AI-Assisted RF Fingerprinting for Identification of User Devices in 5G and FutureG

Aishwarya Jawne
Center for Connected Autonomy & AI,
Florida Atlantic University
ajawne2023@fau.edu

Georgios Sklivanitis
Center for Connected Autonomy & AI,
Florida Atlantic University
gsklivanitis@fau.edu

Dimitris A. Pados
Center for Connected Autonomy & AI,
Florida Atlantic University
dpados@fau.edu

Elizabeth Serena Bentley
Air Force Research Laboratory
elizabeth.bentley.3@us.af.mil

***Abstract*—As 5G networks expand to support increasingly complex and diverse applications, ensuring robust identification and authentication of user devices has become a critical requirement for physical layer security. This paper investigates the potential of machine learning techniques for radio frequency (RF) fingerprinting as a scalable solution for identifying and authorizing access to trusted user devices as well as detecting rogue user devices in 5G networks. Specifically, we evaluate the performance of three prominent deep learning architectures— ResNet, Transformer, and LSTM — across various configurations, including spectrogram and raw IQ slice inputs made from varying packet sizes. The results demonstrate that ResNet models, when paired with spectrogram inputs, achieve the highest classification accuracy and scalability, while effectively addressing challenges such as the Next-Day Effect. Contrary to existing works, which focus on training deep neural networks (DNNs) for device classification, we highlight the critical role of spectrograms in capturing distinct hardware impairments when used to train DNNs for RF fingerprint extraction. These RF fingerprints are then used to distinguish between trusted and rogue 5G devices, as well as for device classification and identification. By identifying the optimal configurations for these tasks and exploring their applicability to real-world datasets collected from an outdoor software-defined radio testbed, this paper provides a pathway for integrating AI-driven radio frequency fingerprinting for authentication of user devices in 5G and FutureG networks as a cornerstone for enhanced physical layer security.**

I. Introduction

The rapid uptake and widespread use of 5G networks has transformed the landscape for Internet of Things (IoT) systems with improved data rates, latency, and support for diverse applications. The benefits of this technology are accompanied by a variety of security threats, including unauthorized access and device impersonation. The current commercial 5G security measures in place fall short in device authentication efforts, especially against adversarial entities exploiting hardware or protocol vulnerabilities [2].

This work was supported in part by the NSF Engineering Research Center for Smart Streetscapes EEC-2133516 and NSF Grant CNS-2117822, and by the Air Force Research Laboratory under Grant FA8750-21-F-1012.

Workshop on Security and Privacy of Next-Generation Networks (FutureG) 2025
24 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-7-3
https://dx.doi.org/10.14722/futureg.2025.23009
www.ndss-symposium.org

Radio Frequency Fingerprinting (RFF) is a swiftly popularizing area of study that is emerging as a promising solution to address device authentication vulnerabilities amongst other network uses. RFF leverages a device's unique hardware deficiencies detected through its signal transmission to verify its true identity. Recurring imperfections in a device's signal transmission resulting from hardware deficiencies are extracted and analyzed using machine learning (ML) or deep learning (DL) techniques so that their unique features can be turned into a standardized "fingerprint" for the device [3],[27].

In this paper, we delve into the efficacy of artificial intelligence (AI) in radio device identification to gain insights into the practicality of using this technology as a reliable part of the device authentication process in 5G and FutureG networks. The contributions of this work are as follows:

1) We evaluate three neural network models: Residual Network (ResNet), Transformer, and Long Short-Term Memory (LSTM), on their ability to extract discerning hardware impairments from 5G signal transmissions.
2) We compare combinations of different inputs and DL architectures to determine the ideal conditions required to output consistently accurate device identification results.
3) We assess the scalability of AI technologies covered in this paper for real-world 5G Radio Access Network (RAN) physical layer security applications, including rogue device detection.

This paper aims to explore an approach for advancing device authentication at the physical layer of a RAN using AI with a comprehensive analysis of different DL architectures, configurations, and simulated physical layer security breaches.

II. Background

The current device authentication measures in place in 5G networks are a collection of secure functions in the 5G Core that ensure only legitimate user equipment (UE) can access the network. The UE makes its initial contact with the next-generation Node B (gNB) or base station by performing the Random Access Channel (RACH) procedure over the Physical Random Access Channel (PRACH). Upon receiving this signal, the base station assigns the UE a temporary identifier and forwards a Registration Request from the UE to the Access and Mobility Management Function (AMF) in the 5G Core. The Registration Request consists of an

encrypted Subscription Permanent Identifier (SUPI), which can be vulnerable to impersonation. The AMF works with the Authentication Server Function (AUSF) to retrieve the authentication credentials for the UE from Unified Data Management (UDM). The UDM generates authentication vectors based on the 5G-AKA (Authentication and Key Agreement) or EAP-AKA' protocols, which are sent back to the AMF via the AUSF. The AMF sends the vectors back to the UE in an Authentication Request, expecting a response back from the UE. The AMF then validates the UE's response against the expected response provided by the AUSF. If the responses at both ends match, the authentication is successful and after establishing security modes and resource allocation, the AMF sends the UE a Registration Accept message. With that the device is authenticated and securely connected to the 5G Network [1],[26].

Even with advanced cryptographic techniques and encryption this process has vulnerabilities making it susceptible to attacks for unauthorized network access. A common method employed by malicious entities to breach this process's security is device impersonation or spoofing, in which an unauthorized device poses as a legitimate one to connect to the network by using another legitimate device's SUPI or International Mobile Subscriber Identity (IMSI). These credentials can be stolen in a variety of ways including SIM cloning, authentication database breaches, signal interceptors, Man-in-the-Middle Attacks, fake base stations, malware, etc. [2],[4],[5]. Although there are many layers of security in place in current authentication protocols, they are mainly focused on verifying credentials put forth by a UE while neglecting its physical identity. This means that a device's credentials may be legitimate but they may not actually belong to that device [6],[26].

