Measure Once, Train Often: Scaling Waveform Classifiers with Channel-Augmented Data

Sage Trudeau*†, Spencer Chun†, Brandon Nguyen*, Kaushik Chowdhury*
*The University of Texas at Austin, Austin, TX, USA
†MIT Lincoln Laboratory, Boston, MA, USA
Emails: strudeau@utexas.edu, spencer.chun@ll.mit.edu, brandon.qt.nguyen@utexas.edu, kaushik@utexas.edu

***Abstract*—Deep neural network (DNN)-based wireless signal classifiers trained on synthetic or lab-collected data often fail in real-world deployments due to unseen and variable channel conditions. Although such models may exceed 95% accuracy on synthetic benchmarks, their performance can degrade sharply in over-the-air settings—a challenge known as the synthetic-to-real gap. This gap arises because collecting labeled data for every signal class across all possible channel conditions is impractical, especially as wireless channels vary dynamically and many channel instances may not occur during waveform transmission. We propose Channel Measurement Augmented Training (CMAT), a lightweight and scalable pipeline that speeds up conventional training dataset generation. CMAT uses a low-cost OFDM sounder to capture a small number of channel impulse responses (CIRs) at the deployment site. These CIRs are then convolved offline with target waveforms to create realistic, site-specific training data without over-the-air transmission of each signal class. We evaluate CMAT across eight modulation types, four DNN architectures, and a multitude of deployment environments. Results show that when DNNs trained with CMAT-based synthetic datasets are tested with real-world data, our approach can recover up to 51% of the synthetic-to-real accuracy loss using only seconds worth of CIR snapshots per site. Principal Component Analysis confirms that a few components capture most of the channel variability. CMAT is open-sourced to support scalable DNN training without the cost-prohibitive collection of class-specific field data.**

***Index Terms*—data augmentation, channel measurement, waveform classification**

I. Introduction

Recent advances in deep learning have accelerated interest in radio-frequency (RF) machine learning (ML) for tasks such as modulation recognition, protocol identification, emitter fingerprinting, and spectrum sensing. While models often achieve over 95% accuracy on benchmark datasets like RadioML [1] when trained and tested under matched synthetic conditions, their performance can degrade sharply in real-world over-the-air deployments. This *synthetic-to-real gap* is driven by *distribution shift*: real received signals differ from training data due to site-specific propagation, hardware impairments, and noise. Writing the received signal as $y = Hx + n$, the key bottleneck is data: collecting labeled y for every waveform class x across the variety of deployment channels H is impractical.

•**Current state-of-the-art.** Two broad strategies are used to mitigate this gap. (i) *Data-centric* augmentation injects simulated channel effects (e.g., Rayleigh/Rician fading, phase noise, IQ imbalance) into training data to improve robustness [2], [3]. However, these statistical models may miss site-specific effects that dominate performance in a particular environment. (ii) *Model-centric* adaptation (e.g., adversarial or contrastive objectives [4], or post-deployment fine-tuning [5]) can improve transfer, but still requires collecting target-domain waveform examples (often labeled) for each new site, limiting scalability.

•**Proposed approach: measure once, train often.** We ask: can we prepare an RF classifier for a deployment environment *without* collecting any over-the-air examples of the target waveform classes at that site? We propose Channel-Measurement–Augmented Training (CMAT), which decouples *environment capture* from *class capture*. For an environment of interest, we measure a reusable bank of channel impulse responses (CIRs) $\{H\}$ once using an OFDM sounder. During training, we generate clean baseband waveforms x offline and synthesize site-specific training samples by convolving x with randomly sampled measured channels H. Thus, new models or new classes can be trained using the same measured $\{H\}$, without revisiting the site.

•**Challenges in designing CMAT.** CMAT must address: **(C1)** stable, reusable channel estimates (naive ZF inversion is ill-conditioned when $|\hat{H}[k]|$ is small); **(C2)** faithful synthesis that preserves constellation geometry and phase/burst structure after filtering by long, frequency-selective CIRs; and **(C3)** validation that models trained only on synthetic waveforms augmented with measured CIRs generalize to real over-the-air test data.

•**Contributions.**

(i) *Waveform-agnostic channel capture and stable synthesis:* We implement an OFDM-based CIR estimation pipeline and a numerically stable forward replay that applies measured $\hat{H}[k]$ to synthetic signals (Section III), avoiding ZF division by small $|\hat{H}[k]|$. (Addresses **C1**.)

(ii) *General-purpose RF synthesis engine:* We generate physically grounded, label-consistent training data by convolving clean baseband waveforms with measured CIRs while preserving key signal structure across modulation families. We plan to release CMAT as open source to enable community dataset generation [6]. (Addresses **C2**.)

