its variance (θ_1) and center frequency (θ_2) is given by

$$f_{(\theta_1,\theta_2)}[n] = n^{(\gamma-1)} e^{-2\pi\theta_1 n} e^{j2\pi\theta_2 n} \tag{4}$$

where $n \in [0, N-1]$ and the 3-dB bandwidth B of the filter is given as

$$\theta_1 = \frac{B}{2}\sqrt{\frac{1}{2^{1/\gamma}-1}}. \tag{5}$$

In both (4) and (5), γ is the order of the filter and it is used as a fixed parameter rather than a learnable parameter of the filter and is set to four as done in [30].

4) Complex Ricker Wavelet Filter: Ricker wavelet is the negative normalized second derivative of a Gaussian function. Introduced in [26], it is also known as the Mexican Hat Wavelet. The time-domain formulation of a complex Ricker filter with respect to variance (θ_1) and center frequency (θ_2) is given by

$$f_{(\theta_1,\theta_2)}[n] = \frac{2}{\sqrt{3\theta_1}\pi^{1/4}}\left(\frac{n^2}{\theta_1^2} - 1\right) e^{-\frac{n^2}{2\theta_1^2}} e^{j2\pi\theta_2 n} \tag{6}$$

where $n \in [-(N/2), ((N-1)/2)]$. As shown in Fig. 1, the Ricker filter in the frequency domain has two peaks. The 3-dB bandwidth (θ_2) is determined by measuring the frequency range from the first 3-dB point of the initial peak to the second 3-dB point of the subsequent peak. The relationship between θ_1 and the 3-dB bandwidth, B, is expressed as follows:

$$\theta_1 = \frac{2}{\pi B}. \tag{7}$$

Since the learned filters have finite lengths, they will have ripples and sidelobes in the frequency domain. To mitigate such effects, windowing is a popular solution in signal processing. Hence, we apply windowing as well by multiplying the parameterized filters with a fixed Hamming window as

$$f_{\boldsymbol{\theta}}^{\omega}[n] = f_{\boldsymbol{\theta}}[n] \cdot \omega[n] \tag{8}$$

where the window function $\omega[n]$ is fixed and do not contain any learnable parameters as specified by

$$\omega[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) \tag{9}$$

where N is the length of both the window and the learned filters. The time- and frequency-domain representations of each of the parameterized filters after windowing are shown

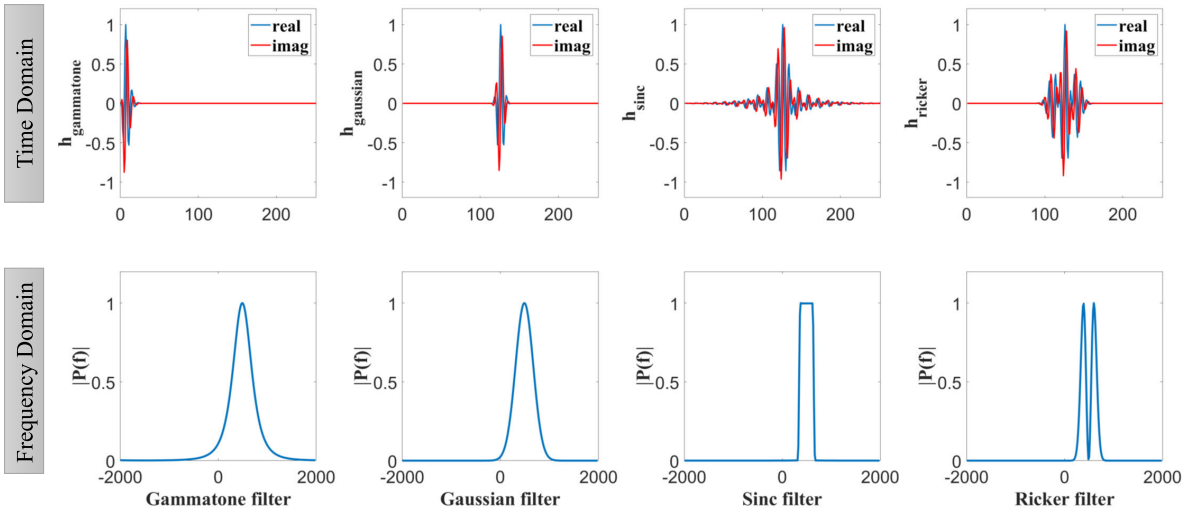


Fig. 1. Examples of different types of filters after windowing in the time and frequency domains with a center frequency of 500 Hz, a bandwidth of 300 Hz, and a sampling frequency of 4000 Hz.

in Fig. 1. While the time domain real and imaginary components of the filters follow the underlying functions, the frequency-domain representations are different bandpass filters.

Since our goal is to develop the model to learn interpretable parameters of structured filters, we used the window function as fixed; however, it is possible to design the window functions with learnable parameters to jointly learn both the filters and the windowing functions. Consequently, the only learnable parameters in the complex learnable filter layer, as depicted in Fig. 2, are $\theta_{1,j}$ and $\theta_{2,j}$ for each learned filter $j \in [0, J-1]$. For CV operations, the bandwidth related to θ_1 is initialized over the entire frequency range $[-f_s/2, f_s/2]$, where f_s is the sampling frequency. In the real-valued case, θ_1 is initialized in the range $[0, f_s/2]$. During training, the neural network updates $\boldsymbol{\theta}_j \, \forall j$ in the learnable filter layer together with other parameters of the whole model.

