Finding a Needle in a (Spectrum) Haystack: Multi-Band Multi-Device Radio Fingerprinting

Ildi Alla[a,b], Milin Zhang[c], Jonathan Ashdown[d], Valeria Loscri[a], Francesco Restuccia[c]

[a] *Inria Lille-Nord Europe, France*
[b] *SnT, University of Luxembourg, Luxembourg*
[c] *Institute for the Wireless Internet of Things at Northeastern University, United States*
[d] *Air Force Research Laboratory, United States*

Abstract

As the spectrum becomes increasingly crowded, quick and reliable authentication of wireless devices is critical to avoid harmful interference to incumbents of the spectrum. *Radio fingerprinting* achieves fast waveform-level authentication by distinguishing devices based on unique hardware imperfections in the radio circuitry. However, existing approaches can fingerprint only one signal in a specific band, making them inapplicable in real-world scenarios where multiple signals coexist in spectrum bands. This paper introduces *Multi-band Multi-device Radio Fingerprinting* (`M2RF`) to address this challenge. Specifically, we propose a learning-driven segmentation algorithm to directly process in-phase/quadrature (I/Q) samples coming from the receiver and assign each I/Q sample to a specific radio. In contrast to existing approaches, `M2RF` simultaneously identifies and locates in the spectrum multiple devices that emit overlapping signals and avoids the burden of processing data, making the overall approach with reduced overhead and faster. Our approach can be generalized to different channels and signal bandwidths without retraining, making it scalable. Experiments in three different spectrum scenarios under 2 transmission conditions and with 15 radio transmitters demonstrate the effectiveness of `M2RF`, achieving up to 99.56% of F1-score,

Email addresses: `ildi.alla@uni.lu` (Ildi Alla), `zhang.mil@northeastern.edu` (Milin Zhang), `jonathan.ashdown@us.af.mil` (Jonathan Ashdown), `valeria.loscri@inria.fr` (Valeria Loscri), `f.restuccia@northeastern.edu` (Francesco Restuccia)

This article has been accepted for publication in *Computer Networks* (Elsevier). This is the author's accepted manuscript.

and 92.44% detection rate of malicious users with only a 2.72% mean Miss Rate (MR). A demo video of M2RF is available (M2RF – Demo Video).

Keywords: RF Fingerprinting, Multi-Device Authentication, Semantic Spectrum Segmentation, Dynamic Spectrum Sharing, Anomaly Detection

1. Introduction

The sheer growth of the Internet of Things (IoT) is quickly saturating unlicensed spectrum bands [1]. As unlicensed bands become saturated, *spectrum sharing* will become one of the very few options to sustain the IoT growth in the years to come [2, 3, 4, 5]. The key issue is that today, IoT operators that want to share spectrum with licensed users – also called *incumbents* – must contact database systems located in the cloud, which determine if the spectrum is available based on geographical coordinates [6]. *This centralized manual approach lacks scalability and does not allow for fine-grained real-time spectrum management.* Conversely, a scalable and effective solution would be to let IoT devices opportunistically discover which spectrum sub-bands are currently available among ongoing licensed transmissions, provided they do not cause harmful interference to incumbents [7].

It is easy to observe that dynamic spectrum access systems will create fundamentally new security challenges where incumbents must be protected by secondary users not abiding by spectrum rules. To prevent such issues, spectrum must be *continuously monitored* to make sure only authorized devices are using the spectrum. Traditional wireless authentication systems such as WPA for Wi-Fi [8] or 5G-AKA for cellular networks [9] are based on cryptography or password-based authentication. As such, they operate primarily on the network or application layers, failing to meet the real-time requirements for spectrum sharing [10], [11]. In addition, these methods are proven insufficient against various attacks, such as spoofing, replay, and impersonation attacks [12, 13, 14].

In recent years, *radio fingerprinting* has emerged as a viable approach to spectrum-level authentication. Specifically, radio fingerprinting leverages the inherent hardware imperfections present in every radio circuitry [15, 16, 17, 18] to form a unique and unforgeable "fingerprint" that can authenticate devices [19]. By exploiting these characteristics, radio fingerprinting offers a security solution that is resistant to attacks such as MAC address spoofing and identity cloning [20].

Existing work – discussed in details in Section 2 – has a series of core limitations that make it unable to perform real-time spectrum-level authentication. Specifically, Figure 1 shows the fundamental difference between prior work and our proposed approach. **First**, current approaches only classify one signal in a given channel of interest. Conversely, multiple signals are usually overlapping in adjacent bands making the classification problem harder. **Second**, conventional methods assume prior knowledge of operating frequency of transmitters and only classify signals in that specific frequency band. However, signals may be partially observed by the receiver, e.g., because they are partially outside the operating bandwidth.

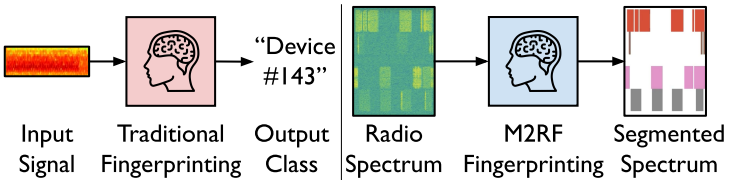


Figure 1: Traditional Radio Fingerprinting vs M2RF.

This paper changes the current state of the art by proposing the first ever spectrum-level authentication system named *Multi-band Multi-device Radio Fingerprinting* (M2RF), where multiple devices are located and identified in the same spectrum band using *spectrum segmentation*. The right side of Figure 1 shows at a very high level the main objective of M2RF. The proposed approach directly operates on unprocessed in-phase/quadrature (I/Q) inputs coming from the radio receiver front-end, thus eliminating pre-processing steps. The proposed spectrum segmentation model, based on a Deep Neural Network (DNN), has been specifically designed to handle dynamic signal and channel bandwidths through the integration of a non-local block, which captures long-range dependencies across frequency and distinguishes subtle differences in RF signals via a self-attention mechanism. In addition, M2RF incorporates a combined loss function that integrates both local-level and region-level features, further enhancing its ability to learn intricate signal features while maintaining consistent accuracy. An aggregation block is built to support wideband classification by combining predictions across overlapping frequency bands, allowing the model to span and accurately identify signals across different frequency segments.

The unique features of our solutions are very promising on different open problems. From one side, it is possible to detect in real-time the *intrusion*

tentative from malicious nodes. Another really interesting perspective provided by our approach is the capacity to "locate" in the spectrum the activity of malicious devices, posing the fundamentals of advanced anti-jamming solutions, targeting with high precision the "malicious" operating frequencies. From another perspective, our approach permits to better manage the shared resources, due to the real-time information of who (device) is where (in the spectrum).

Summary of Novel Contributions

• We propose a real-time radio fingerprinting approach named `M2RF` that can simultaneously fingerprint multiple devices coexisting in the shared spectrum. `M2RF` includes (i) a scalable dataset generation pipeline that can represent real-world spectrum conditions such as overlapping signals, and (ii) an energy-efficient DNN optimized for resource-constrained devices. To the best of our knowledge, this is the first work proposing a simultaneous multi-device radio fingerprinting system;

• We introduce a new anomaly detection mechanism to detect adversary and interference in spectrum sharing scenarios. We leverage Total Variation (TV) analysis to identify attacks by detecting irregularities in the DNN output. Specifically, it exploits the fact that the DNN produces noisy and randomized outputs when fed with an unseen signal. This means that real-time detection is realized without prior knowledge of the specific attacks strategy;

• We evaluate the performance of `M2RF` using a comprehensive 82 GB dataset of over-the-air (OTA) data from 15 identical Wi-Fi cards, which represents the worst case for radio fingerprinting as identical devices may have closer fingerprints [21]. In addition, we have collected data via wired connection to have data unaffected by the wireless channel [22]. To simulate real-world threats, we consider both informed and uninformed adversaries. For *informed adversary*, we collected data from additional identical Wi-Fi devices having full knowledge of the authentication approach. For *uninformed adversary*, we collected data from different Wi-Fi cards. Moreover, we collect other wireless technologies as interference (e.g., BLE, LTE, Zigbee) to evaluate the `M2RF` performance in congested multi-technology environments;

• Our experimental results show that `M2RF` achieves F1-score of 94.99% and Intersection over Union (IoU) of 90.54% with over-the-air non-overlapping signals. In the challenging scenario of overlapping signals, `M2RF` achieves F1-score of 77.06% and IoU of 63.39% without retraining and/or fine-tuning. Moreover, `M2RF` detects adversaries with an accuracy of 92.44%, demonstrat-

ing resilience against both informed and uninformed attacks. When other technologies are present, M2RF achieves an overall accuracy of 81.52%. A demo video of M2RF is available (M2RF – Demo Video).

2. Background and Motivation

Radio fingerprinting is a technology that authenticates wireless devices based on unique characteristics inherent in their transmitted radio signals [23]. The key idea is based on the fact that each radio device has its unique hardware imperfections in its circuitry, which manifest as subtle yet measurable differences in signal transmission. Compared to conventional cryptography-based methods, radio fingerprinting offers a more robust authentication mechanism since these physical properties are inherently unclonable. Furthermore, by operating directly at the physical layer, radio fingerprinting provides greater agility and computational efficiency without requiring full-stack protocol operations.

With the rapid development of IoT, there is an increasing need of wireless communication service to connect massive devices to network. To maximize spectrum efficiency in massive connectivity scenarios, dynamic spectrum management has been proposed to enable opportunistic signal transmission in available sub-bands [24]. To this end, a scalable and rapid authentication method is required to identify massive IoT devices in real time. The computational agility inherent in waveform-level operations makes radio fingerprinting a promising candidate for this purpose.

However, existing radio fingerprinting approaches fail to address the following challenges in spectrum sharing:

• **Dynamic Operating Frequency**. Existing approaches involve band filtering and pre-processing to remove interference and channel effect before classification [25], which assumes prior knowledge of the bandwidth and operating frequency of transmitted signal. However, in the spectrum sharing system, devices can dynamically select their operating frequencies based on spectrum availability. Therefore, the target signal may not operate at the same frequency as assumed by the radio fingerprinting method, or may even fall partially outside the filter bandwidth, causing the algorithm to fail in identifying the target device.

• **Simultaneous Transmissions**. In spectrum sharing, multiple devices can transmit simultaneously within the same spectrum. This creates a significant

scalability challenge for existing radio fingerprinting methods, as current approaches can only authenticate one device at a time [22]. To classify multiple signals in real time, these methods must iteratively process multiple signal instances across different operating frequencies, which introduces considerable computational overhead and latency.

• **Uncontrolled Interference**. Another significant challenge in spectrum sharing is that the uncontrolled spectrum environment can have considerable noise and interference which can compromise the accuracy of fingerprinting. As the spectrum is an open resource, interference may exist intentionally or unintentionally. Current fingerprinting approaches designed to work in controlled environments and with minimal interference can fail to generalize to the complex and varying environment [26].

These limitations motivate us to create a brand-new radio fingerprinting design that can simultaneously identify multiple signals in the spectrum in real time. Specifically, we aim to address the following research questions:

• **RQ1 – How to properly model and process the complex spectrum environment?**

A typical data pre-processing pipeline in conventional radio fingerprinting involves shifting signals to the operating frequency, removing noise outside the band of interest, and processing the signal within the band for feature extraction [22]. This approach ensures that the data is controlled and purified to improve classification performance but requires prior knowledge of operating frequency and can only identify one radio at a time, which does not meet the requirements of spectrum sharing environments. To overcome this limitation, a new pipeline is needed to model and process complex spectrum environments where multiple radio transmissions do not match the expected operating frequency or are only partially observable by the receiver.

To address **RQ1**, we create a new data pre-processing pipeline – described in Section 4.1 – that can simulate complex spectrum conditions with controlled data transmission and pre-processing. This pipeline, similar to conventional radio fingerprinting approaches, creates a purified signal repository that enables the algorithm to effectively learn useful features in the signals of interest. However, it is distinct from other radio fingerprinting methods by augmenting and stitching multiple purified signals to simulate real spectrum conditions, where multiple signals coexist, overlap, or are partially observable.

• **RQ2 – How can simultaneous radio device authentication be achieved in this spectrum environment?**

A naive approach to achieving multiple device authentication would be to extend current radio fingerprinting methods to iteratively process each signal in the spectrum. However, this approach requires additional complexity to identify the center frequency of each radio transmission, as signals are transmitted dynamically throughout the spectrum. Moreover, the computational burden increases significantly when massive transmissions occur within the same spectrum, resulting in substantial latency. Therefore, developing a new fingerprinting paradigm that can achieve simultaneous multi-device authentication is a critical research challenge in dynamic spectrum sharing.

In Section 4.2, we address **RQ2** by leveraging a novel DNN solution based on "semantic spectrum segmentation". The neural network is trained to take I/Q samples as input, and directly segment waveforms in the frequency domain. This way, it removes the complexity of per-signal processing and hence achieving a real-time multi-device authentication.

• RQ3 – Is the new framework generalizable to different frequency and environment?

Spectrum sharing requires real-time monitoring of an ultra-wide spectrum band which is typically larger than the observable bandwidth of the spectrum sensor. For example, [27] considered a scenario covering the frequency range from 400 MHz to 6 GHz. It is infeasible to create a single fingerprinting algorithm to monitor the entire 6 GHz spectrum due to hardware constraints. A viable approach is to divide the entire spectrum into multiple channels and leverage fingerprinting to rapidly scan these channels. However, existing radio fingerprinting methods developed in controlled environments often struggle to generalize effectively to different environments and scenarios. For instance, [28] reported an 82% accuracy drop when testing in real-world scenarios. To ensure the approach is practical in real-world applications, it is critical to validate whether the proposed framework is generalizable.

To answer this question, we design adaptive bandwidth processing by sweeping and aggregating fingerprinting results across multiple channels. This approach enables the fingerprinting algorithm to be tested across different channel bandwidths and signal bandwidths. We describe this adaptive processing in Section 4.3.

• RQ4 – Can the algorithm detect interference and adversaries in the spectrum?

As the spectrum is an open resource, interference may occur intentionally or unintentionally. For example, an unauthorized device may attempt to occupy a channel by intentionally mimicking the behavior of authorized ra-

dios through spoofing or replay attacks. In addition, other unknown signals may transmit within the band and cause unintentional interference. Such interference may cause misclassification of the fingerprinting system, hence compromising the spectrum management. Therefore, detecting interference and adversaries in the spectrum is as important as identifying the authorized devices.

To address **RQ4**, we propose a post-processing method based on total variation that can effectively detect anomaly in the spectrum. The key idea is that interference will have lower certainty in the inference results, which can be detected by checking the consistency of inference. A detailed discussion of this method is provided in Section 4.4.

3. Threat Model and System Overview

As spectrum is an open resource, malicious traffic poses significant threats to legitimate users. For example, adversarial devices may attempt to authenticate by cloning legitimate user behavior, while unintentional interference can occur when signals are transmitted on the same channel. Both adversaries and interference can severely degrade the quality of service for authorized users. Therefore, an effective spectrum management requires the algorithm not only to authenticate legitimate users but also to detect malicious traffic in the spectrum. In this section, we outline these potential threats in the spectrum, as well as how `M2RF` is structured to defend against these challenges in a high level.

3.1. Threat Model

Figure 2 overviews the threat model in the dynamic spectrum management. Alice, the authorized user, will access the network through a specific channel with its unique hardware characteristics. On the other hand, Eve, the adversary tries to access the network through the same channel by cloning Alice's behavior. Bob, the authenticator, continuously monitors the spectrum to authenticate Alice and detect Eve using radio fingerprinting techniques. In this scenario, Eve can perform different attack strategies:
① *Spoofing.* This scenario involves an attacker (Eve) emulating the credentials of an authorized device (Alice) by cloning identifiers such as MAC addresses. Here, Eve's goal is to deceive the authentication system by masquerading as a legitimate device without replicating the hardware-specific imperfections that are unique to Alice.

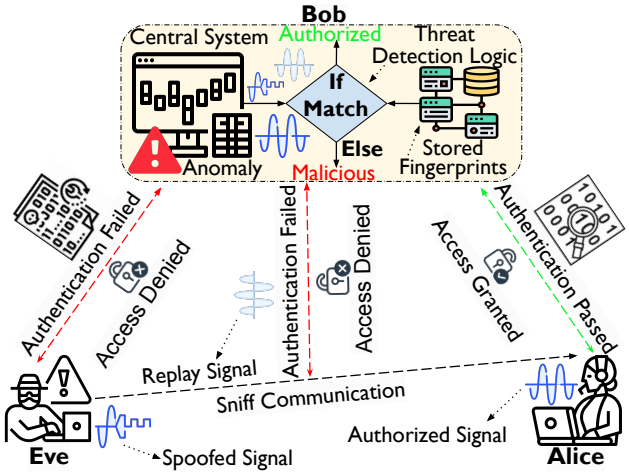


Figure 2: Overview of interactions among Alice (authorized device), Eve (attacker), and Bob (authenticator) to detect and prevent spoofing and replay attacks through radio fingerprinting.