(iii) *Synthetic-to-real generalization across architectures and environments:* Across 8 modulation classes, 4 model architectures (CNN, RNN, T-PRIME, MCformer), and 4 test

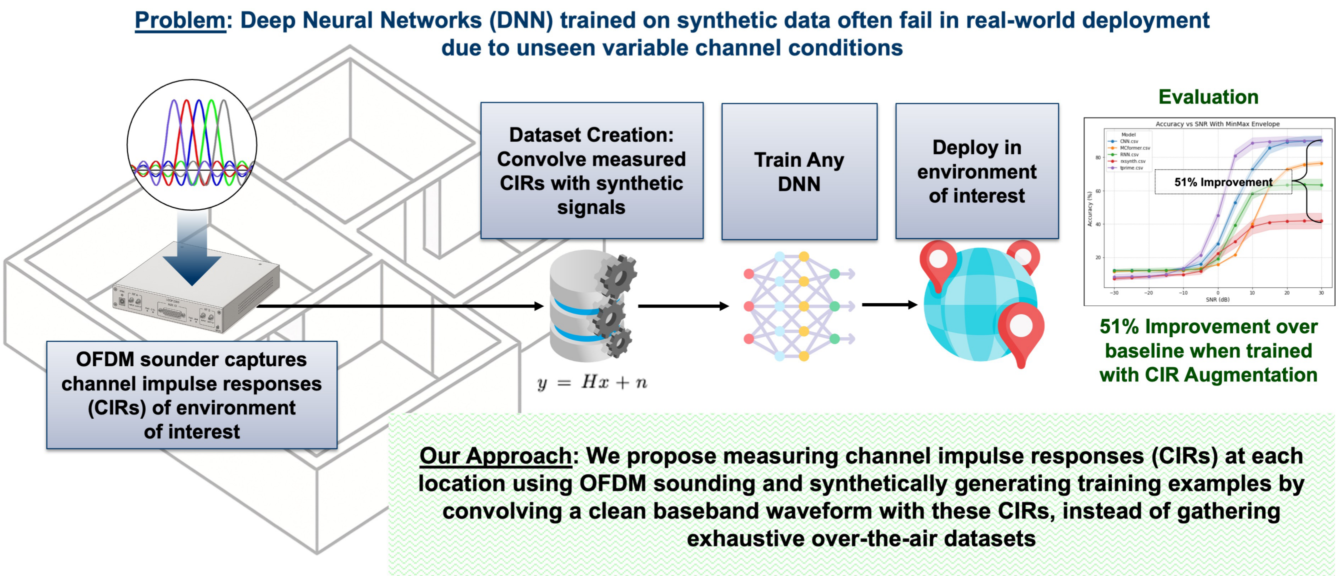


Fig. 1. **CMAT workflow (measure once, train often)**. At a deployment site, we collect *only* an OFDM sounding waveform to estimate a reusable bank of channel impulse responses (CIRs) $\{H\}$. Training data for *any* waveform class is generated offline by convolving clean synthetic baseband x with sampled CIRs H (no over-the-air recordings of the target classes are required). Over-the-air waveform recordings are used *only* for evaluation/benchmarking in this paper.

environments (Wired, OTA2, OTA10, Indoor-Lab), CMAT consistently reduces the synthetic-to-real gap; in our study, CMAT recovers up to 51% of the synthetic-to-real accuracy loss when evaluated on over-the-air test data, while requiring no over-the-air class recordings for training data generation. (Addresses **C3**.)

(iv) *Measurement efficiency and collection guidance:* We analyze how performance scales with the number of measured CIR snapshots via a PCA analysis of CIR variability (Section IV-C). This yields guidance on how many channel measurements are needed per site. (Addresses **C1, C3**.)

II. Related Work

We group prior work into four categories—data-centric augmentation (A), measured-channel emulation (B), analogous methods in adjacent domains (C), and domain adaptation (D)—and contrast each with our proposed CMAT pipeline.

A. Data-Centric Augmentation for RF ML

Data augmentation is widely used in RF ML tasks such as modulation and waveform classification to simulate channel effects and expand training diversity. Classical techniques include additive white Gaussian noise (AWGN), Rayleigh and Rician fading [7], and frequency-selective channels modeled via random finite impulse response (FIR) filters [3]. The latter approach is analogous to classical channel emulators that shape white noise through FIR filters to produce the desired fading spectra, similar to the Jakes model for Rayleigh fading [8]. However, these methods rely on statistical abstractions rather than physical measurements, often limiting their realism and environment-specific fidelity.

In contrast, CMAT uses real measured channel impulse responses H to generate site-specific augmentations. It synthesizes all waveform classes as if they had propagated through a measured environment, enabling more faithful and physically grounded data generation. This approach bridges the gap between general-purpose statistical augmentation and environment-specific deployment needs.

B. Measured Channels, Emulation, and Equalization

In classical communications, measured channel models are used for physical-layer testing. Standardized modern channel profiles like the 3GPP Technical Specification [9] include aspects such as the Extended Pedestrian A model (EPA), Extended Vehicular A model (EVA), and Extended Typical Urban model (ETU), which are based on empirical delay spreads. Recent developments in *digital twin* simulations leverage site-specific measurements and ray tracing to emulate propagation conditions for virtual prototyping [10], offering a promising path for ML training data generation.

In RF ML, measured channels are typically used post hoc via equalization. For example, ORACLE [11] uses preamble-based equalization before classification, while Fu *et al.* [12] combine Least-Mean-Square (LMS) equalization with feature alignment for device identification. However, these techniques assume access to received signals collected directly from the target environment and depend on pilot symbols—constraints that are incompatible with blind or signal-agnostic tasks. CMAT differs by using measured H directly for data synthesis, without requiring labeled waveforms or pilots. It also avoids per-packet equalization during inference, instead building robustness through offline exposure to thousands of real CIRs.