The model sequence following the initial learnable filter layer includes pooling, normalization, ReLU activation, and dropout operations, all using complex operations, as illustrated in Fig. 2. This sequence forms the learnable filter block, which serves as the initial layer of the proposed DNN architecture.

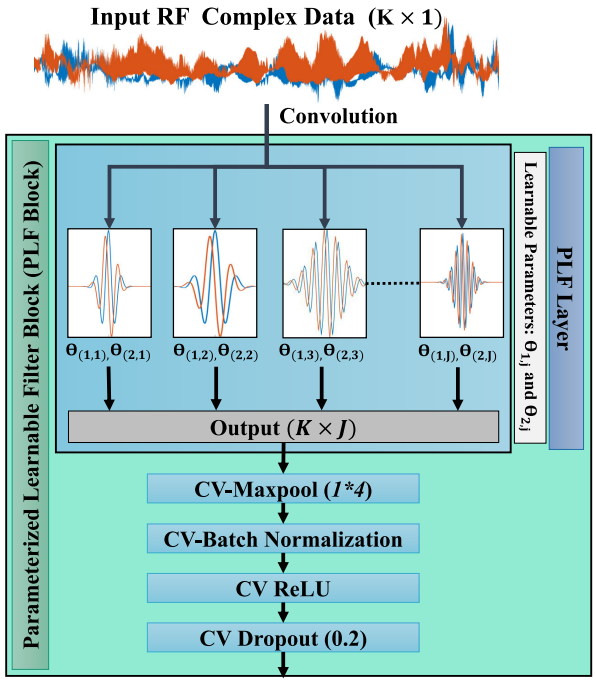


Fig. 2. PLF block.

B. Benefits of the Proposed PLF Layer

1) Interpretability: One key benefit of the proposed approach is its interpretability. The model works directly with complex RF signal spectra, and the parameters it learns have clear physical meanings. This allows a better understanding of which frequencies and spectral bands are important for RF sensing. By revealing these critical components, the model provides insights that are easier to interpret and understand compared with traditional convolution layers.

2) Small Number of Parameters: Another significant advantage is the reduced number of parameters required by the model. For example, a 1-D Gaussian filter layer with J filters only needs $2J$ trainable parameters, while a 1-D convolutional layer with J filters needs $N \times J$ trainable parameters, where N is the filter length. This means the model is smaller and requires less number of parameters, which is advantageous in terms of computational resources and training efficiency.

C. Complete Proposed Architecture

The proposed PLF-based CV-CNN architecture, PLFNet, is developed to process the 1-D complex raw RF data. The first layer of the proposed architecture comprises J CV PLFs, each with a length of $N = 251$. This layer is then followed by a complex max-pooling layer of size 1×4, batch normalization, ReLU activation, and dropout of 0.2. This sequence of

layers comprises the PLF block, which is shown in Fig. 2. The PLF block is followed by a series of convolutional block. Each convolutional block has a convolutional layer with 64 filters and a kernel size of 1×5 for each filter. This is followed by a max-pooling layer of size 1×3, batch normalization, ReLU activation, and dropout of 0.2. The outputs of the convolutional blocks are flattened before being input to a dense layer with a size of Z. After the application of dropout of 0.2, the network uses a softmax layer for classification. All the layers mentioned were developed to work on CV data. The CV operations for convolution, fully connected layers, activation functions, pooling, and normalization were detailed in our previous work [28]. These operations were implemented by representing complex numbers as their real and imaginary components, allowing for efficient computations within a standard deep learning framework. For better visualization, the flow diagram of the full proposed architecture is shown in Fig. 3.

III. Performance Evaluation of the Proposed Method

A comprehensive evaluation of the classification performance of the proposed model under both synthetic and real-world RF data has been made compared with several state-of-the-art techniques. Evaluations are done for two different applications: RF waveform modulation recognition using a synthetic RF dataset and American Sign Language (ASL) recognition using an experimental dataset. For both the applications, the proposed PLFNet is compared with the following approaches.

1) *CNN-2D:* The complex raw radar is first transformed to a 2-D spectrogram image and then processed by a conventional 2-D real CNN.
2) *CNN-1D:* A real 1-D CNN that takes raw radar data as input with two separate channels for real and imaginary components.
3) *CVCNN-1D:* A complex 1-D CNN that directly processes complex raw radar data with complex neural network operations.
4) *Real-Valued Learnable Filters:* This is the real-valued version of the proposed PLFNet to evaluate the benefit of CV process. For the formulation of real-valued PLFs, the real parts of each complex filter were used. Learnable filters are applied to raw radar data using two channels for real and imaginary components separately.

Sections III-A1 and III-A2 will discuss the details of each compared architecture and their training process.

A. Detailed Descriptions of Compared DNNs

1) CNN-2D Architecture: To use the CNN-2D architecture, first a TF transformation is done over the raw RF signals. We compute the spectrogram of input RF data—the square modulus of the windowed STFT across the RF data $x(t)$ as

$$S(n, \omega) = \left| \int_{-\infty}^{\infty} h(t-n)x(t)e^{-j\omega t}dt \right|^2 \tag{10}$$

where $h(t)$ is the window function.