Physically, all devices are attributed with minuscule differences through the manufacturing process, even if their designs and systems are identical. These differences show up in subtle ways through a device's manner of transmitting signals such as nonlinearities, in-phase and quadrature (IQ) imbalance, noise, timing, and frequency offset caused by slight hardware imperfections [7],[8],[26]. These features can be found through the transmitter's raw IQ data (extractable from the receiving gNB's physical layer) but they are imperceptible by simple analysis, for which we turn to AI [27].

In this paper, we explore three types of neural networks renowned for their ability to detect and extract intricate features from input data:

1) ResNet: a neural network known for its residual connections, which effectively address the vanishing gradient problem and enable the training of very deep networks, making it highly effective at extracting hierarchical features from complex data [9].
2) Transformer: a neural network that excels in capturing long-range dependencies and relationships within input data using self-attention mechanisms, making it particularly adept at identifying intricate patterns across sequences or spatial dimensions [10].
3) LSTM: a neural network designed to capture temporal dependencies and context over long sequences with its gated architecture, which allows it to selectively retain and forget information, making it ideal for sequential data analysis [11].

In its current monolithic design, the physical layer of the 5G RAN and the raw IQ data exchange within, are relatively inaccessible. Understanding the value of analyzing raw IQ data and implementing RFF for device identification will guide the design of 5G and FutureG Open RANs which could integrate such functions for increased security and resiliency.

III. Related Work

Despite substantial progress, many gaps remain in understanding the comparative performance of different DL architectures on 5G data and their respective ability to generalize across diverse network conditions. This paper seeks to address these gaps by systematically evaluating multiple architectures and input representations.

Feature extraction in radio frequency (RF) focuses on identifying unique device characteristics embedded in transmitted signals. DL techniques have revolutionized this process by reducing the need for manual preprocessing and leveraging raw IQ samples or spectrograms for robust representation [27]. Device-specific fingerprints are a direct result of manufacturing variations in RF circuitry components such as power amplifiers, frequency mixers, and oscillators [7],[8]. Convolutional neural networks (CNNs) and transformer-based models have demonstrated the ability to classify these fingerprints with high precision for WiFi and LoRa devices [12]-[18].

In the context of 5G networks, RFF-based techniques provide viable solutions for secure device identification and localization [19],[26]. In a recent publication, PRACH signal analysis and differential constellation trace figure (DCTF) representation have shown promising results in identifying devices under varying channel conditions [20]. Additionally, there are examples of IQ data from 5G transmissions from Software Defined Radios (SDRs) being successfully classified using CNNs and DL in [21], [28]. However, significant challenges remain, including channel-induced variability and the Next-Day Effect (also referred to as "Day-After-Tomorrow" effect in related literature), which refers to the degradation of identification performance over time [22].

While numerous studies on RFF focus on DL applications, few systematically compare different architectures or input representations in the specific context of 5G networks. Additionally, the lack of standardized datasets poses a significant barrier to reproducibility and cross-study evaluations. Furthermore, challenges related to scalability and real-world deployment arising from environmental variability and potential adversarial conditions require further exploration [27].

IV. Methods for Device Authentication

A. Device Identification via AI Classification

In this part of the paper, we consider the three kinds of neural networks described in the previous section to carry out device identification/classification. The architecture of the ResNet, Transfomer, and LSTM models (Figure 1) remain constant throughout this portion of the experiment to capture

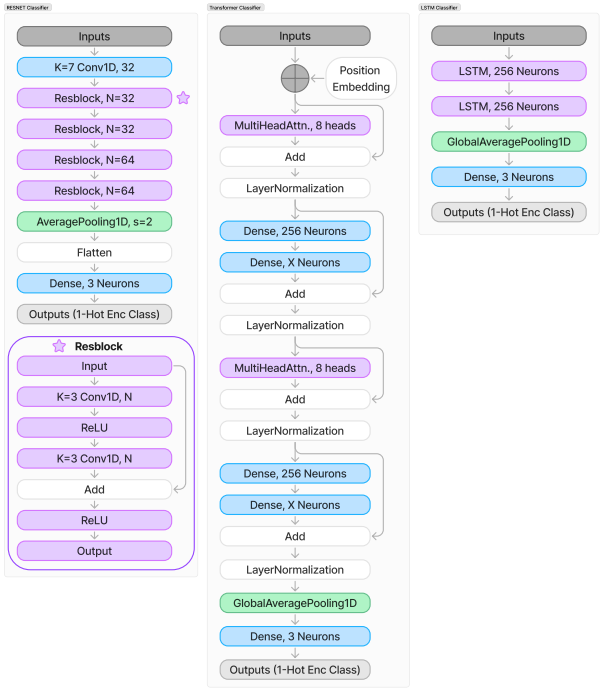

Fig. 1. Classifier models (left to right): ResNet, Transformer, & LSTM. [1]

the effect of manipulating model inputs. The set of inputs consists of 10 versions of the signal dataset used for training and classification, where each version is a different representation of the data as raw IQ slices or spectrograms made using one of the 5 packet sizes (see Section V-B for more details). These variations allow for a comparison of the models' capabilities in handling different representations of the same data. The output for both experiments is a soft-max classification result, represented in a one-hot encoded format to clearly identify distinct classes, which in this case is a fixed set of 3 devices. In this section, the model is trained to minimize identity loss directly from the model output to focus it as strictly a classification task. This is changed in the second part of the paper described in the next subsection. Holistically, this setup aims to demonstrate the potential of a specific neural network architecture to effectively extract meaningful features from 5G data, irrespective of whether the input is in raw time-series format or spectrogram representation. The models are assessed on their ability to accurately interpret and contextualize these features, ultimately establishing a foundation for their performance in differentiating between various UEs in a 5G network. Since this set of experiments is not directly related to the process of RFF, it does not need to be recreated in order to carry out the second section of the study.