C. Analogous Approaches in Adjacent Domains

In audio and speech ML, it is common to convolve clean audio with measured room impulse responses (IRs) to simulate reverberation [13]. Similarly, convolving audio with measured device-specific IRs arising from the frequency responses of different microphones improves generalization in device-robust acoustic scene classification [14]. These successes in other fields validate the core insight behind CMAT: using measured propagation CIRs improves generalization across waveforms.

D. Model-Centric Domain Adaptation

When augmentation is insufficient, domain adaptation techniques aim to learn channel-invariant features. Early methods fine-tuned synthetic-trained models on limited over-the-air data, improving generalization by up to 10% [7]. More recent approaches leverage adversarial training to learn model shortcomings ahead of time [15], or contrastive learning to bridge domain gaps [4], yielding 5–15% gains in unseen environments. These methods, however, still require access to target-domain waveforms.

III. CMAT Design: Channel Estimation

CMAT follows Fig. 1: (i) measure a reusable bank of phase-consistent channel impulse responses (CIRs), (ii) generate site-specific synthetic training data by replaying clean baseband waveforms through measured channels, and (iii) train a classifier. This section describes (i)–(ii) and the core design goals: **C1** stable, waveform-agnostic channel capture; **C2** faithful offline replay for diverse waveforms; and **C3** reuse of a measured channel bank across models/classes with generalization validated on OTA test sets (Section IV-A).

CMAT decouples *environmental capture* from *class capture*. For a deployment environment e, we measure a channel bank $\mathcal{H}^{(e)} = \{H_i\}_{i=1}^{M}$ once using an OFDM sounder. To train a classifier over classes $c \in \mathcal{C}$, we generate clean baseband waveforms $x \sim p_{\text{wave}}(x \mid c)$ offline and synthesize training samples by drawing $H \sim \text{Unif}(\mathcal{H}^{(e)})$ and applying a forward replay

$$\tilde{y} = \text{Replay}(x, H) \approx H * x + n.$$

No over-the-air recordings of the target classes in the environment e are required to generate training data. The same measured $\mathcal{H}^{(e)}$ can be reused across architectures and label spaces; adding a new class only requires regenerating synthetic x, not revisiting the site.

A. Wideband OFDM Channel Sounding (C1,C2)

We use a wideband OFDM packet with a sync preamble and a dense reference symbol (RS) enabling per-subcarrier channel estimation. In our implementation, we use an FFT size N=2048 at F_s=30.72 Msps (15 kHz spacing), a CP length N_{CP}=512, and an occupied RS bandwidth of $\approx$ 18 MHz (1201 active subcarriers with a DC null and wide guards). The key requirement for CMAT is that the sounding bandwidth exceeds the occupied bandwidth of all downstream waveform classes, ensuring that replay does not truncate spectra (**C2**).

We repeatedly transmit the sounding packet and record synchronized IQ samples (USRP X310 at 915 MHz in our setup). To preserve phase consistency across subcarriers (**C1**), we perform timing synchronization and coarse CFO correction using a CP-based CFO estimator, averaging multiple symbol-level estimates within a burst to reduce variance. The result is a stable per-packet channel frequency response (CFR) estimate $\hat{H}[k]$ that can be exported and reused.

B. Per-Subcarrier Channel Estimation

Let $X[k]$ and $Y[k]$ denote the known RS and its received spectrum on pilot subcarriers. We compute an LS CFR estimate:

$$\hat{H}[k] = \frac{Y[k]\, X^*[k]}{|X[k]|^2} \;\approx\; Y[k]\, X^*[k] \quad \text{(unit-modulus pilots)}, \tag{1}$$

and estimate the noise power from pilot residuals,

$$\hat{N}_0 = \text{mean}_k \left| Y[k] - \hat{H}[k]\, X[k] \right|^2. \tag{2}$$

We store $\hat{H}[k]$ (and optionally its CIR via IFFT) for each received packet to form the environment-specific bank $\mathcal{H}^{(e)}$.

C. Stable Forward Replay for Dataset Synthesis (C1,C2)

CMAT *replays* synthetic waveforms through measured channels rather than attempting to invert the channel. This avoids the well-known ill-conditioning of ZF-style division when $|\hat{H}[k]|$ is small (deep fades), which can amplify estimation error and noise (**C1**).

Given a synthetic time-domain example $x_{\text{synth}}[n] \in \mathbb{C}^{1024}$, we map it onto the sounder FFT grid (size N_f=2048 in our setup) and apply the measured CFR on the measured band. Let $\mathcal{K}$ denote the set of measured subcarriers (1201 bins in our implementation). We compute

$$Y_{\text{sim}}[k] \;=\; \hat{H}[k]\, X_{\text{interp}}[k], \qquad k \in \mathcal{K}, \tag{3}$$

optionally add band-limited AWGN $W[k] \sim \mathcal{CN}(0, \hat{N}_0)$ on $k \in \mathcal{K}$, and synthesize the replayed time-domain signal by inserting the simulated band and taking an IFFT:

$$\tilde{y}[n] \;=\; \text{IFFT}\{Z[k]\}, \qquad Z[k] = \begin{cases} Y_{\text{sim}}[k], & k \in \mathcal{K}, \\ X_{\text{interp}}[k], & \text{otherwise}. \end{cases} \tag{4}$$

We then downsample $\tilde{y}[n]$ back to the original sample rate/length used by the classifier. Fig. 2 shows that the replayed spectrum closely matches a true over-the-air reception, validating that replay transfers the measured passband ripple/selectivity while preserving waveform structure.