TABLE I
Model Sizes of Compared Models

Layer	Blocks	CVCNN-1D	CNN-1D	RV-PLF	PLFNet
PLF	(256, 251)	-	-	1024	1536
Conv_input	(256, 251)	130,048	129,280	-	-
CB 1	(64, 5)	164,480	82,240	82,240	164,480
CB 2	(64, 5)	41,600	20,800	20,800	41,600
CB 3	(64, 5)	41,600	20,800	20,800	41,600
CB 4	(64, 5)	41,600	20,800	20,800	41,600
CB 5	(64, 5)	41,600	20,800	20,800	41,600
CB 6	(64, 5)	41,600	20,800	20,800	41,600
Dense 1	256	66,048	49,408	49,408	66,048
Softmax	100	51,400	25,700	25,700	51,400
Total Parameters		619,976	390,628	262,372	491,464

After generating the TF spectrograms, they are saved as images of size 128×128 for both the synthetic and real-world datasets. These images are then used as input for the CNN-2D model. Drawing from prior work [31], a CNN architecture with four convolutional blocks was designed. Each block contains two convolutional layers, followed by a 2×2 max-pooling layer, batch normalization, ReLU activation, and a dropout rate of 0.2. The convolutional blocks use 32 filters each with a kernel size of 3×3. Following the convolutional blocks, the tensor is flattened and processed through a dense layer with the dimension of 256×1. A dropout rate of 0.5 is applied before passing the output to a softmax classifier as detailed in Fig. 4.

2) Other Compared Architectures: For real-valued networks, a CNN-1D and a PLF-based CNN-1D, were used to process the 1-D complex RF data. Both the real-valued networks separate the raw complex RF data stream into its real and imaginary parts, forming a real input representation of size $L \times 2$, where L is 4000 and 13 050 for RF waveforms and ASL data, respectively. The real-valued PLFNet follows the same architecture as the CV one, described in Section II-C. The principal difference is that the model only computes real-valued operations in each layer instead of CV ones.

In lieu of the PLF block, the CNN-1D architecture uses a convolutional block of the same length as the PLFs as its first layer, after which its architecture is the same as that of the PLFNet. Similarly, for the CVCNN-1D, the PLF block is replaced with a convolutional block with complex operations. Details about all the architectures and a layer-by-layer comparison of the number of parameters for each network are given in Table I.

B. Classification of RF Waveform Modulation Types

1) Dataset Generation: To facilitate the evaluation of the proposed methodology, one of the applications we tested is to classify RF waveform modulation. For this purpose, it is necessary to generate a dataset of raw RF data with known transmitted waveform types. For the initial validation of the proposed approach, we chose to generate a purely simulated dataset consisting of synthetic noisy waveform receptions.

The generated dataset consists of five common waveform modulations: linear frequency modulation (LFM), rectangular, Costas code, Barker code (binary phase code), and Frank code (polyphase code). We include 500 samples of each of these modulations for a total of 2500 waveforms with the

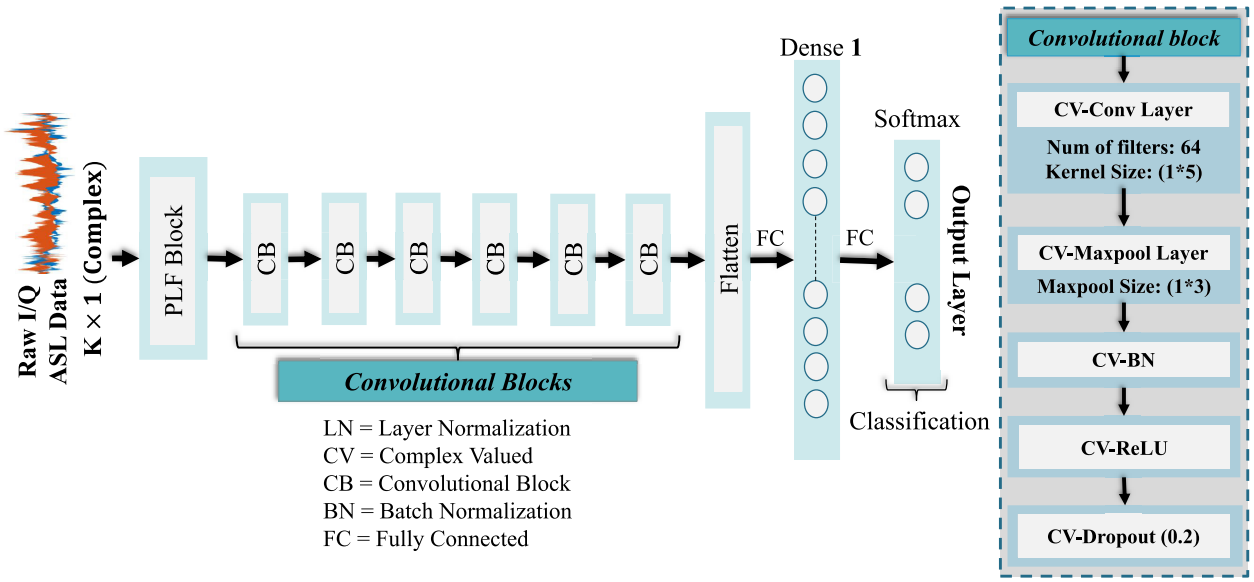


Fig. 3. Flow diagram of the proposed PLFNet architecture.

