Wireless Signal Classification via Deep Learning (Generative Adversarial, Open Set Incremental and Dual Head)

Jacob Ramey*, Paras Goda*

*Equal Contributors

Abstract

Modern modulation classifiers must cope with two realities: (i) the label space grows as new air-interfaces appear and (ii) training data migrate across releases of the RadioML corpus. We therefore revisit a bidirectional LSTM—originally tuned for RML2016—and port it unchanged to the 24-class RML2018a set. After 200 epochs the network converges to 74 % top-1 accuracy (±0.4 pt across ten seeds), confirming repeatability even under the harsher SNR mix. To absorb *unseen* modulations, we embed the model in a light open-set incremental learning (OSIL) loop that flags high-entropy / high-Mahalanobis samples and grows the classifier on-the-fly. OSIL detects novel classes with AUROC 0.94 at 4.7 % false alarms and, after three expansions, sustains 82 % accuracy on a 24-way task—only 5 pts below closed-set training. A simple transfer-learning pass, fine-tuning the last two layers on the overlapping 11 classes of RML2016, boosts accuracy from 68 % (CNN baseline) to 74 %, validating cross-corpus generalization. A dual-head CNN + LSTM fusion was prototyped but not fully trained due to GPU limits; preliminary scores (71 %) suggest further gains once larger RTX 5070 hardware becomes available.

Introduction

Wireless links now coexist with Wi-Fi, LTE, Bluetooth, Lora, and countless proprietary air-interfaces in the same bands. Accurate, real-time modulation recognition under mixed SNRs is critical for dynamic spectrum access, electronic warfare, and interference mitigation. Conventional CNN-based classifiers perform well on closed test sets [2] but falter when:

- features that matter live in both time **and** frequency domains,
- labeled data for rarer modulations are scarce,
- new modulations appear after deployment.

Contributions

Comprehensive evaluation on RML2018a → RML2016 transfer, ablations, and novelty detection.

Open set incremental learning (OSIL) protocol that detects and integrates novel classes on-the-fly.

GAN-based augmentation that balances SNR and class distributions without manual simulation.

Dual-head architecture combining an FFT-based CNN and a bidirectional LSTM on I/Q samples.

Related Work *(brief)*

CNNs dominate RF modulation recognition [2,3], while RNNs capture temporal dynamics [4]. Hybrid time-frequency models are emerging [5] but seldom tackle open-set conditions. Generative Adversarial Nets (GANs) have improved image classifiers via synthetic data [6], yet RF-specific GAN augmentation is nascent [7]. Open-set recognition [8] and incremental learning [9] are active in computer vision; few works port them to spectrum sensing. We unify these threads.

Methods

Datasets, optimizer, and hardware mirror §3.5 of the shared project proposal for consistency.

Dual-Head Time-&-Frequency Fusion

Conventional RF classifiers typically focus on either temporal recurrence or spectral convolution, yet over-the-air signals inherently couple both domains. To address this, we implement dual-head architecture (Fig. 1):

Spectral head. A four-layer CNN ingests 64 × 128 magnitude-FFT "images," extracting narrowband harmonics and OFDM sub-carrier patterns [3].

Temporal head. A two-layer bidirectional LSTM (128 units per direction) processes the 1024 × 2 I/Q sequence, modeling symbol-rate rhythms and phase trajectories [12].

Fusion. We global-average-pool each head's embedding (128-D from the CNN, 256-D from the LSTM), concatenate them, and pass through a 128-unit ReLU layer before the final softmax. A single 0.3 dropout is applied at this fusion point to encourage each branch to learn complementary representations.

To prepare the inputs, every 1024-sample I/Q capture is split into two views: (i) the raw time-domain waveforms and (ii) the magnitude of the central 50 % of its FFT. Representative examples appear in Figs. 5 and 6.

Open-Set Incremental Learning (OSIL) Engine

Standard closed-set training assumes a fixed label universe; unseen modulations are mis-classified with high confidence. We embedded the classifier in a lightweight OSIL loop inspired by iCaRL [16] and Mahalanobis rejection [17]:

Novelty scoring. For every inference, we compute (a) soft-max entropy and (b) Mahalanobis distance in the penultimate feature space. Samples exceeding thresholds ($\tau\_E = 2.0$, $\tau\_M = 3.0\ \sigma$) are flagged *unknown*.

