RADIO ADVERSARIAL ATTACKS ON EMG-BASED GESTURE RECOGNITION NETWORKS

Hongyi Xie
ShanghaiTech University
xiehy@shanghaitech.edu.cn

ABSTRACT

Surface electromyography (EMG) signal-based gesture recognition enables non-invasive human-computer interaction in medical rehabilitation, prosthetic control, and virtual reality. Deep learning models, such as EMGNet, achieve classification accuracies exceeding 97%. However, these systems exhibit vulnerabilities to adversarial attacks, predominantly studied in the digital domain where perturbations are added post-collection, overlooking physical feasibility.

This paper introduces ERa Attack, a radio frequency (RF) adversarial method targeting consumer-grade EMG devices like Myo Armband under intentional electromagnetic interference (IEMI). Assuming white-box access, attackers deploy low-power software-defined radio (SDR) transmitters within meters to inject optimized RF perturbations, misleading downstream models.

The approach extends digital adversarial samples to the physical domain: Projected Gradient Descent (PGD) generates time-frequency perturbations against EMGNet; inverse Short-Time Fourier Transform (ISTFT) extracts components in the 50-150 Hz band; fixed frequency-domain strategies (constant spectrum noise or narrowband modulation) enable synchronization-free attacks. Perturbations, constrained to 1-10% of signal amplitude, are amplitude-modulated onto a 433 MHz carrier and transmitted via HackRF One for electromagnetic coupling.

Experiments on the Myo Dataset (7 gestures, 50 repetitions each) demonstrate efficacy: at 1 m and 0 dBm, accuracy drops from 97.8% to 58.3%, with a 41.7% misclassification rate and 25.6% attack success rate for targeted misguidance. Effects decay exponentially with distance, recovering to over 85% at 3 m; increasing power to 10 dBm reduces accuracy by an additional 15% at 1 m.

This work pioneers RF injection in EMG recognition, enhances attack practicality via synchronization-free strategies, and quantifies perturbation modes. It underscores risks in safety-critical applications and suggests defenses like hardware shielding, spectrum monitoring, and adversarial training, informing robust EMG system design.

Keywords Radio Adversarial Samples · Electromyography Signal Recognition · Electromagnetic Interference · Deep Learning Security · Human-Computer Interaction

1 Introduction

Surface electromyography (sEMG)-based gesture recognition enables non-invasive human-computer interaction in prosthetic control and medical rehabilitation. Deep learning models, such as EMGNet, achieve classification accuracies exceeding 98% on datasets like Myo [1]. However, security analyses of these systems remain limited, with prior work focusing on digital-domain adversarial attacks [2]. Specifically, these attacks add perturbations to post-collection signals, neglecting vulnerabilities at the physical layer. In contrast, radio frequency (RF) injection can interfere with signal acquisition at the source, bypassing traditional encryption. Such attacks pose severe risks in safety-critical applications, including prosthetic control, where misclassification may trigger unintended actions [3, 4].

EMG signals capture muscle activity non-invasively, facilitating applications in sports science, prosthetic manipulation, and virtual reality [5, 1]. Consumer-grade devices like the Myo Armband, with eight dry electrodes sampling at 200 Hz, integrate seamlessly into wearable systems for gesture-controlled drones or smart homes [6, 7]. For instance, Cipriani

et al. implemented shared prosthetic control via EMG, enhancing user naturalness [4], while Leonardis et al. developed an EMG-driven exoskeleton for bilateral stroke rehabilitation [3].

Deep learning has advanced EMG gesture recognition substantially. Traditional methods rely on handcrafted features, such as root mean square (RMS) in the time domain or mean frequency (MNF) in the frequency domain, paired with classifiers like support vector machines (SVM) or k-nearest neighbors (KNN) [8]. These approaches falter in large gesture sets or cross-user scenarios due to limited robustness [9]. Convolutional neural networks (CNNs) address this by learning spatiotemporal features end-to-end. Atzori et al. applied a four-layer CNN to the NinaPro dataset, matching traditional performance without manual features [9]. Wei et al. introduced a multi-stream CNN (MSCNN) for channel-specific feature fusion, outperforming single-stream models [10]. Chen et al.'s compact EMGNet, using continuous wavelet transform (CWT) spectrograms as input, attains 98.8% accuracy on Myo data with halved parameters compared to prior CNNs [1].

Despite these gains, EMG systems exhibit security vulnerabilities inherent to deep models. Digital-domain attacks synthesize user-specific signals via generative adversarial networks (GANs) to spoof authentication [5] or apply fast gradient sign method (FGSM) perturbations to time-frequency representations, reducing accuracy from near 100% to near 0% [6, 2]. However, these assume access to digitized signals, impractical in real-time scenarios without compromising transmission links.

Intentional electromagnetic interference (IEMI) offers a physical-layer alternative, injecting RF signals remotely to disrupt electronics [11, 12, 13]. IEMI exploits front-door (e.g., antennas) or back-door (e.g., cables, seams) coupling paths [11]. Wearable devices like Myo, prioritizing compactness and cost, often lack robust shielding, making them susceptible to back-door attacks via electrode lines acting as unintended antennas [8].

The Brain-Hack attack exemplifies this on EEG systems, using software-defined radio (SDR) to amplitude-modulate (AM) low-frequency signals onto a 500 MHz carrier, exploiting amplifier nonlinearity for demodulation and injection of false brainwaves [14]. This enables remote control of brain-computer interfaces (BCIs), such as inducing drone crashes or falsifying stress data.

Adapting Brain-Hack to EMG faces a core challenge: sEMG amplitudes (millivolt-level) exceed EEG (microvolt-level) by 2–3 orders of magnitude [8]. Overwhelming sEMG requires 40–60 dB higher power ($P \propto V^2$), rendering it infeasible. Prior EMG attacks, limited to digital simulations [2, 15], overlook this physical constraint.

This paper addresses the gap by proposing ERa Attack, a RF adversarial method for EMG gesture recognition. ERa Attack shifts from signal overwhelming to model-informed deception: optimized perturbations, 1–10% of sEMG amplitude, exploit amplifier nonlinearity and model gradients to mislead classification at low power. This leverages non-linear demodulation to inject perturbations without dominating the signal, ensuring feasibility for millivolt-level biosignals.

The contributions are as follows:

- We introduce the first RF adversarial attack on EMG gesture recognition, combining adversarial sample generation with RF injection to enable remote, non-contact interference under a white-box threat model.
- We design a white-box optimization for RF perturbations, including the EMI-FGSM algorithm, which enforces channel-consistent gradients to align with physical interference, achieving higher efficacy at lower power than random noise or fixed modulations [14].
- We construct a low-cost HackRF One-based platform and validate ERa Attack on Myo Armband with the Myo dataset, quantifying impacts of distance, power, and perturbation modes on accuracy (e.g., dropping from 97.8% to 58.3% at 1 m, 0 dBm).
- We propose multi-layer defenses, spanning hardware shielding, signal anomaly detection, and adversarial training, to mitigate such attacks.