B. Device RF Fingerprinting via AI Feature Extractors

In this part of the paper, we consider the same neural network architectures as before but this time used for RFF

[1]For parts of the experiment where spectrogram inputs are used, all the 1D layers change to 2D versions with the rest remaining the same (for example, AveragePooling1D turns into AveragePooling2D). In all Transformer models X = 2 with slice inputs and X = 200 when the input is spectrograms.

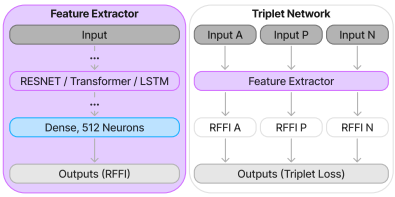

Fig. 2. Feature extractor modification & Triplet network.

extraction which will then be used for device identification. Keeping practical applications in mind, the goal is to distinguish between trusted and non-trusted devices, successfully identify a rogue device masquerading amongst a set of legitimate devices on the basis of being able to distinguish each device based solely on its IQ data. The experiments described in the prior section lend the assumption that ResNet, Transformer, and LSTM models are able to decipher patterns of discerning hardware impairments to varying degrees. With this idea, three feature extractor models are created based on the ResNet, Transformer, and LSTM neural networks from the previous section with the main change in architecture being the output of a 512-element vector representing the radio fingerprint of the input (Figure 2).

The feature extractor is part of a triplet network that receives batches of anchor and positive inputs from a particular device's set of packets and negative inputs from a different device. These 3 inputs are all processed through a feature extractor to obtain anchor, positive, and negative radio fingerprints and the feature extractor is trained by minimizing the triplet loss of the 3 radio fingerprints (Figure 2). This setup allows us to compare classification and rogue device detection outputs of each configuration by manipulating the inputs (10 versions of the same IQ signal data with varying representation methods and packet sizes like in the previous section).

Rogue device detection and device classification both use the k-Nearest-Neighbor (kNN) algorithm on radio fingerprints produced using a trained feature extractor model. IQ data from a set of enrolled or legitimate devices is preprocessed based on the choice of raw IQ slice or spectrogram and packet size and then it is processed through a trained feature extractor to produce radio fingerprints. These legitimate device radio fingerprints and their associated labels are used to initiate and train a kNN classifier. For device classification, a set of unlabeled IQ data from the same devices used in training the kNN classifier is preprocessed and fingerprinted in the same way for the trained kNN classifier to predict device labels. The classification accuracy score is based on kNN label prediction and this is the main metric used to determine the performance of the feature extractor and input configuration as a whole. Rogue device detection is done in a similar way up to and including the point of enrolling legitimate device radio fingerprints into the kNN classifier. For testing the rogue detection function, unlabeled data from both legitimate and rogue devices are combined together and a number of nearest enrolled neighbors to each of the test fingerprints is found using the trained kNN classifier. A detection score is calculated

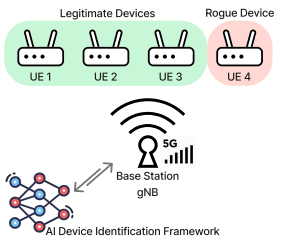

Fig. 3. Hardware setup for device authentication experiment.

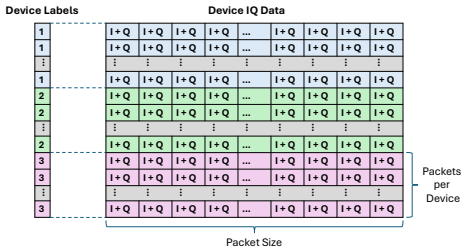

Fig. 4. Data & label matrix layout for experiment inputs.

as the mean distance to the nearest 15 neighbors of each test device sample. Based on this detection score, the test set labels are assigned as 1 for a legitimate device and 0 for a rogue device. The performance of the rogue device detection function is determined mainly from its Receiver Operating Characteristic (ROC) curve, which is based on the False Positive Rate (FPR) and True Positive Rate (TPR) of the classifier. The area under the ROC curve (AUC) is the main metric used for ranking the rogue device detection performance for each combined feature extractor and input configuration [26].

The objective of this part of the study is to evaluate the efficacy of creating RFFs using 3 kinds of DL models with different representations of IQ data. This idea is driven by the need for scalability. Without relying solely on direct classification, we could explore the development of RFF databases that can be continuously accessed and modified by the network through device authentication. RFF feature extraction and device classification model training could also take place offline independently of real-time device authentication processes to avoid creating a system bottleneck. The results of this experiment yield a baseline understanding of how AI could be integrated into future RANs to enhance security.

V. Experimental Setup

A. Data Source

The various datasets created for use in this experiment are sourced from the published PAWR dataset by the POWDER Platform. This PAWR dataset contains continuous streams of 5G New Radio (NR) IQ data from 4 different SDRs transmitting to a central SDR base station acting as their gNB in an outdoor university campus setting (Figure 3). Each UE radio is configured to transmit 5G NR standard-complaint frames that were generated using MATLAB's 5G toolbox. The incoming signals were sampled in the gNB at a rate of 7.69 mega-samples per second (MS/s) at a center frequency of 2.685 GHz, corresponding to 5G NR in LTE Band 7. The contents of the PAWR dataset consist of all 4 UEs' data. Each UE has 5 sets of 5.3 million consecutive IQ data points, with each set collected approximately 10 seconds apart. Although the original dataset also included WiFi and LTE signals, we focus only on the 5G NR transmission data [21].