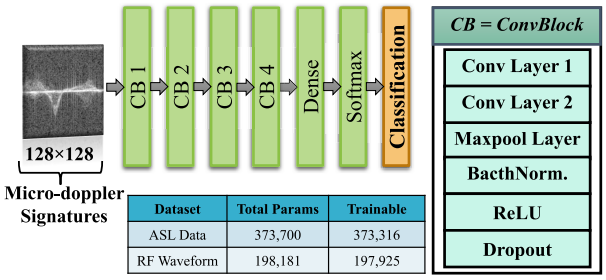


Fig. 4. CNN architecture for μD-spectrogram classification.

TABLE II
PARAMETER RANGES FOR DIFFERENT TYPES OF WAVEFORM

Types	Parameters	Range of Values
ALL	f_c (center freq.)	$U\left[-\frac{f_c}{3}, \frac{f_c}{3}\right]$
	A (Amplitude)	$U[1, 10]$
	f_s (sampling freq.)	4000 Hz
	SNR	$U[0, 20]$
LFM	B (Bandwidth)	$U\left[\frac{f_c}{10}, f_c\right]$
Costas	f_{step} (num of freq steps)	$U[5, 20]$
	step_size	f_{step}/f_s
Barker	L_B (barker code length)	$\{7, 11, 13\}$
Frank	M (phase_steps)	$\{6, 7, 8\}$

corresponding ground-truth labels. Each of these waveform types has several variables that are randomly varied to generate 2500 distinct waveforms to ensure variation within the dataset. The sampling frequency, f_S, is 4000 with a duration of 1 s. Table II offers a full accounting of the parameters available for modulation, as well as the extent of their variation. Each data sample is added random white Gaussian noise with a signal-to-noise ratio (SNR) selected uniformly random between 0 and 20 dB, as listed in Table II.

TABLE III
CLASSIFICATION PERFORMANCE COMPARISON OF RF WAVEFORM MODULATION

Network	Testing Accuracy	Precision	Recall	F1 Score
CNN2D	95.2	95.56	96.19	95
CNN-1D	64.6	66.9	66.61	63.76
CVCNN-1D	82.2	77.1	83.6	77.53
Sinc	84.2	90.81	81.05	77.95
CV-Sinc	95.6	95.48	95.64	95.52
Gaussian	90.78	90.55	90.67	90.65
CV-Gaussian	96.8	96.22	96.52	96.67
Gammatone	83.4	88.42	83.32	79.86
CV-Gammatone	94.2	94.61	94.41	94.06
Ricker	86.4	88.9	87.07	86.48
CV-Ricker	95.2	95.9	95.32	95.24

2) Results for Waveform Classification: For this application, input RF data size is 1×4000 and the number of filters in the PLF block was chosen to be $J = 128$, with four following convolutional blocks and a dense layer size of 128. Table III shows a comparison of classification performance across different architectures. The PLFNets with learnable filters—using sinc, Gaussian, Gammatone, and Ricker functions—performed well, achieving testing accuracies of 95.6%, 96.8%, 94.2%, and 95.2%, respectively, Gaussian filter providing the best performance among tested learnable filters. In contrast, the CVCNN-1D model, which does not have learnable filters in its first layer, only reached an accuracy of 82.2%. This difference highlights the importance of PLFs for building an enhanced feature set for the rest of the convolutional layers.

In addition, CV CNNs consistently outperformed real-valued CNNs, with an average accuracy improvement of about 10.5%. The CNN-2D model, which works not on the raw IQ data but on the TF spectrogram images, achieved a strong accuracy of 95.2%, similar to the PLFNet models, with much

higher computational complexity as we will demonstrate next on the experimental dataset case.

C. Classification Performance Over Real-World Dataset

to evaluate the performance of compared methods in a real-world HAR scenario, a challenging dataset consisting of 100 ASL signs was used. The details about the experimental radar system and setup can be found in [32].

The dataset comprises 2500 radar sign samples from five participants, with each class having 25 samples. Each sign was recorded five times per participant, each recording lasting 4 s, followed by a 2-s inter-stimulus interval. This resulted in 28 s of data collection per class from each participant, capturing a total of 89 600 chirps. For each 4-s sample, 12 800 chirps were collected. To account for minor inconsistencies between the start and end times of each sample, an additional 125 chirps were included at both ends. Each chirp consists of 256 fast-time samples, resulting in a 2-D CV raw radar data size of $256 \times 13\,050$ for a single sign. After range processing, the range bins corresponding to the target location are summed, creating a 1-D slow-time complex raw radar data vector of $1 \times 13\,050$, which will be input to compared models.