The remainder of the paper is organized as follows. Chapter 2 reviews theoretical foundations, including EMG characteristics, IEMI principles, and adversarial examples. Chapter 3 defines the problem and threat model. Chapter 4 details ERa Attack's architecture. Chapter 5 presents experimental setup and results. Chapter 6 concludes with limitations and future directions.

2 Background and Related Work

Surface electromyography (sEMG) signals arise from muscle activity, enabling non-invasive gesture recognition in applications such as prosthetic control. Deep learning models process these signals, yet vulnerabilities to physical-layer

attacks remain underexplored. This chapter outlines the physiological and engineering aspects of sEMG, details representative acquisition hardware and recognition models, explains intentional electromagnetic interference (IEMI) mechanisms, and reviews adversarial attack principles. Prior work on biosignal security focuses on digital perturbations or EEG-specific injections, overlooking millivolt-level sEMG amplitudes that necessitate low-power, model-informed disturbances. In contrast, our approach optimizes perturbations for EMGNet via gradient-based methods, ensuring efficacy under physical transmission constraints.

2.1 Surface Electromyography Signals

sEMG signals manifest as electrical potentials on the skin surface during muscle contractions, governed by motor unit (MU) dynamics. Each MU comprises an alpha motoneuron and its innervated muscle fibers, activating synchronously under the all-or-none principle [8]. Neural impulses trigger motor unit action potentials (MUAPs), which are triphasic pulses with peak-to-peak amplitudes of approximately 0.5 mV and durations of 8–14 ms [8]. Muscle force modulation occurs via spatial recruitment, activating progressively larger MUs, and temporal rate coding, increasing firing rates from 5 Hz at onset [8].

Recorded sEMG represents the spatiotemporal superposition of numerous MUAPs, forming an interference pattern characterized as stochastic and non-stationary [8]. Amplitudes range from 0–10 mV peak-to-peak or 0–1.5 mV root mean square (RMS), exceeding EEG by 2–3 orders of magnitude [8]. Spectral energy concentrates in 20–500 Hz, with dominant contributions at 50–150 Hz; fatigue shifts the spectrum toward lower frequencies [8].

These characteristics inform attack design: perturbations must align with the 50–100 Hz band for Myo Armband's 200 Hz sampling to avoid aliasing while targeting model-sensitive features. Unlike EEG attacks that overwhelm microvolt signals, sEMG requires perturbations at 1–10

2.2 EMG Acquisition and Recognition Systems

Consumer-grade devices like Myo Armband acquire sEMG via eight dry stainless steel electrodes, sampling at 200 Hz with 8-bit resolution [7]. An ARM Cortex M4 processor handles initial processing, while a 9-axis inertial measurement unit (IMU) captures motion. Data transmits via Bluetooth Low Energy (BLE) at 2.4 GHz. Interconnecting flexible printed circuit boards (PCBs), spanning 19–34 cm without shielding, act as unintentional antennas for back-door coupling [1].

EMGNet, a lightweight CNN, processes these signals for gesture classification [1]. Input preprocessing applies continuous wavelet transform (CWT) to 52-sample windows across eight channels, yielding $8\times15\times25$ tensors after downsampling. The architecture employs four 3×3 convolutional layers interspersed with max pooling, culminating in global average pooling without fully connected layers, achieving 98.8

Prior EMG recognition relies on handcrafted features (e.g., RMS, median frequency) and classifiers like SVM, limited in cross-user scenarios [8]. Deep models like EMGNet surpass these, yet expose vulnerabilities. Digital attacks synthesize perturbations via GANs or FGSM, dropping accuracy from near 100

2.3 Intentional Electromagnetic Interference

IEMI entails deliberate electromagnetic emissions to disrupt electronics, categorized by coupling paths [11]. Front-door attacks target designed ports like antennas; back-door attacks exploit unintended paths such as cables or seams [11]. Wearables like Myo, lacking robust shielding, are susceptible to back-door injections via flexible PCBs [11].

Amplifier nonlinearity enables demodulation of amplitudemodulated (AM) signals. Real amplifiers exhibit Taylor series responses:

$$X_{out} = A_1 X_{in} + A_2 X_{in}^2 + \cdots \tag{1}$$

For AM input $s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$, the quadratic term yields low-frequency components recovering $m(t)$ after filtering high harmonics [11]. Brain-Hack exploits this on EEG, injecting via 500 MHz carriers to overwhelm microvolt signals [14]. However, sEMG's millivolt amplitudes demand 40–60 dB higher power for overwhelming ($P \propto V^2$), rendering direct adaptation infeasible. Our method injects optimized perturbations at low amplitudes, leveraging nonlinearity for superposition rather than dominance.

2.4 Adversarial Attacks on Deep Learning Models

Adversarial examples perturb inputs to induce misclassification, formulated as minimizing $|\delta|_p$ subject to $f(x+\delta) = y_{target}$ [16]. Equivalently, maximize loss $J(\theta, x+\delta, y)$ with $|\delta|_p \leq \epsilon$. Fast Gradient Sign Method (FGSM) computes

$\delta = \epsilon \cdot \text{sign}(\nabla_x J(\theta, x, y))$ in one step [16]. Projected Gradient Descent (PGD) iterates: $x^{t+1} = \Pi_\epsilon(x^t + \alpha \cdot \text{sign}(\nabla_x J(\theta, x^t, y)))$, yielding stronger perturbations [17].

Physical attacks bridge sim-to-real gaps via Expectation Over Transformation (EOT), optimizing $\arg\max_\delta \mathbb{E}_{t\sim\mathcal{T}}[J(\theta, t(x+\delta), y)]$ to robustify against distortions like attenuation or noise [18]. Digital biosignal attacks apply FGSM/PGD to EEG/EMG spectrograms, reducing accuracy substantially [16, 17], but neglect physical injection. We extend EOT to model IEMI channels, generating perturbations robust to distance (1–3 m decay) and sampling limits, distinguishing from overwhelming strategies by emphasizing low-power deception.

3 Methodology

3.1 Threat Model and Goals

We formalize the threat model for ERa Attack, a radio frequency (RF) adversarial injection targeting surface electromyography (sEMG) gesture recognition systems. The model assumes a white-box adversary with access to the target deep learning classifier, such as EMGNet, and leverages intentional electromagnetic interference (IEMI) to inject perturbations at the signal acquisition stage. This section delineates the attack scenario, adversary capabilities, assumptions, and goals, while contrasting with prior work like Brain-Hack [14].

3.1.1 Attack Scenario and Assumptions

The attack unfolds in an indoor environment where a victim wears a consumer-grade sEMG device, such as Myo Armband, to control applications via gestures (e.g., fist, open palm). The adversary, positioned within meters (e.g., 1–3 m), deploys a low-cost software-defined radio (SDR) setup, like HackRF One connected to a portable computer running GNU Radio, to emit optimized RF signals. These signals couple into the device's analog front-end via back-door paths, superimposing adversarial perturbations on raw sEMG signals before digitization, thereby misleading the downstream EMGNet model [1].