All UEs utilized bit-similar USRP X310 radios as transmitters, while the gNB was implemented using a USRP B210 radio. The spatial configuration of the UEs and the gNB varied, with distances ranging from 300 meters to 1 kilometer. The entire data collection process was repeated on the next day to provide temporal variability for evaluation. We leverage both days of data to gain insights on the Next-Day Effect. This effect in RFF is known as a pattern of device classification efficacy of a DL model gradually declining over the days following model training due to channel variability, i.e., a model or system trained with Day 1 data will classify Day 1 data better than Day 2 data even if Day 2 data is from the exact same devices as Day 1 [22].

B. Data Preprocessing

The raw IQ data streams from the PAWR dataset are repackaged and further processed before they are used as inputs for the DL systems. Raw IQ data gathered over a single day from all devices is segmented into consecutive packets of a determined size and combined in a single table as shown in Figure 4. In this experiment, we want to observe the effect of packet size on device identification, the packet sizes used in this experiment are: 76900, 38450, 7690, 3845, and 769 IQ samples per packet. A separate version of each data collection day's training data set, enrollment data set, test data set, and rogue device data set is created in each of the 5 packet sizes, which totals to 40 files used as inputs for this experiment.

The series of packet sizes for testing are determined based on the length of the data frame of the 5G transmission being used in this study. A typical 5G data frame is approximately 10 ms long and since the transmission is sampled at 7.69 MS/s, each data frame contains approximately 76900 samples [23]. The starting packet size coincides with the size of a whole data frame. This is also referred to as a full-resolution packet size throughout the rest of the paper. The other packet sizes 38450, 7690, 3845, and 769, coincide with 50%, 10%, 5%, and 1% of the full-resolution packet size respectively.

For the parts of the paper utilizing IQ slices, the complex value contents of the input files are split so that each packet has 2 rows of the length of the experimenting packet size, the first row containing the real values of the IQ samples and the second row containing its respective imaginary values (Figure 6). This is the processed input received by each device classifier and feature extractor model (via the triplet loss network) for IQ slice experiments.

For the parts of the paper that use spectrogram inputs, the complex IQ data values of the input files are normalized in each packet and then converted into spectrograms. The

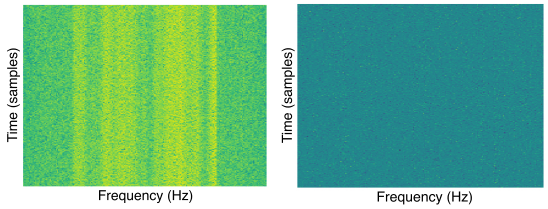

Fig. 5. Spectrogram (left) and channel isolated spectrogram (right) examples, respectively.

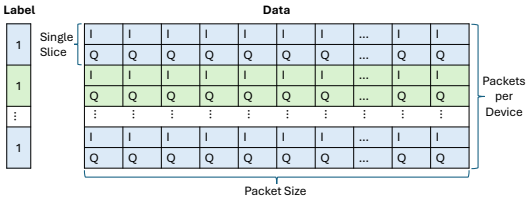

Fig. 6. Example of an IQ slice.

spectrograms are created out of each packet using Short Time Fourier Transform (STFT) with the following specifications: boxcar window, 200 frequency bins (length of segment in number of samples), 175 overlapping samples, and 200 points for the Fourier Transform. Then, to make the spectrograms channel-independent, values in each channel of the spectrogram are divided by its preceding channel's values with the assumption that what is left behind are the hardware impairments [25]. With memory limitations in mind, each spectrogram is then cropped to only use the center 40% along the time axis while keeping the entire frequency range intact (Figure 5). Cropped spectrograms are the input received by both the device classifier and feature extractor models to test device identification from visual IQ data representation.

C. Training and Evaluation Process

The hyperparameters for both the classifier models and feature extraction models were configured as follows: the learning rate was set to RMSProp 1e-3, and the batch size was set to 20. For triplet loss, where applicable, the margin was set to 0.1. For kNN classification, the number of neighbors was fixed at 15 for all kNN functions.

Training for all models is done using 267 packets per legitimate device. This limitation arises from the minimum number of packets available across all packet sizes. Of these, 10% of the training data is allocated for validation during the training process. For testing, 34 packets per legitimate device were used in both parts of the study. In the radio fingerprint extraction experiment, the kNN classifiers were trained using extracted radio fingerprints of the enrollment dataset containing 69 packets from each legitimate device (unused packets from the same devices used in training). For the rogue detection experiment, the test dataset combined 34 packets from each legitimate device with 34 packets from the rogue device. We considered 3 legitimate devices for all experiments and included 1 rogue device for the rogue detection process.

The experiments were conducted using 5 different packet sizes, cycling through all 4 combinations of training on Day 1 or Day 2 data and evaluating on Day 1 or Day 2 data. For each run of the experiment (for a particular combination of packet size, training day data, and evaluation day data) the model (classifiers and feature extractors) trains and evaluates (classifies devices or classify radio fingerprints and detects rogue devices for the triplet network) 5 times using the average of 5 runs to represent the performance metric of a certain configuration. This setup ensures a comprehensive evaluation of the models' performance under various temporal and data conditions while accounting for variability introduced by different packet sizes and cross-day training and evaluation.