For compared models, in the initial layer, the number of filters is chosen to be $J = 256$, the network has a series of six convolutional blocks before the softmax, and the dense layer has a length of 256 nodes. To evaluate performance, we split the dataset with 80% allocated for training and 20% for testing. All THE models were trained to converge over 150 epochs. Posttraining, the models were assessed using the same test set, and confusion matrices were generated for each model. The performance metrics—accuracy, precision, recall, and $F1$ scores—were derived from these confusion matrices and are summarized in Table IV along with top-N accuracy values measuring how well a model can predict the correct class within its top N predictions. From the achieved results, we can point out some key conclusions.

1) Among the methods using 1-D raw radar inputs, both complex and real-valued PLFNets performed better than the CNN models, CNN-1D and CVCNN-1D. Using PLFs as the first layer of the network and capturing frequency-domain information improved performance compared with directly using random convolutional filters on raw radar data.
2) CV networks showed better classification accuracy, about 6.9% higher on average, compared with real-valued ones. This highlights the importance of using CV networks and designing complex PLFs suitable for complex radar data.
3) Our proposed PLFNet especially with CV Gaussian filters achieved higher accuracy compared with CNN-2D, which is the common approach in the literature [31] and processes μD signatures after TF transformation. Therefore, PLFNet's ability to provide similar or slightly better results without the computational burden of explicitly calculating μD signatures is very significant.

As PLFNets with Gaussian filters demonstrated the best performance among other compared PLFs, following computational and interpretability evaluations will focus on Gaussian filter case only.

TABLE IV
CLASSIFICATION PERFORMANCE OF THE COMPARED MODELS IN TERMS OF EVALUATION METRICS FOR 100-CLASS ASL DATA

Network	Testing Accuracy			Prec.	Recall	F1
	Top1	Top3	Top5			
CNN-2D	63.95	81.46	89.92	66.27	62.98	63.21
CNN-1D	10.82	21.79	29.16	10.87	9.98	10.08
CVCNN-1D	17.63	27.94	38.89	16.81	17.76	17.15
Sinc	56.26	75.03	81.92	61.5	56.78	55.95
CV-Sinc	64.8	80.32	86.23	69.69	64.85	64.41
Gaussian	55.7	75.73	81.64	60.03	55.6	55.26
CV-Gauss	65.97	83.33	89.76	70.04	65.8	65.27
Ricker	58.22	76.58	84.6	64.99	58.35	57.81
CV-Ricker	64.14	81.01	86.07	68.31	64.25	63.37
Gammatone	56.75	74.68	80.16	58.63	56.75	55.52
CV-Gamma	63.71	79.74	86.29	67.76	63.65	63.53

TABLE V
COMPUTATIONAL TIME REQUIRED FOR EACH NETWORK

Network	Total training time	Pre-processing time	Avg. inference time	Total latency
CNN-2D (on μDs)	0.16 Hrs	2601 ms	32 ms	2633 ms
CNN-1D	0.68 Hrs	377 ms	167 ms	451 ms
CVCNN-1D	0.76 min	377 ms	211 ms	588 ms
GaussNet	2.07 hrs	377 ms	252 ms	629 ms
CV-GaussNet	2.62 hrs	377 ms	275 ms	652 ms

1) Computational Latency: In addition to the performance of the approaches, another key metric is the overall latency of a technique, which is the time from data acquisition to generate a prediction. To achieve near real-time classification, methods that combine low latency with high performance are the most suitable. We assessed the computational needs of each approach by measuring the average times for preprocessing, training, and the time required for the network for prediction (inference time), as detailed in Table V.

For the experiments, both data processing and network training were conducted on an Alienware m15 R7 laptop with an NVIDIA 3060 GPU, Intel 11th Gen CPU, and 32 GB of memory. The total training time for the CNN-2D model was much shorter compared with other methods. The 1-D CNNs, having a larger network, took longer to train than the standard CNN-2D working on μD-spectrogram images. The preprocessing time mentioned in Table V is the duration it takes for one data sample, from acquisition to being ready as an input to a DNN. After the 2-D radar data matrix is obtained from the corresponding TX–RX pair of the radar, range processing is applied to the fast-time samples, and 1-D slow-time data are obtained by summing the range bins for each slow-time sample. These 1-D data are the input for all networks except CNN-2D. For CNN-2D, additional moving target identification (MTI) filtering and STFT operations are needed to generate μD-spectrogram images. A 4096-point FFT, a window size of 256, and an overlap of 240 points

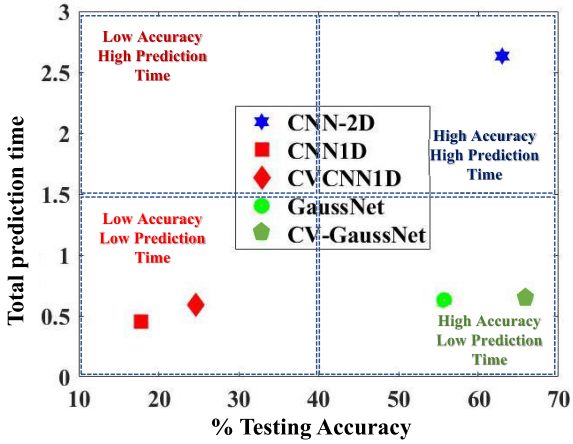


Fig. 5. Testing accuracy versus prediction time.