In summary, replaying clean waveform exemplars through a measured CIR bank produces label-consistent training data that reflects the target environment's propagation statistics, enabling "measure once, train often" across models and evolving class sets.

Algorithm 1: CMAT

Input: Measured CFR $\hat{H}[k]$ on measured subcarriers $\mathcal{K}$ ($\mathcal{K} = \{424, \ldots, 1624\}$ in our implementation); synthetic IQ $x_{\text{synth}}[n]$ with occupied BW $<$ sounding BW; FFT size N_f matching the sounder FFT; optional noise-power estimate $\hat{N}_0$ from Eq. (2).

Output: Simulated receive $\tilde{y}[n]$

1: **Map to N_f-bin grid:** $X_{\text{interp}} \leftarrow \text{FFT}_{N_f}\{\text{upsample\_to}_{N_f}(x_{\text{synth}})\}$.

2: **Initialize composite spectrum:** $Z \leftarrow X_{\text{interp}}$.

3: **Play through measured channel: for** $k \in \mathcal{K}$ **do**

  $Y_{\text{sim}}[k] \leftarrow \hat{H}[k]\, X_{\text{interp}}[k]$

4: **(Optional) Add band-limited complex AWGN in frequency: if** $\hat{N}_0$ *provided* **then**

  for $k \in \mathcal{K}$ **do**

    $Y_{\text{sim}}[k] \leftarrow Y_{\text{sim}}[k] + W[k], \quad W[k] \sim \mathcal{CN}(0, \hat{N}_0)$

5: **Insert simulated band: for** $k \in \mathcal{K}$ **do**

  $Z[k] \leftarrow Y_{\text{sim}}[k]$

6: **Synthesize time signal:** $\tilde{y}[n] \leftarrow \text{IFFT}_{N_f}\{Z\}$.

7: **Return to original length:** $\tilde{y}[n] \leftarrow \text{downsample}(\tilde{y}[n])$.

Note: Steps 3–6 avoid division by small $|\hat{H}[k]|$ (mitigating **C1**) and preserve constellation geometry and phase continuity (**C2**). As an alternative to Step 4, complex white noise can be added in time after Step 6: $\tilde{y}[n] \leftarrow \tilde{y}[n] + w[n], \quad w[n] \sim \mathcal{CN}(0, \sigma^2)$.

IV. Evaluation

In this section, we evaluate CMAT under deployment distribution shift, where the training data is synthetic and the test data is over-the-air. We (i) quantify architecture-agnostic gains across CNN, RNN, T-PRIME, and MCformer in four environments; (ii) benchmark CMAT against common statistical augmentations (AWGN, Rayleigh/Rician, RandomFIR, and a hardware-impairment model) using both default and environment-tuned parameterizations; (iii) study measurement efficiency by varying the number of CIR snapshots and relate learning curves to a PCA analysis of the CIR bank; and (iv) demonstrate adding a new class (802.11b) without recollecting class-specific OTA training data by reusing the same measured CIR library.

In all experiments, training uses only (i) clean synthetic waveforms and (ii) measured CIR banks from the OFDM sounder; no over-the-air recordings of the target classes are used for training data generation. Over-the-air class recordings are collected only for testing to quantify the synthetic-to-real gap. Unless noted otherwise, we hold training budgets fixed and report mean accuracy with variability over multiple evaluations (25×1000 examples per class per SNR).

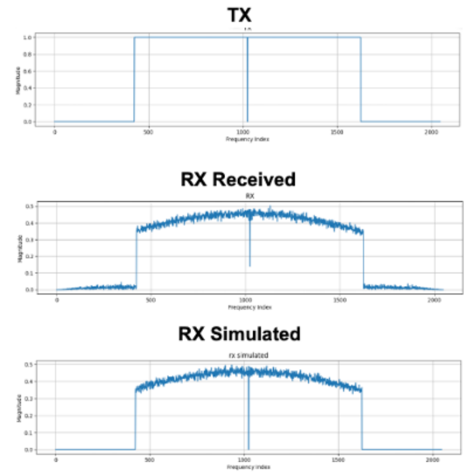


Fig. 2. **Verification of synthesis.** Top: TX spectrum of a synthetic waveform. Middle: True over-the-air RX spectrum measured. Bottom: RX simulated by applying the measured $\hat{H}[k]$ to the synthetic TX via Eq. (3). The close match in passband shape and ripple validates the CMAT play-through.

A. Architecture-Agnostic Gains with CMAT

Goal: To quantitatively analyze whether CMAT's benefits persist across different model families, we evaluate four representative architectures: a lightweight convolutional network (CNN) [7], a recurrent neural network (RNN) [16], a transformer variant for 1-D IQ input (T-PRIME) [17], and the multi-channel transformer (MCformer) [18]. We use open-source implementations and wireless datasets, making only minimal adjustments to load data in sigmf format. For each model, we train a baseline with no augmentation and a CMAT-trained version using synthetic signals augmented with measured CIRs. Measurements are taken once per site and then reused for all classes across all models. Models are evaluated in four test environments (Wired, OTA2, OTA10, Indoor-Lab) using 25 batches of 1000 test examples per class to capture run-to-run variance. Performance is reported as average accuracy versus SNR (Figure 3, where solid lines indicate CMAT, dotted lines indicate no augmentation, and shaded regions represent $\pm 1\sigma$).

