Anti-Sensing: Defense against Unauthorized Radar-based Human Vital Sign Sensing with Physically Realizable Wearable Oscillators

Md Farhan Tasnim Oshim[1], Nigel Doering[2], Bashima Islam[3], Tsui-Wei Weng[2], and Tauhidur Rahman[2]

***Abstract*— Recent advancements in Ultra-Wideband (UWB) radar technology have enabled contactless, non-line-of-sight vital sign monitoring, making it a valuable tool for healthcare. However, UWB radar's ability to capture sensitive physiological data, even through walls, raises significant privacy concerns, particularly in human-robot interactions and autonomous systems that rely on radar for sensing human presence and physiological functions. In this paper, we present Anti-Sensing, a novel defense mechanism designed to prevent unauthorized radar-based sensing. Our approach introduces physically realizable perturbations, such as oscillatory motion from wearable devices, to disrupt radar sensing by mimicking natural cardiac motion, thereby misleading heart rate (HR) estimations. We develop a gradient-based algorithm to optimize the frequency and spatial amplitude of these oscillations for maximal disruption while ensuring physiological plausibility. Through both simulations and real-world experiments with radar data and neural network-based HR sensing models, we demonstrate the effectiveness of Anti-Sensing in significantly degrading model accuracy, offering a practical solution for privacy preservation.**

I. Introduction

Recent advancements in contactless sensing technologies, particularly using Ultra-Wideband (UWB) radar, have enabled various applications such as vital signs monitoring [1], [2], [3], [4], [5], [6] and gesture recognition [7], [8], [9] without physical contact. These technologies offer convenience and efficiency but also raise significant privacy concerns due to their ability to capture sensitive personal information in public spaces without consent. Radar's high penetration capability allows it to sense through walls, posing a potential threat to privacy. In various settings, from homes to public spaces like transportation stops and waiting rooms, it can expose individuals' vital signs, gestures, and behavioral data, leading to the possible misuse of personal information. Unauthorized estimation of vital signs, particularly heart rate, poses significant privacy concerns, as it can disclose sensitive health information, such as stress levels, without consent, leading to discrimination, health profiling, and surveillance, ultimately infringing on personal autonomy and security.

*This work is in part supported by the National Science Foundation under grant 2320678 (PI Rahman) and start up grant support from Halıcıoğlu Data Science Institute – UC San Diego, and Manning College of Information & Computer Sciences – UMass Amherst.

[1]Md Farhan Tasnim Oshim is with Manning College of Information and Computer Sciences, University of Massachusetts Amherst, MA, USA farhanoshim@cs.umass.edu

[2]Nigel Doering, Lily Weng, and Tauhidur Rahman are with Halıcıoğlu Data Science Institute, University of California San Diego, CA, USA nfdoerin@ucsd.edu, lweng@ucsd.edu, trahman@ucsd.edu

[3]Bashima Islam is with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, MA, USA bislam@wpi.edu

While the benefits of contactless radar sensing are evident, the security and privacy implications remain underexplored. Adversarial attacks, a growing concern in machine learning and computer vision, pose a novel threat to radar-based sensing. Adversarial attacks typically employ white-box methods such as FGSM [10], IFGSM [11], PGD [12], JSMA [13], DeepFool [14] , and C&W [15], or black-box techniques like ZOO [16], GenAttack [17], and Boundary Attacks [18], [19], [20]. Although most attacks have focused on the digital domain, the exploration of perturbations in the physical domain to deceive such systems remains limited. While some research exists on physical-world attacks in the context of computer vision [21], [22], [23], [24], the radar system domain has not been similarly investigated. In particular, the susceptibility of radar data to physically realizable perturbations, such as deliberate, imperceptible modifications to sensor inputs, has not been adequately addressed. These physical perturbations can serve as a defense mechanism for our setting, where they could be used to mislead radar-based sensing systems, resulting in inaccurate vital sign estimations, thereby protecting an individual's privacy.

This paper proposes a novel physically realizable perturbation technique, Anti-Sensing, a pipeline designed to effectively disrupt and deceive unauthorized radar sensing models. Our approach utilizes physical perturbations in the form of oscillating devices, for example, motors as a defense against unauthorized vital sign monitoring, heart rate (HR) in particular. The oscillating frequencies and span of these devices are optimized using a gradient-based defense algorithm that we designed to ensure maximum loss between model predictions and the ground truths. These perturbations mimic legitimate cardiac motion and sufficient noise components, thereby misleading radar-based recognition models into producing inaccurate heart rate predictions. By strategically introducing oscillatory signals that simulate natural human heart rate, our method aims to protect individuals' privacy by thwarting unauthorized radar sensing attempts. We validate our system's success through simulated perturbations on a real dataset and neural network architectures, and subsequently validate it using real collected data with physical devices. Through experiments and analysis, we demonstrate that our proposed perturbation method can effectively deceive radar sensors, making it a practical solution for safeguarding personal privacy in public and private settings. The key contributions of our work can be summarized as follows:

- We present Anti-Sensing, a novel perturbation technique that uses oscillatory motions optimized by an