VI. Experimental Results

A. Slice and Spectrogram Classification

First, we evaluated the general ability of the 3 neural networks with direct one-hot classification of 3 UEs, which revealed many insights with respect to the choice of model, packet size, and input form.

1) Effect of Input Configuration: Based solely on the classification accuracy, a general trend observed across all models and input representations is that inputs to any model made from a larger packet-size database would yield better accuracy scores that decrease proportionally as the packet sizes get smaller. However, there is also a trade-off in time and resources spent for training and classifying with larger packet sizes. We observe proof of that through the training and classification of each model type increasing exponentially with packet sizes. In cases where the models are working with slices instead of spectrograms, this effect is much more noticeable. Although, the longer training times are associated with larger packet sizes, larger packet sizes do not necessarily ensure higher accuracy. This pattern is also observed in overall classification times of each configuration.

In comparing the outcomes of using different representations of the same transmission for direct device classification, spectrograms emerge as the clear best method (Figure 7). Using slices, each model's ability to distinguish each UE vary from mid-range to poor across any combination of train-day data and classification-day data. Using spectrograms enables the use of all three neural network models with large packet sizes to produce excellent classification results. This indicates that spectrograms are more compatible with all model architectures with respect to computing power and time limitations.

Transformers models run out of memory in the process of carrying out tensor math and positional encoding from a full resolution slice input due to their high memory and computational requirements. In our experiment, the Transformer could only accept slices in packet sizes of 769 for training and classification but it performed according to the general trend of improved classification accuracy with larger packet sizes for spectrogram inputs.

LSTM models fail to classify slices effectively, achieving only 33% accuracy across all packet sizes of slices but also perform according to the trend of better classification accuracy with spectrograms from larger packets. Additionally, using IQ

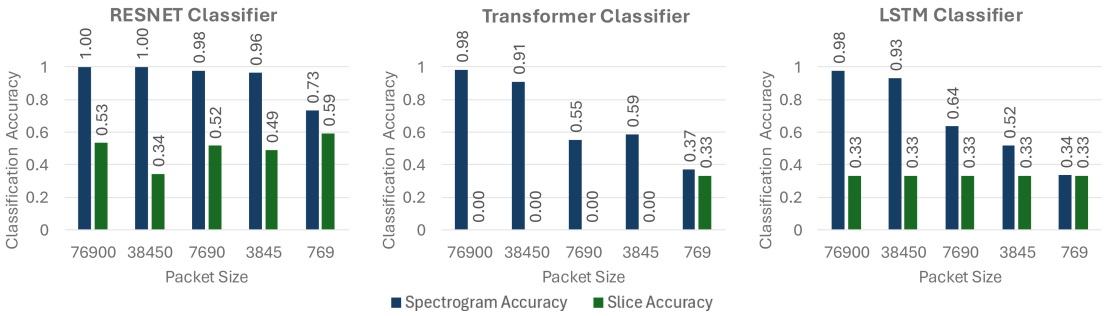

Fig. 7. Comparing average classification accuracy of the ResNet, Transformer, and LSTM models.

slices as inputs for LSTMs, results in excessively long training times with large packets with no improvement in accuracy.

ResNet models contradict the trend of larger packet sizes when using raw IQ slices. With slice inputs we observe that classification scores are getting lower as the packet size increases. However with spectrogram inputs, decreasing the packet size used for creating spectrograms leads to a faint but noticeable decline in classification accuracy. Training times are nearly halved when using IQ slice inputs over spectrograms but the accuracy also significantly decreases. ResNets exhibit greater consistency in performance, especially with smaller packet sizes, for both spectrogram and slice inputs.

2) The Next-Day Effect: Models trained on Day 1 data experience a notable drop in classification accuracy when classifying Day 2 data. However, this accuracy gap is less pronounced when classifying Day 1 data using a model trained on Day 2 data. Same-day training and classification consistently yield high accuracies. In comparing the effect of IQ representation for training, specifically with ResNet, we noticed that using spectrograms improves classification ability by up to 15%.

3) Overall Outcomes: The combination of ResNet and spectrogram inputs achieves the best classification performance, delivering the highest accuracy with significantly lower training and classification times compared to other models. A ResNet model trained with full-resolution packet sizes, particularly spectrogram inputs, emerges as the most effective approach for distinguishing 5G devices based on their transmission features. However, ResNet does not outperform other models in mitigating the Next-Day Effect with large packet spectrogram inputs. LSTMs and Transformers struggle with producing results on par with ResNet, especially using IQ slice data. Additionally, they are computationally demanding and need a higher degree of input optimization to avoid excessive training time and space or out-of-memory issues.

B. RF Fingerprint Extraction

This part of the paper evaluates the same three DL models from the previous section repurposed for extracting RFF as part of an architecture design inspired by real-world device authentication in 5G and FutureG. The objective is to determine whether AI can effectively capture discernible features to distinguish between enrolled trusted user devices and rogue devices. This approach aims to establish an understanding of scalable and standardized authentication mechanisms for 5G networks with potential to adapt to FutureG networks with appropriate adjustments.

1) Effect of Input Configuration: With respect to packet sizes, larger packet sizes are consistently associated with improved classification and rogue detection accuracy. However, computational and memory constraints limit the performance of Transformer networks when processing large packet slices. When spectrogram inputs are used, larger packet sizes lead to consistently better results across all three models, albeit with increased training, classification, and rogue detection times.

In terms of classification performance, LSTMs and Transformers both perform slightly better on average as classifiers compared to raw RFF extractors when spectrograms are used as inputs. Transformers show performance comparable to LSTMs when using spectrogram inputs but are challenged by raw IQ slice inputs, suffering from computational inefficiencies and out-of-memory issues beyond the packet size of 7690. Both LSTMs and Transformers exhibit slower training and inference times for classifying and detecting rogue devices.