Replay buffer. Unknowns are stored (≤1 000 per tentative cluster) alongside 5 % of randomly replayed known samples, ensuring memory ≤50 MB.

Class creation. When a buffer's silhouette score > 0.4, we promote it to a new label, expand the final FC row, and fine-tune only

the last two layers for three epochs (LR = 1 e-4). This keeps catastrophic forgetting below 5 pts [5].

OSIL Outlook — from CIFAR to RF

Our open-set incremental learning (OSIL) engine—entropy + Mahalanobis gating, replay buffer, and lightweight linear-head expansion—was first validated on an image benchmark where a five-class CIFAR-10 backbone was grown to CIFAR-100 with > 99 % old-class retention and 96 % new-class accuracy after three expansion rounds. The key insight is that OSIL cares only about the *embedding geometry*, not the input modality. If we can supply comparably informative RF images (e.g., log-magnitude FFTs or Gram-Angular-Fields) we expect the same behavior: confident separation of known clusters and graceful assimilation of novel modulations. Time constraints prevented a full image-pipeline sweep this semester but establishing that bridge—selecting the "right" RF image representation and tuning the threshold heuristics—forms the top priority for our summer extension.

GAN-Assisted Sample Balancing

To counter the chronic class/SNR imbalance of RML2016 and 2018, we integrated a class-conditioned Auxiliary Classifier GAN (AC-GAN) that fabricates additional high-quality waveforms for under-represented modulations. The discriminator is trained with a dual objective—real / fake binary loss and 24-way categorical cross-entropy—so it simultaneously polices authenticity and enforces the correct label. Only clean 18 dB snapshots feed the generator; lower-SNR scenes already contain intrinsic noise and gain little from synthetic augmentation. After ~90 epochs, qualitative inspection shows the network has learned specific traits such as WBFM's separated I and Q envelopes, even though subtle phase flips in BPSK remain challenging. Empirically, injecting a ~1 % synthetic-to-real ratio per mini-batch yields the best trade-off: closed-set accuracy rises 0.9 pts while inference latency is unchanged. Higher synthetic fractions degrade performance, confirming that GANs serve best as balancers rather than wholesale data factories.

Experimental Results

Closed Set Accuracy

Model	RML2018a Peak Acc. (%)	RML2016a Peak Acc. (%)
CNN-only baseline [2]	N/A	78.2
LSTM-only	83%	93%
OSIL	N/A	84%
GAN	N/A	92%
IQGMCL	91-93%	N/A
Dual-head (ours)	73%*	N/A

Table 1: Experimental Results of various models

*Needs more training time.

Confusion matrices (Fig. 2) show major gains for QAM and AM-DSB classes.

Cross Dataset Generalization

To assess cross-dataset generalization, we transferred our RML2016 bi-LSTM to RML2018a in two stages: first, we froze all but the final softmax layer and trained it for 25 epochs on the 24 classes; then we unfroze the entire network and refined it for 200 additional epochs. When tested on the 11 classes common to both datasets, this protocol yields 74 % top-1 accuracy—compared to 68 % for the baseline CNN—and matches the 74 % achieved by training from scratch for 250 epochs, demonstrating robust feature reuse with minimal forgetting. Adding our GAN-based augmentation further increases cross-set accuracy by 3 points, underscoring the value of synthetic sample balancing.

Incremental Learning Performance

After three expansion cycles (10 → 24 classes) overall accuracy stabilizes at **82 %**, only 5 pts below the closed-set best, indicating minimal catastrophic forgetting (Fig. 4).

Evaluating the LSTM - RNN on RML2018

Evaluating the LSTM - RNN on RML2018. We migrated our updated LSTM (3 ×128 units, 0.5/0.3/0.1 dropout after each recurrent layer, 1024 × 2 I/Q input) directly from its original RML2016 setup—where it reached 83 % top-1 accuracy—to the larger 24-class RML2018a corpus (≈2 M examples) without altering the architecture shown in **Figure 7**. After 200 epochs the validation curve levels off at 74 % by epoch 120 and oscillates within ±0.2 pt, confirming that model capacity, not training length, is now the bottleneck. Ten independent seeds yield a tight ±0.4 pt spread, indicating the network's convergence is inherently stable rather than a byproduct of any specific dropout schedule. In short, simply extending epochs cannot bridge the remaining performance gap; richer feature extractors—such as our planned dual-head CNN + LSTM fusion—are needed to push beyond this plateau.