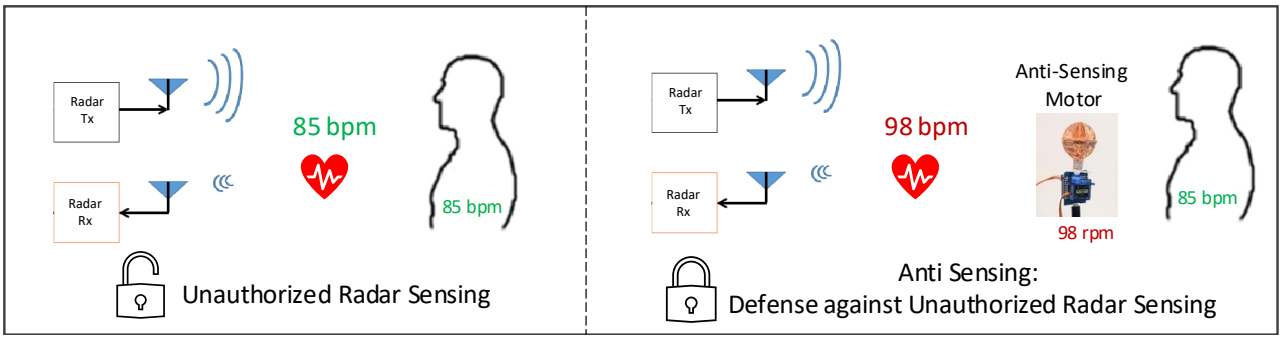


Fig. 1: Anti-Sensing Scenario: The left panel shows a radar system detecting an individual's heart rate (85 bpm) without permission. In contrast, the right panel illustrates our anti-sensing solution, where a motor generates a false heart rate signal (98 bpm), effectively blocking the radar from sensing the person's true heart rate (85 bpm) and ensuring privacy.

appropriate gradient-based attack objective to deceive radar sensing models for regression tasks. This approach introduces a new paradigm in adversarial defense by creating physical perturbations that mimic legitimate human heart rates.

- We demonstrate that our attacks are physically realizable using oscillating devices, which can be easily worn on the wrist. We use a programmable servo motor to generate variable frequency motion for defense against unauthorized heart rate estimations. These devices allow for precise, reproducible physical perturbations that can effectively counteract radar-based sensing mechanisms.
- We validate our anti-sensing approach through simulated perturbations applied to a real dataset collected at a sleep clinic and to state-of-the-art neural network-based radar sensing models, followed by testing with data collected using the anti-sensing oscillating device. Our results show that the proposed perturbations are effective in disrupting unauthorized radar sensing models, offering a practical solution for safeguarding individuals' heart rate information.

II. Related Works

Most adversarial perturbations against contactless sensing have primarily focused on the digital domain, involving white-box or black-box attacks on the collected data. For example, Xie et al. [25] demonstrated universal targeted adversarial attacks in the digital domain against mmWave-based Human Activity Recognition (HAR), effectively deceiving different models, including voxel-based and heatmap-based, while remaining entirely in the digital domain. Similarly, Ozbulak et al. [26] showed the vulnerability of radar-based CNNs for human activity recognition to both white-box and black-box adversarial attacks, revealing that even minimal perturbations, such as those applied only to input padding, can significantly alter model predictions.

Staat et al. [27] proposed IR-Shield, a countermeasure using intelligent reflecting surfaces (IRSs) to obfuscate wireless channels, achieving detection rates of 5% or less in advanced Wi-Fi-based human motion attacks. RF-Protect [28] presents a hardware reflector coupled with a generative mechanism to produce realistic human trajectories aimed at enhancing privacy by introducing artificial human reflections into FMCW radar data to guard against unauthorized through-wall monitoring. However, these existing approaches are ineffective against attacking UWB radar-based sensing, particularly for vital signs, as UWB relies on precise time-of-flight measurements, and unlike FMCW or mmWave systems, delay and frequency in UWB systems are uncorrelated.

III. Background of Radar-based Sensing

Ultra-wideband (UWB) radar is a non-invasive sensing technology that emits nanosecond-duration electromagnetic pulses across a broad frequency spectrum (3.1–10.6 GHz), enabling precise detection of both macro-scale movements, such as gestures, and micro-scale physiological activities, including breathing and heartbeats. Its wide bandwidth provides high spatial resolution, making it well-suited for capturing detailed motion data. Additionally, UWB radar's broad frequency range and pulse characteristics enable non-line-of-sight sensing, allowing detection through walls and other obstacles due to its superior penetration properties. By analyzing the time delay of reflected signals, it tracks a target's movement, while advanced signal processing removes clutter from static objects, ensuring accurate monitoring in dynamic environments - particularly in healthcare, human-object interaction, and human-robot interaction applications.

A received signal at the UWB receiver can be written as the following equation,

$$r(t) = \sum_{j=1}^{L} a_j s(t - \tau_j) + w(t) \tag{1}$$

Here, $s(t)$ is the transmitted pulse, a_j and τ_j represent the amplitude and the propagation delay of the j^{th} multipath component, $w(t)$ is the additive noise from the channel, and L is the total number of reflected paths. The time delay τ_j, also known as the time of flight (ToF), can be used to calculate the target's distance d_j using the relation $d_j = \frac{c \cdot \tau_j}{2}$, where c is the speed of light. For simplicity, we will use d_j in the equation directly to refer to the target distance range bin.