The ResNet model emerged as the most effective for both device classification and rogue detection (Figure 8). ResNet demonstrates high accuracy with both spectrograms and raw IQ slices, particularly with larger packet sizes. Notably, the ResNet model is able to train, classify, and detect rogue devices more efficiently with spectrogram inputs. Even though training is significantly longer than the other models with both forms of input, ResNet-based classification and rogue device detection takes approximately the same amount of time. When using IQ slices, ResNet exhibits slightly lower classification accuracy, and the Next-Day Effect is more pronounced.

2) The Next-Day Effect: Despite this, the Next-Day Effect was significantly reduced by using RFF extraction methods compared to plain device classification. In some cases, especially with ResNet models trained on spectrograms derived from large packets, the Next-Day Effect was nearly eliminated when comparing classification accuracies (Figure 9).

3) Overall Outcomes: In summary, the optimal configuration for device identification and rogue device detection using RFF extraction involves ResNet models with larger packets and spectrogram inputs. For the highest classification accuracy and robust rogue detection metrics, full-resolution spectrograms are preferred. However, medium packet sizes can also be reliably supported. When the number of training

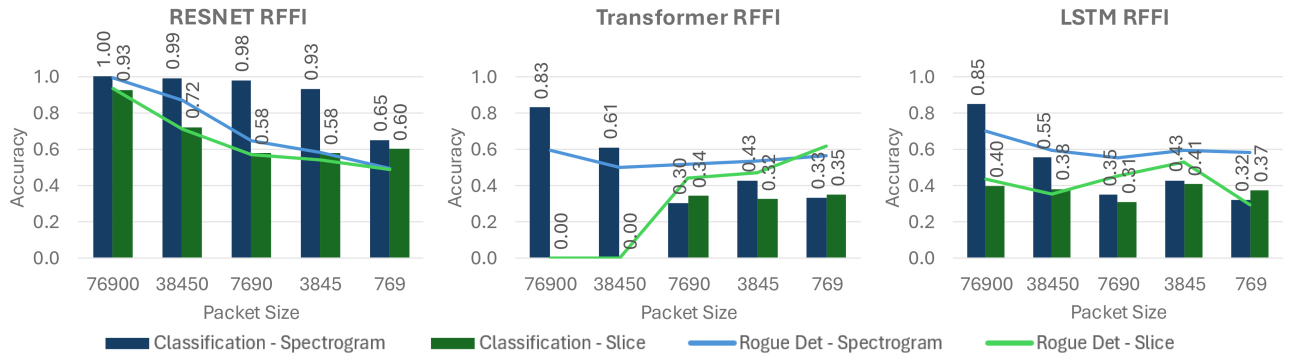

Fig. 8. Comparing trusted and rogue device identification capabilities of the RFF triplet networks using ResNet, Transfomer, and LSTM feature extractor models.

packets is limited, spectrograms remain the best input choice for ResNet due to their efficiency and accuracy. For faster classification and rogue device detection with near-perfect accuracy, offline training with a greater number of large packet slices is recommended. In contrast, LSTMs and Transformers are less capable of achieving comparable quality and speed when working with similar inputs and packet size constraints. These findings underscore the potential of ResNet models for scalable and efficient RFF in 5G user device authentication.

VII. Discussion

A. Analysis of Results

Although generating spectrograms introduces additional complexity in terms of computational and memory requirements, they remain the most effective medium for RFF. Spectrograms provide a clearer representation of hardware deficiencies compared to raw IQ samples, as evidenced by the substantial mitigation of the Next-Day Effect when spectrograms are used instead of slices. Furthermore, spectrograms offer greater reliability due to their superior classification accuracy, particularly when combined with ResNet models operating at full-packet-size configurations (Figure 10). This combination achieves near-perfect accuracy compared to slices under identical conditions. Spectrograms are also advantageous for scalability. By maintaining a consistent number of features or frequency bins in the spectrogram, both larger and smaller frame sizes can be supported for device authentication, provided the full bandwidth of the signal is preserved. Although training and classification with spectrograms require more time compared to slices, they consistently produce better results, even with limited data. This superior performance can be attributed to the effectiveness of time-frequency representations in capturing distinct transmitter features, as opposed to time-series data. This finding strongly supports the idea that hardware impairments in transmitters are best characterized in the time-frequency domain, and since models achieve optimal performance at full-frame packet sizes, these impairments likely recur at specific frequencies but could appear at any point along the frame. Consequently, spectrograms are the most reliable input for creating "unique RFFs" for each device.

However, the use of spectrograms does come with trade-offs. Their computational demands are higher than those of slices, even when using the same packet size. To make spectrogram-based RFF feasible for practical applications, passive and time-intensive processes, such as model training and device enrollment, may need to be conducted offline. Additionally, accelerating rogue device detection would require hyperparameter optimization and potentially the use of dedicated high-performance hardware.

LSTM and Transformer networks are less suitable for RFF applications. To achieve performance comparable to ResNet models, these networks would need significantly increased complexity, such as deeper architectures, which would lead to higher computational costs and longer processing times. This makes LSTM and Transformer networks impractical for real-world deployment compared to the efficiency and reliability offered by ResNet when using spectrograms.

B. Limitations

This paper faced several limitations that could influence the generalization and robustness of the results. First, the experiments are based on a public dataset, therefore, lack verification of the exact controlled signal transmitted by the UEs to the base station. Ensuring that the same signal is transmitted by each UE on all days would better isolate hardware deficiencies and minimize the risk of inadvertently capturing features specific to the transmitted signal itself.