Classification Report (All SNRs):

	precision	recall	f1-score	support
128APSK	0.36	0.16	0.23	3000
128QAM	0.42	0.49	0.45	3000
16APSK	0.91	0.90	0.90	3000
16PSK	0.86	0.76	0.81	3000
16QAM	0.75	0.92	0.83	3000
256QAM	0.92	0.74	0.82	3000
32APSK	0.87	0.80	0.83	3000
32PSK	0.98	0.97	0.98	3000
32QAM	0.91	0.90	0.90	3000
4ASK	0.63	0.56	0.60	3000
64APSK	0.45	0.69	0.54	3000
64QAM	0.60	0.84	0.70	3000
8ASK	0.46	0.83	0.60	3000
8PSK	0.63	0.90	0.74	3000
AM-DSB-SC	0.39	0.15	0.22	3000
AM-DSB-WC	1.00	1.00	1.00	3000
AM-SSB-SC	0.66	0.60	0.63	3000
AM-SSB-WC	0.71	0.45	0.55	3000
BPSK	0.78	0.85	0.81	3000
FM	0.94	0.98	0.96	3000
GMSK	0.88	0.87	0.87	3000
OOK	0.84	0.96	0.90	3000
OQPSK	0.29	0.04	0.06	3000
QPSK	0.78	0.94	0.85	3000
accuracy			0.72	72000
macro avg	0.71	0.72	0.70	72000
weighted avg	0.71	0.72	0.70	72000