Findings: CMAT consistently improves accuracy across all architectures and SNR ranges, with the largest gains occurring in the 0–15 dB region where real-world performance typically transitions from chance to reliable. T-PRIME performs best overall, with CMAT producing a steep accuracy rise of 51%, suggesting a strong synergy between attention mechanisms and physically grounded channel variability. CNN models also benefit substantially: despite their smaller capacity, CMAT lifts their high-SNR accuracy to competitive levels, making them attractive for resource-constrained deployments. MCformer and RNN follow similar improvement patterns, escaping low-accuracy plateaus typical of non-augmented training and continuing to improve with SNR. However, the RNN shows the smallest relative benefit from CMAT, moving from the highest baseline performance to the lowest augmented performance among the four models.

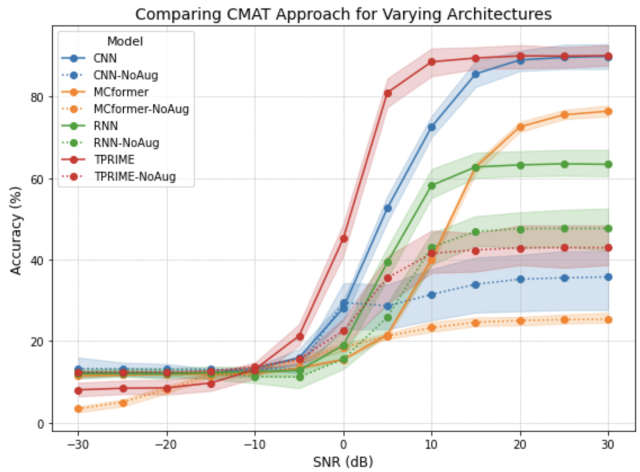


Fig. 3. Average accuracy vs. SNR across Wired, OTA2, OTA10, and Indoor-Lab for multiple architectures. Solid lines use CMAT; dotted lines are No-Aug baselines. Shaded bands show variability across environments. CMAT consistently raises performance, with the largest gains near the 0–15 dB transition region.

CMAT-Configuration Takeaways: The comparison demonstrates that CMAT is architecture-agnostic and consistently narrows the synthetic-to-real gap regardless of whether the model is convolutional, recurrent, or transformer-based. In practice, CMAT should be paired with the strongest available model, such as T-PRIME, when absolute accuracy is the top priority. At the same time, compact CNNs trained with CMAT deliver robust performance even when compute or latency constraints limit model complexity.

B. Comparison to State of the Art RF Data Augmentation Techniques

Goal: To isolate the effect of augmentation, we select one of the above-mentioned architectures (T-PRIME), with the associated training budget and datasets, and vary only the data–augmentation strategy used for training. We compare CMAT against common practice baselines: AWGN, Rayleigh [19], Rician [19], RandomFIR [3], Hardware Impairments [20], and a No-Aug baseline. We evaluate on real OTA captures from our indoor-lab environment ("OTA2") across an SNR sweep. Shaded bands in the plots indicate the minimum and maximum of the accuracy across 25 runs at each point for statistically correct results.

Findings: With default MATLAB parameters (not shown due to space), CMAT performs the best across most of the SNR range. The Impairments baseline is the most competitive statistical approach at moderate to high SNR, but it saturates well below CMAT. In contrast, Rayleigh, Rician, and RandomFIR cluster together and plateau at mid-range accuracies even at high SNR, while the AWGN baseline performs only slightly better than No-Aug beyond low SNR.

When environment-tuned parameters are used, as shown in Figure 4, statistical baselines improve significantly. The retuning involves adjusting the number of taps and tap power normalization for RandomFIR to match delay-spread, as well as modifying K-factor and flatness ranges for Rician and Rayleigh fading. This tuning shifts RandomFIR, Rayleigh, and Rician upward across the SNR sweep and leads to a substantially higher high-SNR plateau, closing much of the gap to CMAT. Even after tuning, however, CMAT maintains a consistent advantage in the high-SNR regime. We attribute this to site-specific phenomena captured by the measured channels' fine-grained delay spectra, and hardware or propagation interactions that are difficult to reproduce with a small number of statistical parameters. While tuned statistical models approximate the macro statistics of OTA2, CMAT transfers the micro-structure that the classifier ultimately exploits.

A surprising outcome arises from the AWGN baseline. Contrary to prior experience in controlled or near-flat channels such as coax or anechoic setups, AWGN augmentation underperforms in OTA2. Our hypothesis is that OTA2 exhibits pronounced frequency selectivity and other distortions that AWGN cannot emulate, and that AWGN encourages the model to rely on sharp features that become smeared under multipath, thereby harming generalization in reflective indoor spaces. Despite these results, AWGN remains a valuable tool for regularization, and we plan to continue including it in future model development.