In the context of 5G, the PRACH process used during network attachment includes various formats and preamble lengths. The impact of different PRACH formats and preambles on model training and classification was not explored. Future work should investigate whether it is necessary to include all combinations of PRACH configurations to standardize the training process or whether certain models, such as ResNet, remain unaffected by such variations.

Lastly, the diversity of transmitter models used in the experiment was limited. To improve the robustness of the findings, future research should involve a larger set of UEs and a broader variety of transmitter models. This would help ensure that the models generalize well across diverse hardware configurations and are applicable to real-world scenarios. These limitations highlight opportunities for further investigation to enhance the applicability and scalability of RFF techniques.

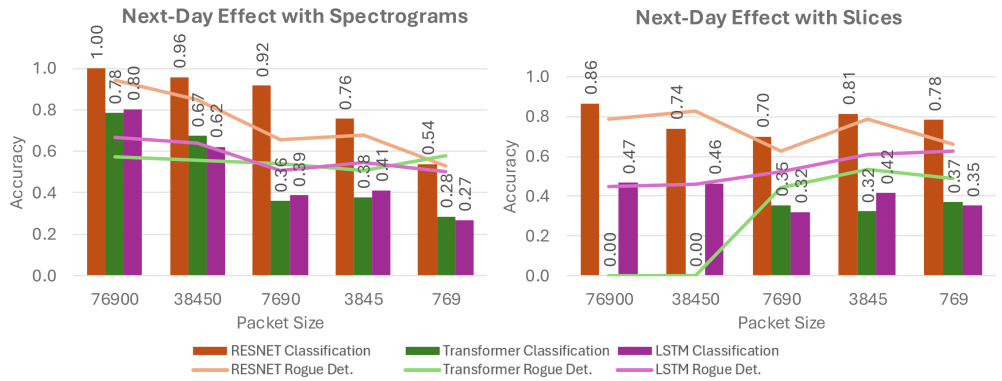

Fig. 9. Effects of training feature extractor models with Day 1 data for classification and rogue device detection with Day 2 data - Next-Day Effect.

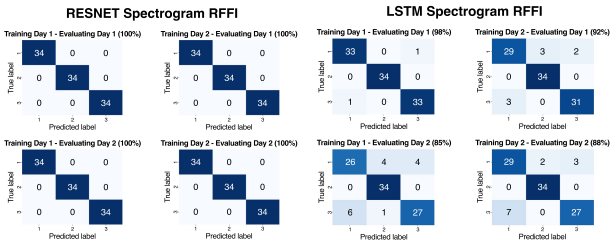

Fig. 10. Comparing classification consistency in RFF triplet network models using ResNet and LSTM.

VIII. Conclusions

A. Summary of findings

The findings of this study highlight the superiority of ResNet models over Transformers and LSTMs in terms of device identification and rogue device detection, particularly when working with limited input data. ResNet demonstrates the ability to effectively discern transmitter features, making it the most suitable model for RFF tasks. Additionally, spectrogram inputs significantly outperform IQ slices in device recognition, emphasizing that transmitter-specific features are more distinct in the time-frequency domain than in the time domain. At full resolution, channel independent spectrograms yield such reliable results that the Next-Day Effect -a common challenge in temporal data consistency- is nearly resolved.

While larger packet sizes generally improve training and inference accuracy, ResNet models can perform reasonably well with less than full-resolution packets, further demonstrating their adaptability and efficiency. As a result, ResNet could be a practical and effective solution for real-world RFF extraction. Moreover, the classification of extracted RFFs is shown to scale well and be robust to the Next-Day Effect.

Although Transformers and LSTMs currently lag behind in performance, their potential can be realized by constraining inputs to large packet spectrograms and incorporating substantial hardware upgrades to handle the computational complexity required for efficient training and inference. These findings underscore the critical role of input types and DL architectural efficiency in achieving high-performing, scalable solutions for 5G device authentication.

B. Future Directions

This paper utilized channel-independent spectrograms as the sole visual representation of raw IQ data. Future work could explore alternative visual representations, such as constellation plots, which might better capture hardware impairments and transmitter characteristics. Such representations could enhance the discriminatory power of the models and improve overall performance of device identification.

In this experiment, the largest packet size corresponded to the full length of a single data frame. Future experiments could investigate the impact of varying packet sizes, such as multiples of the data frame length (e.g., 1(data frame), 2(data frame), 3(data frame), etc.) to evaluate the trade-offs between accuracy, computational efficiency, and data resolution.

Further development of the architectures used in this study is also warranted. More complex configurations, such as deeper architectures with additional layers, attention heads, or other modifications, could be evaluated to improve performance. Additionally, experimenting with alternative classifiers beyond kNN may offer valuable insights and potentially better results for certain applications.

Finally, a key avenue for future exploration is the adaptation of this methodology as an authentication protocol or as part of an app in the Open RAN stack [24]. This adaptation would require ensuring that the process meets stringent time constraints while maintaining the high accuracy and scalability demonstrated in this study [27]. A potential experimental setup to evaluate the practicality of this method involves deploying an Open RAN gNB with custom applications and interfaces to authenticate the physical identity of multiple 5G UEs (instead of SDRs with simulated 5G signals) using RFF. Such research could pave the way for practical deployment of RFF in real-world 5G and FutureG network configurations.

IX. Acknowledgment

The authors thank the POWDER Platform for publicly providing their datasets, enabling this study. Furthermore, the authors acknowledge the use of Open AI's ChatGPT 4o in generating text used to aid content drafting for this paper based on the authors' own experimental findings and analysis.