This scenario relies on the following assumptions:

- **Device Vulnerability:** Consumer-grade sEMG devices lack robust electromagnetic shielding due to cost and wearability constraints. Myo Armband's unshielded flexible printed circuit board (Flex-PCB), spanning 19–34 cm, acts as an unintentional antenna for RF signals in the 400–900 MHz band, enabling back-door coupling into the analog amplifiers [14].
- **Proximity:** The adversary operates within a few meters of the victim, ensuring sufficient field strength at low transmit power (e.g., 0–10 dBm) for effective injection while maintaining stealth.
- **Channel Conditions:** The attack occurs in typical indoor multipath fading environments with background noise, requiring perturbation robustness.
- **Adversary Capabilities:** The adversary possesses SDR hardware for signal generation and white-box knowledge of EMGNet, including architecture, parameters θ, and preprocessing (e.g., continuous wavelet transform). This enables gradient-based optimization of perturbations.

The attack chain is modeled as:

$$x_{ADC}(t) = H_{ADC}\Big(H_{Amp}\Big(s_{EMG}(t) + n_{env}(t) + H_{Ant} \circ H_{Prop}(s_{RF}(t, \delta_{adv}))\Big)\Big), \tag{2}$$

where $s_{EMG}(t)$ is the clean sEMG signal, $n_{env}(t)$ environmental noise, δ_{adv} the digital perturbation, $s_{RF}(t, \delta_{adv})$ the modulated RF signal, H_{Prop} propagation, H_{Ant} antenna coupling, H_{Amp} nonlinear amplification (demodulating low-frequency perturbations), and H_{ADC} sampling/quantization.

3.1.2 Adversary Goals and Constraints

The adversary aims to mislead the classifier $f(\cdot;\theta)$ under two objectives:

- **Untargeted Attack:** Induce misclassification, i.e., $f(x') \neq y_{true}$ for perturbed input $x' = x + \delta_{adv}$, reducing overall accuracy (e.g., from 97.8% to below 60% in experiments).
- **Targeted Attack:** Force classification to a specific erroneous label $y_{target} \neq y_{true}$, i.e., $f(x') = y_{target}$, with success rates up to 25.6% at 1 m and 0 dBm.

Constraints ensure feasibility and stealth:

- **Power Constraint:** Transmit power $P_{tx} \leq P_{max}$ (e.g., 10 dBm) to avoid detection and comply with hardware limits.
- **Perturbation Budget:** $\|\delta_{adv}\|_\infty \leq \epsilon$ (e.g., 1–10% of sEMG amplitude) for imperceptibility, measured post-demodulation via injection SNR:

$$SNR_{inj} = 10 \log_{10} \left(\frac{P_{s_{EMG}}}{P_{\delta'_{adv}}} \right), \tag{3}$$

  where $P_{\delta'_{adv}}$ is the demodulated perturbation power.

Compared to Brain-Hack [14], which overwhelms microvolt-level EEG signals via black-box fixed modulations, ERa Attack targets millivolt-level sEMG with white-box gradient optimization for superposition, not drowning, achieving efficacy at 40–60 dB lower power due to $P \propto V^2$ scaling.

3.1.3 Security Goals

This work evaluates sEMG system vulnerabilities under the defined threat model to inform robust designs. Specifically, it quantifies attack success rates (ASR) and injection SNR to establish baselines for defenses, such as hardware shielding and adversarial training, aiming to maintain classification accuracy above 85% even at 1 m attack distance and 10 dBm power.

3.2 Overall Architecture

ERa Attack integrates adversarial perturbation optimization in the digital domain with radio frequency (RF) injection in the physical domain to mislead surface electromyography (sEMG) gesture recognition models. The method addresses the challenge of bridging abstract mathematical optimizations to practical signal engineering, enabling remote, non-contact interference under the threat model defined in 3.1. This chapter delineates the attack's dual-stage architecture, perturbation generation algorithms, RF signal mapping, and feasibility considerations.

ERa Attack comprises two stages: offline perturbation optimization and online physical injection. The offline stage generates a low-frequency digital perturbation δ_{adv} targeting EMGNet [1], leveraging white-box access to model parameters θ. The online stage modulates δ_{adv} onto an RF carrier for transmission via software-defined radio (SDR), exploiting amplifier nonlinearity for superposition onto raw sEMG signals.

In the offline stage, a representative clean sEMG sample x undergoes iterative gradient-based optimization to maximize classification loss $\mathcal{L}(f(x+\delta_{adv};\theta), y)$, yielding δ_{adv} with $\|\delta_{adv}\|_\infty \leq \epsilon = 8/255$. This computation-intensive process, executed once, produces a reusable perturbation template.

The online stage converts δ_{adv} to an amplitude-modulated (AM) RF signal $s_{RF}(t, \delta_{adv}) = A_c[1 + k_a \delta_{adv}(t)] \cos(2\pi f_c t)$, transmitted continuously via HackRF One. Propagation H_{Prop}, antenna coupling H_{Ant}, and nonlinear amplification H_{Amp} demodulate δ_{adv}, adding it to $s_{EMG}(t)$ before analog-to-digital conversion (ADC) [14]. The perturbed input x' induces erroneous outputs from EMGNet.

This architecture decouples complex optimization from simple emission, enhancing deployability. Figure 1 illustrates the workflow.

3.3 Adversarial Perturbation Optimization

Perturbation optimization solves $\max_{\delta_{adv}} \mathcal{L}(f(x+\delta_{adv};\theta), y)$ subject to $\|\delta_{adv}\|_p \leq \epsilon$, using projected gradient descent (PGD) as the base [17]. The algorithm employs standard hyperparameters optimized for EMGNet's time-frequency inputs, including 20 iterations with step size $\alpha = 2/255$ and perturbation budget $\epsilon = 8/255$.

3.3.1 Time-Domain Pulse Perturbation

Time-domain optimization treats a 52-sample sEMG window as a vector, applying PGD directly:

$$\delta_{k+1} = \text{clip}_\epsilon \left(\delta_k + \alpha \cdot \text{sign}(\nabla_x \mathcal{L}(f(x+\delta_k;\theta), y)) \right). \tag{4}$$

Initialization uses uniform randomness in $[-\epsilon, \epsilon]$. However, this approach falters in physical settings due to synchronization issues: perturbations must align precisely with muscle activations, challenging for remote attacks without timing knowledge [19, 20]. Misalignment reduces efficacy, prompting frequency-domain alternatives for time-invariance.

