Adversarial learning attack on the CBRS Spectrum Sharing system

James Onyejizu, Chibuikem Ezemaduka
Department of Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY, 12180
{onyejj, ezemac}@rpi.edu

Abstract

The Citizens Broadband Radio Service (CBRS) represents a paradigm shift in spectrum sharing, offering enhanced bandwidth and connectivity for next-generation wireless devices through the utilization of the underexploited 3.5 GHz frequency band. This band, historically reserved for US government radar systems (primary users), is now accessible to commercial wireless operators(secondry users) in the absence of primary users. The smooth operation of CBRS depends heavily on its precision in detecting radar system activity, where deep learning classifiers have shown promising results. Despite considerable advancements in classification performance, the susceptibility of these systems to adversarial learning attacks remains under-investigated. This study addresses this gap by implementing and evaluating the potential of adversarial attacks to compromise the CBRS radar detection system. Such attacks could render the classifier oblivious to radar operations, precipitating a transmission collision and consequent system breakdown. Our findings reveal the limitations of adversarial influence, with attacks achieving efficacy predominantly at low Signal to Noise Ratios (SNR). While classifier accuracy remains stable above 10 dB SNR, there is a notable degradation in model confidence up to 40 %, underscoring the need for enhanced defense strategies against adversarial attacks in CBRS operations.

1 Introduction

Given the widespread of wireless communication systems, the efficient allocation of the radio frequency spectrum has become a matter of great importance.The traditional wireless spectrum allocation struggles to accommodate the explosive growth of data-hungry mobile devices, leading to what is known as the artificial spectrum scarcity problem. This issue arises from the limitations of traditional spectrum leasing, where a frequency band is auctioned to an operator for a fixed period. As a result, even though bands may remain idle when their primary owner is not active, secondary wireless devices are not allowed to use the band for their services. A proposed solution to this challenge is spectrum sharing or dynamic spectrum access, where secondary users (SUs) are permitted to utilize frequency bands primarily owned by the original spectrum owners, termed primary users (PUs), provided their transmissions do not cause harmful interference [1]. The Citizens Broadband Radio Service (CBRS) emerged as an innovative framework for spectrum sharing, promising to optimize the use of the 3.5 GHz band—a precious and previously underutilized service. Notably, the CBRS framework is predicated on the accurate detection of PU activity to prevent disruptive interference. State-of-the-art deep learning (DL) classifiers are at the vanguard of this detection process, offering high accuracy in discerning the presence of radar signals. Substantial research has been devoted to enhancing these classifiers. Currently, however there exists a paucity of exploration into their resilience against adversarial learning attacks. Such an attack could manipulate the classifier

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

to overlook radar activity, potentially leading to transmission collisions and consequent failures within the CBRS system. Adversarial learning, a concept originally developed within the domain of artificial intelligence, has found a widespread application in the domain of wireless communications especially in systems security. The CBRS spectrum sharing system, reliant on sophisticated sensing and decision-making algorithms, is not immune to such subversions. These adversarial attacks do not merely represent a potential nuisance; they pose a real and present danger to the integrity of communications and the efficacy of dynamic spectrum access. As such, there exists a need to fill this research void by investigating the feasibility of adversarial attacks on these DL classifiers. By implementing such attacks, we aim to evaluate their impact on the CBRS radar classifier system comprehensively. To the best of our knowledge, this work provides the first investigation into the feasibility of adversarial attacks on deep learning classifiers used for radar detection.

2 Background

CBRS marked a revolutionary stride in the utilization of the 3.5 GHz radio frequency band in the United States. Prioritizing adaptability and efficiency, the CBRS framework has been a catalyst for wireless innovation. It operates on a three-tiered access system designed to accommodate a diverse array of users—from federal incumbents to small private enterprises—thereby democratizing the availability of spectrum and fostering a level playing ground for new wireless technologies.US government radar systems - PUs share its frequency band with commercial wireless operators (Secondary user).However, SUs can only use the frequency channel when the radar system is inactive. This operational paradigm is illustrated in Fig. 1, where the classifier plays a crucial role in detecting the activity status of the radar, ensuring harmonious spectrum sharing between the PUs and SUs (LTE transmiter).

Spectrum sharing in CBRS works on the principle of dynamic access, where users must seamlessly enter and exit the spectrum without disrupting incumbent operations. To manage this, Spectrum Access Systems (SAS) and Environmental Sensing Capability (ESC) networks employ advanced accurate sensing mechanisms to predict usage patterns and detect the presence of primary users.