② *Replay Attacks.* Eve intercepts transmissions from Alice and replays them, aiming to deceive the system and gain unauthorized access without the need to directly imitate Alice's radio signal characteristics. Replay attacks exploit captured communication sessions, assuming they will appear legitimate upon retransmission.
③ *Device Impersonation.* Eve manipulates signal characteristics to closely imitate Alice's fingerprint. By using similar devices or attempting software-based modifications [29], Eve aims to create a sufficiently close match to bypass RF fingerprint detection. This approach assumes that Eve has knowledge of Alice's signal characteristics and attempts to mimic them. Still, the unique hardware imperfections inherent to Alice's device cannot be acquired by Eve.

3.2. Type of Malicious Traffic

In our system, we consider three different type of malicious traffic based on its knowledge model for the legitimate device:
Informed Adversary. Eve possesses detailed knowledge of Alice's hardware features, the radio fingerprinting algorithm and how the authentication is performed by Bob. This knowledge allows Eve to adopt more sophisticated techniques to approximate Alice's signal characteristics.

Uninformed Adversary. Eve has basic knowledge about the system such as the wireless technology (e.g. Wi-Fi) but lacks specific knowledge about Alice's hardware imperfections and Bob's detection mechanisms. Eve may attempt standard spoofing or basic replay methods without insight into the physical layer defense, relying on generic attack methods.
Interference. Eve has no knowledge about the system. It unintentionally occupies the channel and creates malicious interference to the legitimate user. This scenario represents a common shared spectrum situation where different wireless technologies operate in the same frequency band (e.g. Bluetooth vs Wi-Fi).

3.3. Defense Mechanism

In a high level, the proposed framework `M2RF` achieves robust authentication with following strategies:
① *Real-Time Monitoring.* `M2RF` continuously monitors spectrum of interest, adapts to different frequency bands and bandwidths with scanning and aggregation as described in Section 4.3 for authentication of legitimate users and detection of malicious traffic.
② *Multi-Device Authentication.* As discussed in Section 4.2, `M2RF` leverages a Deep Learning (DL)-driven semantic segmentation algorithm to directly label each waveform data (I/Q samples) in the frequency domain based on their waveform features, which results in a simultaneous labeling of all waveforms in the channel.
③ *Anomaly Detection.* `M2RF` detects malicious traffic across multiple frequency bands based on fingerprint consistency check. Compared to legitimate signals, malicious signals show increased randomness in the inference results, which can be detected by total variation as detailed in Section 4.4.

4. The `M2RF` Framework

To address the research questions outlined in Section 2, we propose `M2RF`, a new radio fingerprinting framework for spectrum sharing. Figure 3 overviews the main components of `M2RF`. The process begins with acquiring I/Q data, followed by a novel pre-processing pipeline to address **RQ1**. These signals are then used to train a DNN for "semantic spectrum segmentation", a new approach which effectively address **RQ2**. During inference phase, adaptive

bandwidth processing and anomaly detection modules are proposed to address **RQ3** and **RQ4** respectively, ensuring the proposed framework is practical in real-world scenarios. We explain each component of `M2RF` in the following subsections.

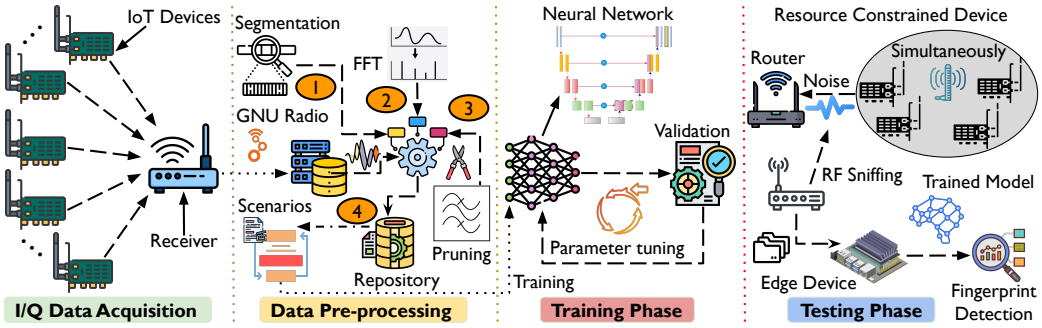


Figure 3: Overview of `M2RF`, from I/Q data acquisition through data pre-processing, training, and testing in real-world scenarios.

4.1. Data Pre-processing

Data collected in controlled environments often fails to generalize to complex real-world spectrum conditions. However, directly collecting spectrum data in open environments presents its own difficulties: unknown signals may be transmitting simultaneously, creating interference that compromises algorithm performance. Additionally, labeling data containing these unknown signals is inherently challenging, as their sources and characteristics cannot be readily identified.

To address **RQ1**, we introduce a novel data pre-processing pipeline that effectively simulates comprehensive spectrum conditions through controlled data collection. This pipeline comprises two components: controlled data collection to establish a comprehensive signal repository, and a data augmentation process that simulates real-world scenarios using signals from the controlled environment. Note that this pipeline is only applied during the training phase. During inference, the fingerprinting algorithm operates directly on real-world spectrum data.

4.1.1. Signal Repository

We first build a curated collection of individual, high-quality radio signals captured under controlled conditions. These signals are recorded sequentially,

ensuring that only one transmission occurs at a time, with known center frequencies f_c and bandwidths B. This controlled environment ensures that each signal is free from external interference or overlapping transmissions, capturing the characteristics necessary for subsequent processing and accurate radio fingerprinting. Once collected, each radio signal $s(t)$ undergoes the following procedure:

❶ **Segmentation**: The continuous time-domain signal $s(t)$ is segmented into smaller, fixed-length portions to capture individual signal instances. We divide the continuous recording into segments of duration T, which corresponds to a fixed number of I/Q samples, to isolate relevant transmissions. The segmentation process can be represented as:

$$s_{\text{seg}}[n] = s[n] \cdot w[n], \tag{1}$$

where $w[n]$ is a rectangular windowing function defined in discrete time as:

$$w[n] = \begin{cases} 1 & \text{for } 0 \leq n < N \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where N is the fixed number of I/Q samples in each segment and $s_{\text{seg}}[n]$ represents the n-th sample of the segmented signal. The segmented signal $s_{\text{seg}}[n]$ contains a fixed duration of the transmission, ready for frequency domain processing.

❷ **Fast Fourier Transform (FFT)**: The segmented time-domain signal $s_{\text{seg}}[n]$ is then converted to the frequency domain using the Fast Fourier Transform (FFT). This transformation yields the frequency spectrum $S_{\text{fft}}(f)$, which represents the signal's frequency components:

$$S_{\text{fft}}(f) = \text{FFT}\{s_{\text{seg}}[n]\}. \tag{3}$$

❸ **Pruning of Unwanted Frequency Components**: To focus on the signal band of interest and eliminate out-of-band noise, we apply frequency pruning. This step retains only the frequency components within the bandwidth B around the center frequency f_c, effectively isolating the relevant spectral portion for radio fingerprinting. The pruned signal $S_{\text{pruned}}(f)$ is obtained by applying a binary mask in the frequency domain:

$$S_{\text{pruned}}(f) = S_{\text{fft}}(f) \cdot M(f), \tag{4}$$

where $M(f)$ is a frequency mask defined as:

$$M(f) = \begin{cases} 1 & \text{for } f_c - \frac{B}{2} \leq f \leq f_c + \frac{B}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

❹ **Storage in Signals Repository**: The pruned frequency-domain I/Q samples $S_{\text{pruned}}(f)$ are then stored in the signals repository. This repository serves as a comprehensive and clean dataset of individual radio fingerprints in the frequency domain. It is designed for subsequent use in analysis, scenario simulation, and training of the fingerprinting model, ensuring the necessary data quality for accurate device identification. The pre-processing pipeline, based on our implemented steps, can be summarized as the transformation:

$$s(t) \to s_{\text{seg}}[n] \to S_{\text{fft}}(f) \to S_{\text{pruned}}(f) \to \text{sig}_{\text{repo}}. \tag{6}$$

4.1.2. Simulation of Scenarios

Scenario generation is a crucial step that simulates a wide variety of real-world environments. By integrating multiple signals into a "stitched" wideband signal, this approach reduces the need for extensive real-world data collection. The spectrum stitching process operates by placing individual signals into an extended temporary buffer and then extracting the central portion. Let $\mathbf{S}_i \in \mathbb{R}^{W_i \times 2}$ denote the i-th pruned signal with bandwidth W_i bins, where $W_i = \lceil n_{iq} \cdot w_i / B \rceil$ for signal bandwidth w_i in Hz. Here, n_{iq} is the number of I/Q samples (frequency bins) used to discretize the observable bandwidth B. We create an extended buffer $\mathbf{T} \in \mathbb{R}^{(n_{iq}+W_{\max} \cdot 2-2) \times 2}$ where $W_{\max} = \max_i W_i$ is the maximum signal bandwidth in bins.

For each signal $\mathbf{S}_i$, we determine its placement location ℓ_i according to the scenario configuration:

$$\ell_i = \begin{cases} \lfloor |\mathbf{T}|/2 - W_i/2 \rfloor & \text{with probability } p_{\text{centered}} \\ \text{Uniform}(0, |\mathbf{T}| - W_i) & \text{otherwise} \end{cases}, \tag{7}$$

where $|\mathbf{T}|$ denotes the length of the temporary buffer, and p_{centered} controls the probability of centered placement. The stitching operation accumulates signals into the temporary buffer:

$$\mathbf{T}[\ell_i : \ell_i + W_i, :] \mathrel{+}= \mathbf{S}_i, \tag{8}$$

where the += operator allows overlapping signals to sum naturally, simulating realistic spectral interference. After all signals (n_{signals}) are placed, we extract the central n_{iq} samples:

$$\mathbf{X} = \mathbf{T}[W_{\max} - 1 : W_{\max} + n_{iq} - 1, :] + \mathbf{N}, \tag{9}$$

where $\mathbf{N} \in \mathbb{R}^{n_{iq} \times 2}$ is background noise sampled from the signal repository. This extended buffer approach ensures that signals placed near the edges of the observable bandwidth are not artificially truncated, maintaining realistic spectral characteristics. The corresponding label matrix $\mathbf{L} \in \{0, 1\}^{C \times n_{iq}}$ is constructed by tracking which device classes occupy each frequency bin through boolean masks that follow the same placement and extraction procedure.

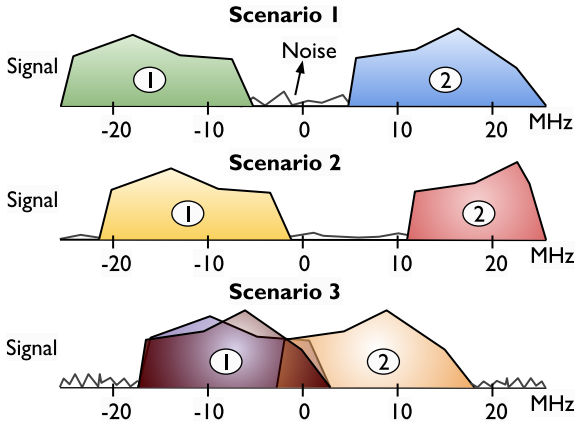


Figure 4: Visual representation of scenarios: (Top) Scenario 1 with non-overlapping signals, (Middle) Scenario 2 with randomly positioned signals, and (Bottom) Scenario 3 showcasing partial or full overlapping signals, all within a 50 MHz bandwidth.

By using such stitching procedure made with individual signal widths, our pipeline dynamically generates numerous training samples that can work with different signal bandwidths. Figure 4 shows an example of scenario generation with 50 MHz observable bandwidth and 20 MHz wide signals. The scenarios are generated with a procedure described in Algorithm 1, which assembles signals into a complete training sample via 'spectrum stitching'.

The algorithm first determines if the observable bandwidth will be empty, based on the probability $prob_{empty}$. If not empty, it randomly selects $n_{signals}$

Algorithm 1 Sample generation using spectrum stitching augmentation.

Require: sig_{repo} (signal repository), n_{iq} (number of I/Q samples), B (observable bandwidth), $max_{signals}$ (maximum simultaneous signals), $prob_{empty}$ (probability of empty bandwidth), $prob_{centered}$ (probability of centered placement)

Decide if the bandwidth is empty based on $prob_{empty}$
if bandwidth is not empty **then**
  Select randomly $n_{signals} \sim \text{Uniform}(1, max_{signals})$
  Randomly choose $n_{signals}$ transmitters from sig_{repo}
  for each signal $\mathbf{S}_i$ **do**
    Extract $\mathbf{S}_i$ from sig_{repo}
    Determine placement location ℓ_i using Equation 7
    if Scenario 1 **then**
      Sequential placement: $\ell_i = \ell_{i-1} + W_{i-1}$
    else if Scenario 2 **then**
      Random placement without overlap
    else if Scenario 3 **then**
      Random placement allowing overlaps (Equation 8)
    end if
    Update label matrix $\mathbf{L}$ and buffer $\mathbf{T}$
  end for
end if
Add background noise $\mathbf{N}$ from sig_{repo}
Extract final sample $\mathbf{X}$ using Equation 9
return $\mathbf{X}$ and $\mathbf{L}$

between 1 and $max_{signals}$ from the singal repository sig_{repo}. Placement within the bandwidth is guided by Equation 7, where parameter $prob_{centered}$ controls the distribution of signal positions. For Scenario 1, signals are placed sequentially, for Scenario 2, random placement without overlap is enforced, and for Scenario 3, overlapping is permitted through the accumulation operation in Equation 8. The final algorithm step adds background noise sourced from the sig_{repo} to the stitched signal to simulate realistic conditions. The resulting sample $\mathbf{X}$ is then stored with its label matrix. The resolution of scenarios, defined by the frequency sub-band size into which the observable bandwidth is divided, is given by: $resolution\ (R) = \frac{B}{n_{iq}}$. This resolution sets the granularity for analyzing and classifying the signal spectrum. The label matrix $\mathbf{L}$, structured as $C \times n_{iq}$ (where C is the number of classes), enables fine-grained classification across the bandwidth.

4.2. Spectrum Fingerprinting

As discussed in Section 2, conventional fingerprinting algorithms that classify one device at a time cannot scale to multi-device authentication. To address **RQ2**, we propose an approach based on semantic segmentation. This approach directly takes wideband RF data as input and labels each I/Q waveform in the frequency domain, enabling simultaneous localization and detection of multiple devices across the spectrum. The following subsections outline the multi-label semantic segmentation methodology, the structure of the DNN model, its adaptability to varying input sizes, generalization strategies, and the scalable processing techniques we implemented. Additionally, we provide a detailed discussion of the specific adversarial detection technique used in this approach. A detailed explanation of the various loss functions is provided in Appendix A.

4.2.1. Multi-Label Signal Segmentation

Our approach utilizes semantic segmentation, a technique originally developed for computer vision tasks. The key idea is to segment objects from the background by labeling each pixel belonging to those objects based on their semantic information within the frame. Similarly, we apply this idea to spectrum sharing tasks, identifying target signals within noisy spectrum environments.

Specifically, we transform the captured waveform into the frequency domain and divide it into multiple sub-channels. A DL-based semantic segmentation algorithm is then applied to detect signals across the bandwidth. Similar to image-based semantic segmentation, our signal segmentation approach labels each sub-channel based on waveform-level features. This enables the simultaneous detection and classification of multiple overlapping signals within the bandwidth.

One significant difference between the image segmentation and signal segmentation is the multi-label nature of the radio environment. In an image, the object in the behind will be blocked by the object in the front by its non-transparent nature. In contrast, different signals can coexist within the same frequency band without occluding each other in the radio environment. As a result, each frequency bin can be assigned to multiple classes simultaneously. Therefore, we extend the semantic segmentation algorithm to output a binary segmentation map for each class, where each map indicates the presence or absence of the corresponding class within the given frequency bin. The

final segmentation output is a matrix where each row corresponds to a class and each column corresponds to a frequency bin.

4.2.2. DNN Model Architecture

Our backbone is inspired from U-Net, which was initially proposed for biomedical image segmentation [30]. We adapted this architecture for radio fingerprinting by replacing the 2D convolutional layers with 1D convolutions to process I/Q samples effectively. As illustrated in Figure 5, the architecture comprises five encoding and five decoding blocks. The encoding path systematically downsamples the input data, capturing features at varying levels of abstraction through 1D convolutional layers, batch normalization, and ReLU activations. Max pooling layers are employed within each encoding block to reduce the spatial dimensions.