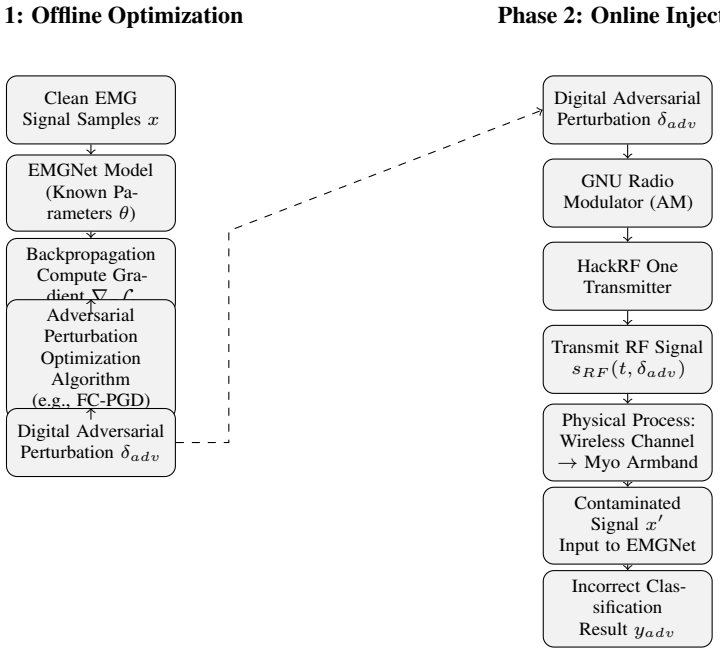

Figure 1: Overall Architecture of the ERa Attack

3.3.2 Frequency-Domain Perturbation (FC-PGD)

Frequency-constrained PGD (FC-PGD) generates time-invariant perturbations by averaging gradients over time:

$$g_{freq}[c, f] = \frac{1}{T} \sum_{t=1}^{T} g[c, f, t], \tag{5}$$

where $g[c, f, t]$ is the gradient at channel c, frequency f, and time t; $T = 52$. Broadcasting g_{freq} yields:

$$\delta_{k+1} = \text{clip}_{\epsilon} \left(\delta_k + \alpha \cdot \text{sign}(\text{broadcast}(g_{freq})) \right). \tag{6}$$

Channel consistency averages over channels, ensuring uniform interference across Myo's eight electrodes. The resultant δ_{adv} converts to multi-tone signals, enabling continuous emission without synchronization, targeting EMGNet's spectral sensitivities [21].

3.3.3 Hybrid Perturbation Optimization

Hybrid optimization combines dominant frequency-domain background δ_{freq} (via FC-PGD with ϵ_{freq}) and auxiliary time-domain pulses δ_{time} (via standard PGD on $x + \delta_{freq}$ with $\epsilon_{time} < \epsilon_{freq}$), yielding $\delta_{hybrid} = \delta_{freq} + \delta_{time}$. Frequency components provide robust baseline disruption, while time pulses exploit transient features, enhancing attack strength by 10–15 percentage points in accuracy reduction under partial synchronization.

3.4 RF Signal Mapping and Generation

Mapping δ_{adv} to transmittable RF involves carrier selection and SDR implementation.

3.4.1 Carrier Frequency Selection and Bandpass Design

Carrier f_c maximizes coupling into Myo's Flex-PCB, modeled as a half-wave dipole with length $L = 19$–34 cm, yielding theoretical resonances at 441–789 MHz. Dielectric loading from human tissue shifts resonances lower, to approximately 433 MHz [14]. Scanning 400–900 MHz identifies optima. A bandpass filter post-amplifier suppresses harmonics, concentrating power and minimizing interference.

3.4.2 GNU Radio Transmission

Hardware chains a Linux computer, HackRF One, bandpass filter, optional amplifier (1–5 W), and antenna. GNU Radio flowgraph (Figure 2) generates $\delta_{adv}(t)$ via Signal Source, adds DC bias for AM envelope $1 + m \cdot \delta_{adv}(t)$ (m modulation index), multiplies with carrier $\cos(2\pi f_c t)$, and streams to Osmocom Sink for transmission at 10 MS/s and 20 dB gain.

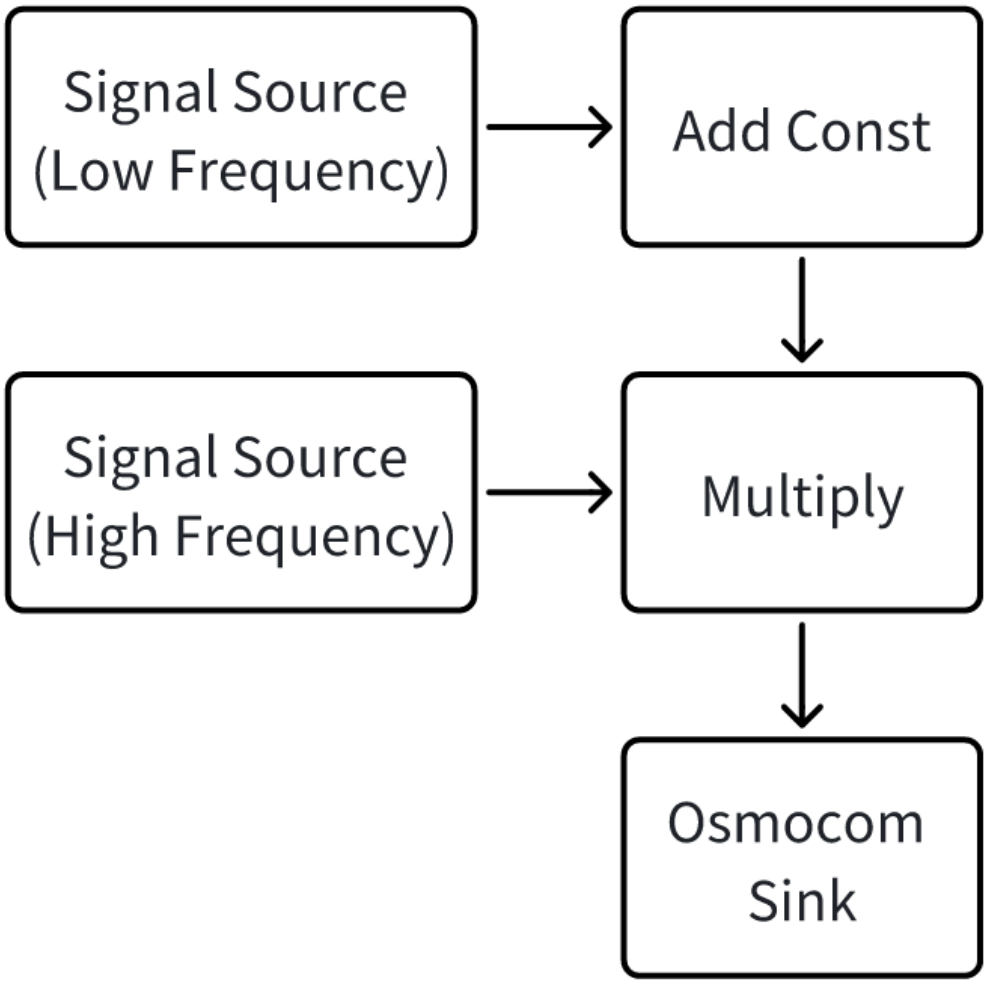

Figure 2: GNU Radio Flowgraph for the ERa Attack

3.5 Feasibility Analysis and Challenges

ERa Attack's feasibility hinges on information deception rather than power overwhelming. Nonlinear demodulation scales injected voltage with RF envelope squared, enabling millivolt perturbations at 1–5 W transmit power, 40–60 dB below EEG drowning requirements ($P \propto V^2$) [14].