Synthetic Vital Sign Motion Generation

This subsection demonstrates the process of synthesizing vital sign motion based on ultra-wideband (UWB) radar principles. UWB pulses can be modeled as Gaussian-modulated sinusoidal signals, and a single radar scan of a point target at a fixed range bin d_j can be expressed as:

$$s(t_i, d_j) = \exp\left(-\left(\frac{t_i - d_j}{\omega_0/T_s}\right)^2\right) \cdot \exp\left(i \cdot 2\pi \cdot f_0 \cdot T_s(t_i - d_j)\right) \tag{2}$$

where,

$$d_j = \mathrm{A} \cdot \sin\left(2\pi \frac{f_{\text{osc}}}{60 \cdot F_s} \cdot x_i\right) + \text{offsets}[k] \tag{3}$$

and,

- T_s: Fast-time sampling period
- ω_0: Width of the Gaussian pulse
- f_0: Frequency of the radar pulse in Fast Time
- f_{osc}: Frequency of the sinusoidal motion a.k.a. frequency of the Vital Sign (HR) in rpm
- F_s: Sampling frequency of the radar in Slow Time.
- M: Total number of radar scans
- N: Total number of range bins
- A: Spatial amplitude
- offsets: Target locations
- x_i: Scan indices from 1 to M
- d_j: Range bin positions modulated by the synthetic vital sign motion from 1 to N.

Offsets determine the initial target location. If there are multiple targets, synthetic sinusoids are generated for each of them and then summed up together to form a single radar scan. By stacking the Gaussian-modulated pulses over all scans and offsets, we get a 2D radargram for a specific observation window. The x-axis of the radargram, also known as the fast-time axis, denotes distance or range, while the y-axis, also known as the slow-time axis, indicates time. To demonstrate how closely our simulation matches reality, we present Figure 2, a side-by-side comparison of a pendulum with a metal bob oscillating at 90 rpm (1.5 Hz) positioned at one meter from the radar.

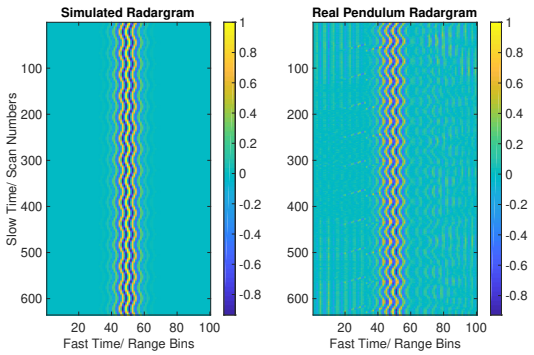


Fig. 2: Comparison of synthetic radargram (left) and real radargram (right) of a point target (a pendulum with a metal bob) oscillating at a frequency of 90 rpm (1.5 Hz).

IV. Proposed Anti-sensing Method:

Our goal is to develop a defense mechanism that deceives radar sensing models by introducing deliberate, physically realizable perturbations to radar inputs. Specifically, we aim to mislead radar systems, such as those used for vital sign estimation (e.g., heart rate), into generating inaccurate predictions, thereby protecting individual privacy. This problem is framed as an attack on radar sensing models to induce erroneous estimations. In our proposed algorithm, we optimize oscillatory motions through a gradient-based attack to deceive the vital sign estimation model. Since the perturbations act as a defense mechanism to safeguard privacy, we refer to our algorithm as Sinusoidal Defense Algorithm, a white-box targeted attack requiring knowledge of the vital sign model, as illustrated in Algorithm 1.

A typical white-box classification adversary with a targeted attack optimizes the following objective, where $h_\theta(\cdot)$ is the trained model and y is the target class, typically different from the true class:

$$\underset{\delta \in \Delta}{\text{minimize}}\, \ell(h_\theta(x + \delta), y) \tag{4}$$

subject to $\Delta = \{\delta : \|\delta\|_\infty \leq \epsilon\}$, where ϵ is the attack budget. However, using machine learning models to monitor contactless vital signs primarily falls under the realm of regression tasks where the labels are continuous values. When dealing with radar signals, it is essential to note that x in Equation 4 represents radargrams rather than images. To generate adversaries for vital sign monitoring, we propose additional conditions that are imposed on the potential perturbation δ. Here are two key design considerations for optimizing perturbations:

- **Localized Perturbations:** Rather than applying perturbations to the entire image or radargram, we advocate adding δ exclusively to the range bins where the target of interest is situated. For instance, if the target resides on the n^{th} range bin, then the vital sign would predominantly span across the neighboring range bins of n. Consequently, perturbations should also be introduced to these pertinent range bins. This design consideration is ensured in the algorithm's process-step 4 through d_j in Equation 6, specifically through the *offsets* parameter in the expression of d_j as shown in Equation 3.
- **Structured Perturbations:** The aforementioned approach of perturbation addition in Equation 4 operates on a per-pixel or per-element basis within the radargram, lacking structure and specificity tailored to deceive vital sign estimation models. As an alternative strategy, we propose the introduction of a periodic signal, specifically a sinusoid, atop the genuine periodic vital sign signature. This augmentation aims to ensure that the perturbation sinusoid is just enough to overwhelm the original vital sign, rendering it challenging for the vital sign estimation model to accurately predict true heart rate. This design consideration is realized in the algorithm introducing sinusoid with frequency f_{osc} in the first term of d_j as shown in Equation 3.