CMAT, which uses measured channel responses H, achieves the highest accuracy and site fidelity and requires only a brief on-site sounding sweep without class-specific OTA data collection. RandomFIR, Rayleigh, and Rician, when tuned, provide strong general-purpose baselines that can operate across sites without measurements. However, their default parameters can be misleadingly weak, as they capture only broad multipath statistics and miss site-specific channel structure. The Impairments method is functionally similar to RandomFIR depending on parameterization and is particularly valuable when hardware variability is a dominant factor or when known hardware characteristics are involved. AWGN can still be useful in near-flat or highly controlled channels if used as the sole data-augmentation strategy, though it works best in combination with other methods for regularization. Additional strategies such as curriculum learning or fine tuning can prevent accuracy loss at high SNR while improving low-SNR performance.

CMAT-Configuration Takeaways: Off-the-shelf defaults materially understate the capability of statistical augmentation, whereas environment-aware tuning yields large performance gains. Despite these gains, CMAT retains a high-SNR advantage, consistent with its goal of teaching the model site-specific channel information. When a brief sounding sweep of an environment is feasible, CMAT offers the best accuracy-to-effort trade-off. When such measurements are not possible, tuned statistical models provide a practical alternative.

C. Scaling Number of Channel Measurements

Goal.: Measuring channels is inexpensive compared to full OTA class collection, but it is not free. We therefore ask how many CIR snapshots are needed to unlock most of CMAT's benefit. Because this study is compute-intensive, requiring a separate train+test cycle for each point, we evaluate three

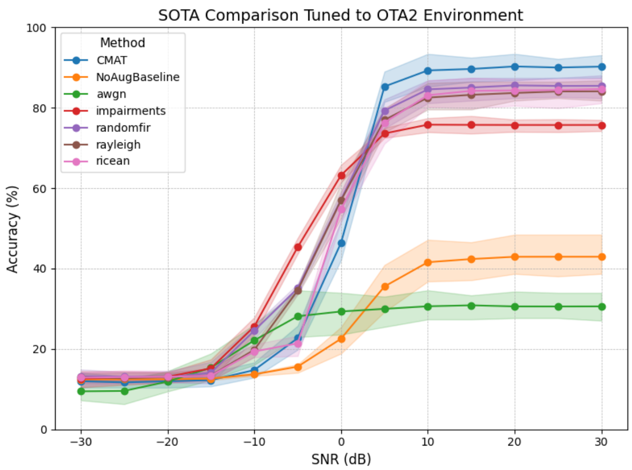
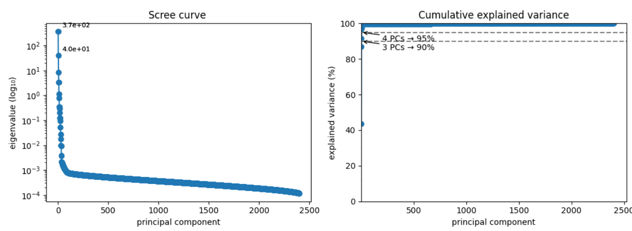
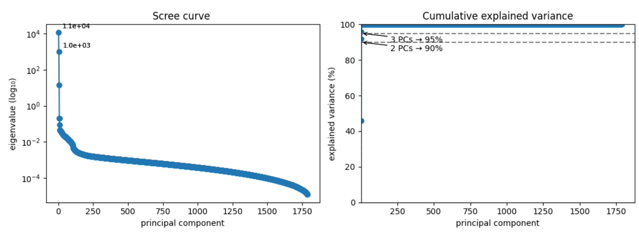


Fig. 4. SOTA comparison in OTA2 after *tuning* RandomFIR, Rayleigh, and Rician parameters to roughly match indoor-lab statistics. CMAT still leads at high SNR.



(a) PCA Analysis: Indoor OTA lab



(b) PCA Analysis: Wired reference link

Fig. 5. PCA eigen–spectra of the measured CIR banks. Both plots use a log-scaled y-axis. Dashed lines mark 90% and 95% cumulative-variance thresholds.

representative scenarios (Fig. 6): (i) T-PRIME in a Wired link to represent a simple, nearly deterministic channel; (ii) T-PRIME in an Indoor-Lab environment to characterize a complex multipath setting; and (iii) CNN in the Indoor-Lab to examine architecture sensitivity under the same propagation conditions. For each x-axis value, we train a model using exactly that many measured CIRs: `0` corresponds to purely synthetic data (baseline), `1` uses a single randomly selected CIR, `100` draws from hundred random CIRs. During training, each example is stochastically convolved with one of the available CIRs.

Findings.: The first few CIRs deliver the largest gains. In the Indoor-Lab with T-PRIME, accuracy rises from the synthetic baseline of roughly 40% to approximately 60–70% with only one to five CIRs, after which performance increases more gradually. The propagation environment has a greater influence than the model architecture: the Wired curve (T-PRIME) starts high and saturates near the low-90% range by about one hundred CIRs, while the Indoor-Lab curves saturate lower and require more snapshots before improvements diminish (T-PRIME reaches roughly 80% accuracy by about 100–250 CIRs, while CNN reaches about 65% after several hundred CIRs). Variance at very low CIR counts is noticeable, consistent with sensitivity to which specific multipath realizations are selected.

To explain these front-loaded gains and the environmental dependence, we applied PCA to the measured CIR banks (Fig. 5). For the Indoor-Lab, the first three principal components dominate the spectrum, already explaining 90% of the variance ($d_{90} = 3$), and adding a fourth increases coverage to 95% ($d_{95} = 4$). By contrast, the coaxial Wired link collapses even further: a single basis captures most energy ($\lambda_1 \approx 1.1 \times 10^4$), adding a weak echo ($\lambda_2 \approx 10^3$) that reaches 90%, and a third completes 95% coverage ($d_{90} = 2$, $d_{95} = 3$).