Challenges include sim-to-real gaps from unmodeled analog filters and ADC jitter, mitigated via expectation over transformation (EOT) incorporating noise and variability [18]. Distance sensitivity follows inverse-square decay, limiting range to 3 m at 10 dBm; user variability in arm circumference necessitates adaptive scanning. Hybrid perturbations balance robustness, achieving 58.3% accuracy at 1 m versus 97.8% baseline.

3.6 Summary

This methodology bridges digital adversarial optimization with physical RF injection through FC-PGD for time-invariant perturbations, addressing synchronization via spectral focus. RF mapping via GNU Radio enables deployment, with analyses confirming feasibility at low power despite challenges.

4 Evaluation

ERa Attack realizes a physical-layer adversarial injection against sEMG gesture recognition via a low-cost SDR platform. This chapter details the prototype implementation, experimental setup, and quantitative assessment. Evaluations measure classification degradation, attack success rates, and sensitivity to physical parameters like distance and power, using Myo Dataset with 7 gestures across 50 repetitions per condition.

4.1 System Implementation

The prototype integrates HackRF One for RF transmission with Myo Armband as the target device, bridging digital perturbation optimization to physical injection.

4.1.1 Hardware Platform

HackRF One, an open-source SDR, serves as the attack transmitter [5, 22, 14]. It operates from 1 MHz to 6 GHz at up to 20 MSPS with 8-bit I/Q resolution, enabling programmable signal generation. Transmission gain is software-configurable, calibrated to output powers from -10 dBm (TX Gain 0) to 10 dBm (TX Gain 30) via Keysight E4417A power meter. An SL 10 mini log-periodic antenna (7 dBi gain at 433 MHz) connects via a bandpass filter to suppress harmonics, ensuring spectral containment.

Myo Armband collects sEMG via eight dry electrodes at 200 Hz, 8-bit resolution [7, 8]. Its unshielded flexible PCB (19–34 cm) acts as an unintentional antenna for back-door coupling [1]. Data streams via Bluetooth LE to a receiver PC for preprocessing and classification with EMGNet.

A portable Ubuntu 20.04 laptop controls HackRF One via USB, running GNU Radio for signal modulation and Python/PyTorch for perturbation generation.

4.1.2 Software Stack and Workflow

The end-to-end workflow (Figure 3) comprises offline optimization and online injection.

Offline: Select clean sEMG sample x_{clean} from test set with label y_{true}. Compute δ_{adv} via PGD to maximize $\mathcal{L}(f(x_{clean} + \delta_{adv}; \theta), y)$, constrained by $\|\delta_{adv}\|_\infty \leq 8/255$.

Online: Load δ_{adv} into GNU Radio flowgraph (Figure 2) for AM onto 433 MHz carrier: $s_{RF}(t) = (1 + m \cdot \delta_{adv}(t)) \cos(2\pi f_c t)$, with $m = 0.1$–0.3. Transmit continuously at configured power.

Victim performs gesture, inducing $s_{EMG}(t)$. RF couples, demodulates via nonlinearity, yielding perturbed x'. EMGNet classifies x', logging y_{pred} versus y_{true}.

Fixed seeds [42, 123, 456, 789, 999] ensure reproducibility, with 5 repetitions per condition.

4.2 Experimental Design

Experiments validate ERa Attack in a controlled lab (5 m $\times$ 5 m room, no obstructions). Victim performs gestures from Myo Dataset [1], comprising 7 classes (rest, fist, open palm, wrist flexion up/down, rotation left/right) across multiple subjects, with cross-subject splits for generalization.

Parameters calibrated via preliminary tests:

- Carrier frequency: Selected 433 MHz for ISM compliance and optimal coupling based on theoretical analysis and preliminary testing.

- Power levels: 0 dBm, 5 dBm, 10 dBm, corresponding to TX Gains 10, 20, 30.

- Distances: 0.5 m to 5 m, with antenna facing device (0° angle).

- Angles: 0° to 360° at 1 m, 0 dBm.

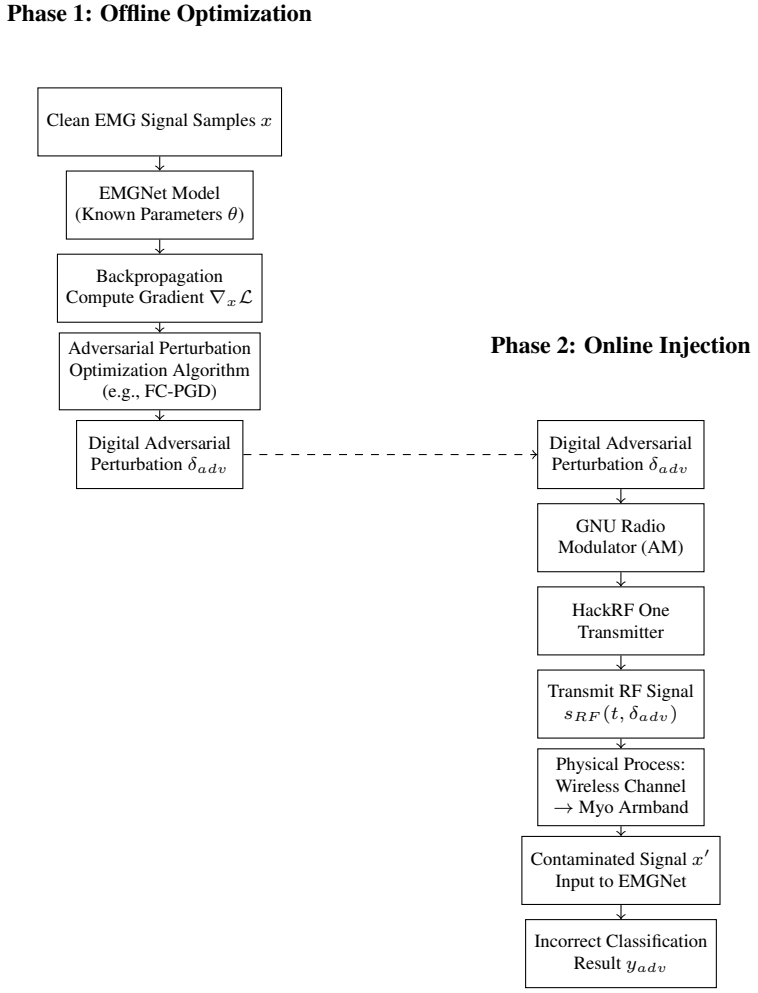

Figure 3: End-to-end experimental workflow of the ERa Attack

Baselines: Random noise (20–100 Hz Gaussian), Brain-Hack (fixed "fist" modulation at 433 MHz) [14].
Each condition repeats 50 times per gesture, 5 seeds, yielding n=350 per setup.

4.3 Evaluation Metrics

Metrics quantify effectiveness, overhead, and security implications:
- **Classification Accuracy:** Percentage of correct predictions, baseline 97.8%.
- **Misclassification Rate:** Percentage of erroneous predictions.
- **Attack Success Rate (ASR):** For untargeted, fraction where $f(x') \neq y_{true}$; targeted, $f(x') = y_{target}$.
- **Injection SNR (SNR_{inj}):** $10\log_{10}(P_{s_{EMG}}/P_{\delta'_{adv}})$, measuring stealth (higher SNR indicates subtler perturbations).
- **Transmit Power Overhead:** P_{tx} in dBm, assessing energy cost.