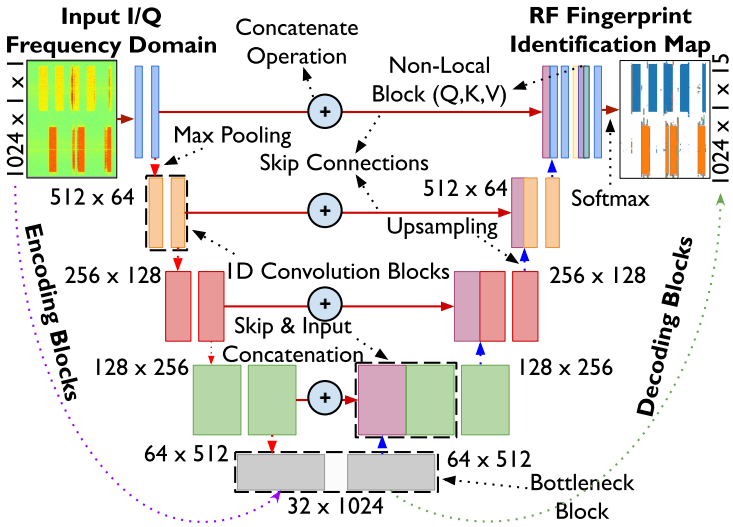


Figure 5: Adapted U-Net architecture for radio fingerprinting. The left side shows the encoding path for feature abstraction, the middle includes the bottleneck and non-local block for long-range dependencies, and the right side represents the decoding path.

The decoding path mirrors the encoding process, progressively reconstructing the data to its original size using upsampling layers. Skip connections between corresponding encoding and decoding blocks ensure that spatial information, crucial for accurately identifying device-specific characteristics in radio signals, is preserved throughout the DNN. The final layer

applies a 1x1 convolution to produce a multi-channel output, resulting in a $C \times n_{iq}$ matrix, where each channel corresponds to a different class in the multi-label segmentation task.

4.2.3. Integration of Non-local Block

The standard U-Net architecture relies entirely on convolutional operations with limited receptive fields, which can struggle to capture long-range dependencies across the frequency spectrum [31]. This limitation is particularly problematic for RF fingerprinting where signal characteristics may span large bandwidths and exhibit correlations across distant frequency bins. To address this challenge, we integrate a non-local block [32] that employs self-attention mechanisms to model global dependencies across the entire spectrum.

The non-local block computes a weighted aggregation of features from all positions in the feature map, enabling each frequency bin to attend to every other frequency bin regardless of distance. Given an input feature map $\mathbf{X} \in \mathbb{R}^{F \times n_{iq}}$ where F is the number of channels and n_{iq} is the number of frequency bins, we employ a bottleneck design with intermediate dimension $F' = F/2$ to reduce computational complexity. The block computes three transformed representations through 1×1 convolutions:

$$\boldsymbol{\theta}(\mathbf{X}) = \mathbf{W}_\theta \mathbf{X}, \quad \boldsymbol{\theta} \in \mathbb{R}^{F' \times n_{iq}}, \tag{10}$$

$$\boldsymbol{\phi}(\mathbf{X}) = \text{MaxPool}(\mathbf{W}_\phi \mathbf{X}, 2), \quad \boldsymbol{\phi} \in \mathbb{R}^{F' \times n_{iq}/2}, \tag{11}$$

$$\mathbf{g}(\mathbf{X}) = \text{MaxPool}(\mathbf{W}_g \mathbf{X}, 2), \quad \mathbf{g} \in \mathbb{R}^{F' \times n_{iq}/2}, \tag{12}$$

where $\mathbf{W}_\theta, \mathbf{W}_\phi, \mathbf{W}_g \in \mathbb{R}^{F' \times F}$ are learnable projection matrices implemented as 1×1 convolutional layers. The MaxPool operation with stride 2 spatially downsamples $\boldsymbol{\phi}$ and $\mathbf{g}$ by half, reducing computational complexity from $\mathcal{O}(F' \cdot n_{iq}^2)$ to $\mathcal{O}(F' \cdot n_{iq} \cdot n_{iq}/2) = \mathcal{O}(F' \cdot n_{iq}^2/2)$.

The self-attention mechanism computes pairwise affinities between all frequency positions. For each position $i \in [1, n_{iq}]$, the attention weights to all downsampled positions $j \in [1, n_{iq}/2]$ are:

$$\mathbf{f}_{i,j} = \frac{\exp(\boldsymbol{\theta}(\mathbf{x}_i)^\top \boldsymbol{\phi}(\mathbf{x}_j))}{\sum_{k=1}^{n_{iq}/2} \exp(\boldsymbol{\theta}(\mathbf{x}_i)^\top \boldsymbol{\phi}(\mathbf{x}_k))}, \tag{13}$$

where $\mathbf{f}_{i,j}$ represents the normalized attention weight between full-resolution position i and downsampled position j. This results in an attention matrix

$\mathbf{f} \in \mathbb{R}^{n_{iq} \times n_{iq}/2}$. The asymmetric dimensions, full resolution for queries ($\boldsymbol{\theta}$) and half resolution for keys and values ($\boldsymbol{\phi}, \mathbf{g}$), reduce computational cost while preserving the ability to capture global context at the full output resolution. This allows the network to learn which frequency positions are most relevant for classifying each location, effectively capturing long-range correlations that are essential for distinguishing overlapping RF signals with similar spectral characteristics.

The final output of the non-local block aggregates information from all downsampled positions using the computed attention weights:

$$\mathbf{y}_i = \mathbf{W}_z \left(\sum_{j=1}^{n_{iq}/2} \mathbf{f}_{i,j}\, \mathbf{g}(\mathbf{x}_j) \right) + \mathbf{x}_i, \tag{14}$$

where $\mathbf{W}_z \in \mathbb{R}^{F \times F'}$ projects the aggregated features back to the original channel dimension, followed by batch normalization with zero-initialized weights and biases. The residual connection $+\mathbf{x}_i$ ensures stable gradient flow and allows the block to learn incremental refinements. The zero initialization ensures that the non-local block initially acts as an identity mapping, allowing the network to progressively learn global dependencies during training without disrupting already-learned local features [32].

The computational complexity of the non-local block is $\mathcal{O}(F' \cdot n_{iq}^2/2)$ due to the pairwise affinity computation in Equation 13. While this enables powerful global reasoning, applying non-local blocks at multiple stages would result in prohibitive computational costs. Therefore, we strategically position a single non-local block after the final decoding layer, where the feature dimension has been reduced to C (number of classes), minimizing computational overhead while still capturing critical long-range dependencies for the final classification decision.

4.2.4. Noise Normalization

We implement adaptive noise normalization during inference to ensure robust generalization across varying signal-to-noise ratio (SNR) conditions encountered in real-world deployments. During training, we estimate the noise floor N_{floor} by recording the minimum values of the smoothed signal power spectrum across the training dataset. Specifically, for each training sample $\mathbf{s}$, we compute its power spectrum $P(\mathbf{s})$ and apply a moving average

filter to obtain the smoothed power $\bar{P}(\mathbf{s})$. The noise floor is estimated as:

$$N_{\text{floor}} = \min_{\mathbf{s} \in \mathcal{D}_{\text{train}}} \left\{ \min_{f} \bar{P}_f(\mathbf{s}) \right\}, \tag{15}$$

where $\mathcal{D}_{\text{train}}$ is the training dataset and $\bar{P}_f(\mathbf{s})$ denotes the smoothed power at frequency bin f.

During inference, we normalize the input I/Q samples by the estimated noise level to maintain consistent SNR characteristics. For an input sample $\mathbf{x}_{\text{in}} \in \mathbb{R}^{2 \times n_{iq}}$, the normalized input is computed as:

$$\mathbf{x}_{\text{norm}} = \frac{\mathbf{x}_{\text{in}}}{\sqrt{\rho \cdot N_{\text{floor}}}}, \tag{16}$$

where ρ is a correlation factor that accounts for the effective noise bandwidth of the receiver. This normalization ensures that the model receives inputs with similar statistical properties to those encountered during training, regardless of absolute power levels or noise conditions in the deployment environment, thereby enhancing generalization to unseen channels and SNR conditions.

4.3. Adaptive Signal Bandwidth and Channel Bandwidth Processing

As discussed in **RQ3**, spectrum sharing often requires the system to monitor a broader spectrum than their observable channel bandwidth B. Additionally, transmitted signals may dynamically adjust their signal bandwidth W based on available spectrum resources and throughput requirements. To address this challenge, we implement adaptive signal and channel bandwidth processing by scanning channels and aggregating the results.

Our key intuition is that hardware imperfections are intrinsic to the physical components of the device and are thus independent of the signal bandwidth. Thus, when a signal with a bandwidth larger than W is received, the `M2RF` divides it into smaller segments, each matching the W for which the model was trained. Similarly, when faced with a signal that spans a larger observable bandwidth $\tilde{B} > B$, `M2RF` divides the larger bandwidth into smaller, partially overlapping segments, each of size B, as illustrated in Figure 6. Each segment is processed individually by the DNN and the outputs are combined to form a final output that covers the entire bandwidth $\tilde{B}$. After processing, the predictions from these individual segments—whether divided by signal width or observable bandwidth—are aggregated to form a cohesive understanding of the entire wider signal $\tilde{W}$ or bandwidth $\tilde{B}$.

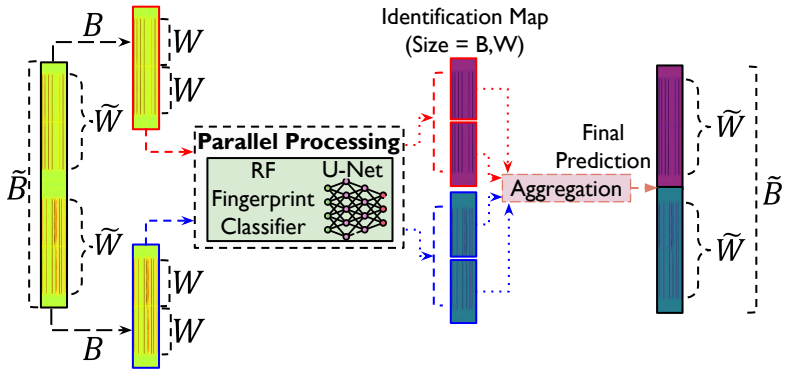


Figure 6: Adaptive wide-band and channel bandwidth processing pipeline: The input signal covering a larger bandwidth $\tilde{B}$ or wider channel width $\tilde{W}$ is divided into overlapping segments of size B or W, processed by the U-Net model. The outputs are then aggregated to produce the final prediction across the full signal or bandwidth.

Let $\mathcal{S} = \{S_1, S_2, \ldots, S_K\}$ denote the set of K overlapping segments, where each segment S_k spans frequencies $[f_k^{\text{start}}, f_k^{\text{end}}]$ with overlap ratio Ω. The model processes each segment independently to obtain individual predictions $\hat{\mathbf{y}}_k \in \mathbb{R}^{C \times n_{iq}}$ for segment S_k. To combine predictions from overlapping segments, we employ weighted averaging in the overlap regions. For each frequency bin f covered by multiple segments, the final prediction is computed as:

$$\hat{y}_f = \frac{1}{|\mathcal{K}_f|} \sum_{k \in \mathcal{K}_f} \hat{y}_{k,f}, \tag{17}$$

where $\mathcal{K}_f = \{k \mid f \in S_k\}$ is the set of segments that contain frequency bin f, and $\hat{y}_{k,f}$ is the prediction for bin f from segment k. For non-overlapping regions, $|\mathcal{K}_f| = 1$ and the prediction is used directly. The aggregation strategy ensures smooth transitions between segments and reduces boundary artifacts, enabling accurate fingerprinting across bandwidths without model retraining. This capability highlights the scalability and portability of our approach, making it highly versatile for deployment across various RF environments.

4.4. Anomaly Detection

Beyond identifying legitimate devices, we are also interested in detecting interference and malicious traffic in the spectrum. To address **RQ4**, we in-

troduce a novel anomaly detection approach by leveraging the uncertainty in the DNN output. During training, only legitimate signals are used, which results in confident predictions for authorized devices. Conversely, a malicious signal or interference not seen during training will be less confident, thus enabling `M2RF` to detect adversaries by evaluating the randomness of the spectrum map.

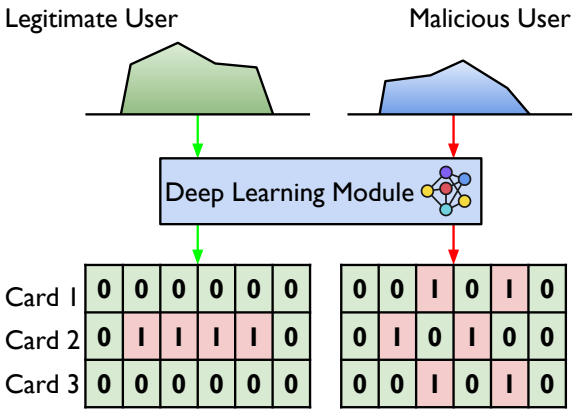


Figure 7: An example of DNN output for a legitimate user vs malicious user. The inference map presents high randomness while legitimate user output has more consistency.

To quantify adversarial activity, we apply total variation, which evaluates the consistency of the DNN output across the frequency domain. Higher total variation indicates a higher likelihood of a malicious signal. For a 1D vector x, the total variation is defined as:

$$TV(x) = \sum_{i=0}^{N-1} |x_{i+1} - x_i|, \tag{18}$$

where x_i is the i-th element in vector x while N denotes the dimension of the input. While total variation is first introduced for denoising [33, 34], the element-wise distance $|x_{i+1} - x_i|$ evaluates the consistency in DL module output in our case, making it a good metric to detect the malicious user who constantly has a noisier output than legitimate user. For example, the total variation of the legitimate user output in Figure 7 is 2 while the malicious user output has a total variation of 12. By increasing the resolution in frequency domain (e.g., in our experiment we use 4096 as the input and output size),

the difference of total variation between legitimate and malicious users will increase significantly.

By comparing the TV values for legitimate and malicious signals, we set a detection threshold λ_m:

$$TV(x) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_m, \tag{19}$$

where H_0 denotes the hypothesis that signal x is not an adversary and H_1 denotes the alternative hypothesis that x is considered as adversary.

5. Experimental Setup

Our data collection setup captures radio fingerprints under two distinct scenarios—wireless and wired—using sophisticated hardware to ensure accuracy and reliability. As shown in Figure 8, the configuration includes a Multiple-Input Multiple-Output (MIMO) system for the simultaneous transmission of 15 PCI-E wireless LAN cards, all identical in model and version (802.11ac/ax). This choice of identical devices creates a challenging test scenario, generating highly correlated signals to rigorously test `M2RF`'s ability to distinguish between identical transmitters. The ASUS RT-AX86U router is used as the primary receiver, and I/Q data is captured via an USRP X310 and USRP B200mini, each equipped with VERT2450 and L-com antennas.

We created two testbeds to evaluate radio fingerprinting performance under both wireless and wired data collection methods. The wireless setup, where radio transmissions from the PCI-E cards are captured by the USRP radios through antennas, simulates a realistic, uncontrolled environment, typical of actual deployments. This setup provides insight into radio fingerprint behavior in dynamic conditions affected by interference, multi-path effects, and environmental variability. Conversely, the wired setup provides a controlled baseline with a higher signal-to-noise ratio (SNR) of around 20-25 dB, compared to 15-20 dB in the wireless setup. This comparison highlights the robustness of our radio fingerprinting approach across varying SNR levels and operating conditions.