Adversarial learning exploits the adaptability of learning algorithms to undermine their performance, turning strength into vulnerability. It involves manipulating the input to a machine learning system in such a way that it makes a mistake—like a chess grandmaster misled by a novice's unconventional move. These attacks can be particularly insidious as they often require minimal but precisely targeted alterations to be effective.

Within the context of CBRS, the sophistication of machine learning-based spectrum management systems opens up the potential for adversarial attacks. An adversary, by understanding the learning models that govern SAS and ESC, could devise inputs that mimic incumbent signals or otherwise falsely trigger the protective mechanisms of the spectrum sharing system. Such actions could result in denial of service, unauthorized spectrum access, or other forms of malicious interference, which in turn could undermine trust in the entire CBRS framework.

3 Related Work

Spectrum sharing has been extensively studied as a strategy to mitigate artificial spectrum scarcity by allowing more flexible use of available spectral resources. A prime example of this approach in the United States is the CBRS, which introduces a model for sharing the 3.5GHz band, historically reserved for government radar systems, with commercial wireless operators. This innovative system allows operators to utilize the band when the radar system is not active, contingent upon the accurate detection of the radar's activity status to prevent harmful interference [2].

Recent advancements in this area have focused on the development of deep learning-based classifiers designed to accurately detect the presence of radar signals in measurement samples collected from monitoring sensors. These sensors periodically survey the wireless environment, gathering data on active transmissions within the spectrum area. The effectiveness of these classifiers is critical to the success of systems like CBRS, as they enable the dynamic allocation of spectrum resources based on real-time usage and radar activity [3], [4].

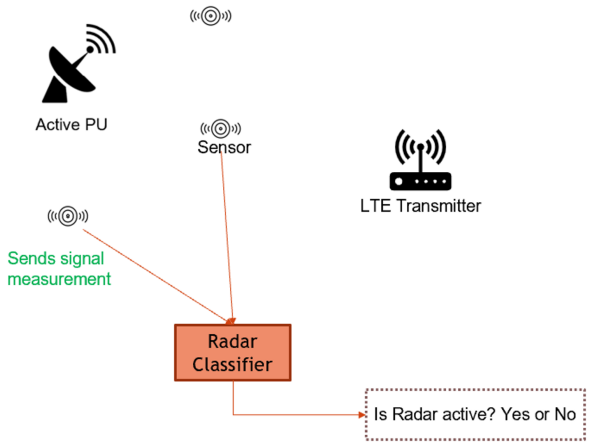

Figure 1: Schematic representation of CBRS spectrum sharing: the primary user (radar system) is active, while the secondary user (LTE transmitter) remains idle. The diagram also depicts the deployment of a classifier tasked with identifying radar presence.

Further examination reveals notable related works [6], [7] that have identified possible attack surfaces against communication systems through adversarial learning techniques. Particularly, work [6] outlines potential adversarial attacks on spectrum sharing within 5G systems, including scenarios that closely mirror the CBRS system. Both our study and [6] explore the impact of adversarial attacks on spectrum sharing, specifically focusing on deceiving classifiers about radar presence, which is crucial for managing access to the spectrum.

While there are similarities in examining the vulnerabilities of spectrum sharing to adversarial attacks, our work diverges significantly from [6] in depth and application. Unlike [6], which primarily outlines potential attack scenarios without delving into practical implementation, our research extends to the actual execution of these attacks, evaluating their effectiveness, and understanding their limitations within the CBRS framework. This hands-on approach offers a more granular understanding of the threats and defenses within spectrum sharing systems.

Additionally, while [7] investigates adversarial attacks in a different context—Wi-Fi-based behavior recognition systems, referred to as CBRS—these studies underscore the broad applicability and varying objectives of adversarial learning in wireless systems, which differ significantly from our focus on spectrum sharing's radar detection.

In summary, the closest related work to our project is [6], from which our practical, implementation-focused approach marks a significant advancement in the field. This distinction highlights the novelty of our research and its contribution to enhancing the security and reliability of spectrum sharing technologies.

Despite the progress in enhancing classifier accuracy, there remains a significant gap in the literature regarding the robustness of these systems against adversarial machine learning attacks. Such attacks could potentially deceive the classifiers into incorrectly identifying the presence of radar signals, leading to unauthorized use of the spectrum and consequent interference. While adversarial attacks have been explored in various contexts within wireless communications and machine learning, their impact on real-world spectrum sharing systems, particularly those employing deep learning techniques like the CBRS, has yet to be thoroughly investigated.