TABLE VI
CV-GAUSSNET PERFORMANCE V. # GAUSSIAN FILTERS

Num of filters	Num of parameters	Training time	Testing accuracy	Avg. prediction time
8	331,256	0.37 Hrs	35.65%	65 ms
16	336,424	0.45 Hrs	48.31%	71 ms
32	346,760	0.58 Hrs	58.44%	87 ms
64	367,432	0.79 Hrs	62.76%	114 ms
256	491,464	2.62 Hrs	65.97%	275 ms

were used for STFT resulting in a higher preprocessing time for CNN-2D.

The testing accuracy of the approaches and the corresponding total latency for prediction are shown in Fig. 5. While conventional 1-D CNNs take the least time, their accuracy is very low. While the standard 2-D CNN working on μD-images has high accuracy, it also has a very high latency. The proposed PLFNet provides both high accuracy with low latency. Hence, it can be seen that the proposed PLFNet is a viable solution for near real-time and high-accuracy radar-based classification problems.

A detailed analysis of PLFNet with Gaussian filters was performed to understand how the number of learned Gaussian filters affects accuracy and latency. The model was trained with different numbers of PLFs, ranging from 8 to 256. The resulting testing accuracy and average prediction times for each setup are shown in Table VI. Reducing the number of filters led to faster prediction times, but too few filters also reduced accuracy. Using 64 filters resulted in about an 81% decrease in total prediction time while maintaining almost the same performance as the CNN-2D.

D. Interpretability of PLFNets

A key feature of the proposed PLFNet architecture is its enhanced interpretability. Each PLF depends only on parameters with clear physical meanings, such as spectral frequencies and bandwidths. To identify which spectral bands are favored by the PLFNet, we calculated the weights of the convolutional layer that follows the PLF block. By examining the normalized weight magnitudes for each of the 256 kernels, we can see

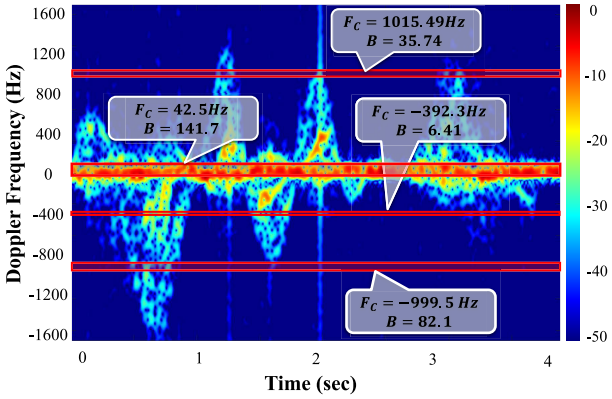


Fig. 6. Four most important filters projected on a μD-spectrogram.

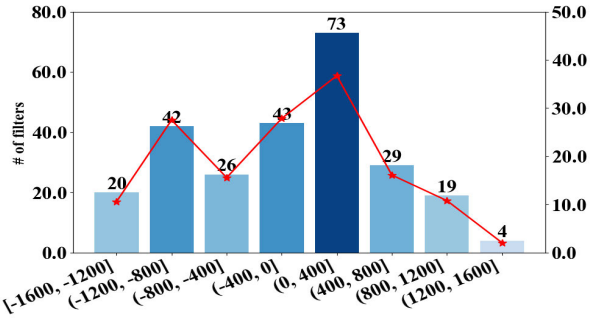


Fig. 7. Filter and weight distribution of the PLFNet with Gaussian filters.

that the model gives different levels of importance to specific kernels. For better visualization, the top four Gaussian kernels were projected onto a μD-spectrogram image as seen in Fig. 6 along with their central frequency and bandwidth values. It is interesting to note that one of the top kernels mainly focuses on the lower Doppler frequency region which has most of the signal energy, while another on the lower and upper frequency tips, and the fourth on a band centered at around -400 Hz.

The histogram and cumulative weight distribution of all the 256 learned Gaussian filters in the frequency domain is shown in Fig. 7. Most kernels were learned with center frequencies falling into the 0-400-Hz range, suggesting that the model prioritizes this spectral segment using more kernels and weights here. Interestingly, the distribution of kernels across frequencies is not symmetrical. Certain frequency regions, like [1200, 1600), are given less importance, showing minimal activity, while other bands, such as $(-1200, -800]$ or [400, 800), have more learned kernels and weights. We also evaluated other PLF-based approaches and saw similar trends in filter distribution. This makes sense because these bands contain most of the Doppler components related to hand movements in sign language.

IV. CONCLUSION AND FUTURE WORK

In this study, we presented a novel approach using CV parameterized learnable filters (PLFs), enabling the network to develop more interpretable filters that directly process complex RF data signals. To optimize our model's efficiency and

compatibility with the raw complex RF data stream, we crafted a complex PLF layer. This layer allows the model to learn filters characterized by trainable parameters of the defined filter functions such as the center frequency and bandwidth. Parameterized filters such as sinc, Gaussian, Gammatone, and Ricker wavelets have been developed and tested. The remaining architecture used complex versions of standard DNN components, including convolution, max-pooling, batch normalization, ReLU, and dropout.