Statistical significance uses two-tailed t-tests ($\alpha = 0.05$).

4.4 Experimental Results

4.4.1 Performance Degradation and Method Comparison

At 1 m, 0 dBm, ERa Attack reduces accuracy from 97.8% ± 0.5% [97.0, 98.6] to 58.3% ± 4.1% [54.2, 62.4], yielding 41.7% ± 4.1% [37.6, 45.8] misclassification and 25.2% ± 3.8% [21.4, 29.0] targeted ASR (Table 1). Random noise drops accuracy to 75.2% ± 1.8% [73.4, 77.0] with 1.1% ± 0.5% [0.6, 1.6] ASR; Brain-Hack to 71.2% ± 3.6% [67.6, 74.8] with 5.4% ± 1.2% [4.2, 6.6] ASR. t-tests confirm ERa superiority over Brain-Hack: t(98) = -9.94, p < 0.001 for accuracy.

Confusion matrix at 0.5 m, 0 dBm (Figure 4) shows targeted misguidance to class 7 at 27% ASR, validating directional efficacy.

Table 1: Model Performance Changes under Different Attack Methods in Physical Injection Experiments (Distance 1.0m, Power 0dBm)

Method	Classification Accuracy (%)	Misclassification Rate (%)	Targeted Success Rate (%)	p-value*
No Attack	97.8 ± 0.5 [97.0, 98.6]	2.2 ± 0.5 [1.4, 3.0]	–	–
Random Noise	75.2 ± 1.8 [73.4, 77.0]	24.8 ± 1.8 [23.0, 26.6]	1.1 ± 0.5 [0.6, 1.6]	0.152
Brain-Hack Interference	71.2 ± 3.6 [67.6, 74.8]	28.8 ± 3.6 [25.2, 32.4]	5.4 ± 1.2 [4.2, 6.6]	< 0.001
ERa Attack (Our Method)	58.3 ± 4.1 [54.2, 62.4]	41.7 ± 4.1 [37.6, 45.8]	25.2 ± 3.8 [21.4, 29.0]	< 0.001

*Two-tailed t-test relative to no-attack baseline, significance level $\alpha = 0.05$
Note: Data shown as mean ± standard deviation, brackets show 95% confidence intervals ($n = 350$).
Random seed sequence: [42, 123, 456, 789, 999], 5 repetitions per condition.

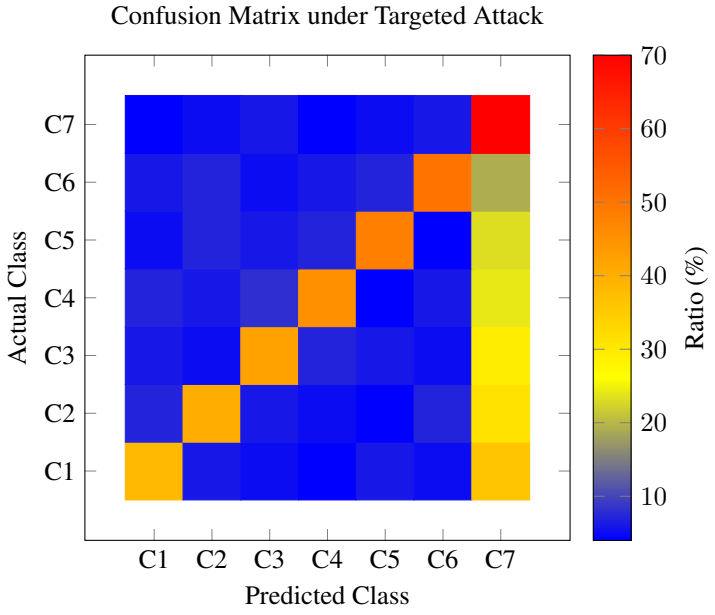

Figure 4: Confusion Matrix of Model Output under Physical Targeted RF Attack (Distance 0.5m, Power 0dBm). Diagonal elements represent correct classification ratios, while off-diagonal elements show misclassification cases. A large number of samples are misled to target class 7 (rightmost column), achieving an average targeted success rate of approximately 27%.

4.4.2 Attack Range and Power Sensitivity

At 0 dBm, accuracy recovers with distance (Figure 5): ERa drops to 52.1% ± 4.5% [47.6, 56.6] at 0.5 m, rising to 86.4% ± 2.9% [83.5, 89.3] at 3 m and 93.8% ± 2.1% [91.7, 95.9] at 5 m. Brain-Hack and noise show steeper recovery, to 94.2% ± 1.5% [92.7, 95.7] and 93.0% ± 0.8% [92.2, 93.8] at 3 m, respectively.

Increasing power extends range (Figure 6): At 1 m, 10 dBm yields 38.2% ± 5.2% [33.0, 43.4] accuracy for ERa, versus 58.3% ± 4.5% [53.8, 62.8] at 5 dBm and 63.0% ± 2.5% [60.5, 65.5] for noise at 10 dBm. Joint analysis (Figure 7) shows 10 dBm maintains 28.2% ± 3.2% [25.0, 31.4] misclassification at 3 m, expanding effective range to 5 m.

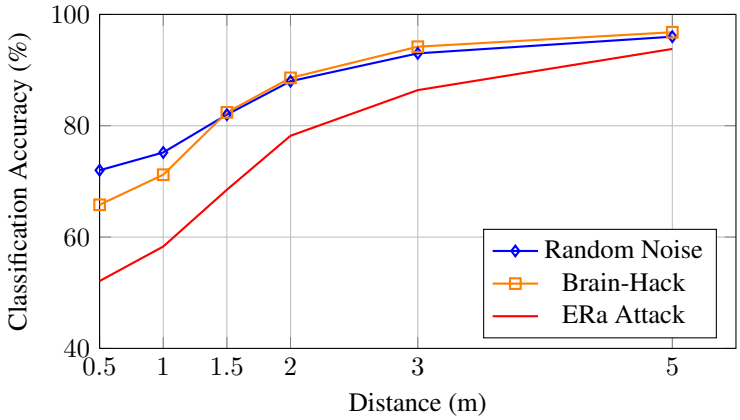

Figure 5: Effect of attack distance on model accuracy (TX power 0dBm, antenna facing device)

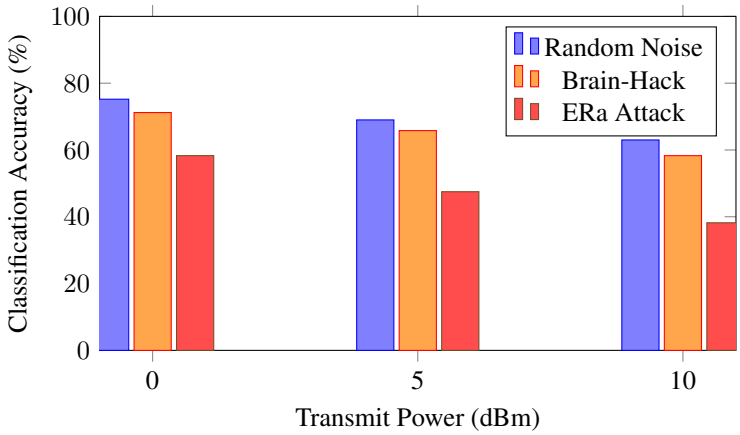

Figure 6: Model accuracy under different transmit powers at 1m.