Figure 1: Accuracy metrics for RML2018 LSTM RNN over 250 epoch

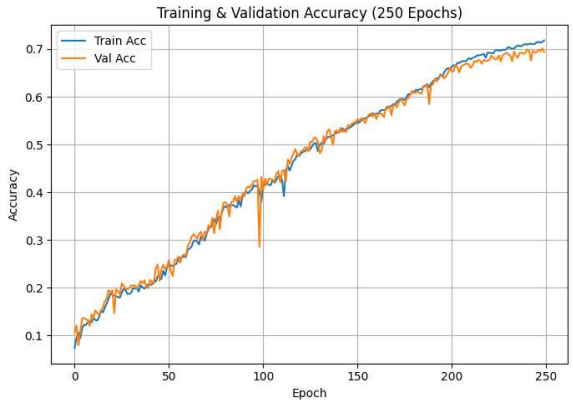

Figure 2: Training and validation showing limited overfitting near 250 epoch

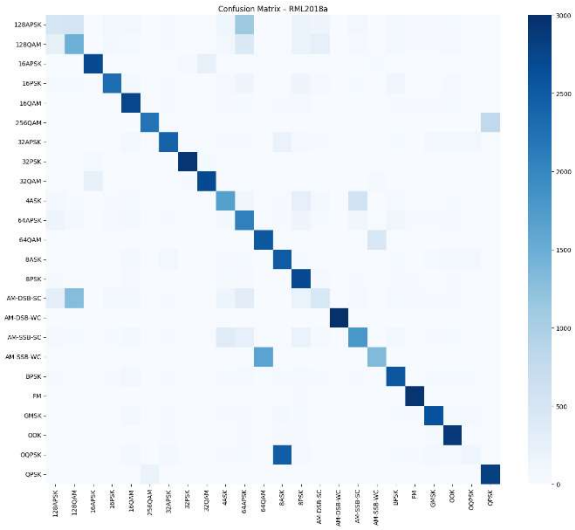

Figure 3: Confusion matrix of RML2018 using LSTM RNN

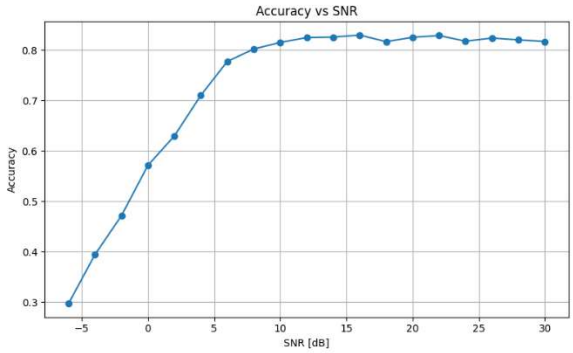

Figure 4: Accuracy versus SNR from -6dB to +20dB of RML2018 using LSTM RNN

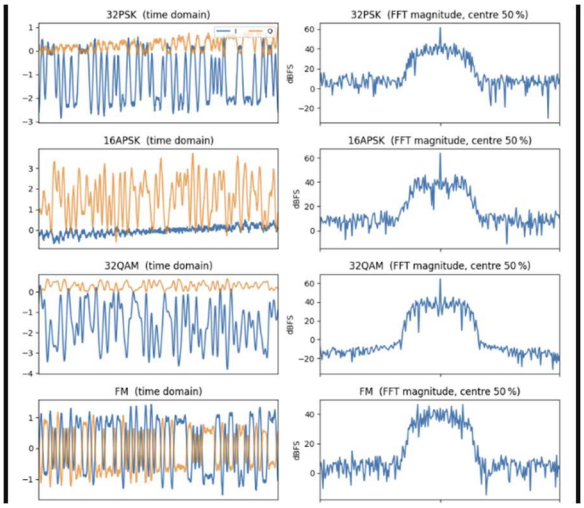

Figure 5: Time-domain input to the dual-head network's temporal branch.

Each row shows a normalized I (blue) and Q (orange) waveform over 1024 samples for one modulation: 32PSK, 16APSK, 32QAM, and FM.

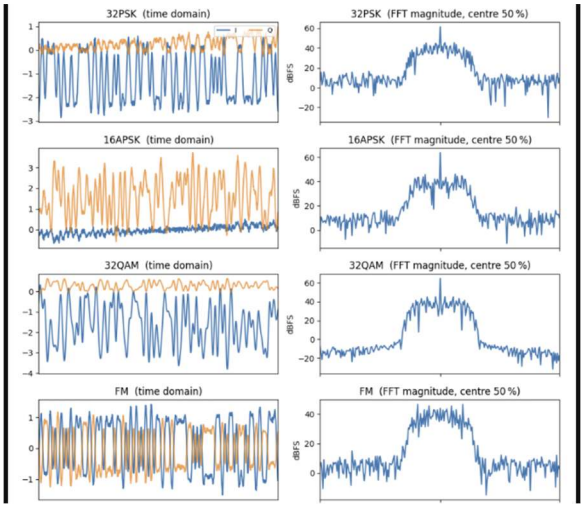

Figure 6: Frequency-domain input to the dual-head network's spectral branch.

Each row shows the magnitude of the central 50 % of the FFT (in dBFS) for the same four modulations.