The ratio of independent principal components to the total number of channel measurements is small in both environments, indicating that only a modest set of representative CIR snapshots is required to span the dominant variability. As a practical rule of thumb, collecting on the order of $5 \times d_{90}$ snapshots provides a good minimal solution for resource-constrained settings, with additional CIRs offering diminishing returns. This conclusion aligns with the accuracy–vs–CIR-count learning curves in Fig. 6.

Taken together with the PCA analysis, several collection guidelines emerge. Starting small, with five to twenty-five CIRs, already captures most day-to-day variability in typical rooms and yields substantial improvements over purely synthetic training. Using a minimum of roughly $5 \times d_{90}$ snapshots ensures adequate coverage of the dominant channel subspace, and if resources permit, collecting more is generally beneficial — we found about two thousand snapshots to be a conservative "one-size-fits-most" number. Collection effort should scale with environmental complexity: simple or wired quasi-static sites reach performance saturation with about 50–100 CIRs, whereas cluttered or dynamic indoor spaces benefit from roughly 100–250 CIRs, particularly when supporting higher-order modulations.

CMAT-Configuration Takeaways.: Only a handful of CIR snapshots are required to achieve the majority of CMAT's benefits because measured channels often occupy a low-dimensional subspace that can be well approximated with a small representative bank. The intrinsic dimensionality d_{90} of an environment can be quickly estimated via PCA on a pilot collection, and gathering about $5d_{90}$ snapshots provides a strong baseline. Beyond this level returns diminish, but if resources are not constrained, larger collections (hundreds to thousands of CIRs) provide an extra safety margin and robustness.

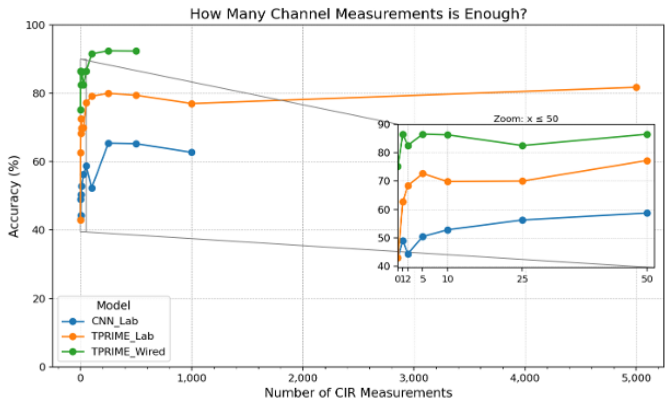


Fig. 6. number of CIR used during model training vs accuracy in target environment.

D. Adding New Classes Without Field Recollection

Goal: A practical barrier to deploying RF classifiers is the need to recollect labeled data in a target environment whenever a new signal class is introduced. CMAT sidesteps this cost: once a site's channel impulse responses are measured, any new waveform class can be synthesized offline by filtering clean baseband exemplars with the measured channels, requiring no additional over-the-air class collection at that site.

Findings: To demonstrate this capability, we add 802.11b as a ninth class to our modulation-recognition task. This class is intentionally different from the eight originals (64QAM, 16QAM, 8PSK, QPSK, BPSK, PAM4, GFSK, CPFSK): our 802.11b implementation employs direct-sequence spread spectrum (DSSS) with DBPSK at 1 Mb/s. We generate clean 802.11b baseband, convolve it with the Indoor lab environment CIR library, and retrain using the same T-PRIME architecture and budget. We then evaluate by capturing new unseen test data in our indoor lab environment and mixing it with the previous test dataset. Across the SNR sweep (not shown due to space), overall accuracy remains essentially unchanged relative to the original 8-class setting while enabling recognition of the new 802.11b class in the same environment.

Figure 7 shows the confusion matrix at 30 dB SNR for the 9-class model. The new 802.11b class achieves strong recall (e.g., ∼82%) with limited leakage to other classes, while the existing classes preserve their characteristic error patterns (e.g., residual ambiguity between high-order QAMs).

CMAT-configuration takeaways: CMAT converts one-time channel measurements into a reusable asset: after measuring an environment once, practitioners can expand the label space by adding new waveform classes while still obtaining strong performance on unseen OTA data from that site. This property is attractive for deployments where the set of relevant signals evolves over time while the location remains constant.

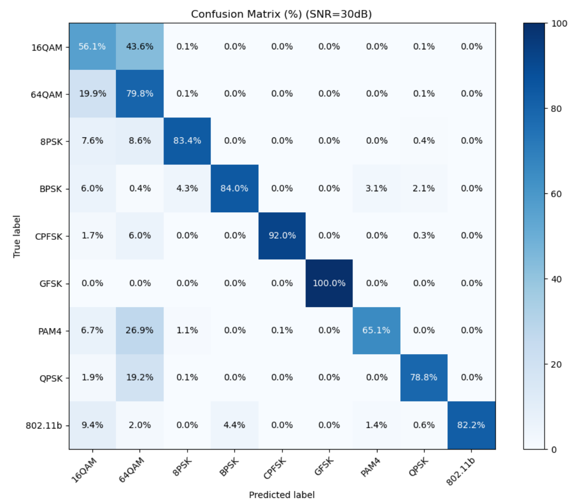


Fig. 7. Confusion matrix (%) at 30 dB SNR for the 9-class model. The 802.11b row/column shows high recall with limited confusion, and the existing classes retain their prior structure. CMAT adds classes without broad degradation and without re-collection of class specific training data.