REFERENCES

[1] 3GPP, "5G System security architecture," 3GPP TS 33.501 V19.0.0, Release 19, 2024.

[2] N. Ludant, P. Robyns, and G. Noubir, "From 5G sniffing to harvesting leakages from privacy-preserving messengers," in *Proc. IEEE Symp. Security Privacy (SP)*, May 2024.

[3] S. Chen *et al.*, "Deep RF fingerprinting for secure device authentication," in *Proc. IEEE INFOCOM Workshops*, Apr. 2019, pp. 976–981.

[4] W. Li, N. Wang, L. Jiao, and K. Zeng, "Physical Layer Spoofing Attack Detection in MmWave Massive MIMO 5G Networks," *IEEE Access*, vol. 9, pp. 60419–60430, 2021, doi: 10.1109/ACCESS.2021.3073115.

[5] M. Lichtman, R. Rao, V. Marojevic, J. Reed, and R. P. Jover, "5G NR Jamming, Spoofing, and Sniffing: Threat Assessment and Mitigation," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, Kansas City, MO, USA, 2018, pp. 1–6, doi: 10.1109/ICCW.2018.8403769.

[6] R. Reising, M. A. Temple, and J. A. Jackson, "Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1180–1192, Jun. 2015.

[7] B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Computing Surveys*, vol. 45, no. 1, Article 6, pp. 1–29, Dec. 2012, doi: 10.1145/2379776.2379782.

[8] T. J. O'Shea *et al.*, "Radio machine learning dataset generation with GNU radio," in *Proc. GNU Radio Conf.*, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770–778, Jun. 2016.

[10] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] G. Shen, J. Zhang, A. Marshall, M. Valkama, and J. R. Cavallaro, "Toward Length-Versatile and Noise-Robust Radio Frequency Fingerprint Identification," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2355–2367, 2023, doi: 10.1109/TIFS.2023.3266626.

[13] G. Shen, J. Zhang, X. Wang, and S. Mao, "Federated Radio Frequency Fingerprint Identification Powered by Unsupervised Contrastive Learning," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 9204–9215, 2024, doi: 10.1109/TIFS.2024.3469820.

[14] G. Shen, J. Zhang, A. Marshall, M. Valkama, and J. Cavallaro, "Radio Frequency Fingerprint Identification for Security in Low-Cost IoT Devices," *arXiv preprint arXiv:2111.14275*, Nov. 2021.

[15] H. Li, K. Gupta, C. Wang, N. Ghose, and B. Wang, "RadioNet: Robust Deep-Learning Based Radio Fingerprinting," in *2022 IEEE Conference on Communications and Network Security (CNS)*, pp. 190–198, 2022, doi: 10.1109/CNS56114.2022.9947255.

[16] S. Hanna, S. Karunaratne, and D. Cabric, "Deep Learning Approaches for Open Set Wireless Transmitter Authorization," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2020, doi: 10.1109/SPAWC48557.2020.9154295.

[17] 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. 165–176, Mar. 2020, doi: 10.1109/TCCN.2019.2949308.

[18] A. Al-Shawabka *et al.*, "Exposing the Fingerprint: Dissecting the Impact of the Wireless Channel on Radio Fingerprinting," *Institute for the Wireless Internet of Things, Northeastern University*, Boston, MA, USA, 2024.

[19] N. Hou, Y. Cheng, X. Ji, and W. Xu, "Towards Safer Connections: Secure Authentication in 5G Networks Leveraging Radio Frequency Fingerprinting," in *2024 9th International Conference on Computer and Communication Systems (ICCCS)*, pp. 277–284, 2024, doi: 10.1109/ICCCS61882.2024.10603286.

[20] H. Fu, H. Dong, J. Yin, and L. Peng, "Radio Frequency Fingerprint Identification for 5G Mobile Devices Using DCTF and Deep Learning," *Entropy*, vol. 26, no. 1, p. 38, Dec. 2024, doi: 10.3390/e26010038.

[21] G. Reus-Muns, D. Jaisinghani, K. Sankhe, and K. R. Chowdhury, "Trust in 5G Open RANs through Machine Learning: RF Fingerprinting on the POWDER PAWR Platform," in *Proceedings of the 2020 IEEE Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6, doi: 10.1109/GLOBECOM42002.2020.9348261.

[22] S. AlHazbi, S. Sciancalepore, and G. Oligeri, "The Day-After-Tomorrow: On the Performance of Radio Fingerprinting Over Time," in *Proceedings of the Annual Computer Security Applications Conference (ACSAC '23)*, Austin, TX, USA, Dec. 4–8, 2023, pp. 1–12, doi: 10.1145/3627106.3627192.

[23] 3GPP, "NR; Physical channels and modulation (Release 15)," 3GPP TS 38.211 V18.4.0, Sep. 2024.

[24] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "dApps: Distributed Applications for Real-time Inference and Control in O-RAN," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, Oct. 2021.

[25] 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, doi: 10.1109/TIFS.2022.3152404.

[26] Xie et al., "A Survey of Physical-Layer Authentication in Wireless Communications," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 282–315, First Quarter 2021.

[27] S. Al-Hazbi, A. Hussain, S. Sciancalepore, G. Oligeri, and P. Papadimitratos, "Radio Frequency Fingerprinting via Deep Learning: Challenges and Opportunities," in *Proc. IWCMC*, 2024.

[28] Y. Lin, H. Wang, and H. Zha, "The technology of radio frequency fingerprint identification based on deep learning for 5G application," *Security and Safety*, vol. 3, 2024, Art. no. 2023026, doi: 10.1051/sands/2023026.