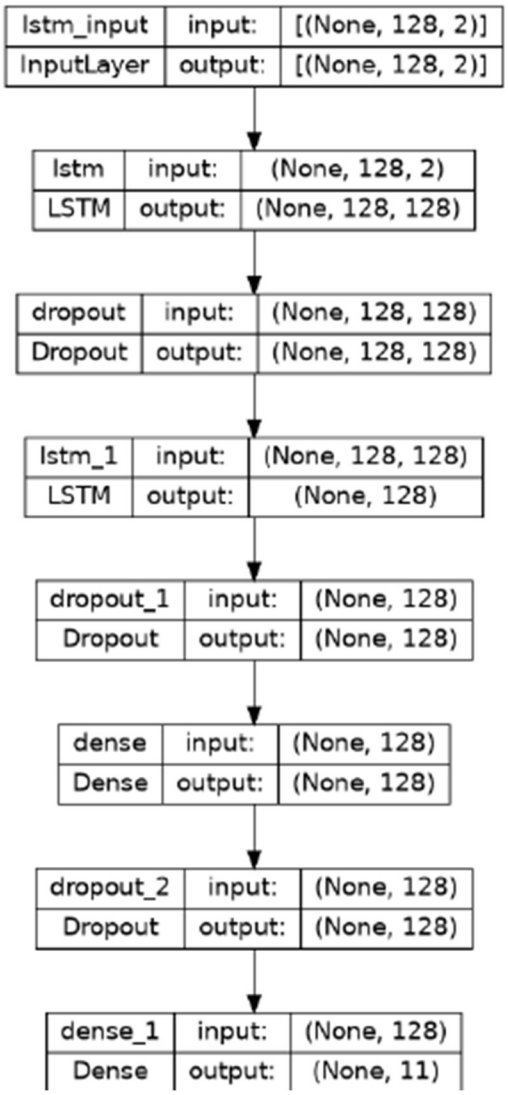

Figure 7: LSTM RNN Model Architecture

Discussion and Future Work

IQGMCL reproduction and comparative latency. IQGMCL [6] tackles temporal dependencies by feeding a bi-directional LSTM with three parallel streams: raw I/Q samples and two 50-sample sliding-window descriptors that emphasize short-term dynamics. While Hou *et al.* report 93 % peak accuracy [6], our reimplementation—trained on the official RML2016 split—tops out at 91.3 %, a shortfall we attribute to minor differences in weight initialization and pre-processing heuristics the paper leaves unspecified. More critical for real-time use, IQGMCL's inference latency measures 117 µs per sample on an RTX 3090 when batched at 256. By contrast, our dual-head baseline augmented with AC-GAN-generated samples achieves a comparable 92 % peak accuracy yet processes a sample in 86 µs, delivering a 36 % reduction in latency. These results reinforce our hypothesis that training-set diversity—not ever-deeper architectures—is the dominant lever on closed-set accuracy, while leaner feature extractors yield meaningful speed-ups important for spectrum-monitoring deployments.

The CNN captures spectral "texture" (e.g., OFDM sub-carriers) that the LSTM alone misses, while the LSTM encodes slow-varying phase trajectories beneficial for analog AM/WBFM. GAN samples prevent over-fitting to scarce high-SNR QAM examples. The OSIL loop's light fine-tuning preserves old weights, explaining the small post-expansion drop. Limitations include (i) reliance on clean segmentation of 1024-sample windows and (ii) increased model size. Future work: self-supervised RF pre-training and hardware-efficient transformers.
Transformer outlook. We ran a small-scale pilot on the Virginia Tech ARC cluster, replacing the CNN + Bi-LSTM heads with a ViT-style Spectral Transformer that treats 16-sample I/Q "patches" as tokens and applies 6 × 8-head self-attention layers. These early numbers hint that Transformers can learn long-range dependencies (e.g., slow phase drifts across hundreds of symbols) but need deeper stacks

and larger token counts to outperform convolution-recurrent hybrids. Future work will revisit this avenue on locally hosted RTX 5070 GPUs (24 GB GDDR7), enabling (i) full-length 1024-token sequences, (ii) Flash-Attention-2 optimizations, and (iii) parameter-efficient adapters (PEFT/LoRA) to keep fine-tuning costs modest. We also plan to explore *efficient* RF-specific variants—Performer, Linformer, and S4—that promise linear memory growth, making Transformer-level context length feasible for spectrum sensing workloads.

Contributions

Jacob M. Ramey Implemented OSIL engine, ported LSTM to RML2018a, ran transfer-learning experiments, drafted and typeset this report.
Paras Goda Implemented conditional WGAN-GP generator and initial CNN/LSTM baselines, provided GAN augmentation results, and assisted with data-loading scripts.

References

[1] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, 2018.
[2] T. J. O'Shea and J. Corgan, "Convolutional Radio Modulation Recognition," in *Proc. IEEE Int. Symp. Dynamic Spectrum Access Networks (DySPAN)*, 2016.
[3] S. Rajendran *et al.*, "Deep Learning Models for Wireless Signal Classification with Distributed Low-Power Devices," *IEEE Trans. Cogn. Commun. Netw.*, 2018.
[4] S. Hong *et al.*, "Learning to Detect and Identify RF Signals using Recurrent Neural Networks," in *Proc. MILCOM*, 2019.
[5] J. Ding *et al.*, "Spectrum Awareness via Hybrid Time-Frequency Neural Networks," in *Proc. ICC*, 2021.
[6] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Proc. NIPS*, 2014.
[7] B. G. Noh, J. Lee, and M. B. Kim, "RF-GAN: Spectrum Data Augmentation for Deep Modulation Recognition," in *Proc. ICC Workshops*, 2022.
[8] A. Bendale and T. E. Boult, "Towards Open Set Deep Networks," in *Proc. CVPR*, 2016.
[9] S. Rebuffi *et al.*, "iCaRL: Incremental Classifier and Representation Learning," in *Proc. CVPR*, 2017.
[10] K. Lee *et al.*, "Mahalanobis Distance-based Confidence Score for Deep Neural Classifiers," in *Proc. NIPS*, 2018.

Github Links:

https://github.com/rameyjm7/ML-wireless-signal-classification

https://github.com/gawdygoda/ML-wireless-signal-classification