Data were collected over three days within a laboratory setting to capture a wide range of signal conditions. This approach accounted for environmental factors like temperature fluctuations and electromagnetic interference. Data was collected across two specific frequency bands – 5.5 GHz (channel 100) and 5.6 GHz (channel 120) – within a 50 MHz observable bandwidth, with each

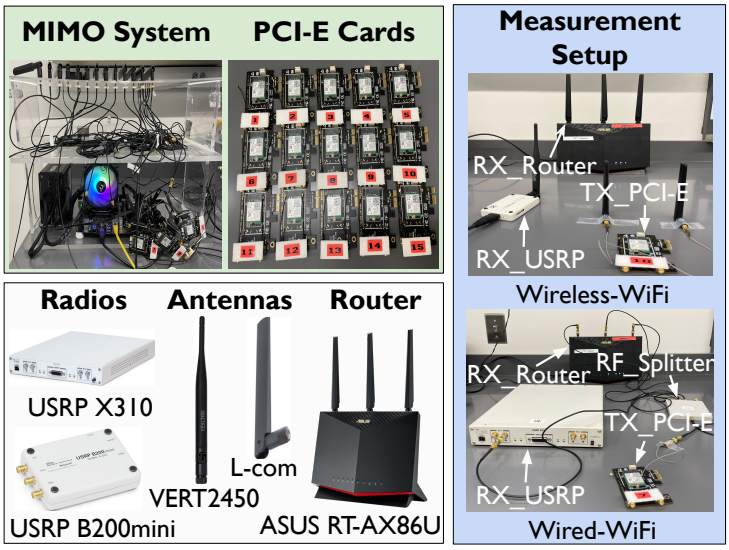


Figure 8: Overview of the data collection testbeds; (Top Left): MIMO system with PCI-E cards for simultaneous transmission; (Top Right): Individual PCI-E cards; (Bottom Left): Radios, antennas, and router setup; (Right): Measurement setup for wireless and wired data collection.

Wi-Fi card transmission occupying a 20 MHz bandwidth. To prevent signal overlap and ensure distinct RF fingerprints, each PCI-E card's transmission was captured separately, achieving the precision necessary to differentiate between devices with nearly identical hardware profiles.

5.1. Hardware Characteristics

DNN training was conducted on a system featuring 4 NVIDIA A100 80GB PCIe GPUs, 512 GB RAM, and dual Intel Xeon Silver 4410Y processors, ensuring efficient processing for large-scale DNN tasks. After training, the DNNs were run on the Jetson Orin Nano module, powered by a 6-core ARM Cortex-A78AE 64-bit CPU. The system is equipped with 8 GB of 128-bit LPDDR5 memory and a 1024-core NVIDIA Ampere architecture GPU with 32 Tensor Cores, capable of delivering up to 40 TOPS.

5.2. Experimental Dataset and Training

Experiments were conducted in both wireless and wired modes, primarily focusing on signals within a 50 MHz bandwidth. The primary dataset includes signals from 15 authorized devices, comprised 1.25 million samples,

with 80% allocated for training and 20% for testing across three distinct scenarios: non-overlapping, overlapping, and partially observed signals. For training, we used the Adam optimizer with a StepLR learning rate scheduler, starting at 0.001 and reducing by a factor of 0.1 every 30 epochs. The DNN was trained for 100 epochs with a batch size of 1024, with early stopping applied after 30 epochs of no improvement to prevent overfitting.

To rigorously test the `M2RF` resilience against potential attacks, we prepared additional testing datasets for specific attack scenarios:

• **Adversary:** For informed adversary, we collected Wi-Fi data from unauthorized devices identical in hardware to the authorized devices but excluded these samples from training. For uninformed adversary, we collected data using Wi-Fi protocol but from devices having different hardware characteristics than those used during training;

• **Interference**: We collected data with devices using Bluetooth Low Energy (BLE) with bandwidth 1 MHz, Long Term Evolution (LTE) with bandwidth 10 MHz, Zigbee with bandwidth 2 MHz as interference, as well as with additional Wi-Fi devices in the 2.4 GHz band.

6. Performance Evaluation

6.1. Performance Across Input Sizes and Scenarios

6.1.1. Wireless Mode

We started with wireless data, and compared F1-score for different input sizes (1024, 2048, 4096) between the three scenarios defined in Figure 4 to analyze how well `M2RF` performs. As depicted in Figure 9, the F1-scores for all scenarios significantly improve as we increase the input size. For scenario 1, the F1 score approaches as high as 86.65% with 1024 input size but increases nearly perfectly to a value of 94.99% to even an input size of 4096, showing that the trained model without overlap predicts very well between each signal class. In scenario 2, a similar trend is observed, where F1-scores increases from 82.11% at 1024 to reach the value of around 90.45% at 4096. The significant improvement is in scenario 3, where the F1-score improves from 69.71% at 1024 to 77.06% at 4096, demonstrating the `M2RF` robustness against overlapping signals. On the other hand, larger input sizes imply higher processing latency.

Similarly, the IoU metric is a good indicator for segmentation accuracy. As expected, the IoU reaches 90.54% for an input size of 4096 in scenario 1, while scenario 2 slightly drops to 82.75%, due to the complexity introduced

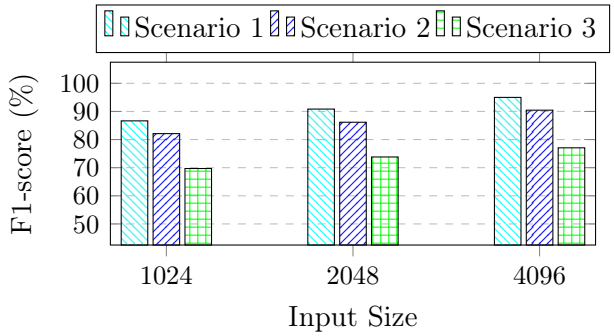


Figure 9: F1-score comparison for various input sizes across scenarios using wireless data.

with random signal placements. In scenario 3 IoU decreases to 63.39% as signals now overlap and become harder to localize within the spectrum correctly. However, these results highlight the adaptability and robustness of the `M2RF` in varying RF environments. Detailed metrics for each input size across scenarios are provided in Appendix B.

6.1.2. Wired Mode

We further evaluate the robustness of `M2RF` in a wired setup, using the same input size of 4096 to ensure consistency with the wireless mode. Figure 10 shows that the wired setup achieved notably higher scores, which is expected given the absence of interference.

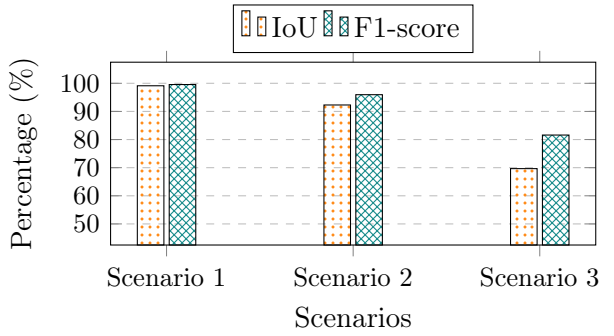
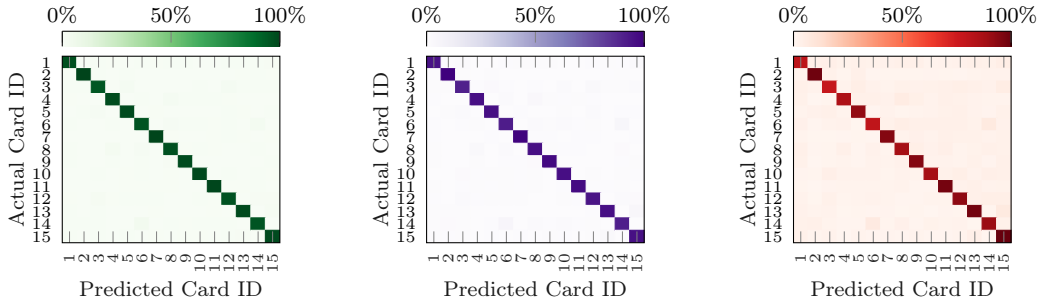


Figure 10: IoU and F1-score for different scenarios with an input size of 4096 using wired data.

In scenario 1 where the signals do not overlap, `M2RF` obtained a IoU of 99.12% and an F1-score of 99.56%, implying almost no misclassification. This



(a) Scenario 1. (b) Scenario 2. (c) Scenario 3.

Figure 11: Confusion matrices for three scenarios in wireless mode.

shows the ability of M2RF to identify independent, non-overlapping signal sources. In a moderately challenging conditions (scenario 2) with partially observed signals, we obtain an IoU of 92.29% and F1-score of 95.95%, demonstrating the resilience of M2RF. The wired setup resulted in 69.66% IoU and 81.6% F1-score, even under scenario 3 – the worst-case conditions with both fully and partially overlapping signals.

6.1.3. Confusion Matrices Analysis

Deeper insights into the M2RF's classification accuracy were gained by analyzing the confusion matrices for each scenario, using wireless data with an input size of 4096. These matrices offer a detailed breakdown of true and false identifications across all 15 devices. Further analysis of confusion matrices for the wired mode is provided in Appendix C.

In scenario 1 (Figure 11a), the confusion matrix of non-overlapping signals provides that M2RF achieves a high classification accuracy, and most devices get an accuracy greater than 90%. For instance, card 1 achieves 94.1%, and card 2 reaches 97.2%, reflecting the system's effectiveness in distinguishing devices when signals are isolated and clearly separated.

Scenario 2, shown in Figure 11b, introduces an increased level of complexity with randomly placed signals within the bandwidth. This setup represents a more realistic real-world scenario, where signals can be partially or fully present into the observable bandwidth, generating signal glitches on duration and shifting. Although M2RF achieves high performance for multiple devices (e.g., card 2 at 94.8%) there is a small drop in accuracy for others, such as card 3 (85.2%). The variation in this performance illustrates the M2RF's ro-

bustness to non-ideal conditions where the signals are not perfectly isolated, thus making the classification problem harder and M2RF learning to generalize across unpredictable scenarios.

Scenario 3 is shown in Figure 11c which is full or partial overlapping of signals. At this point, accuracy rates further decrease, where card 1 gets only 68.1% and card 3 gets only 65.9%. Such a decrease is intuitive as if the signals overlap in the same frequency range, finding the characteristics of signal would be difficult. These obstacles notwithstanding, the M2RF still performs well in accurately classifying most devices and shows it could operate in congested RF environments where overlaps are frequently observed.

The distinguishing of devices in scenario 3 by the M2RF also suggests its ability to test against jamming attack and even detect it. In those cases, we may actually have intentional disruptions within the spectrum represented by overlapping signals. The M2RF's strong localization of these overlapping signals would enable the development of a target countermeasure where the M2RF can detect and suppress jamming in real-time with minimal impact to other neighboring communications. This capability adds a critical dimension to the M2RF 's utility in dynamic, interference-prone environments, where prompt and accurate signal identification is essential for maintaining security and integrity.

6.2. Generalization

This section evaluates M2RF's ability to generalize beyond training conditions across three dimensions: bandwidth and channel width variations, temporal channel variations through cross-day validation, and carrier frequency offset (CFO) robustness. These experiments address whether M2RF learns intrinsic hardware imperfections that remain consistent across diverse deployment conditions, rather than memorizing training-specific patterns.

6.2.1. Bandwidth and Channel Width Generalization

The DNN was initially trained with a 50 MHz bandwidth and 20 MHz-wide signals and then tested under varying bandwidths to assess whether the hardware imperfections it leverages remain consistent. These evaluations provide insights into M2RF's adaptability to changing spectral conditions. Key cases are illustrated in Figures 12 and 13, using data collected on a different day with unseen signal conditions, which typically vary due to environmental factors.

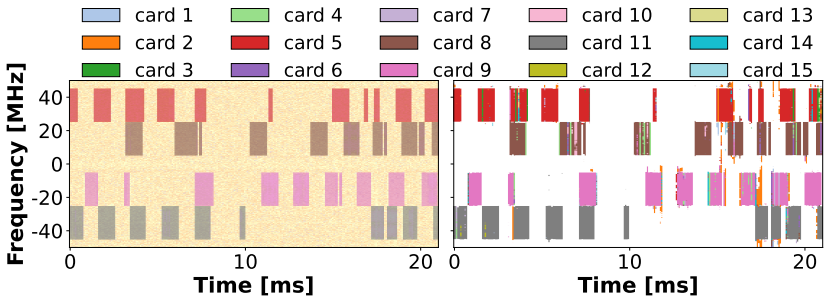


Figure 12: 100 MHz bandwidth, 4 signals 20 MHz wide. (Left) Ground Truth, (Right) Model Prediction.

In the first test, the M2RF system was evaluated with a 100 MHz bandwidth containing four devices, each occupying 20 MHz. As shown in Figure 12, M2RF successfully distinguishes between these signals, achieving an F1-score of 90.22% and an IoU of 82.32%. Although performance shows a slight reduction from the 50 MHz baseline, this decrease can be attributed to the increased spectral complexity and channel noise. Nonetheless, M2RF demonstrates high generalization, maintaining effective detection and localization of multiple signals within the broader bandwidth without requiring retraining.

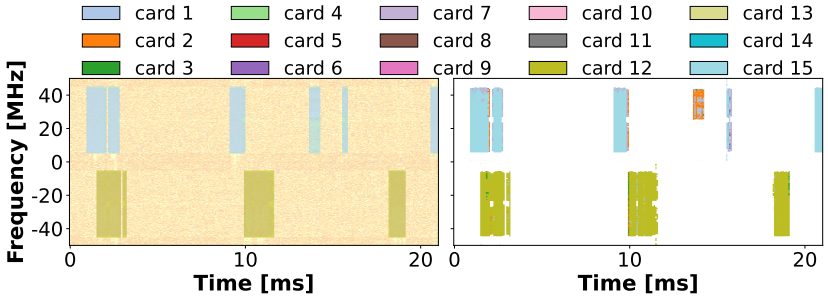


Figure 13: 100 MHz Bandwidth, 2 signals 40 MHz wide. (Left) Ground Truth, (Right) Model Prediction.

In a second test, M2RF was evaluated with two signals that were 40 MHz wide in a 100 MHz bandwidth. M2RF achieved F1-score of 87% and an IoU of 77.29%. The broader signal, along with its associated noise, added complexity to this configuration. Although the accuracy drops a little, M2RF correctly identifies and localizes every signal, as shown in Figure 13. The above result

illustrates the robustness of M2RF to changes in channel width and signal structure. Most importantly, the system performed a correct classification without retraining, making our M2RF approach even more robust.

This establishes that, although we see some degradation in performance at higher bandwidth demand cases, the fundamental approach remains resilient. The signal-recognition capabilities of the M2RF, including its robustness to changing channel width and signal overlap, rely on hardware-induced imperfections that vary consistently across devices. These frequency-specific distortions as a result of hardware imperfection persist across different bandwidths allowing for reliable classification and consequently localization of the signal, yielding a robust fingerprinting approach that can work well with both diverse and dynamic RF environments. Such flexibility is essential for practical applications where spectral conditions will change, and the system has to function properly across a wide range of bandwidths and interference levels.

6.2.2. Channel Independence Validation

We conducted cross-day validation experiments where the model must authenticate devices under channel conditions not encountered during training, to assess whether M2RF learns hardware fingerprints rather than channel characteristics. We collected data across three days with varying propagation conditions, multipath characteristics, and ambient noise levels (Section 5). We design three training configurations: (1) *train on day 1+2, test on day 3 (unseen)*; (2) *train on day 1 only, test on days 2-3 (unseen)*; and (3) *train on all days* as a performance upper bound. All experiments use scenario 3 (overlapping signals), representing the most challenging case.

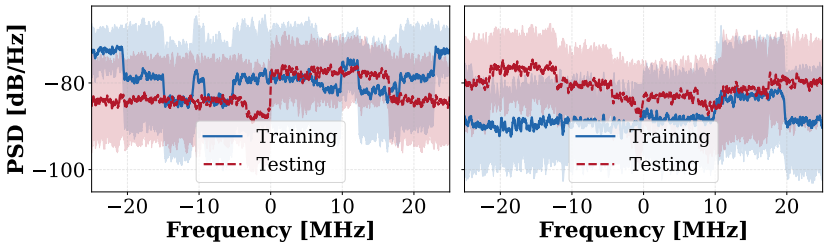


Figure 14: Power spectral density (PSD) comparison between training and testing sets for card 5 (left) and card 11 (right). Shaded regions indicate sample variability. The distinct power levels and spectral characteristics confirm different channel conditions between datasets.

Figure 14 visualizes the channel diversity between training (day 1) and testing conditions (day 3) for two representative devices. The distinct power levels, spectral shapes, and variance confirm that training and testing data experience different channel realizations. Table 1 presents the cross-day generalization results. When training on day 1+2 and testing on the unseen day 3 data, `M2RF` achieves an F1-score of 62.16% and IoU of 47.16%. Comparing this to the all-days baseline performance on Day 3 (76.98% F1-score), the model retains 80.7% of its performance despite never encountering day 3's channel conditions during training. This substantial performance retention on completely unseen channel realizations provides strong evidence that `M2RF` learns intrinsic hardware imperfections rather than channel-specific patterns. If the model were memorizing channel characteristics, we would expect severe performance degradation when tested on day 3, as the channel conditions would be different from training.

Table 1: Cross-day generalization performance demonstrating channel-independent hardware fingerprint learning (scenario 3, 15 devices).