Our proposed Algorithm 1 incorporates the creation of a structured, localized perturbation guided by learned parameters through gradient optimization, reflecting the nature of the vital signature. The following equation represents the proposed defense objective, where y is the target heart rate, typically away from the true heart rate:

$$\underset{\delta \in \Delta}{\text{minimize}}\, \ell(h_\theta(x_{i,j} + \delta_{i,j,k}), y) \tag{5}$$

In our case, $\delta_{i,j,k}$ is a function of optimized frequency and spatial amplitude of the perturbation sinusoid in the radargram, i.e. $\delta_{i,j,k} : f \mapsto f(A_{\text{opt}}, f_{\text{opt}})$

Our attack budget is not defined by a single parameter but rather by a set of constraints. Specifically, it is a combination of several factors where the constraints are represented as $\Delta = \{\delta : \delta_A \leq \epsilon_A, \delta_f \leq \epsilon_f\}$.

Algorithm 1 Sinusoidal Defense Algorithm

Input:

- $h_\theta(\cdot)$: Target (trained) model
- $x \in \mathbb{R}^{M \times N}$: Original radargram
- y: Targeted HR (away from true HR)
- α: Step size
- T: Number of iterations
- $L(\cdot)$: Loss function

Output:

- Optimized frequency f_{opt} and spatial amplitude A_{opt}
- Perturbed radargram $x' \in \mathbb{R}^{M \times N}$

Process:

1: Initialize $x' \leftarrow x$

2: Initialize frequency estimate $f_{\text{opt}} \leftarrow$ random number $\in [f_{\min}, f_{\max}]$

3: **for** $t = 1$ **to** T **do**

4: Generate synthetic radargram perturbation with d_j defined by Equation 3:

$$\delta_{i,j,k}(A_{opt}, f_{\text{opt}}) = \exp\left(-\left(\frac{t_k - d_j}{\omega_0 / T_s}\right)^2\right) \cdot \exp\left(i \cdot 2\pi \cdot f_0 \cdot T_s(t_k - d_j)\right) \tag{6}$$

5: Add the perturbation to the original radargram:

$$x' \leftarrow x + \delta_{i,j}(A_{\text{opt}}, f_{\text{opt}})$$

6: Compute the gradient of the loss w.r.t. f_{opt} and A_{opt}:

$$G_f \leftarrow \nabla_{f_{\text{opt}}} L(h_\theta(x'), y) \;\; \& \;\; G_A \leftarrow \nabla_{A_{\text{opt}}} L(h_\theta(x'), y)$$

7: Update the estimated frequency using gradient descent:

$$f_{\text{opt}} \leftarrow f_{\text{opt}} - \alpha \cdot G_f \;\; \& \;\; A_{\text{opt}} \leftarrow A_{\text{opt}} - \alpha \cdot G_A$$

8: Clip the updated frequency and spatial amplitude to ensure it stays within predefined bounds:

$$f_{\text{opt}} \leftarrow \text{clip}(f_{\text{opt}}, f_{\min}, f_{\max}) \;\&$$

$$A_{\text{opt}} \leftarrow \text{clip}(A_{\text{opt}}, A_{\min}, A_{\max})$$

9: **end for**

10: **return** Optimized f_{opt}, A_{opt}, and x'

A. Constraint on Frequency:

For heart rate estimation, we should ensure that it is recommended that the deviation δ_f from the true frequency be kept within ϵ_{hr}, ensuring that the estimated heart rate falls within the physiological range of 50 to 100 beats per minute, indicative of normal human heart rate. Any data falling outside this range should be classified as noise and disregarded by the system.

B. Constraint on Spatial Amplitude/ Span:

The spatial amplitude of the perturbation signal, also known as the perturbation sinusoid, should be kept within the target's occupancy range bins, such that $\delta_A \leq \epsilon_A$, where $\mid \epsilon_A \mid$ is 25 range bins. This limitation arises from human targets typically occupying an average of 46 cm, which is around 50 range bins at specific time intervals, given the scale of 9 mm per range bin in P440 UWB radar [29].

The algorithm optimizes both frequency f_{opt} and spatial amplitude A_{opt} of the sinusoidal perturbation based on the above constraints to ensure the attack remains within the physiological and spatial limits defined for heart rate estimation and radargram occupancy.

V. Hardware and Measurement Setup

A. UWB Radar Setup

We utilize a monostatic time domain Ultra-WideBand (UWB) Impulse Radar P440 [29] with time windowing capabilities for sensing vital signs. It operates from 3.1 to 4.8 GHz frequency centering at 4.3 GHz.