We validated the effectiveness of our architecture on both synthetic and experimental datasets. For the synthetic dataset, a five-class RF waveform modulation recognition dataset is generated, while for the experimental case, a 100-ASL sign RF dataset is used. Our current approach not only slightly outperforms the widely used state-of-the-art 2-D CNNs that operate on μD-spectrogram images and other CNN variants that handle 1-D radar data in terms of accuracy but also significantly reduces latency, indicating strong potential for near real-time RF data classification. The learned filters rely solely on meaningful physical parameters such as center frequency and bandwidth, enhancing interpretability. Future research will focus on deploying the proposed models on edge computing platforms to showcase near real-time applications. Recognizing multiple signal modulations arriving simultaneously is another challenging problem that could benefit from the proposed approach.

REFERENCES

[1] E. Kurtoğlu, S. Biswas, A. C. Gurbuz, and S. Z. Gurbuz, "Boosting multi-target recognition performance with multi-input multi-output radar-based angular subspace projection and multi-view deep neural network," *IET Radar, Sonar Navigat.*, vol. 17, no. 7, pp. 1115–1128, Jul. 2023.

[2] S. Gurbuz, *Deep Neural Network Design for Radar Applications*. London, U.K.: IET, 2020.

[3] G. Kong, M. Jung, and V. Koivunen, "Waveform classification in radar-communications coexistence scenarios," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[4] Y. Shi, K. Davaslioglu, Y. E. Sagduyu, W. C. Headley, M. Fowler, and G. Green, "Deep learning for RF signal classification in unknown and dynamic spectrum environments," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Nov. 2019, pp. 1–10.

[5] V. Chen, *The Micro-Doppler Effect in Radar*, 2nd ed., Norwood, MA, USA: Artech House, 2019.

[6] S. Biswas and A. C. Gurbuz, "Deep learning based high-resolution frequency estimation for sparse radar range profiles," in *Proc. IEEE Radar Conf. (RadarConf24)*, vol. 32, May 2024, pp. 1–6.

[7] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019.

[8] E. Kurtoglu, A. C. Gurbuz, E. A. Malaia, D. Griffin, C. Crawford, and S. Z. Gurbuz, "ASL trigger recognition in mixed activity/signing sequences for RF sensor-based user interfaces," *IEEE Trans. Hum.-Mach. Syst.*, vol. 52, no. 4, pp. 699–712, Aug. 2022.

[9] B. Debnath, I. A. Ebu, S. Biswas, A. C. Gurbuz, and J. E. Ball, "FMCW radar range profile and micro-Doppler signature fusion for improved traffic signaling motion classification," in *Proc. IEEE Radar Conf. (RadarConf24)*, vol. 22, May 2024, pp. 1–6.

[10] R. Yu, Y. Du, J. Li, A. Napolitano, and J. Le Kernec, "Radar-based human activity recognition using denoising techniques to enhance classification accuracy," *IET Radar, Sonar Navigat.*, vol. 18, no. 2, pp. 277–293, Feb. 2024.

[11] J. Guo et al., "Complex field-based fusion network for human activities classification with radar," in *Proc. IET Int. Radar Conf.*, vol. 2020, Nov. 2020, pp. 68–73.

[12] A. Dahal, S. Biswas, A. C. Gurbuz, and S. Z. Gurbuz, "Comparison between Wi-Fi-CSI and radar-based human activity recognition," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2024, pp. 1–6.

[13] M. Chakraborty, H. C. Kumawat, S. V. Dhavale, and A. B. Raj, "Application of DNN for radar micro-Doppler signature-based human suspicious activity recognition," *Pattern Recognit. Lett.*, vol. 162, pp. 1–6, Oct. 2022.

[14] S. Z. Gurbuz, C. Clemente, A. Balleri, and J. J. Soraghan, "Micro-Doppler-based in-home aided and unaided walking recognition with multiple radar and sonar systems," *IET Radar, Sonar Navigat.*, vol. 11, no. 1, pp. 107–115, Jan. 2017.

[15] S. Biswas, B. Bartlett, J. E. Ball, and A. C. Gurbuz, "Classification of traffic signaling motion in automotive applications using FMCW radar," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2023, pp. 1–6.

[16] S. Biswas, A. Manavi Alam, and A. C. Gurbuz, "HRSpecNET: A deep learning-based high-resolution radar micro-Doppler signature reconstruction for improved HAR classification," *IEEE Trans. Radar Syst.*, vol. 2, pp. 484–497, 2024.

[17] P. Pan, Y. Zhang, Z. Deng, S. Fan, and X. Huang, "TFA-Net: A deep learning-based time-frequency analysis tool," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 9274–9286, Nov. 2023, doi: 10.1109/TNNLS.2022.3157723.

[18] S. Yang, J. L. Kernec, F. Fioranelli, and O. Romain, "Human activities classification in a complex space using raw radar data," in *Proc. Int. Radar Conf. (RADAR)*, 2019, pp. 1–4.

[19] Z. Li, J. Le Kernec, Q. Abbasi, F. Fioranelli, S. Yang, and O. Romain, "Radar-based human activity recognition with adaptive thresholding towards resource constrained platforms," *Sci. Rep.*, vol. 13, no. 1, p. 3473, Mar. 2023.