Training Configuration	**Test Set**	**F1 (%)**	**IoU (%)**	**Precision (%)**	**Recall (%)**
Day 1+2	Day 1 (Seen)	75.77	61.67	83.06	69.78
	Day 2 (Seen)	75.05	60.81	82.26	69.18
	Day 3 (Unseen)	**62.16**	**47.16**	**68.50**	**57.74**
Day 1 Only	Day 1 (Seen)	75.33	61.11	82.60	69.38
	Day 2 (Unseen)	**64.02**	**48.41**	**70.34**	**59.13**
	Day 3 (Unseen)	**57.84**	**43.03**	**63.53**	**53.97**
All Days (Day 1+2+3)	Day 1 (Seen)	76.85	63.08	84.32	70.80
	Day 2 (Seen)	77.35	63.78	84.82	71.33
	Day 3 (Seen)	76.98	63.34	84.42	70.94
	Different Days (Unseen)	**73.05**	**58.68**	**80.42**	**67.35**

The day 1 only experiments further validate this finding. When trained on a single day and tested on days 2-3 (both unseen), the model achieves F1-scores of 64.02% and 57.84% respectively, demonstrating that even limited exposure to channel variations during training enables some degree of generalization. The performance gap between single-day and multi-day training highlights the value of diverse training conditions, while the non-zero performance on unseen days confirms hardware-based learning. Most critically, when testing on entirely different collection days with significantly varied environmental conditions (different temperature, humidity, and electromagnetic environment), the model trained on all days achieves 73.05% F1-score and 58.68% IoU. This represents 94.9% performance retention compared to

the day 3 baseline (76.98% F1-score), despite the substantial differences in collection conditions. This result definitively demonstrates that M2RF generalizes across diverse channel realizations by learning device-specific hardware characteristics that remain invariant across varying propagation environments.

6.2.3. Robustness to Carrier Frequency Offset

Beyond bandwidth and temporal generalization, we validate M2RF's robustness to carrier frequency offset (CFO). CFO arises from two primary sources: **(1)** frequency mismatch between transmitter and receiver local oscillators due to manufacturing tolerances, and **(2)** Doppler shifts in mobile environments [35, 36, 37]. For a baseband signal $s(t)$ with carrier frequency f_c, CFO Δf manifests as a frequency-domain circular shift in the received signal spectrum. Mathematically, a CFO of Δf introduces a time-domain phase rotation:

$$r(t) = s(t)e^{j2\pi\Delta f t}, \tag{20}$$

where $r(t)$ is the received signal affected by CFO. In the frequency domain after FFT, this corresponds to a circular shift of the spectrum:

$$R[k] = S[(k - k_\Delta) \bmod N], \tag{21}$$

where $k_\Delta = \lfloor N \cdot \Delta f / B \rfloor$ is the shift in frequency bins, N is the FFT size, and B is the bandwidth. For IEEE 802.11 WLAN devices operating at 5.5 GHz with oscillator tolerance of ±20 ppm [38, 39, 40], typical CFO ranges from −220 kHz to +220 kHz.

To evaluate M2RF's robustness to CFO, we apply controlled frequency shifts ranging from −500 kHz to +500 kHz to our scenario 3 test dataset. Frequency shifts are implemented as circular shifts in the frequency domain, where a shift of Δf MHz corresponds to $k_\Delta = \lfloor 4096 \cdot \Delta f / 50 \rfloor$ bins. We test 41 uniformly spaced shift values and measure performance using 10 000 test samples per shift point. Figure 15 presents performance metrics across the tested CFO range. Across the entire ±500 kHz range, F1-score varies from 73.42% to 77.41%, representing only 3.99% maximum degradation despite frequency shifts exceeding IEEE 802.11 specifications by a factor of 2.3. This remarkable stability demonstrates that M2RF is robust to CFO perturbations not explicitly modeled during training.

The periodic performance variations exhibit a distinctive triangular pattern with peaks occurring at approximately 200 kHz intervals. This behavior

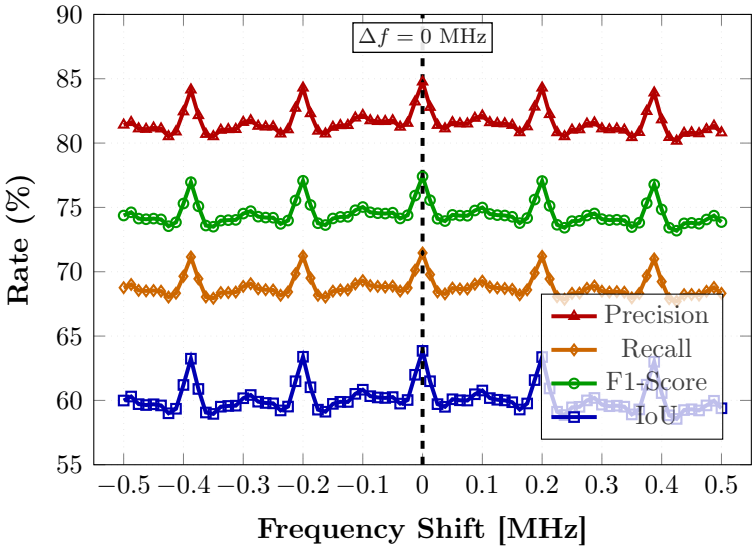


Figure 15: Performance metrics versus CFO. `M2RF` maintains stable performance across ±500 kHz frequency shifts, with only 3.99% F1-score degradation. The periodic performance variations correlate with signal alignment relative to the 20 MHz Wi-Fi channel boundaries, while the overall robustness validates that `M2RF` generalizes to frequency errors beyond those encountered during training.

is not an artifact but rather scientific evidence that our CFO simulation correctly interacts with the inherent frequency structure of IEEE 802.11 signals. In OFDM-based systems such as Wi-Fi, the transmitted signal consists of multiple orthogonal subcarriers with a characteristic spacing determined by the system bandwidth and FFT size. When the applied CFO aligns with integer multiples or harmonics of this subcarrier structure, the frequency-shifted spectrum maintains better alignment with the original signal's periodic features, resulting in performance peaks. Conversely, when the CFO falls between these alignment points, the spectral features experience maximum distortion relative to the training distribution, causing the observed performance valleys. Nevertheless, even at worst-case shifts, the `M2RF` retains over 95% of its baseline performance, validating that our scenario generation approach, despite using controlled signal captures, produces training data that generalizes effectively to frequency perturbations encountered in real deployments.

6.3. Scalability

In addition, we evaluate the scalability of M2RF as the number of devices increases. We conducted a systematic analysis with 5, 10, and 15 transmitters under scenario 3 (overlapping signals), which represents the most challenging authentication environment. The M2RF architecture maintains a fixed parameter count of 11.52 M parameters with a model memory footprint of 44.01 MB across all configurations. The U-Net architecture uses fixed hyperparameters $\alpha = 1$ (channel width multiplier) and $\beta = 5$ (network depth with 5 encoder-decoder levels), while only the output layer adapts to the number of device classes. This design ensures consistent model capacity and computational requirements independent of the number of monitored devices. Table 2 presents the scalability results measured on the Jetson Orin Nano edge device. The analysis reveals that M2RF maintains consistent computational efficiency across device counts, with inference time remaining nearly constant at approximately 15.9 ms and GPU memory usage stable at around 126 MB, demonstrating excellent computational scalability.

Table 2: Performance and computational metrics for different numbers of devices in scenario 3.

Devices	Precision (%)	Recall (%)	F1 (%)	IoU (%)	Inference (ms)	Memory (MB)
5	91.85	78.86	84.71	74.29	15.87	126.06
10	86.00	72.68	78.68	65.68	15.92	126.36
15	84.56	70.97	77.06	63.39	15.95	126.69

As expected, classification performance exhibits graceful degradation as the number of devices increases from 5 to 15, which is inherent to the increased complexity of distinguishing among more identical hardware fingerprints. Specifically, the F1-score decreases from 84.71% for 5 devices to 77.06% for 15 devices, a reduction of 7.65% across a three times increase in device count. Despite this degradation, M2RF maintains robust performance even with 15 devices, achieving 77.06% F1-score and 63.39% IoU in the challenging overlapping scenario. The minimal impact on computational resources (inference time increases by only 0.08 ms and memory by 0.63 MB when scaling from 5 to 15 devices) confirms that M2RF can efficiently scale to authenticate multiple devices simultaneously without requiring proportional increases in computational budget. This computational stability, combined with acceptable performance degradation, demonstrates the practical viabil-

ity of M2RF for real-world spectrum sharing deployments where the number of authorized devices may vary dynamically.

6.4. System Defense under Malicious Activity

We rigorously evaluated M2RF's defense capabilities against both *informed* and *uninformed adversaries* within Wi-Fi networks, as well as interference from other wireless technologies in congested spectrum environments.

In the informed adversary scenario, the attacker has full knowledge of the authentication method and ML model used, as well as access to identical hardware as the legitimate devices. On the other hand, the uninformed adversary represents a more realistic scenario in which the attacker uses potentially different hardware versions unknown to the system at training time. To evaluate the robustness of M2RF in crowded spectrum environments, we also performed inference detection with respect to signals from other non-Wi-Fi technologies like BLE, LTE and Zigbee. This multi-technology challenge tests the capability of M2RF to detect unauthorized transmissions in various signal types, reflecting real-world conditions in densely populated RF environments.

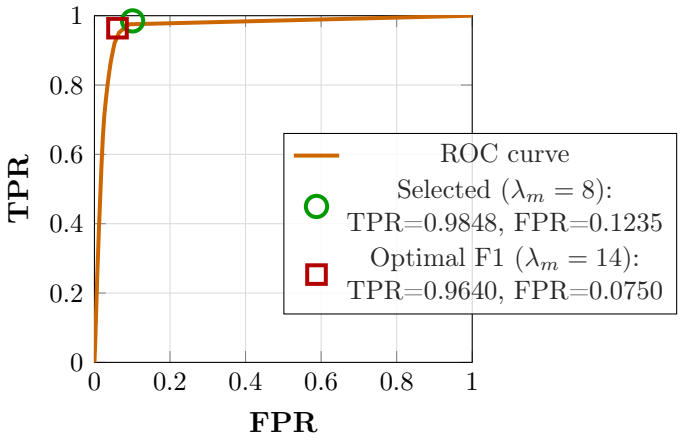


Figure 16: ROC curve for TV-based anomaly detection showing AUC of 0.9711. The optimal threshold ($\lambda_m = 14$) and selected threshold ($\lambda_m = 8$) are marked, demonstrating the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR).

The selection of detection threshold λ_m in Equation 19 directly impacts the trade-off between security and operational efficiency. To determine the

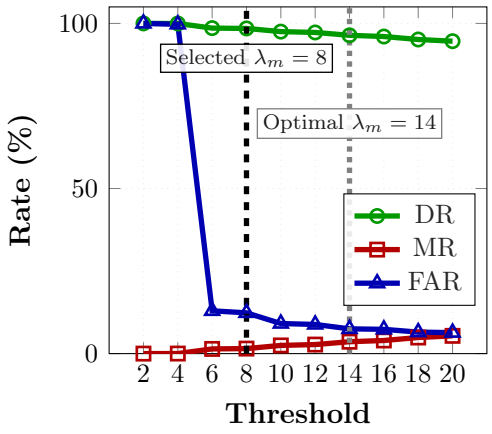


Figure 17: Threshold sensitivity analysis showing the impact of different TV thresholds on Detection Rate (DR), Miss Rate (MR), and False Alarm Rate (FAR). The selected threshold $\lambda_m = 8$ prioritizes high DR (98.48%) while maintaining acceptable FAR (12.35%).

optimal threshold, we conducted a comprehensive analysis using 100 000 test samples, evenly split between authorized and malicious transmissions. Figure 16 shows the ROC curve for our TV-based detection mechanism, achieving an AUC of 0.9711, which demonstrates excellent discriminative power. The statistical analysis reveals that TV distributions of authorized (mean = 8.75, std = 21.22) and malicious signals (mean = 101.91, std = 56.04) are significantly different (t-test: $p < 0.001$, Cohen's $d = 2.20$), confirming the effectiveness of TV as a discriminator.

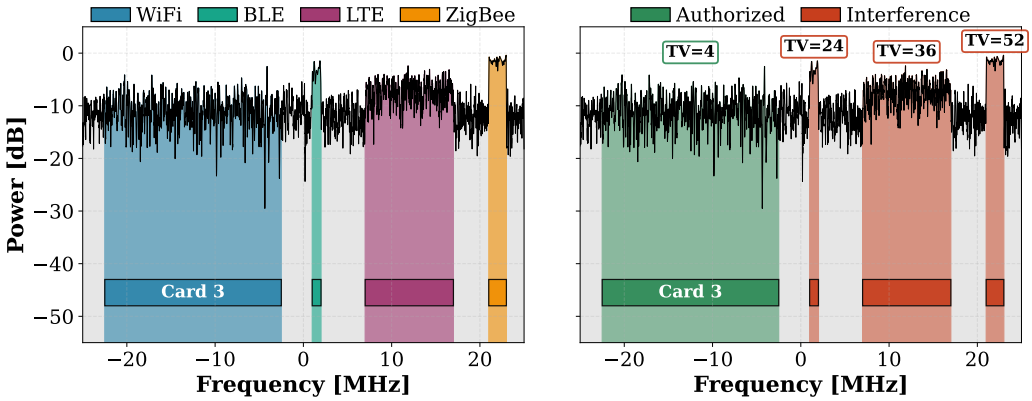
Figure 17 presents a sensitivity analysis across multiple threshold values. While the optimal F1-score of 94.57% occurs at $\lambda_m = 14$, we selected $\lambda_m = 8$ to prioritize security in spectrum sharing scenarios. This choice achieves a detection rate (DR) of 98.48% with a miss rate (MR) of only 1.52%, compared to 96.40% DR and 3.60% MR at $\lambda_m = 14$. The trade-off is a moderate increase in false alarm rate (FAR) from 7.50% to 12.35%, which is acceptable since false alarms can be resolved through secondary verification, whereas missed detections directly compromise spectrum security. This conservative approach ensures that unauthorized transmissions are detected with high probability, which is critical for protecting incumbent users in dynamic spectrum access systems.

As shown in Table 3, `M2RF` consistently performed well in distinguishing authorized from malicious signals using a TV-based threshold of $TV = 8$.

Table 3: M2RF performance under attacks in Wi-Fi networks.

Attack Type	User	P (%)	R (%)	F1 (%)	MR (%)	FAR (%)
Uninformed	**Authorized**	98.29	87.66	92.67	12.35	1.52
	Malicious	88.90	98.48	93.44	1.52	12.35
Informed	**Authorized**	95.71	87.51	91.43	12.49	3.91
	Malicious	88.51	96.09	92.14	3.91	12.49
					Overall Accuracy:	**92.44%**

In uninformed attack scenario, the system reached an F1-score of 92.67% for authorized devices and 93.44% for malicious devices, with a mean MR of 2.72% across both attack types, demonstrating reliable detection of unauthorized signals even when attackers use different hardware. For informed attack, where the attacker's hardware matches that of the authorized devices, M2RF maintained a high F1-score of 91.43% for authorized devices and 92.14% for malicious devices, showing robustness against highly sophisticated adversaries, without requiring retraining on specific attack data.



(a) Input spectrum. (b) Detection results.

Figure 18: Cross-protocol interference detection demonstration. (a) Composite spectrum showing an authorized Wi-Fi transmitter (card 3) alongside interference sources from heterogeneous wireless technologies: BLE, LTE, and ZigBee. (b) M2RF detection results using the TV metric.

Figure 18(a) shows a composite spectrum containing an authorized Wi-Fi transmitter (card 3) alongside interference from BLE, LTE, and ZigBee tech-

nologies, each exhibiting its characteristic spectral signature. Figure 18(b) presents the corresponding detection results based on the TV metric. The authorized Wi-Fi device, which was included in the training set, produces stable predictions resulting in a low TV value (TV=4). In contrast, the interference signals from other technologies, never seen during training, generate unstable predictions with significantly elevated TV values (TV=24, 36, and 52 for BLE, LTE, and ZigBee, respectively). This clear separation between low TV (authorized) and high TV (interference) enables reliable detection of foreign signals regardless of their underlying protocol.

Table 4 quantifies M2RF's performance in this multi-technology evaluation, where BLE, LTE, and ZigBee signals are treated as interference to be detected and rejected. The system achieves an overall accuracy of 81.52% in distinguishing authorized Wi-Fi devices from heterogeneous interference sources. The authorized class achieves 87.47% recall, indicating that most legitimate transmissions are correctly identified, while the interference class attains 85.72% precision, demonstrating reliable rejection of foreign signals. These results confirm M2RF's effectiveness in congested spectrum environments such as the 2.4 GHz band, where authorized devices must coexist with various wireless technologies. Notably, this detection capability requires no prior knowledge or training on the specific interference signal types, relying instead on the observation that signals from unknown sources produce unstable fingerprint predictions.