B. Programmable Servo Motor

As an anti-sensing mechanism, we propose using an off-the-shelf SG90 servo motor programmed via an ESP32 WROOM Mini microcontroller to generate variable frequency motion tailored to specific requirements. By optimizing the perturbation f_{opt} within predefined constraints (normal heart rates range from 50 bpm to 100 bpm), this setup allows for precise, reproducible frequency motion that can effectively counteract radar sensing mechanisms. A 3D-printed octahedral reflector, with a diameter of 4.3 cm and wrapped in copper tape, is attached to the motor to improve reflectivity and enhance the signal-to-noise ratio (SNR). Figure 3 illustrates the setup of the anti-sensing device, which is compact enough to be worn on the wrist.

VI. Results

A. Anti-Sensing on Deep Learning-based Vital Sign Estimation Models

1) Dataset: To validate our proposed anti-sensing defense algorithm, we used a sleep dataset [3] collected in a sleep laboratory. Two participants underwent overnight full polysomnography (Siesta, COMPUMEDICS), which included electrocardiography and respiratory inductance plethysmography to measure chest and abdominal wall motion along with simultaneous contactless UWB radar data collection.

Dataset	Setting	HR Model	MAE (Without Anti-Sensing)	MAE (With Anti-Sensing)
Tasnim et al. [3]	Sleep Lab	ResNet - 18 [30]	2.67 bpm	5.35 bpm
		ResNet - 50 [30]	2.37 bpm	5.63 bpm
		CNN 1D+2D [31]	5.63 bpm	12.91 bpm
		Vision Transformer (ViT) [32]	3.28 bpm	9.35 bpm

TABLE I: HR Sensing Model Performance with and without Anti-Sensing on Sleep Dataset

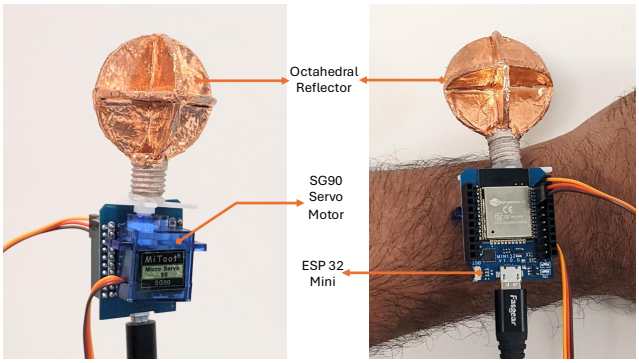


Fig. 3: A programmable servo motor paired with a ESP 32 Mini. A 3D-printed octahedral reflector wrapped with copper tape is attached as a load to the motor to enhance reflectivity and increase the signal-to-noise ratio (SNR).

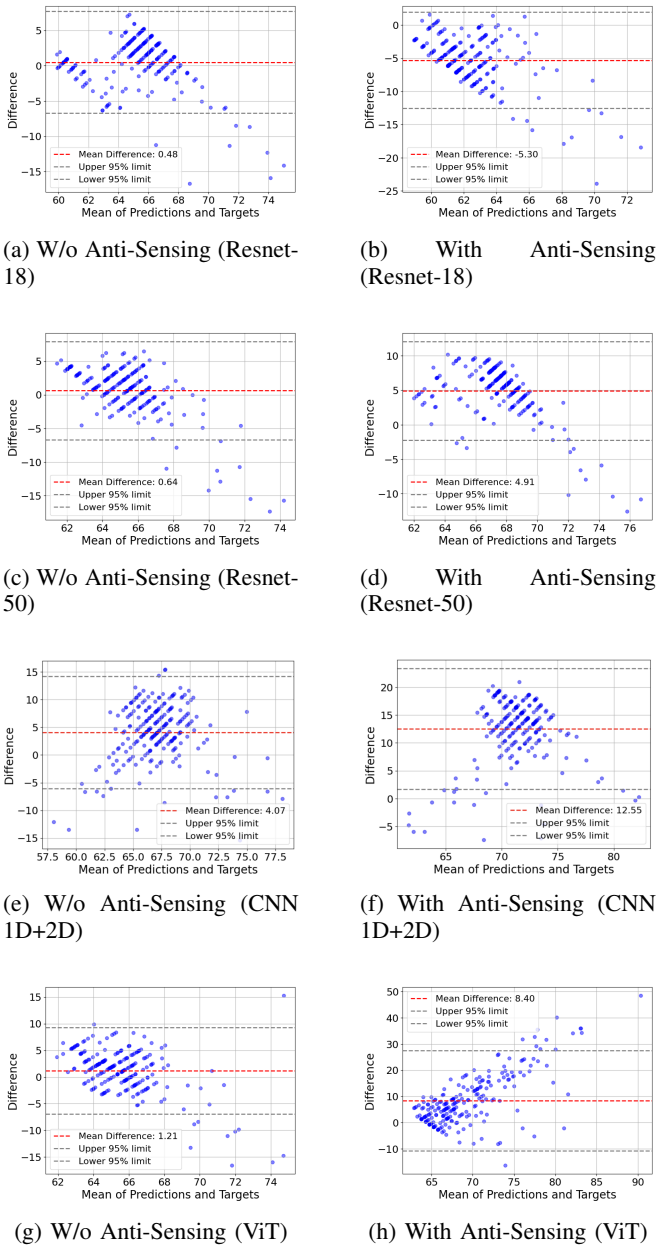
2) Models: We employed pre-trained ResNet-18, ResNet-50 [30], and Vision Transformer (ViT) [32] models and fine-tuned them on the sleep dataset for regression task of heart rate estimation. Additionally, we used a CNN-based model [31] that combines both 1D and 2D signal extraction approaches.