4 Method

We motivate the results of our experiments with a detailed description of the proposed attack model and the method used for our adversarial attack.

4.1 Attack Model

In our model, a primary user (PU), such as a radar system, periodically transmits in its designated frequency band. A secondary user (SU), like an LTE base station, is allowed to utilize this band only when the radar is inactive. To ascertain the radar's activity, we employ a sensor tasked with monitoring the spectral environment. This sensor measures active transmissions and converts these into spectrograms, which are then processed by a trained classifier to detect radar signals. The classifier's role is critical—if it detects radar presence, the SU must suspend its transmission to avoid interference. Should the classifier not detect radar, the SU may continue its operation.

We propose an attack where an intelligent adversary, aware of the classifier's parameters and the training data and introduces a stealthy adversarial signal into the environment. This adversarial signal is designed to blend with the LTE and radar signals in a manner that produces a misleading spectrogram, causing the classifier to miss the radar's presence and allowing the SU to transmit inadvertently during radar activity, as illustrated in Fig. 2.

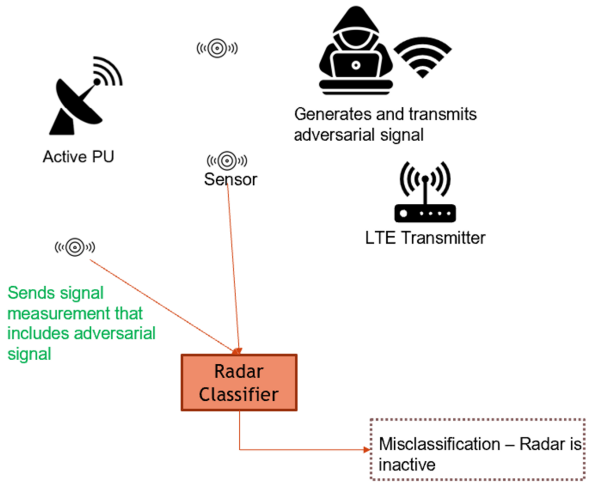

Figure 2: Illustration of adversarial interference in the CBRS spectrum sharing context. The adversarial signal is crafted to ensure the classifier erroneously interprets an active radar signal as absent.

The adversary's signal is carefully calibrated to maintain a low signal-to-noise ratio (SNR), avoiding detection by the sensor and not disrupting the SU's function. Let x_1 and x_2 denote the signals received at the sensor from the LTE transmitter and radar, respectively. Let g_1, g_2, and g_3 represent the channel gains from the LTE transmitter, radar, and adversary transmitter to the sensor, respectively, and let n symbolize the receiver noise. The combined signal measured when the attacker transmits is given by:

$$x' = x_1 g_1 + x_2 g_2 + b g_3 + n, \tag{1}$$

where b is the adversarial signal (perturbation). The task of generating the adversary's signal can be formulated as solving the optimization problem:

$$\max_b L\left(x', y, \theta\right) \quad s.t. \quad \min \|b\| \leq \varepsilon, \tag{2}$$

where y is the true label, ε is the upper bound on the adversarial signal's power (corresponding to the added noise power in this context), θ denotes the trained classifier weights, and $L()$ is the loss function quantifying the discrepancy between the classifier's output and y.

4.2 Adversarial Attack Implementation

Our implementation of an adversarial attack on the CBRS leverages the Basic Iterative Method (BIM) [8] to craft inputs that deceive the radar signal classifier. By iteratively modifying an input sample

x', the BIM approach generates an adversarial example capable of leading the classifier to a false negative result.

The adversarial sample is updated as follows:

$$x' = x' - \alpha \cdot sign\left(\frac{\partial L(x', T)}{\partial x'}\right), \tag{3}$$

with x' as the modified input, α as a predefined step size of 0.2, L as the loss function, and T indicating the target class 'no radar'. The iterative modifications halt upon meeting stopping criteria, either reaching 20 iterations or when the adversarial perturbation exceeds the limit ϵ. As a prospective avenue for further research, we aim to explore and assess the performance of various other adversarial attack methods within the context of the CBRS spectrum sharing system.

5 Experimental Validation

In addition to the aforementioned attack model, we provide a detailed description of the spectrum area simulation settings and proceed to analyze the outcomes subsequent to the implementation of our proposed approach.