Table 4: M2RF performance under interference from other technologies.

User	**P (%)**	**R (%)**	**F1 (%)**	**MR (%)**	**FAR (%)**
Authorized	78.23	87.47	82.59	12.53	24.45
Interference	85.72	75.55	80.31	24.45	12.53
				Overall Accuracy: 81.52%	

6.5. Comparison with Existing OOD Detection Methods

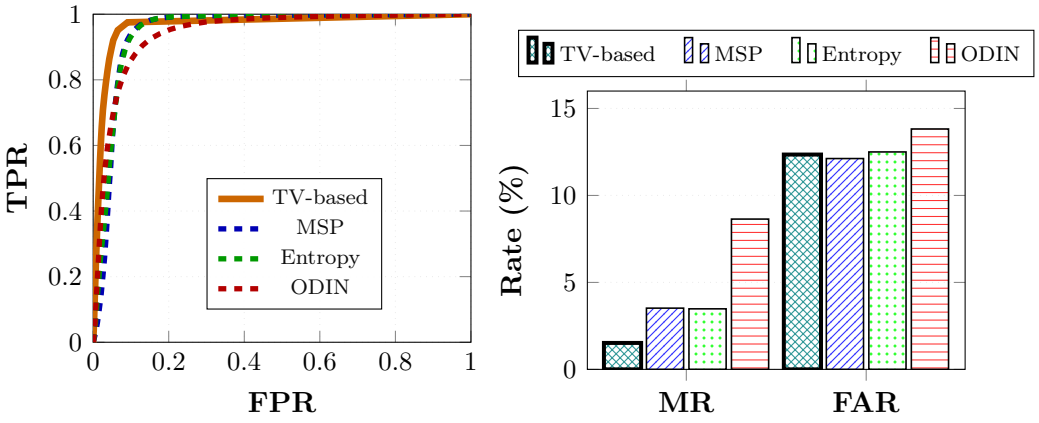
To validate the effectiveness of our TV-based anomaly detection approach, we conducted a systematic comparison with three established out-of-distribution (OOD) detection methods: Maximum Softmax Probability (MSP) [41], Entropy-based detection [42], and ODIN (Out-of-Distribution detector for Neural networks) [43]. These methods represent the state-of-the-

art in uncertainty estimation for DL models and have been widely adopted for OOD detection in computer vision and other domains.

For fair comparison, we evaluated all methods on the same test dataset of uninformed attack, using our pre-trained U-Net model. While our TV-based method uses a threshold of $\lambda_m = 8$ as justified in Section 6.4, we determined optimal thresholds for baseline methods using Youden's J statistic [44], which maximizes the difference between TPR and FPR:

$$J = \max_{\theta} \left(\text{TPR}(\theta) - \text{FPR}(\theta) \right), \tag{22}$$

where θ represents the detection threshold. This approach ensures that each baseline method operates at its best possible operating point, providing a conservative and fair comparison.



(a) ROC curves comparison.

(b) MR and FAR comparison.

Figure 19: Comparison of anomaly detection methods at optimal thresholds: (a) ROC curves showing our TV-based approach achieves the highest AUC of 0.9711; (b) Key metrics at optimal thresholds, demonstrating that TV-based detection achieves the lowest MR (1.52%) while maintaining acceptable FAR.

As shown in Figure 19a, our TV-based approach achieves the highest AUC of 0.9711, outperforming MSP (AUC=0.9486), Entropy (AUC=0.9520), and ODIN (AUC=0.9430). The superior performance stems from TV-based method's object-level aggregation strategy. While pixel-level uncertainty methods (MSP, Entropy, ODIN) evaluate confidence at individual frequency

bins, our approach aggregates predictions across the entire spectrum to evaluate signal-level consistency. This design choice proves particularly effective in RF environments where even authorized signals exhibit high pixel-level uncertainty due to inherent noise and channel variability, which causes pixel-level methods to generate excessive false alarms when relying solely on local confidence scores.

Table 5 presents detailed performance metrics at optimal thresholds for each method. As illustrated in Figure 19b, TV-based detection achieves the lowest MR of 1.52%, compared to MSP (3.52%), Entropy (3.48%), and ODIN (8.64%), while maintaining low FAR (12.35%). The substantial advantage over ODIN, reducing MR by 7.12%, demonstrates the robustness of object-level aggregation for spectrum security applications. These results confirm that TV-based detection achieves superior DR (98.48%) and F1-score (93.44%), making it particularly suitable for spectrum security applications where missing malicious transmissions poses greater risks than occasional false alarms.

Table 5: Performance comparison of OOD detection methods at optimal thresholds.

Method	θ	**AUC**	**DR (%)**	**MR (%)**	**FAR (%)**	**F1 (%)**	**Acc (%)**
TV-based (Ours)	8.0	**0.9711**	**98.48**	**1.52**	**12.35**	**93.44**	**93.08**
Max Softmax Prob	0.0116	0.9486	96.48	3.52	12.12	92.52	92.19
Entropy	0.0441	0.9520	96.52	3.48	12.50	92.37	92.02
ODIN	0.9366	0.9430	91.36	8.64	13.82	89.08	88.78

6.6. Ablation Study

We evaluated the impact of the non-local attention block and the spectrum stitching augmentation strategy to isolate and quantify the contributions of key design choices in `M2RF`. All experiments were performed using scenario 3 with 15 devices. This analysis provides empirical evidence for the effectiveness of each component and validates our architectural decisions.

6.6.1. Impact of Non-Local Block

The non-local block enables the network to capture long-range dependencies across the frequency domain through self-attention mechanisms, which is critical for distinguishing overlapping RF signals. Table 6 presents a comparative analysis of model performance with and without the non-local block. The integration of the non-local block yields substantial performance improvements: IoU increases from 39.75% to 63.39%, F1-score improves from

55.28% to 77.06%, and precision increases from 60.68% to 84.56%. These significant gains demonstrate that the non-local block effectively addresses the limitation of standard CNNs in capturing global context, enabling the model to leverage dependencies across the entire spectrum rather than relying solely on local receptive fields.

Table 6: Impact of non-local attention block on performance and computational cost (scenario 3, 15 devices).

Configuration	IoU (%)	P (%)	R (%)	F1 (%)	GFLOPs	Inference (ms)	Latency (ms)
Without Non-Local	39.75	60.68	51.23	55.28	12.21	13.86	15.55
With Non-Local	**63.39**	**84.56**	**70.97**	**77.06**	**12.49**	**15.95**	**17.44**

However, this performance enhancement comes with a computational trade-off. The non-local block introduces additional matrix multiplications over spatial and feature dimensions, increasing the computational cost from 12.21 GFLOPs to 12.49 GFLOPs (2.2% increase), increasing inference time from 13.86 ms to 15.95 ms (15.1% increase), and end-to-end latency from 15.55 ms to 17.44 ms (12.2% increase). While the non-local block could theoretically be applied to multiple encoding and decoding stages for further performance gains, such an approach would incur prohibitive computational costs that could render real-time operation infeasible. Therefore, we strategically position a single non-local block between the final decoding layer and the output layer, achieving an optimal balance between classification performance and computational efficiency suitable for real-time spectrum monitoring applications.

6.6.2. Impact of Spectrum Stitching Augmentation

The spectrum stitching augmentation strategy enables robust generalization to diverse signal placements and overlapping configurations. To evaluate its effectiveness, we compared two training approaches: (i) a baseline model trained on simple, centered signals following traditional RF fingerprinting methodology where each training sample contains a single signal placed at the center frequency [45, 35, 46, 47], and (ii) our proposed model trained with spectrum stitching augmentation that synthesizes realistic scenarios with multiple signals at varying positions and overlaps. Both models were evaluated on the same challenging scenario 3 test set containing overlapping signals.

Table 7 demonstrates the substantial advantage of spectrum stitching augmentation. The baseline model, trained only on centered single-signal

Table 7: Impact of spectrum stitching augmentation strategy (scenario 3, 15 devices).

Training Strategy	IoU (%)	Precision (%)	Recall (%)	F1 (%)	Δ F1 (%)
Baseline (No Stitching)	32.85	52.38	44.04	47.50	–
With Stitching (Ours)	**63.39**	**84.56**	**70.97**	**77.06**	**+29.56**

samples, achieves an F1-score of merely 47.50% and IoU of 32.85% when tested on overlapping signals, revealing severe overfitting to the simplified training distribution. In stark contrast, the model trained with spectrum stitching achieves F1-score of 77.06% and IoU of 63.39%. The precision improvement from 52.38% to 84.56% is particularly notable, indicating that stitching augmentation dramatically reduces false positives by training the model to recognize signal characteristics across diverse spectral positions rather than relying on fixed frequency locations.

Figure 20 provides qualitative evidence of the stitching strategy's effectiveness through spectrum visualization of two Wi-Fi cards transmitting 20 MHz signals within a 50 MHz observable bandwidth in scenario 1. The raw spectrogram and ground truth labels show two devices occupying distinct frequency bands. The baseline model output exhibits severe misclassifications throughout the spectrum, including erroneous predictions in empty spectral regions. This degradation occurs because the baseline model has only encountered centered signals during training and cannot generalize to signals at arbitrary frequency positions. Conversely, the proposed model trained with spectrum stitching accurately identifies and localizes both devices despite their non-centered placements, demonstrating robust generalization enabled by diverse augmentation. The stitching strategy exposes the model to signals across the entire bandwidth during training, forcing it to learn position-invariant hardware fingerprints rather than memorizing specific frequency locations. This ablation study confirms that spectrum stitching augmentation is essential for `M2RF`'s ability to perform reliable multi-device authentication in dynamic spectrum sharing environments where signal placements are unpredictable.

6.7. Energy-Latency Trade-off for Different Input Sizes on an Edge Device

In our experiments on energy-latency trade-offs, we measured the performance of different input sizes—1024, 2048, and 4096—on both GPU and CPU through Jetson Orin Nano device. The primary metrics were mean

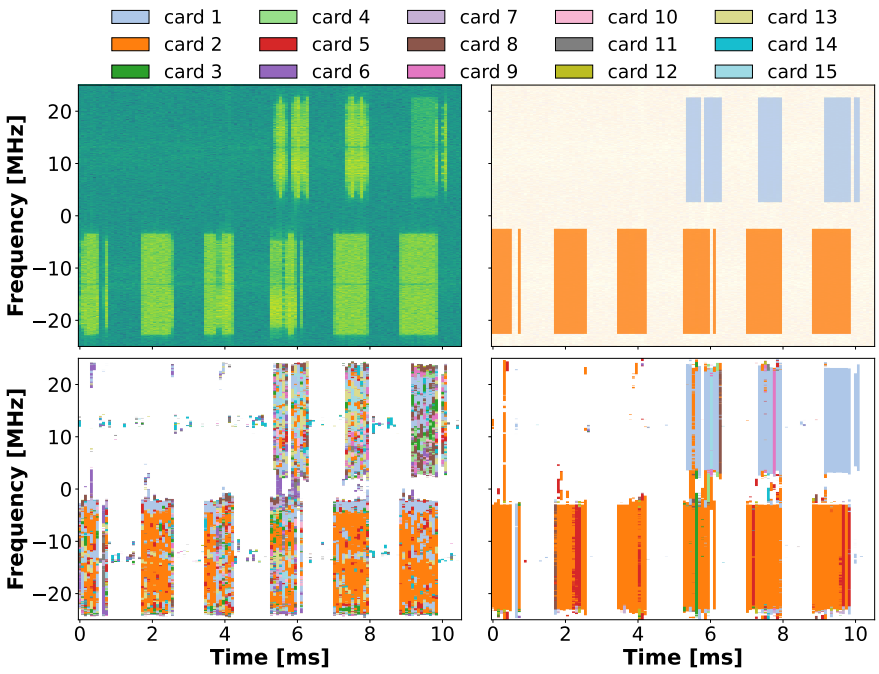


Figure 20: Ablation study visualization comparing model outputs trained with and without spectrum stitching augmentation on OTA data with two Wi-Fi devices (20 MHz signals, 50 MHz bandwidth, scenario 1): (Top Left) Raw spectrogram showing signal activity; (Top Right) Ground truth labels identifying two devices at distinct frequency bands; (Bottom Left) Baseline model output trained without stitching augmentation; (Bottom Right) Proposed model output trained with spectrum stitching augmentation.

inference time (MIT) and mean energy consumption (MEC), which are crucial for evaluating real-time `M2RF` efficiency. Despite the CPU having lower mean power consumption (MPC), the significantly longer MIT leads to much higher total energy consumption compared to the GPU. For example, at an input size of 4096, the inference time on the GPU was 15.95 ms compared to 391.62 ms on the CPU, demonstrating the GPU's substantial speed advantage for time-critical applications.

MEC was calculated by measuring the energy consumption of each individual inference and computing the mean:

$$\text{MEC (mJ)} = \frac{1}{N} \sum_{i=1}^{N} \left(\text{P}_i(\text{W}) \times \text{T}_i(\text{ms}) \right), \tag{23}$$

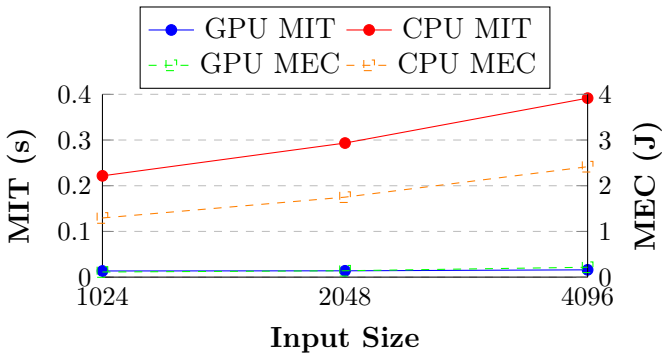


Figure 21: Energy-latency trade-off for different input sizes.

where P_i is the average power during inference i, estimated as $P_i \approx \frac{1}{2}\left(P_i^{\text{start}} + P_i^{\text{end}}\right)$, T_i is the inference time, and N is the number of inferences. This approach highlights that, although the CPU consumes less power at 6.17 watts (W) for an input size of 4096, its much longer inference time results in a significantly higher energy consumption of 2416.30 millijoules (mJ), compared to the GPU's energy consumption of 216.37 mJ for the same input size. Thus, the GPU's slightly higher power draw (13.56 W) is more than offset by its superior processing speed, making it far more energy-efficient in total energy consumption (11 times lower). Figure 21 further illustrates this trade-off, showing how both MEC and MIT increase with input size, but at a much steeper rate for the CPU than for the GPU.

7. Limitations and Discussion

Generalization Across Varying Environments. A key challenge for real-world deployment of radio fingerprinting systems is generalization across diverse environments. Our system was tested in controlled and semi-controlled environments, but real-world conditions involve more complexity, such as unexpected interference and mobility. While the model showed resilience to noise and overlapping signals, further testing in dynamic settings like urban or industrial environments is necessary to fully validate its robustness.

Impact of Dataset Size and Diversity. The scalability of radio fingerprinting depends heavily on the size and diversity of the training data. Although our dataset achieved high accuracy, it may not fully capture the variety of device characteristics encountered in real-world applications. Expanding the dataset to cover a wider range of devices, communication proto-

cols, and environmental conditions would strengthen the model's ability to generalize effectively. Future efforts should emphasize the collection of OTA data from diverse sources, which would improve the model's adaptability to real-world conditions.

Computational Constraints in Edge Devices. Our approach is designed as a centralized framework, with authentication managed by a server equipped for high computational loads. We tested the system on an edge device to evaluate performance under limited resources. In a distributed deployment, where nodes independently manage authentication or face constraints due to low-power edge devices, additional optimization is necessary. Other optimizations could be performed through model quantization, pruning or edge-cloud cooperation to maintain flexibility on central and distributed settings.

Adversarial Attacks. Although our system effectively detects informed and uninformed attacks, it may still be susceptible to advanced adversarial techniques. An attacker could exploit vulnerabilities through perturbations to penetrate the system [48, 49]. Future work should strengthen the defenses of the system without requiring retraining on specific adversarial examples and develop real-time detection mechanisms that adapt to the patterns of evolved attacks.

Manual Tuning and Real-World Deployment. Some components, particularly threshold for detecting malicious users, required manual tuning during experiments. This approach may be less effective in dynamic environments. Adaptive thresholding mechanism, based on real-time feedback, could enhance detection without manual intervention, making the system more practical for real-world use.