Figure 4 shows the Bland-Altman plots comparing the pre-trained and fine-tuned ResNet-18, ResNet-50, CNN 1D+2D, and Vision Transformer (ViT) models on sleep clinic data with and without anti-sensing perturbations. Without anti-sensing, the mean differences were $0.48, 0.64, 4.07$, and 1.21 for ResNet-18, ResNet-50, CNN 1D+2D, and ViT respectively, indicating a close match between predictions and ground truth. However, with anti-sensing applied, the mean differences increased to $-5.30, 4.91, 12.55$, and 8.40, respectively, highlighting the algorithm's effectiveness in disrupting model accuracy.

Figure 5 further illustrates the growing disparity between predicted and actual heart rates under anti-sensing, while Table I shows a significant increase in MAE after the attack: 2.68 bpm for ResNet-18, 3.26 bpm for ResNet-50, 7.28 bpm for CNN 1D+2D, and 6.07 bpm for ViT.

B. Evaluating our Anti-Sensing Device

Figure 6 shows a comparative analysis of a sample human heart rate measurement with and without the anti-sensing motor running at 98 RPM positioned in front of the participant. Without anti-sensing (bottom left), the FFT plot displays the ground truth heart rate of 86 bpm as a prominent peak. Conversely, the FFT plot for the case with anti-sensing (bottom right) shows the anti-sensing motor's frequency as the highest peak, demonstrating the impact of the anti-sensing motor on the measurement. The harmonics



(a) W/o Anti-Sensing (Resnet-18)

(b) With Anti-Sensing (Resnet-18)

(c) W/o Anti-Sensing (Resnet-50)

(d) With Anti-Sensing (Resnet-50)

(e) W/o Anti-Sensing (CNN 1D+2D)

(f) With Anti-Sensing (CNN 1D+2D)

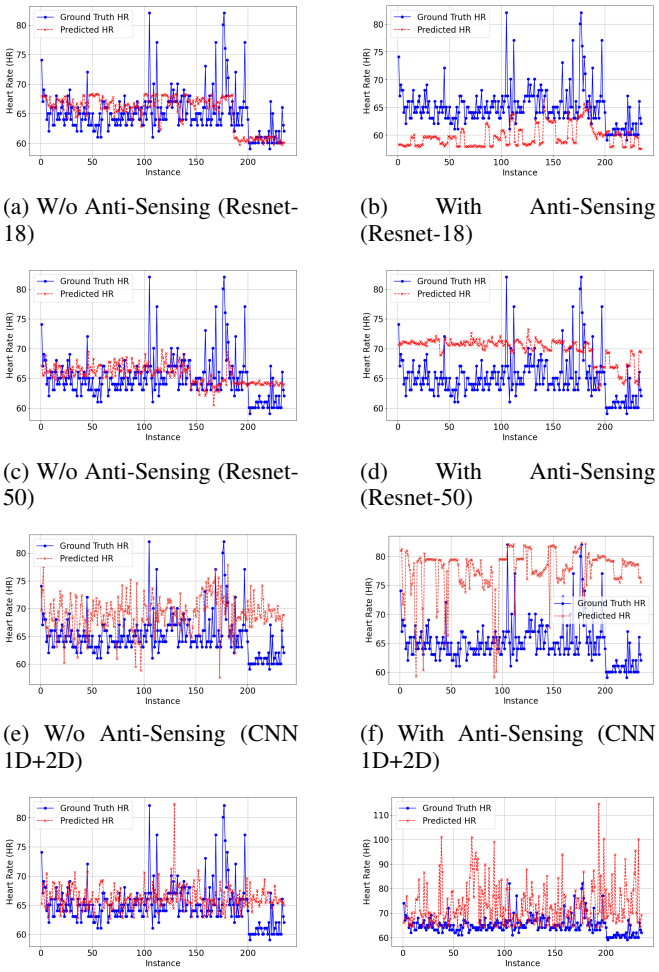
(g) W/o Anti-Sensing (ViT)

(h) With Anti-Sensing (ViT)

Fig. 4: Comparison of Bland Altman Plots from Resnet-18, Resnet-50, CNN 1D+2D, and Vision Transformer (ViT) models for HR estimation without (left column) and with (right column) anti-sensing perturbation applied on sleep dataset.

of the breathing rate (BR) at 60 bpm are evident in each FFT, as the participant's breathing rate was 20 bpm.

To evaluate the performance of our anti-sensing device, we



(a) W/o Anti-Sensing (Resnet-18)

(b) With Anti-Sensing (Resnet-18)

(c) W/o Anti-Sensing (Resnet-50)

(d) With Anti-Sensing (Resnet-50)

(e) W/o Anti-Sensing (CNN 1D+2D)

(f) With Anti-Sensing (CNN 1D+2D)

(g) W/o Anti-Sensing (ViT)

(h) With Anti-Sensing (ViT)

Fig. 5: Comparison of Resnet-18, Resnet-50, CNN 1D+2D, and Vision Transformer (ViT) model performance on HR prediction without (left column) and with (right column) anti-sensing perturbation applied on sleep dataset.