5.1 Simulation Settings

We evaluate the efficacy of attack model using a comprehensive simulation framework developed in MATLAB [9]. The simulation entailed the generation of LTE signals through MATLAB's LTE Toolbox [10], with transmission parameters randomized to reflect a realistic radio environment. Additionally, radar signals were synthesized using the National Institute of Standards and Technology (NIST) signal generation toolbox, ensuring adherence to standard radar signal characteristics.We summarize the different radal signal characteristics[11] on Table 1. Also, we provide a pictorial representation of our setup in Fig. 3. To facilitate the detection and classification of these signals,

Table 1: Radar Signal Characteristics

Radar Type	**Pulse Width (μs)**	**PRR (Hz)**	**Chirp Width (MHz)**	**Pulses per Burst**	**Burst Length (ms)**
PoN1	0.5–2.5	900–1100	NA	15–40	13–44
PoN2	13–52	300–3000	NA	5–20	1–66
Q3N1	3–5	300–3000	50–100	8–24	2–80
Q3N2	10–30	300–3000	1–10	2–8	0.6–26
Q3N3	50–100	300–3000	50–100	8–24	2–80

we generate a spectrogram at the sensor node for each configuration of the signal settings. This spectrogram serves as the input to the subsequent machine learning model. The radar detection algorithm was implemented using TensorFlow. We trained our model on a dataset comprising 1000 samples of spectrograms, which were labeled according to the presence of a radar signal—'1' indicating the presence and '0' signifying the absence. Figs. 4 and 5 depict the LTE-only and the combined LTE-radar spectrograms, respectively.

5.2 Results

We employ evaluation metrics such as classification accuracy and confidence difference to validate our results. Figs. 6 and 7 display the accuracy results for two different radar configuration settings. It is observed that at an epsilon (ϵ) value of -120 dB, the adversarial attack is effective only at a low Signal to Noise Ratio (SNR) of 10 dB. For each SNR value, we evaluated 10 samples.

In Figs. 8 and 9, we show results for the confidence interval. The confidence difference represents the percentage change in the model's classification confidence between the original and the adversarial signal. Notably, while the accuracy remains stable at SNRs greater than 10 dB, the model's confidence decreases up to 40 % . Further, Fig. 11 and 10 illustrate that at a less stringent epsilon constraint of -90 dB, the model's accuracy begins to decline at higher SNRs, reaching up to 30 dB. This implies

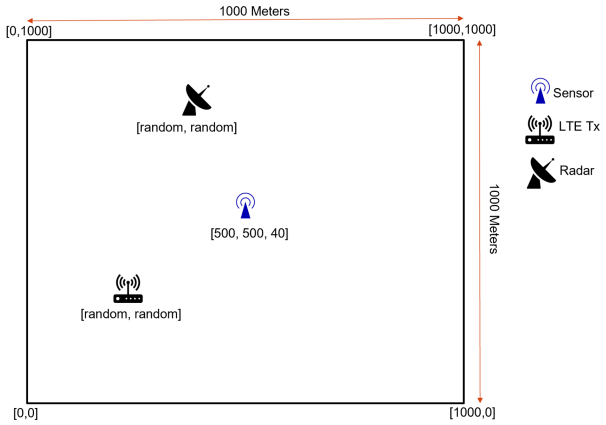

Figure 3: Spectrum area simulation setting

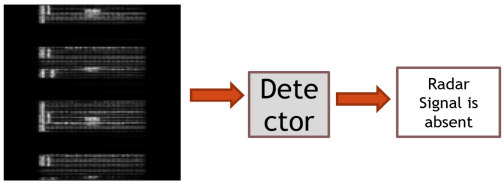

Figure 4: LTE only Spectrogram

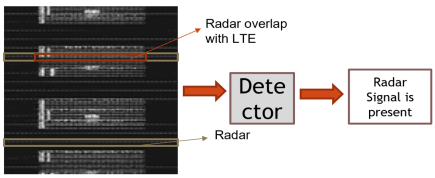

Figure 5: LTE + Radar Spectrogram

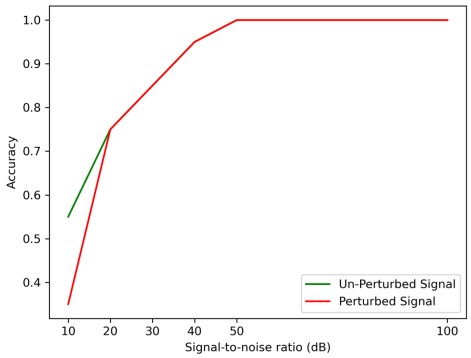

Figure 6: Accuracy results for the PON1 radar configuration.