Frequency-Selective Hardware Imperfections. Our adaptive bandwidth processing assumes that hardware imperfections remain sufficiently consistent across frequency segments, which holds well within our 50–100 MHz operating range, as evidenced by successful generalization. However, for ultra-wideband scenarios spanning multiple GHz, frequency-selective effects such as frequency-dependent I/Q imbalance, power amplifier memory effects, and offset-dependent phase noise may require frequency-aware aggregation mechanisms or separate models trained for different frequency bands. Future work should investigate per-sub-band performance to quantify the impact of frequency-selective impairments and develop adaptive aggregation strategies that account for frequency-dependent variations in hardware fingerprints.

Deployment Factors and Failure Boundaries. While our scenario gen-

eration simulates signal overlaps, random placements, and noise, and our stress testing validates robustness to CFO, certain deployment factors remain outside the current training distribution. These include: (1) time-varying Doppler shifts in mobility scenarios where transmitters or receivers are moving [50], (2) severe multipath propagation causing significant interference in highly reflective or cluttered environments [51], (3) power amplifier saturation introducing spectral regrowth not present in normal operation [35], and (4) extreme temperature-induced oscillator drift beyond tested CFO ranges [52]. Based on our demonstrated generalization across bandwidths, collection days, and frequency shifts, we anticipate graceful degradation rather than catastrophic failure under moderate versions of these factors. However, the failure boundaries under extreme combinations remain to be characterized through field trials in challenging deployment environments.

8. Related Work

Spectrum Sensing. Early spectrum sensing research focused on binary classification to detect whether the frequency band is in use or not [53, 54]. These approaches fall short of supporting sophisticated spectrum policies where devices may require different levels of priority in spectrum access. Recent work proposed to jointly classify multiple signals in the spectrum based on wireless technologies [24] and modulation types [27]. However, [24, 27] fail to provide finer-grained identification of devices.

Radio Fingerprinting. Initial fingerprinting approaches, based on handcrafted features such as phase and amplitude noise to characterize device fingerprints, performed well in controlled environments but suffered from overfitting issues in more complex settings [55, 25, 26]. CNNs and RNNs solve this issue by performing feature extraction over raw I/Q data automatically [56, 57]. However, existing radio fingerprinting methods has a series of limitations making them unpractical in spectrum sharing scenarios. For example, the authors [21] had over 95% accuracy with 16 devices but classified only one signal at time while it was assumed the center frequencies were known. In contrast, our method based on U-Net dynamically classifies transmissions with different, even overlapping center frequencies and brings spectrum localization by segmenting in frequency bins.

Some works, including [58], have demonstrated that CNNs can effectively extract spatial features from RF signals, achieving 90% accuracy for 10 transmitters by transforming ADS-B signals in images [59]. Nonetheless, these ap-

proaches such as hybrid CNNs [37] depend on pre-processing which adds time complexity and is not real-time applicable and fails to deal with overlapping signals or different conditions. We propose an efficient segmentation framework that enhances CNNs for modeling both spatial and temporal inputs using non-local block, while minimizing computational overhead.

In addition, when multi-path fading and interference occur in data from various scenarios, traditional fingerprinting models often struggle to generalize effectively between controlled and real-world environments, leading to frequent model failure and the need for retraining [60, 36]. For instance, [28] reported an 82% accuracy drop when testing in real-world scenarios. With the use of data augmentation methods such as noise and channel augmentation, generalization has improved by reaching an accuracy level of 84.4% [61]. In contrast, we leverage OTA data collection and advanced augmentation, achieving over 90% accuracy in dynamic real-world environments without any retraining.

Adversarial Robustness. Although radio fingerprinting provides strong security, it is still vulnerable to adversarial attacks [49]. Studies have used adversarial training, incorporating clean and perturbed signals to improve robustness [62], like in [63], which improved robustness by 97% for 8 devices under rogue attacks. However, this approach requires extensive retraining with attack-specific data. In contrast, our system uses real-time anomaly detection via TV analysis, identifying signal deviations without prior attack knowledge, achieving 92.44% accuracy under attack conditions.

Energy Efficiency for IoT Applications. Deploying radio fingerprinting on energy-constrained IoT devices is challenging due to high computational demands [64]. Compression methods, like model pruning, reduce costs but often sacrifice accuracy [65, 66, 67]. For example, a study [68] observed a decrease in accuracy of 70.24%, consuming 2250 mJ per inference. Our U-Net-based approach, optimized for edge devices, balances efficiency and accuracy, achieving an inference time of 15.95 ms, an F1-score of 94.99%, and an energy consumption of only 216.37 mJ.

9. Conclusions

This paper demonstrates that radio fingerprinting in multi-device, multi-band environments requires effective management of overlapping signals, diverse bandwidths, and inherent hardware imperfections. We present a U-Net-based model for scalable and robust semantic segmentation of RF signals to

tackle these challenges. We obtain the following results: (i) the combined loss function significantly enhances performance, achieving an IoU of 90.54% and an F1-score of up to 94.99% for non-overlapping signals, and an F1-score of up to 77.06% in overlapping scenario; (ii) the approach maintains reliable detection across varied bandwidths, achieving an F1-score of 90.22% in scenarios with a 100 MHz bandwidth; (iii) the system effectively detects malicious Wi-Fi activities with an overall accuracy of 92.44% and a mean MR of 2.72%, without prior exposure to attack data, and successfully differentiates authorized Wi-Fi devices from non-Wi-Fi technologies with an accuracy of 81.52%, even in congested spectrum environments; and (iv) our system achieves a mean inference time of 15.95 ms and a mean energy consumption of 216.37 mJ on edge device. These results confirm that the `M2RF` has the potential to provide an efficient, scalable solution for IoT security applications with real-time constraints.

Acknowledgments

This research was made possible with the support of the Horizon Europe research and innovation programme of the European Union, under grant agreement number 101092912 (project MLSysOps). This work has also been funded in part by the National Science Foundation under grants ECCS-2229472 and CCF-2218845; in part by the Air Force Office of Scientific Research under contract number FA9550-23-1-0261 and in part by the Office of Naval Research under award number N00014-23-1-2221.

Appendix A. Loss Functions and Optimization

We have investigated a number of loss functions. Here, the loss functions are part of *local-level* and *region-level* metrics. Local-level loss measures accuracy at each frequency bin within the raw I/Q data, capturing fine-grained variations essential for distinguishing signal characteristics. Region-level loss, on the other hand, considers larger segments within the data, promoting consistency across continuous sections and enhancing detection of broader patterns, such as distinct signal regions or transmission boundaries. We summarize them as follows:

Dice Loss (DiL): DiL [69] is a region-based metric used to assess the similarity between predicted labels and ground truth, which is derived from the

Dice coefficient, a widely used measure of similarity. It is defined as:

$$\text{DiL} = 1 - \frac{2 \times \sum_{i=1}^{n} y_i \times \hat{y}_i}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \hat{y}_i + \epsilon}, \tag{A.1}$$

where y_i and $\hat{y}_i$ are the ground truth and predicted values, respectively, and ϵ is a small constant to avoid division by zero. DiL focuses on maximizing the overlap between the predicted and ground truth masks, making it suitable for tasks where precise segmentation is critical.

Intersection over Union Loss (IoUL): The IoUL [70] is another region-based loss function that measures the overlap between the predicted and ground truth. It is defined as:

$$\text{IoUL} = 1 - \frac{\sum_{i=1}^{n} y_i \times \hat{y}_i}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \hat{y}_i - \sum_{i=1}^{n} y_i \times \hat{y}_i + \epsilon}. \tag{A.2}$$

This loss is particularly useful in cases where there is significant class imbalance, as it penalizes both false positives and false negatives.

Cross-Entropy Loss (CEL): The CEL [71] is a loss function for classification tasks and is defined as

$$\text{CEL} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i). \tag{A.3}$$

This loss provides a local-level accuracy of semantic segmentation which is used as a baseline in our experiments and is combined with other loss functions to improve the model performance.

Binary Cross-Entropy Loss (BCEL): BCE [72] is widely used for binary classification tasks and is similar to CEL but adapted for binary output. It is defined as:

$$\text{BCE} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i). \tag{A.4}$$

BCE is effective for multi-label segmentation tasks where each frequency bin can belong to more than one class, making it particularly relevant for radio fingerprinting scenarios with overlapping signals.

Focal Loss (FL): FL [73] is designed to address the issue of class imbalance by down-weighting the contribution of easy examples during training and focusing on hard-to-classify examples. It is defined as:

$$\text{FL} = -\alpha(1 - \hat{y}_i)^{\gamma} \log(\hat{y}_i), \tag{A.5}$$

where α is a balancing factor and γ is the focusing parameter. FL is particularly effective in improving the performance of the DNN on underrepresented classes.

Combined Loss (CL): To leverage the strengths of different loss functions, we have implemented a CL function that integrates both local-level and region-level losses, defined as:

$$\mathrm{CL} = \beta \times \mathrm{CEL} + (1 - \beta) \times \mathrm{IoUL}, \tag{A.6}$$

where β is a weighting factor that balances the contribution of each component. As such, we can optimize both fine-grained accuracy and overall region consistency.

Evaluation of Loss Functions. We discuss a comparative analysis of various loss functions applied to the radio fingerprinting task of non-overlapping signals. Table A.8 shows that the CL function achieves the highest performance across all metrics, with an Intersection over Union (IoU) of 77.37%. This highlights the effectiveness of combining local-level and region-level loss functions.

Table A.8: Performance metrics for different loss functions in scenario 1 with an input size of 1024.

Loss Function	**IoU (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
BCEL	77.08	86.48	86.50	86.49
CEL	77.04	86.44	86.43	86.43
CL	**77.37**	**86.66**	**86.65**	**86.65**
DiL	60.34	83.13	69.75	75.85
FL	76.07	85.79	85.74	85.76
IoUL	59.84	83.11	68.81	75.29

DiL and IoUL show significantly lower IoU values (around 60%). While FL performed better than DiL and IoUL, it does not perform as BCEL, CEL and especially CL. As such, we chose CL for our next experiments.

Appendix B. Detailed Metrics for Each Modality

In Table B.9, we analyze the performance of `M2RF` in detail over various input sizes (1024, 2048, and 4096) with respect to the wireless mode for three different scenarios. We quantify the aspect of device-specific characterization

from each input size, especially in a wireless scenario, when signals are likely to be received and disturbed due to environmental interference. The constant high performance of the 4096 input size shows its appropriateness as a benchmark. Table B.10 presents the corresponding results in wired mode, where the improved signal quality and reduced interference further enhance performance, establishing an "ideal" scenario for comparison.

Table B.9: Metrics for different input sizes and scenarios in wireless mode.

Input Size	**Scenario**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**IoU (%)**
1024	1	86.66	86.65	86.65	77.37
	2	82.21	82.06	82.11	70.68
	3	76.28	64.35	69.71	54.99
2048	1	90.86	90.83	90.84	83.50
	2	86.24	86.10	86.16	76.12
	3	80.87	68.09	73.81	59.45
4096	1	95.02	94.97	94.99	90.54
	2	90.52	90.39	90.45	82.75
	3	84.56	70.97	77.06	63.39

Input Size 1024: When an input has a size of 1024, the metrics show that while the `M2RF` captures features at the device level, it does not perform as well with the lower IoU values, as in scenario 3, where it only achieves an IoU of 54.99%. The small input size limits the signal information that can be used for model processing. This hinders the handling of more complicated signal scenarios with heavy interference or overlaps. Precision, recall, and F1-scores also decrease across scenarios. For example, the F1-score is 86.65% in scenario 1 compared to a much lower F1-score of 69.71% in scenario 3. The aforementioned limitations, however, imply that an input size of 1024 is poor for robust radio fingerprinting within dynamic wireless environments.

Input Size 2048: Increasing the input size to 2048 gives significant improvements on all metrics. In scenario 1, the F1-score is raised to 90.84%, and IoU increases to 83.50%, which demonstrates that it can tell apart device characteristics more easily with a bigger data sample through `M2RF`. On the other hand, in scenario 3, where signals frequently overlap, we observe that the IoU and F1-score are still low at 59.45% and 73.81%, respectively. Though we set the `M2RF` to this input size of 2048 to capture finer intricacies within the signal, it is clear from our results that larger values will be needed in order to obtain strong accuracy under complicated wireless conditions.

Input Size 4096: The M2RF obtains its best performance in terms of all the metrics when having an input size of 4096. Scenario 1 achieves a precision of 95.02% (F1-score of 94.99%, IoU of 90.54%), which suggests that the M2RF is capable of leveraging the broader input size to identify device-specific hardware imperfections in the RF signals. In scenario 3, where it is logical to believe the overlapping signals will challenge the model, we still achieve an F1-score of 77.06% and an IoU of 63.39%, which is a far better performance than using any smaller input size. This means that an input size of 4096 is better for accounting for full radio fingerprinting details, even under wireless mode scenarios in which signal interference occurs and where the transmissions can be more accurately localized within the spectrum.

Table B.10: Metrics for different input sizes and scenarios in wired mode.

Input Size	Scenario	Precision (%)	Recall (%)	F1-Score (%)	IoU (%)	Latency (ms)
1024	1	97.66	97.65	97.66	95.47	14.96
	2	89.81	89.57	89.69	81.81	
	3	79.53	67.17	72.71	58.77	
2048	1	99.34	99.34	99.34	98.70	15.37
	2	94.43	94.09	94.26	89.29	
	3	84.84	71.44	77.47	64.27	
4096	1	99.56	99.55	99.56	99.12	17.44
	2	96.01	95.90	95.95	92.29	
	3	89.41	75.24	81.60	69.66	

Benchmark Results in Wired Mode: Table B.10 presents results for all three input sizes in wired mode, enabling direct comparison with wireless mode and analysis of the performance-latency trade-off under high-SNR conditions. With an input size of 1024, the M2RF achieves an F1-score of 97.66% and IoU of 95.47% in scenario 1, representing an 11% improvement over wireless (86.65%). In scenario 3, the F1-score of 72.71% is notably higher than its wireless counterpart (69.71%), although the performance gap narrows as signal complexity increases. The latency of 14.96 ms makes this the fastest processing option.

Increasing the input size to 2048 yields near-perfect performance in scenario 1, with F1-score and IoU reaching 99.34% and 98.70% respectively. In scenario 2, the M2RF achieves 94.26% F1-score, demonstrating robust handling of randomly positioned signals. In scenario 3, performance reaches 77.47% F1-score, a significant 4.76% improvement over 1024 bins, demonstrating that moderate resolution increases remain beneficial for complex

overlapping scenarios. The modest latency increase to 15.37 ms represents only a 2.7% overhead while providing substantial gains, particularly in complex scenarios.

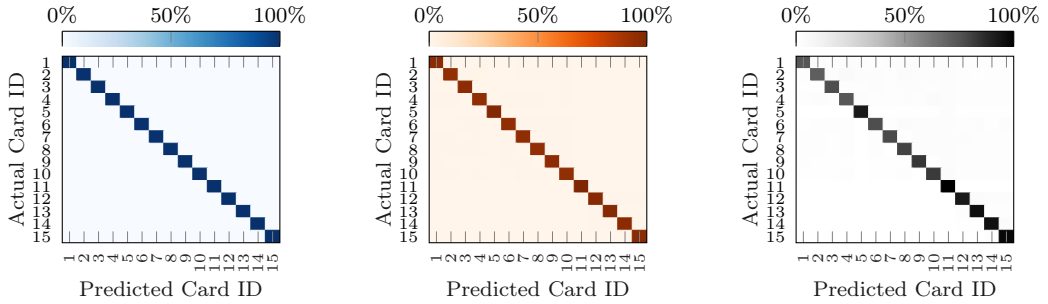
As we can see in Table B.10, input size of 4096 performs better across all scenarios. In scenario 1, the precision, recall, and F1-score were all greater than 99%, with an IoU value of 99.12%, suggesting almost perfect identification accuracy. In scenario 2, the F1-score and IoU are both high (95.95% and 92.29%, respectively), which shows that even in optimal conditions, the `M2RF` accommodates a small portion of signal belonging to a structure within its part of the observable spectrum. When things get difficult in scenario 3 with overlapping signals, the `M2RF` achieves an F1-score of 81.60% and an IoU of 69.66%, also higher than its wireless counterpart.

The wired mode analysis reveals important insights about the performance-latency trade-off and saturation behavior. For non-overlapping signals (scenario 1), F1-score largely saturates by input size 2048 (99.34%), with 4096 providing only marginal improvement (99.56%). However, for overlapping signals (scenario 3), results continues to improve across all input sizes, indicating that higher spectral resolution remains critical for distinguishing overlapping transmissions even under ideal channel conditions. The latency overhead from 1024 to 4096 input size is modest (16.6% increase), while F1-score of scenario 3 improves by 12.2%, demonstrating a favorable performance-latency trade-off.