HR Model	MAE (W/o Motor)	MAE (With Motor)
FFT	2.42 bpm	8.17 bpm
ResNet - 50	2.93 bpm	7.23 bpm
CNN 1D+2D	9.24 bpm	17.56 bpm

TABLE II: Performance on Real Anti-Sensing motor attached to the wrist.

collected data from five individuals wearing it on their wrists, with the wrist positioned close to the chest. The ground truth was collected with a Galaxy smart watch. For each participant, we first executed the sinusoidal defense as depicted in Algorithm 1 for a specific heart rate (HR) model to determine the optimal servo frequency and spatial amplitude. Subsequently, the ESP32 microcontroller was programmed to rotate the servo at the optimized frequency. The spatial amplitude of the system was adjusted by modifying the arm length to which the octahedral reflector was mounted.

Table II shows the impact of the anti-sensing motor on the mean absolute error (MAE) of three heart rate (HR) models.

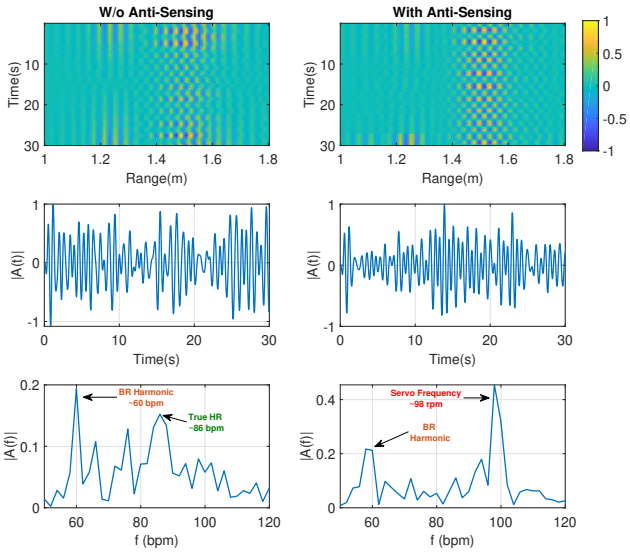


Fig. 6: Sample comparison of HR measurement with (right column) and without (left column) anti-sensing motor running at 98 RPM in front of the participant. The top row displays the corresponding 2D radargrams, the middle row shows the 1D extracted signals, and the bottom row features the FFT plots.

For the FFT model, the MAE increased from 2.42 bpm without the motor to 8.17 bpm with the motor. Similarly, the ResNet-50 model experienced an increase in MAE from 2.94 bpm without the motor to 7.23 bpm with the motor. The CNN 1D+2D model showed a notable rise in MAE from 9.24 bpm without the motor to 17.56 bpm with the motor. The rest of the models perform very poorly, even without anti-sensing; hence, the results are not included. Overall, introducing the anti-sensing motor led to a degradation in performance for all models, with the CNN 1D+2D model exhibiting the most substantial increase in error. The varying degrees of performance degradation across different heart rate models can be attributed to differences in model architectures and the optimized perturbation frequency, which varies across participants.

VII. Conclusion and Future Work

This paper presents Anti-Sensing, a novel defense against unauthorized radar-based heart rate sensing. By introducing physically realizable perturbations via a wearable device, we disrupted radar sensing models, leading to inaccurate heart rate estimations and enhanced privacy protection. Our gradient-based algorithm optimized device oscillations within physiological limits, with experiments validating its effectiveness. Given the growing reliance on radar for human sensing in robotics, this research is crucial for ensuring privacy and security in next-generation robotic systems.

While this work establishes the foundation for physical anti-sensing in the radar domain, future efforts will extend

the technique to more complex tasks, such as gesture recognition, and explore multi-modal defenses. In future, we aim to refine the optimization of the perturbation mechanisms to be real-time, lightweight, compact, and potentially battery-free, further reducing user burden.

REFERENCES

[1] A. Lazaro, D. Girbau, and R. Villarino, "Analysis of vital signs monitoring using an ir-uwb radar," Progress In Electromagnetics Research, vol. 100, pp. 265–284, 2010.

[2] C. E. Goldfine, M. F. T. Oshim, S. P. Carreiro, B. P. Chapman, D. Ganesan, and T. Rahman, "Respiratory rate monitoring in clinical environments with a contactless ultra-wideband impulse radar-based sensor system," in Proceedings of the... Annual Hawaii International Conference on System Sciences. Annual Hawaii International Conference on System Sciences, vol. 2020. NIH Public Access, 2020, p. 3366.

[3] M. F. T. Oshim, T. Surti, C. Goldfine, S. Carreiro, D. Ganesan, S. Jayasuriya, and T. Rahman, "Eulerian phase-based motion magnification for high-fidelity vital sign estimation with radar in clinical settings," in 2022 IEEE Sensors. IEEE, 2022, pp. 1–4.

[4] Z. Xie, H. Wang, S. Han, E. Schoenfeld, and F. Ye, "Deepvs: A deep learning approach for rf-based vital signs sensing," in Proceedings of the 13th ACM international conference on bioinformatics, computational biology and health informatics, 2022, pp. 1–5.