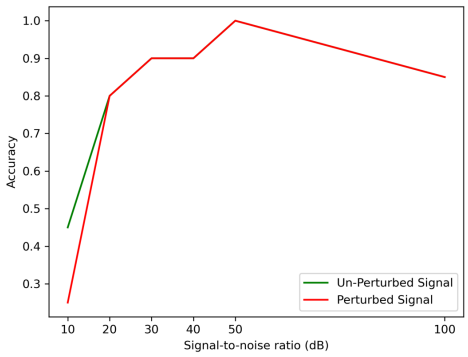

Figure 7: Accuracy results for the Q3N1 radar configuration.

Comparative accuracy analysis of different radar configurations under adversarial attack conditions

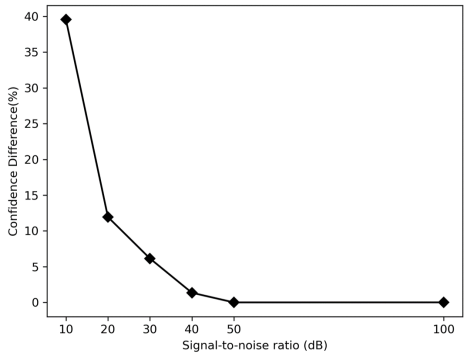

Figure 8: Confidence difference for the PON1 radar after adversarial manipulation.

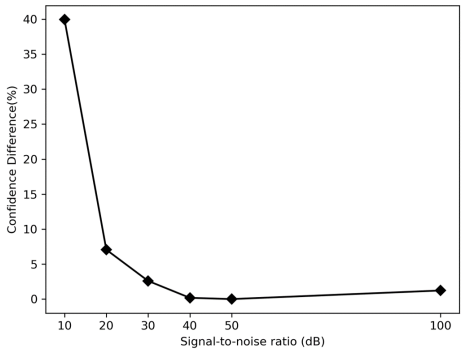

Figure 9: Confidence difference for the Q3N1 radar after adversarial manipulation.

Impact of adversarial attacks on the confidence levels of radar signal classification for different configurations.

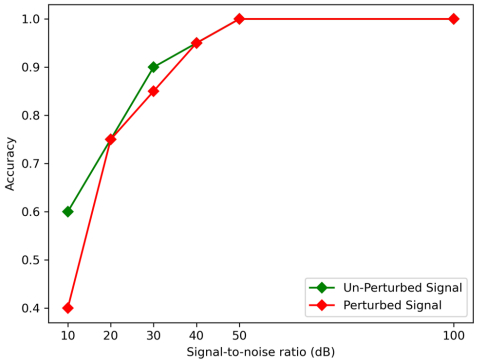

Figure 10: Model accuracy under a less stringent epsilon constraint at different SNRs for the PON1 configuration.

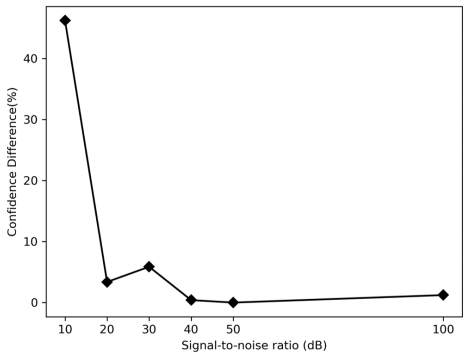

Figure 11: Model confidence variation under different SNR conditions for the PON1 configuration.

Effects of relaxed adversarial constraints(ϵ)=90dB) on accuracy and confidence across different SNR thresholds.

that at a less stringent constraint on the signal power of the adversarial signal, the attack's efficacy could improve, even being effective at higher SNR values.

6 Discussion

This project represents a step in understanding the vulnerabilities of the Citizens Broadband Radio Service (CBRS) to adversarial attacks. By exploiting the precision requirements of deep learning classifiers used in the detection of radar signals, we have demonstrated that these systems are particularly vulnerable to adversarial interference at low Signal to Noise Ratios (SNR). Although the effectiveness of the attacks diminishes with higher SNRs, the notable reduction in model confidence even up to 40 % highlights critical vulnerabilities. This discovery not only extends the current knowledge on the resilience of spectrum sharing systems but also emphasizes the urgency for developing robust adversarial defense mechanisms.