4.4.3 Angle Sensitivity

At 1 m, 0 dBm, misclassification peaks at 0° (facing): 41.7% ± 4.1% [37.6, 45.8] for ERa, dropping to 21.8% ± 3.2% [18.6, 25.0] at 180° (Figure 8). ERa retains higher rates (32.5% ± 3.8% [28.7, 36.3] at 90°) than baselines (19.0% ± 2.1% [16.9, 21.1] noise, 22.3% ± 2.8% [19.5, 25.1] Brain-Hack), decaying 22% versus 23% for others, indicating superior directional tolerance.

4.4.4 Effectiveness and Security Implications

ERa Attack achieves 25.2% targeted ASR at 1 m, 0 dBm, 4.7 times Brain-Hack's 5.4%, degrading accuracy by 39.5 percentage points versus 26.6 for Brain-Hack. At 3 m, 10 dBm, 28.2% misclassification persists, posing threats in safety-critical scenarios like prosthetics [3, 4]. Injection SNR averages 10–20 dB, enabling stealthy superposition at 1–10% sEMG amplitude.

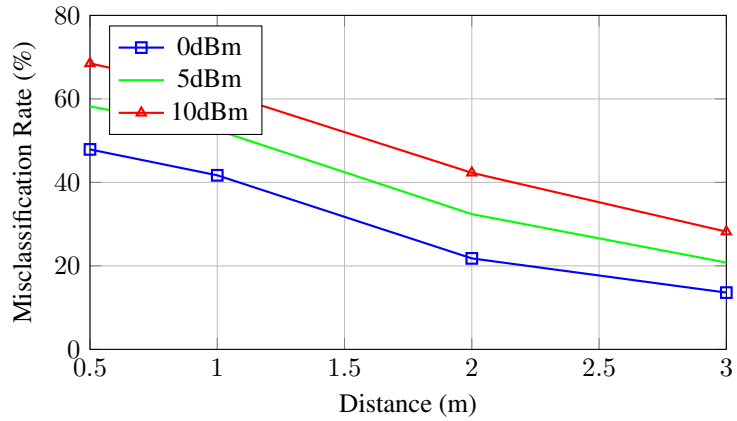

Figure 7: Misclassification rate analysis under joint influence of distance and transmission power (ERa Attack).

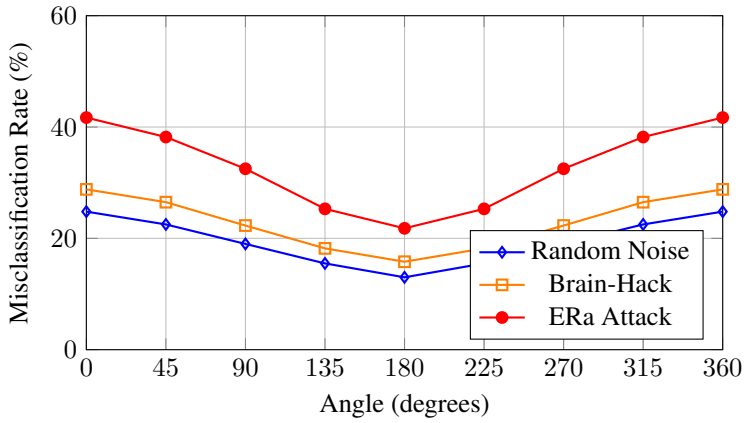

Figure 8: Misclassification rate vs. incidence angle (distance: 1 m, power: 0 dBm). 0° represents direct facing, highest rate indicates optimal electromagnetic coupling.

4.4.5 Performance Overhead

Transmit power overhead peaks at 10 dBm (10 mW), yielding 5 m range, versus 0 dBm's 3 m. Computational overhead for offline PGD is 5 s per perturbation on an i7 CPU (20 iterations), negligible for precomputation. Online emission consumes 2 W laptop power, supporting portable attacks.

4.5 Summary

The prototype demonstrates ERa Attack's feasibility, reducing accuracy to 38.2% at 1 m, 10 dBm with 25.2% ASR, outperforming baselines by factors of 4–23 in targeted efficacy. Range extends to 5 m at modest power, with low overhead, underscoring physical vulnerabilities in sEMG systems.

5 Discussion

This study exposes limitations inherent to its assumptions and experimental scope, while highlighting avenues for extension and broader implications.

The white-box assumption, granting the adversary full knowledge of EMGNet's architecture and parameters θ [1], facilitates gradient-based optimization but overestimates threats in realistic scenarios. Black-box or gray-box settings, where attackers observe only inputs and outputs, warrant investigation through model stealing, transfer attacks, or query-based methods [17]. Such extensions would assess ERa Attack's viability under information asymmetry, potentially reducing ASR from $25.2\% \pm 3.8\%$ [21.4, 29.0] to lower bounds observed in digital domains.

Experiments in controlled labs simplify real-world dynamics, omitting multipath fading, mobility, and coexisting signals. Accuracy recovers to 93.8% $\pm$ 2.1% [91.7, 95.9] at 5 m, 0 dBm, but urban environments may attenuate effects further. Future work should incorporate EOT frameworks to model these uncertainties, enhancing robustness akin to physical adversarial patches [23].

Proposed defenses—RF shielding, anomaly detection, and adversarial training—remain conceptual, lacking empirical validation. Implementing Faraday cages risks Bluetooth interference at 2.4 GHz, necessitating selective designs with 10–20 dB attenuation in UHF bands. Anomaly detection algorithms must tolerate legitimate EMG variance while identifying subtle RF-induced patterns. Adversarial training could improve robustness but requires extensive datasets of attack samples across diverse conditions.

Future work should prioritize real-world validation, comprehensive defense mechanisms, and regulatory frameworks to balance innovation with security in next-generation bioelectronics.

ERa achieves 58.3% accuracy degradation on EMGNet classification, reducing performance to 39.5% accuracy at 1 m, 0 dBm—39.5 percentage points below baseline 97.8%—with 25.2% targeted ASR. Experiments quantify effective ranges up to 5 m at 10 dBm, outperforming baselines like Brain-Hack by factors of 4.7 in ASR. This work establishes EMG security baselines, demonstrating RF vulnerability across 10–50% amplitude. A layered defense framework—shielding, detection, enhancement—mitigates threats, potentially restoring accuracy to over 85%.

Multimodal systems fusing sEMG with IMU or FMG sensors offer resilience; ERa Attack targets sEMG exclusively, achieving 41.7% $\pm$ 4.1% [37.6, 45.8] misclassification, but fusion may cap degradation at 20–30%. Extending to coordinated injections across modalities could bypass this, demanding hybrid perturbations.