V. Conclusion

We introduced Channel-Measurement–Augmented Training (CMAT), a data-centric pipeline that decouples environment capture from class collection. A brief OFDM sounding generates a reusable library of measured CIRs, and any clean waveform can then be rendered through these CIRs offline to produce site-specific training data. Across eight modulation classes, multiple neural architectures, and multiple distinct test environments, CMAT consistently improves synthetic-to-real performance. We further show that CMAT is measurement-efficient: a relatively small number of CIR snapshots can capture dominant channel variability, as supported by PCA analysis and accuracy–vs–CIR learning curves. Finally, CMAT enables extensibility by adding a new class (DBPSK:802.11b) using previously measured channels, without class-specific OTA recollection. At scale, we envision CIR libraries becoming a shareable data asset for training robust RF models, analogous to room impulse responses in speech, enabling rapid adaptation without repeated OTA collections.

Acknowledgment

This work has been supported in part by National Science Foundation grants 2434043, 2431961, and 2526493.

as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

References

[1] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*. Springer, 2016, pp. 213–226.

[2] L. Huang, W. Pan, Y. Zhang, L. Qian, N. Gao, and Y. Wu, "Data augmentation for deep learning-based radio modulation classification," *IEEE access*, vol. 8, pp. 1498–1506, 2019.

[3] N. Soltani, K. Sankhe, J. Dy, S. Ioannidis, and K. Chowdhury, "More is better: Data augmentation for channel-resilient rf fingerprinting," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 66–72, 2020.

[4] M. Du, J. Pan, and D. Bi, "A contrastive learner for automatic modulation classification," *IEEE Transactions on Wireless Communications*, 2025.

[5] Y. Wang, L. Zhou, Z. Yang, L. Wu, Z. Yin, Y. Zhao, and Z. Wu, "An improved modulation recognition algorithm based on fine-tuning and feature re-extraction," *Electronics*, vol. 12, no. 9, p. 2134, 2023.

[6] Trudes808, "Measure-once-train-often," Aug. 2025, gitHub file (README.md); commit d2751c5f9fb250d2a4a94fc0d31aa01380d9da23. [Online]. Available: https://github.com/Trudes808/Measure-Once-Train-Often/blob/d2751c5f9fb250d2a4a94fc0d31aa01380d9da23/README.md

[7] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[8] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[9] 3rd Generation Partnership Project (3GPP), "Evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception," ETSI, ETSI Technical Specification ETSI TS 136 104, May 2024, 3GPP TS 36.104 Release 18. See Annex B: Propagation conditions (EPA/EVA/ETU). [Online]. Available: https://www.etsi.org/deliver/etsi_ts/136100_136199/136104/18.05.00_60/ts_136104v180500p.pdf

[10] H. Luo and A. Alkhateeb, "Digital twin aided compressive sensing: Enabling sitespecific mimo hybrid precoding," in *2024 58th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2024, pp. 17–21.

[11] K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, "Oracle: Optimized radio classification through convolutional neural networks," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 370–378.

[12] H. Fu, L. Peng, M. Liu, and A. Hu, "Deep learning-based rf fingerprint identification with channel effects mitigation," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1668–1681, 2023.

[13] A. Richard, P. Dodds, and V. K. Ithapu, "Deep impulse responses: Estimating and parameterizing filters with deep networks," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3209–3213.

[14] T. Morocutti, F. Schmid, K. Koutini, and G. Widmer, "Device-robust acoustic scene classification via impulse response augmentation," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 176–180.

[15] M. Usama, M. Asim, J. Qadir, A. Al-Fuqaha, and M. A. Imran, "Adversarial machine learning attack on modulation classification," in *2019 UK/China Emerging Technologies (UCET)*. IEEE, 2019, pp. 1–4.

[16] liuzhejun, "Resnet-for-radio-recognition," Jun. 2019, gitHub repository; commit 3818778. [Online]. Available: https://github.com/liuzhejun/ResNet-for-Radio-Recognition

[17] M. Belgiovine, J. Groen, M. Sirera, C. Tassie, A. Y. Yildiz, S. Trudeau, S. Ioannidis, and K. Chowdhury, "T-prime: Transformer-based protocol identification for machine-learning at the edge," *arXiv preprint arXiv:2401.04837*, 2024.

[18] InterDigitalInc, "Fireball: Mcformer (jupyter notebook)," Jul. 2023, gitHub repository; file at commit 8c98a40e6baba489ac9c028aa4fe71b2ae782f79. [Online]. Available: https://github.com/InterDigitalInc/Fireball/blob/8c98a40e6baba489ac9c028aa4fe71b2ae782f79/Playgrounds/MCformer/MCformer.ipynb

[19] *Communications Toolbox*, https://www.mathworks.com/products/communications.html, The MathWorks, Inc., Natick, MA, USA, 1996, [Online; accessed 28-Feb-2025].

[20] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, and K. Chowdhury, "No radio left behind: Radio fingerprinting through deep learning of physical-layer hardware impairments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 165–178, 2019.