Moving forward, the project will delve into more complex scenarios to mirror real-world conditions more closely. The implementation of state-of-the-art (SOTA) radar classifiers[11],[12],[13], despite the lack of accessible source codes, remains a pivotal aspect of our future work. By reconstructing these models, we aim to provide a more thorough evaluation of adversarial tactics under varied operational settings. Moreover, leveraging real-life wireless data will allow us to test the durability of these classifiers against adversarial attacks in more dynamic and unpredictable environments. This

approach will likely yield insights that are crucial for refining the strategies used to secure spectrum sharing technologies.

Additionally, the project plans to address the challenge of misleading multiple sensors simultaneously within a CBRS scenario. This involves not only deceiving individual sensors but carrying out a coordinated attack that affects all sensors in the network concurrently, hence mimicking a sophisticated and realistic threat setting. Developing adversarial examples that can consistently fool multiple detection systems across different settings and environmental conditions will necessitate innovative algorithmic approaches and a deep understanding of signal propagation dynamics. Such efforts are expected to culminate in strategies that enhance the collective security framework of distributed sensor networks, thereby fortifying the overall integrity of the CBRS system.

7 Implementation Plan

Below is a table that outlines our project timeline and delineates responsibilities for each segment of the project.

Table 2: Implementation Plan

Task Description	**Done by**
Conceptualization and design of the CBRS spectrum sharing simulation environment using MATLAB.	James, Chibuikem
Data curation and development/training of a classifier for detecting the presence of radar signals in sensor measurements.	James, Chibuikem
Simulation of the adversarial attack based on the methods described in the report.	James, Chibuikem
Comprehensive evaluation of the attack's efficacy, focusing on compromised classifier accuracy, attack complexity, and other significant insights.	James, Chibuikem
Compilation and writing of the project report, detailing the methodologies, findings, and conclusions.	James, Chibuikem

8 References

1. Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," IEEE Signal Processing Magazine, vol. 24, no. 3, pp. 79–89, 2007.
2. Federal Communications Commission (FCC), "Amendment of the commission's rules with regard to commercial operations in the 3550-3650 mhz band," Report and order and second further notice of proposed rulemaking, 2015.
3. N. Soltani, V. Chaudhary, D. Roy, and K. Chowdhury, "Finding Waldo in the CBRS band: Signal detection and localization in the 3.5 GHz spectrum," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, 2022, pp. 4570–4575.
4. Shamik Sarkar, Milind Buddhikot, Aniqua Baset, and Sneha Kumar Kasera. 2021. DeepRadar: a deep-learning-based environmental sensing capability sensor design for CBRS. In Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom '21). Association for Computing Machinery, New York, NY, USA, 56–68.
5. D. Adesina, C. -C. Hsieh, Y. E. Sagduyu and L. Qian, "Adversarial Machine Learning in Wireless Communications Using RF Data: A Review," in IEEE Communications Surveys and Tutorials, vol. 25, no. 1, pp. 77-100, Firstquarter 2023, doi: 10.1109/COMST.2022.3205184.
6. Sagduyu, Yalin E., Tugba Erpek, and Yi Shi. "Adversarial machine learning for 5G communications security." Game Theory and Machine Learning for Cyber Security (2021): 270-288.

7. Liu, Jianwei, et al. "Physical-world attack towards wifi-based behavior recognition." IEEE INFOCOM 2022-IEEE Conference on Computer Communications. IEEE, 2022.
8. Kurakin, Alexey, Ian J. Goodfellow, and Samy Bengio. "Adversarial examples in the physical world." Artificial intelligence safety and security. Chapman and Hall/CRC, 2018. 99-112.
9. The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.
10. The MathWorks Inc. (2022). LTE Toolbox version: 9.4 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com
11. S. Sarkar, M. Buddhikot, A. Baset, and S. K. Kasera, "Deepradar: a deep-learning-based environmental sensing capability sensor design for cbrs," in Proceedings of the 27th Annual International Conference on Mobile Computing and Networking, pp. 56–68, 2021.
12. Soltani, N., Chaudhary, V., Roy, D., Chowdhury, K. (2022, December). Finding waldo in the cbrs band: Signal detection and localization in the 3.5 ghz spectrum. In GLOBECOM 2022-2022 IEEE Global Communications Conference (pp. 4570-4575). IEEE.
13. Caromi, R., Lackpour, A., Kallas, K., Nguyen, T., Souryal, M. (2021, December). Deep learning for radar signal detection in the 3.5 GHz CBRS band. In 2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN) (pp. 1-8). IEEE.