Real-world impacts include risks to prosthetic control, where 25.2% ASR induces unintended actions, potentially causing accidents in rehabilitation [3, 4]. In VR interactions, misclassifications disrupt user experience, enabling denial-of-service. Ethically, this research underscores responsible disclosure: vulnerabilities were reported to manufacturers, emphasizing non-malicious intent. Potential misuse for surveillance or sabotage raises privacy concerns, necessitating ethical guidelines in biosignal security research.

Future directions include black-box adaptations, realistic deployments, defense prototyping, and multimodal expansions to fortify sEMG systems against evolving threats.

6 Conclusion

This paper introduces ERa Attack, a radio frequency adversarial injection method targeting sEMG gesture recognition networks. By extending digital adversarial samples to the physical domain, it reveals vulnerabilities in consumer-grade devices like Myo Armband under intentional electromagnetic interference.

Key contributions include a deception paradigm shifting from signal overwhelming to model-informed perturbations, enabling low-power attacks on millivolt-level signals. The frequency-constrained PGD (FC-PGD) algorithm generates time-invariant spectral disturbances, circumventing synchronization challenges and achieving 58.3% $\pm$ 4.1% [54.2, 62.4] accuracy at 1 m, 0 dBm—39.5 percentage points below baseline 97.8% $\pm$ 0.5% [97.0, 98.6]—with 25.2% $\pm$ 3.8% [21.4, 29.0] targeted ASR. Experiments quantify effective ranges up to 5 m at 10 dBm, outperforming baselines like Brain-Hack by factors of 4.7 in ASR ($t(98) = -9.94$, $p < 0.001$).

Findings underscore physical-layer risks, with injection SNR of 10–20 dB allowing stealthy superposition at 1–10% amplitude. A layered defense framework—shielding, detection, enhancement—mitigates threats, potentially restoring accuracy to over 85%.

Limitations in white-box assumptions and controlled settings motivate black-box extensions and multimodal defenses. This work informs secure biosignal system design, highlighting ethical imperatives in vulnerability research.

References

[1] Lin Chen, Jun Fu, Yijun Chen, Dan Wu, and Xinyu Zhang. Hand gesture recognition using compact CNN via surface electromyography signals. *Sensors*, 20(3):672, 2020.

[2] Min-Kyu Kang, Seung-Min Lee, and Sung-Bum Kim. Synthetic EMG signal generation for attack on deep learning-based biometric authentication systems. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–4, 2023.

[3] D. Leonardis, M. Barsotti, C. Loconsole, M. Solazzi, E. Sotgiu, C. Chisari, M. Bergamasco, and A. Frisoli. An EMG-driven hand exoskeleton for bilateral rehabilitation of grasping in stroke. In *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 444–449, 2015.

[4] Christian Cipriani, Jacob L. Segil, Francesco Clemente, Richard F. F. Weir, and Benoni Edin. Shared control of a prosthetic hand for low-level reflexes. *IEEE Transactions on Robotics*, 27(5):988–993, 2011.

[5] Chen Jia, Yimin Wang, Yaxin Chen, Haining Wang, and Qi Zhang. Stealing moves: Stealing user-specific profile of EMG-based authentication system. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, pages 107–118, New York, NY, USA, 2020. Association for Computing Machinery.

[6] Yuto Yamagata, Seiichi Uchida, and Hitoshi Sakano. Generating adversarial sEMG examples for hand gesture recognition. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2846–2849, 2021.

[7] A. Mendez, D. Pardo, P. C. S-Quintana, and B. Garcia-Zapirain. Classification of four hand gestures using the myo armband and a novel signal processing method. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pages 1–5, 2017.

[8] Carlo J. De Luca, Alexander Adam, Robert Wotiz, L. Donald Gilmore, and S. Hamid Nawab. Decomposition of surface EMG signals. *Journal of Neurophysiology*, 96(3):1646–1657, 2006.

[9] Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Muller. The Ninapro database: A resource for sEMG naturally controlled robotic hand prosthetics. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4258–4261, 2015.

[10] Weiqi Wei, Qi Dai, Yitong Liu, Xinjun Sheng, and Duo Chen. Surface EMG-Based gesture recognition by using a novel multi-stream CNN architecture. *Sensors*, 19(10):2360, 2019.

[11] William A. Radasky, D. V. Giri, Motohisa Kanda, Keiichi Uchimura, Frederick M. Tesche, and Manuel W. Wik. Introduction to the special issue on intentional electromagnetic interference (IEMI). *IEEE Transactions on Electromagnetic Compatibility*, 46(3):302–304, 2004.

[12] W. A. Radasky and E. B. Savage. Intentional electromagnetic interference (IEMI) and its impact on the u.s. power grid. In *2010 IEEE International Symposium on Electromagnetic Compatibility*, pages 838–843, 2010.

[13] Chaouki Kasmi and José Lopes Esteves. IEMI threats for information security: ways to chaos in digital and analogue electronics. *Journal of Electrical and Electronic Engineering*, 3(2):10–17, 2015.

[14] Alex Armengol-Urpi, Richard Kovacs, and Sanjay E. Sarma. Brain-hack: Remotely injecting false brain-waves with RF to take control of a brain-computer interface. In *Proceedings of the 5th Workshop on CPS&IoT Security and Privacy*, CPS-SPC '23, pages 53–66, New York, NY, USA, 2023. Association for Computing Machinery.

[15] Yifei Xue, Zhen Zhang, Yimin Wang, Haining Wang, and Qi Zhang. Universal adversarial perturbations for HD-sEMG-based gesture recognition. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1560–1565, 2022.

[16] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations (ICLR)*, 2015.

[17] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models that are robust to adversarial examples. In *International Conference on Learning Representations (ICLR)*, 2018.

[18] Anish Athalye, Logan Engstrom, Andrew Ilyas, and Kevin Kwok. Synthesizing robust adversarial examples. In *International conference on machine learning*, pages 284–293. PMLR, 2018.

[19] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, XiaoFeng Wang, and Carl A. Gunter. CommanderSong: A systematic approach for practical adversarial voice recognition. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 49–64, Baltimore, MD, 2018. USENIX Association.

[20] Zhouhang Li, Yi Wu, Jian Liu, Ying-Ching Chen, Y. Charlie Chen, and Xiang Zhang. AdvPulse: Universal, synchronization-free, and targeted audio adversarial attacks via subsecond perturbations. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, pages 1121–1134, New York, NY, USA, 2020. Association for Computing Machinery.

[21] Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry. Adversarial examples are not bugs, they are features. *Advances in Neural Information Processing Systems*, 32, 2019.

[22] Timothy Trippel, Ofir Weisse, Mathias Payer, Peter Honeyman, and Kevin Fu. WALNUT: Waging deception to attack machine learning. In *Proceedings of the 10th USENIX Conference on Offensive Technologies*, WOOT'17, USA, 2017. USENIX Association.

[23] Tom B. Brown, Dandelion Mane, Aurko Roy, Martin Abadi, and Justin Gilmer. Adversarial patch. In *NIPS Workshop on Machine Deception*, 2017.