[20] C. Loukas, F. Fioranelli, J. Le Kernec, and S. Yang, "Activity classification using raw range and I & Q radar data with long short term memory layers," in *Proc. IEEE 16th Intl Conf Dependable, Autonomic Secure Comput.*, Aug. 2018, pp. 441–445.

[21] X. Yao, X. Shi, and F. Zhou, "Human activities classification based on complex-value convolutional neural network," *IEEE Sensors J.*, vol. 20, no. 13, pp. 7169–7180, Jul. 2020.

[22] D. A. Brooks, O. Schwander, F. Barbaresco, J.-Y. Schneider, and M. Cord, "Complex-valued neural networks for fully-temporal micro-Doppler classification," in *Proc. 20th Int. Radar Symp. (IRS)*, 2019, pp. 1–10.

[23] X. Yang, R. G. Guendel, A. Yarovoy, and F. Fioranelli, "Radar-based human activities classification with complex-valued neural networks," in *Proc. IEEE Radar Conf. (RadarConf)*, Mar. 2022, pp. 1–6.

[24] M. Ravanelli and Y. Bengio, "Speaker recognition from raw waveform with SincNet," in *Proc. IEEE Spoken Lang. Technol. Workshop (SLT)*, Dec. 2018, pp. 1021–1028.

[25] S. Sivapatham, A. Kar, and M. G. Christensen, "Gammatone filter bank-deep neural network-based monaural speech enhancement for unseen conditions," *Appl. Acoust.*, vol. 194, Jun. 2022, Art. no. 108784.

[26] H. Khan and B. Yener, "Learning filter widths of spectral decompositions with wavelets," in *Proc. NeurIPS*, vol. 31, 2018, pp. 4601–4612.

[27] A. Alekseev and A. Bobe, "GaborNet: Gabor filters with learnable parameters in deep convolutional neural network," in *Proc. Int. Conf. Eng. Telecommun. (EnT)*, Nov. 2019, pp. 1–4.

[28] S. Biswas, C. O. Ayna, S. Z. Gurbuz, and A. C. Gurbuz, "CV-SincNet: Learning complex sinc filters from raw radar data for computationally efficient human motion recognition," *IEEE Trans. Radar Syst.*, vol. 1, pp. 493–504, 2023.

[29] H. Peic Tukuljac, B. Ricaud, N. Aspert, and L. Colbois, "Learnable filter-banks for cnn-based audio applications," in *Proc. Northern Lights Deep Learn. Workshop*, 2022.

[30] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," Appl. Psychol. Unit, Cambridge, U.K., APU Rep. 2341, 1988.

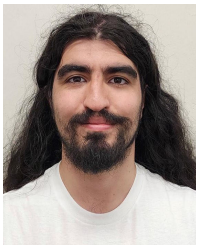
[31] S. Z. Gurbuz and M. G. Amin, "Radar-based human-motion recognition with deep learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, Jul. 2019.

[32] M. M. Rahman, E. A. Malaia, A. C. Gurbuz, D. J. Griffin, C. Crawford, and S. Z. Gurbuz, "Effect of kinematics and fluency in adversarial synthetic data generation for ASL recognition with RF sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 2732–2745, Aug. 2022.

Sabyasachi Biswas (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2019. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Mississippi State University, Mississippi State, MS, USA.

He was a Machine Learning Intern at the High-Performance Computing Collaboratory in Summer 2022, Mississippi State. He is a Research Assistant at the Information Processing and Sensing (IMPRESS) Laboratory, Mississippi State. His research interests include radar signal processing, human activity recognition using radar, camera, and LiDAR, and developing machine learning algorithms for activity classification using raw radar signals.

Mr. Biswas is a member of IEEE Signal Processing Society. He was the Winner and Second Runner-Up of the Graduate Research Symposium in Fall 2022 and 2023, respectively, held at Mississippi State University.



Cemre Omer Ayna (Graduate Student Member, IEEE) received the B.S. degree in electrical and electronics engineering from Koç University, Istanbul, Turkey, in 2021. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Mississippi State University (MSU), Mississippi State, MS, USA.

He has been a Graduate Research Assistant at IMPRESS Laboratory, MSU, working under an NSF CAREER project, since 2021. His main research area is applying machine learning for measurement learning and signal reconstruction problems. He also has works related to image reconstruction and demosaicing, compressive sensing, image segmentation, hyperspectral image processing, RADAR signal processing, model-informed machine learning, and adversarial machine learning.

Ali Cafer 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 in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2005 and 2008, respectively.

He held a Post-Doctoral position at Georgia Tech, in 2009, where he researched compressive-sensing-based computational imaging problems. He held Faculty positions at TOBB University, Ankara, and University of Alabama, Tuscaloosa, AL, USA, from 2009 to 2017, 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. Currently, he is an Associate Professor at the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA, where he is the Director of Information Processing and Sensing (IMPRESS) Laboratory.

Dr. Gurbuz was a recipient of the Best Paper Award for *Signal Processing Journal* in 2013, Turkish Academy of Sciences Best Young Scholar Award in Electrical Engineering in 2014, and the NSF CAREER Award in 2021. He is currently an Associate Editor of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and has served as an Associate Editor for several journals in signal processing and communication areas previously.