In general, the comparison of wireless and wired modes across all input sizes justifies the choice of an input size of 4096 for radio fingerprinting purposes. The wired results exhibit consistent scaling trends similar to wireless mode, with larger inputs improving performance in both conditions while wired mode achieves higher performance due to superior SNR. The results validate that this input size is capable of capturing intricate device-specific imperfections even under difficult conditions, particularly for the challenging overlapping signals that are critical for spectrum sharing applications.

Appendix C. Additional Results for Wired Mode

This part provides an overview of the performance for three scenarios of wired mode under an input size of 4096. This wired connection means that the signals are much cleaner and less susceptible to interference, enabling `M2RF` to attain higher accuracy. The confusion matrix for each of the scenarios demonstrates how well the system performs under various signal conditions.



(a) Scenario 1. (b) Scenario 2. (c) Scenario 3.

Figure C.22: Confusion matrices for three scenarios in wired mode.

Scenario 1: Non-overlapping Signals. In the first wired scenario, we have excellent performance, also seen in the confusion matrix in Figure C.22a, as all card IDs provide accuracy values above 99%. Specifically, card 1 gives us an accuracy of 99.5%, card 3 provides 99.6%, and card 5 reaches 99.8%. Since the signals are non-overlapping in this use case, it is easy for the model to identify each device separately. This scenario underscores the effectiveness of radio fingerprinting under ideal conditions where each device's unique signal characteristics are isolated, thereby minimizing potential misclassification. The high performance in this case exemplifies how hardware imperfections can be effectively leveraged for device identification in a controlled setting.

Scenario 2: Partial Observation with Random Frequency Centers. Scenario 2, illustrated in Figure C.22b, does not degrade accuracy greatly but will degrade it bit by bit more than was the case in scenario 1 due to more complex signals. Note that, for example, cards 1 and 3, we can see the accuracies are 95.3% and 95.7%, but card 5 has a higher accuracy of 98.6%. Partial signals and random frequency centers make it more challenging to separate device-specific features from the data. Nevertheless, the `M2RF` performs reliably well, being able to contend with more complicated signal patterns but still uniquely separating devices due to hardware imperfections inherent in each. These results suggest the robustness of `M2RF` to small changes in signal characteristics, a necessary trait for dynamic RF environments where signal properties experience moderate fluctuations.

Scenario 3: Partial/Full Signals Overlapping. Figure C.22c shows the confusion matrix generated for scenario 3, which again highlights how challenging it is when signals are fully or partially overlapping. The accuracy

is significantly lower in comparison to the previous scenarios. For instance, cards 1 and 3 get lower accuracies of 71.1% and 74% respectively, while good performances are attained on cards 9 through 15 with accuracies above 80%. Because several signals occupy the same or adjacent frequency bins, it becomes challenging for the `M2RF` to discriminate between them and they are more often misclassified. However, improved signal quality in the wired mode helps mitigate some interference, enabling `M2RF` to maintain reasonable performance despite the challenging conditions. In this scenario, the heavy overlap among signals reveals the limitations of radio fingerprinting, while also demonstrating that `M2RF` can still extract discriminative features even in challenging RF environments.

References

[1] Federal Communications Commission (FCC), "Spectrum Crunch," https://www.fcc.gov/general/spectrum-crunch.

[2] L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A Survey of Advanced Techniques for Spectrum Sharing in 5G Networks," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 44–51, 2017.

[3] F. Hu, B. Chen, and K. Zhu, "Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey," *IEEE Access*, vol. 6, pp. 15 754–15 776, 2018.

[4] H. Shokri-Ghadikolaei, F. Boccardi, C. Fischione, G. Fodor, and M. Zorzi, "Spectrum Sharing in mmWave Cellular Networks via Cell Association, Coordination, and Beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 11, pp. 2902–2917, 2016.

[5] L. Lv, J. Chen, Q. Ni, Z. Ding, and H. Jiang, "Cognitive Non-Orthogonal Multiple Access with Cooperative Relaying: A New Wireless Frontier for 5G Spectrum Sharing," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 188–195, 2018.

[6] Linda Hardesty (Fierce Wireless), "What is a CBRS spectrum access system?" https://www.fiercewireless.com/private-wireless/what-a-cbrs-spectrum-access-system, 2020.

[7] L. Zhang, Y.-C. Liang, and M. Xiao, "Spectrum sharing for Internet of Things: A Survey," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 132–139, 2018.

[8] S. Kwon and H.-K. Choi, "Evolution of Wi-Fi Protected Access: Security Challenges," *IEEE Consumer Electronics Magazine*, vol. 10, no. 1, pp. 74–81, 2021.

[9] A. Koutsos, "The 5G-AKA Authentication Protocol Privacy," in *Proceedings of IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2019, pp. 464–479.

[10] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.

[11] L. Baldesi, F. Restuccia, and T. Melodia, "ChARM: NextG Spectrum Sharing Through Data-Driven Real-Time O-RAN Dynamic Control," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2022, pp. 240–249.

[12] T. D. Vo-Huu, T. D. Vo-Huu, and G. Noubir, "Fingerprinting Wi-Fi Devices Using Software Defined Radios," in *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. ACM, 2016, pp. 3–14.

[13] Y. Xu, M. Liu, L. Peng, J. Zhang, and Y. Zheng, "Colluding rf fingerprint impersonation attack based on generative adversarial network," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 3220–3225.

[14] M. Alhaidary and S. M. M. Rahman, "Security vulnerability analysis and corresponding mitigation for password-based authentication using an offline personal authentication device," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE, 2016, pp. 842–849.

[15] Z. Yao, Y. Peng, Y. Wang, C. Xu, J. Wang, Y. Lin, and G. Gui, "A novel radio frequency fingerprint concealment method based on iq imbalance compensation and digital pre-distortion," *IEEE Transactions on Information Forensics and Security*, 2024.

[16] M. R. Khanzadi, D. Kuylenstierna, A. Panahi, T. Eriksson, and H. Zirath, "Calculation of the performance of communication systems from measured oscillator phase noise," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 5, pp. 1553–1565, 2014.

[17] D. Zanetti, S. Capkun, and B. Danev, "Types and origins of fingerprints," *Digital Fingerprinting*, pp. 5–29, 2016.

[18] I. Alla, S. Yahia, V. Loscri, and H. Eldeeb, "Robust device authentication in multi-node networks: Ml-assisted hybrid pla exploiting hardware impairments," in *2024 Annual Computer Security Applications Conference (ACSAC)*, 2024, pp. 1172–1185.

[19] J. Zhang, G. Shen, W. Saad, and K. Chowdhury, "Radio frequency fingerprint identification for device authentication in the internet of things," *IEEE Communications Magazine*, 2023.

[20] T. Jian, B. C. Rendon, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, "Mac id spoofing-resistant radio fingerprinting," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2019, pp. 1–5.

[21] K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, "ORACLE: Optimized Radio Classification through Convolutional Neural Networks," in *Proc. of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2019, pp. 370–378.

[22] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, K. Chowdhury, S. Ioannidis, and T. Melodia, "Exposing the Fingerprint: Dissecting the Impact of the Wireless Channel on Radio Fingerprinting," *Proc. of IEEE Conference on Computer Communications (INFOCOM)*, 2020.

[23] P. Angueira, I. Val, J. Montalban, Ó. Seijo, E. Iradier, P. S. Fontaneda, L. Fanari, and A. Arriola, "A survey of physical layer techniques for secure wireless communications in industry," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 810–838, 2022.

[24] D. Uvaydov, M. Zhang, C. P. Robinson, S. D'Oro, T. Melodia, and F. Restuccia, "Stitching the Spectrum: Semantic Spectrum Segmenta-

tion with Wideband Signal Stitching," *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, 2024.

[25] J. Zhang, R. Woods, M. Sandell, M. Valkama, A. Marshall, and J. Cavallaro, "Radio frequency fingerprint identification for narrowband systems, modelling and classification," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3974–3987, 2021.

[26] N. Soltanieh, Y. Norouzi, Y. Yang, and N. C. Karmakar, "A review of radio frequency fingerprinting techniques," *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 3, pp. 222–233, 2020.

[27] A. Mittal, M. Zhang, T. Gourousis, Z. Zhang, Y. Fei, M. Onabajo, F. Restuccia, and A. Shrivastava, "Sub-6 ghz energy detection-based fast on-chip analog spectrum sensing with learning-driven signal classification," *IEEE Internet of Things Journal*, 2024.

[28] A. Al-Shawabka, P. Pietraski, S. B. Pattar, F. Restuccia, and T. Melodia, "Deeplora: Fingerprinting lora devices at scale through deep learning and data augmentation," in *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, 2021, pp. 251–260.

[29] Y. Xiao, Y. He, X. Zhang, Q. Wang, R. Xie, K. Sun, K. Xu, and Q. Li, "From hardware fingerprint to access token: Enhancing the authentication on iot devices," *arXiv preprint arXiv:2403.15271*, 2024.

[30] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[31] D. W. Romero, D. M. Knigge, A. Gu, E. J. Bekkers, E. Gavves, J. M. Tomczak, and M. Hoogendoorn, "Towards a general purpose cnn for long range dependencies in n d," *arXiv preprint arXiv:2206.03398*, 2022.

[32] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[33] Z. Wang, F. Ma, P. Ji, and C. Fu, "Image denoising based on an improved wavelet threshold and total variation model," in *International Conference on Intelligent Computing*. Springer, 2024, pp. 142–154.

[34] C. Donnat, O. Klopp, and N. Verzelen, "One-bit total variation denoising over networks with applications to partially observed epidemics," *arXiv preprint arXiv:2405.00619*, 2024.

[35] L. Xie, L. Peng, and J. Zhang, "Towards robust rf fingerprint identification using spectral regrowth and carrier frequency offset," in *IEEE INFOCOM 2025-IEEE Conference on Computer Communications*. IEEE, 2025, pp. 1–10.

[36] A. Saeif, S. Savio, and O. Gabriele, "The day-after-tomorrow: On the performance of radio fingerprinting over time," in *Proceedings of the 39th Annual Computer Security Applications Conference*, 2023, pp. 439–450.

[37] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, "Radio frequency fingerprint identification for lora using deep learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2604–2616, 2021.

[38] T.-D. Chiueh and P.-Y. Tsai, *OFDM baseband receiver design for wireless communications*. John Wiley & Sons, 2008.

[39] S. Liao, F. Yang, H. Chen, and Z. Yang, "Enable autonomous backscatter in everyday devices," in *IEEE INFOCOM 2025-IEEE Conference on Computer Communications*. IEEE, 2025, pp. 1–10.

[40] Y. Son, H. Jeon, and J. Kim, "Tackling the coupled frequency offset impairments for ieee 802.11 be wideband wlans," *IEEE Transactions on Wireless Communications*, 2025.

[41] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *International Conference on Learning Representations*, 2017.

[42] L. Smith and Y. Gal, "Understanding measures of uncertainty for adversarial example detection," in *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*. AUAI, 2018, pp. 560–569.

[43] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations*, 2018.

[44] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[45] I. Alla, S. Yahia, V. Loscri, and H. Eldeeb, "Robust device authentication in multi-node networks: Ml-assisted hybrid pla exploiting hardware impairments," in *Annual Computer Security Applications Conference (ACSAC)*, 2024.

[46] J. Zhang, F. Ardizzon, M. Piana, G. Shen, and S. Tomasin, "Physical layer-based device fingerprinting for wireless security: From theory to practice," *IEEE Transactions on Information Forensics and Security*, 2025.

[47] H. Givehchian, N. Bhaskar, A. Redding, H. Zhao, A. Schulman, and D. Bharadia, "Practical obfuscation of ble physical-layer fingerprints on mobile devices," in *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2024, pp. 2867–2885.

[48] Z. Lu, W. Xu, M. Tu, X. Xie, C. Hua, and N. Cheng, "Erasing radio frequency fingerprints via active adversarial perturbation," *arXiv preprint arXiv:2406.07349*, 2024.

[49] S. Karunaratne, E. Krijestorac, and D. Cabric, "Penetrating rf fingerprinting-based authentication with a generative adversarial attack," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[50] G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro, "Towards scalable and channel-robust radio frequency fingerprint identification for lora," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 774–787, 2022.

[51] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, "Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 646–655.

[52] X. Gu, W. Wu, N. Guo, W. He, A. Song, M. Yang, Z. Ling, and J. Luo, "Terff: Temperature-aware radio frequency fingerprinting for smartphones," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2022, pp. 127–135.

[53] D. Uvaydov, S. D'Oro, F. Restuccia, and T. Melodia, "DeepSense: Fast wideband spectrum sensing through real-time in-the-loop deep learning," in *Proc. of IEEE Intl. Conf. on Computer Communications (INFOCOM)*, Vancouver, BC, Canada, May 2021.

[54] C. Liu, J. Wang, X. Liu, and Y.-C. Liang, "Deep CM-CNN for Spectrum Sensing in Cognitive Radio," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2306–2321, 2019.

[55] X. Wang, Y. Zhang, H. Zhang, X. Wei, and G. Wang, "Identification and authentication for wireless transmission security based on rf-dna fingerprint," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 230, 2019.

[56] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, and K. Chowdhury, "No Radio Left Behind: Radio Fingerprinting Through Deep Learning of Physical-Layer Hardware Impairments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 165–178, 2019.

[57] T. Yang, S. Hu, W. Wu, L. Niu, D. Lin, and J. Song, "Conventional neural network-based radio frequency fingerprint identification using raw i/q data," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 8681599, 2022.

[58] W. Wu, S. Hu, D. Lin, and G. Wu, "Reliable resource allocation with rf fingerprinting authentication in secure iot networks," *Science China Information Sciences*, vol. 65, no. 7, p. 170304, 2022.

[59] M. Liu, J. Wang, N. Zhao, Y. Chen, H. Song, and F. R. Yu, "Radio frequency fingerprint collaborative intelligent identification using incremental learning," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 5, pp. 3222–3233, 2021.

[60] A. Elmaghbub and B. Hamdaoui, "No blind spots: On the resiliency of device fingerprints to hardware warm-up through sequential transfer learning," in *Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2024, pp. 134–144.

[61] M. Cekic, S. Gopalakrishnan, and U. Madhow, "Wireless fingerprinting via deep learning: The impact of confounding factors," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021, pp. 677–684.

[62] L. Papangelo, M. Pistilli, S. Sciancalepore, G. Oligeri, G. Piro, and G. Boggia, "Adversarial machine learning for image-based radio frequency fingerprinting: Attacks and defenses," *IEEE Communications Magazine*, pp. 1–7, 2024.

[63] D. Roy, T. Mukherjee, M. Chatterjee, E. Blasch, and E. Pasiliao, "Rfal: Adversarial learning for rf transmitter identification and classification," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 783–801, 2020.

[64] R. C. K. Ellore, R. R. Yakkati, S. R. Yeduri, S. Boppu, and L. R. Cenkeramaddi, "Gsm-based mobile handset identification using rf fingerprints and deep learning on the edge computing devices," *IEEE Sensors Journal*, 2024.

[65] T. Zhang, S. Ye, K. Zhang, J. Tang, W. Wen, M. Fardad, and Y. Wang, "A systematic dnn weight pruning framework using alternating direction method of multipliers," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 184–199.

[66] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, "Learning efficient convolutional networks through network slimming," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2736–2744.

[67] W. Wen, C. Xu, C. Wu, Y. Wang, Y. Chen, and H. Li, "Coordinating filters for faster deep neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 658–666.

[68] T. Jian, Y. Gong, Z. Zhan, R. Shi, N. Soltani, Z. Wang, J. Dy, K. Chowdhury, Y. Wang, and S. Ioannidis, "Radio frequency fingerprinting on the

edge," *IEEE Transactions on Mobile Computing*, vol. 21, no. 11, pp. 4078–4093, 2022.

[69] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*. Springer, 2017, pp. 240–248.

[70] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *International symposium on visual computing*. Springer, 2016, pp. 234–244.

[71] G. Zhao, W. Yang, X. Ren, L. Li, Y. Wu, and X. Sun, "Well-classified examples are underestimated in classification with deep neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 9180–9189.

[72] Y. Huang, J. Qi, X. Wang, and Z. Lin, "Asymmetric polynomial loss for multi-label classification," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[73] T.-Y. Ross and G. Dollár, "Focal loss for dense object detection," in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2980–2988.