[5] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, "Heart rate sensing with a robot mounted mmwave radar," in 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020, pp. 2812–2818.

[6] K. Wu, E. H. Chen, X. Hao, F. Wirth, K. Vitanova, R. Lange, and D. Burschka, "Adaptable action-aware vital models for personalized intelligent patient monitoring," in 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022, pp. 826–832.

[7] S. Ahmed, D. Wang, J. Park, and S. H. Cho, "Uwb-gestures, a public dataset of dynamic hand gestures acquired using impulse radar sensors," Scientific Data, vol. 8, no. 1, p. 102, 2021.

[8] S. Skaria, A. Al-Hourani, and R. J. Evans, "Deep-learning methods for hand-gesture recognition using ultra-wideband radar," IEEE Access, vol. 8, pp. 203 580–203 590, 2020.

[9] J. Park and S. H. Cho, "Ir-uwb radar sensor for human gesture recognition by using machine learning," in 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016, pp. 1246–1249.

[10] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv preprint arXiv:1412.6572, 2014.

[11] K. Alexey, "Adversarial examples in the physical world," arXiv preprint arXiv: 1607.02533, 2016.

[12] A. Mądry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," stat, vol. 1050, no. 9, 2017.

[13] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in 2016 IEEE European symposium on security and privacy (EuroS&P). IEEE, 2016, pp. 372–387.

[14] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2574–2582.

[15] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in 2017 ieee symposium on security and privacy (sp). Ieee, 2017, pp. 39–57.

[16] P.-Y. Chen, H. Zhang, Y. Sharma, J. Yi, and C.-J. Hsieh, "Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models," in Proceedings of the 10th ACM workshop on artificial intelligence and security, 2017, pp. 15–26.

[17] M. Alzantot, Y. Sharma, S. Chakraborty, H. Zhang, C.-J. Hsieh, and M. B. Srivastava, "Genattack: Practical black-box attacks with gradient-free optimization," in Proceedings of the genetic and evolutionary computation conference, 2019, pp. 1111–1119.

[18] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," arXiv preprint arXiv:1712.04248, 2017.

[19] M. Andriushchenko, F. Croce, N. Flammarion, and M. Hein, "Square attack: a query-efficient black-box adversarial attack via random search," in European conference on computer vision. Springer, 2020, pp. 484–501.

[20] C. Guo, J. Gardner, Y. You, A. G. Wilson, and K. Weinberger, "Simple black-box adversarial attacks," in International conference on machine learning. PMLR, 2019, pp. 2484–2493.

[21] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1625–1634.

[22] R. Duan, X. Mao, A. K. Qin, Y. Chen, S. Ye, Y. He, and Y. Yang, "Adversarial laser beam: Effective physical-world attack to dnns in a blink," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16 062–16 071.

[23] Y. Li, Y. Li, X. Dai, S. Guo, and B. Xiao, "Physical-world optical adversarial attacks on 3d face recognition," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 24 699–24 708.

[24] H. Huang, Z. Chen, H. Chen, Y. Wang, and K. Zhang, "T-sea: Transfer-based self-ensemble attack on object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 20 514–20 523.

[25] Y. Xie, X. Guo, Y. Wang, J. Cheng, and Y. Chen, "Universal targeted adversarial attacks against mmwave-based human activity recognition," in Network Security Empowered by Artificial Intelligence. Springer, 2024, pp. 177–211.

[26] U. Ozbulak, B. Vandersmissen, A. Jalalvand, I. Couckuyt, A. Van Messem, and W. De Neve, "Investigating the significance of adversarial attacks and their relation to interpretability for radar-based human activity recognition systems," Computer Vision and Image Understanding, vol. 202, p. 103111, 2021.

[27] P. Staat, S. Mulzer, S. Roth, V. Moonsamy, M. Heinrichs, R. Kronberger, A. Sezgin, and C. Paar, "Irshield: A countermeasure against adversarial physical-layer wireless sensing," in 2022 IEEE Symposium on Security and Privacy (SP), 2022, pp. 1705–1721.

[28] J. Shenoy, Z. Liu, B. Tao, Z. Kabelac, and D. Vasisht, "Rf-protect: privacy against device-free human tracking," in Proceedings of the ACM SIGCOMM 2022 Conference, 2022, pp. 588–600.

[29] "Pulson 440 uwb radar," 2015, https://fccid.io/NUF-P440-A/User-Manual/User-Manual-287844.

[30] U. Saeed, S. Y. Shah, A. A. Alotaibi, T. Althobaiti, N. Ramzan, Q. H. Abbasi, and S. A. Shah, "Portable uwb radar sensing system for transforming subtle chest movement into actionable micro-doppler signatures to extract respiratory rate exploiting resnet algorithm," IEEE Sensors Journal, vol. 21, no. 20, pp. 23 518–23 526, 2021.

[31] S. H. Choi and H. Yoon, "Convolutional neural networks for the real-time monitoring of vital signs based on impulse radio ultrawide-band radar during sleep," Sensors, vol. 23, no. 6, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/6/3116

[32] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.