Multi-Domain Feature Extraction for ML-Based Over-the-Air RF Signal Classification

Achieving 97.72% Accuracy with Efficient Architecture

F. D. Suárez Bonilla, J.M. de la Rosa, G. Liñán-Cembrano

Instituto de Microelectrónica de Sevilla, IMSE-CNM
(CSIC/Universidad de Sevilla)

November 23, 2025

Outline

Introduction

Problem Statement

Challenge

Automatic classification of telecommunication signals in increasingly congested electromagnetic environments

Key Challenges:

- Signal type diversity
- Similar modulation patterns
- Noise and multipath effects
- Real-time processing requirements

Applications:

- Spectrum monitoring
- Electronic defense
- Cognitive radio systems
- Interference detection

Main Contributions

1. **Multi-domain feature extraction** combining time-frequency, wavelet, and frequency domain statistics
2. **High accuracy model** achieving 97.72% on real over-the-air signals
3. **Comprehensive analysis** demonstrating effectiveness across wide SNR range (-20 dB to 18 dB)
4. **Real-time capability** with processing time analysis for practical deployment

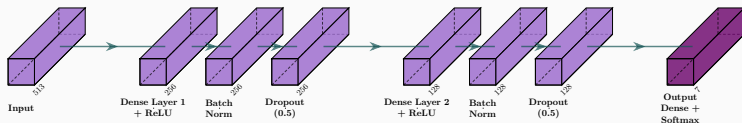# Methodology

Dataset: RadComOta2.45GHz

Table 1: Dataset Specifications

Parameter	Value
Total Records	567,000 labeled signals
Used Records	180,000 (7 classes)
SNR Range	-20 dB to 18 dB
Center Freq.	2.45 GHz
Sampling Rate	10 MS/s
Sample Length	256 samples (I/Q)
Hardware	USRP N210 + VERT2450

Signal Types

- AM Combined (60K samples)
- BPSK SATCOM (20K)
- FMCW Radar Alt. (20K)
- PULSED Air-Ground (20K)
- PULSED Airborne Det. (20K)
- PULSED Airborne Range (20K)
- PULSED Ground Map. (20K)

Modular Pipeline Design

HDF5 Analysis → Preprocessing → Feature Extraction → DNN Classification → Evaluation

Multi-Domain Feature Extraction

Spectral Analysis
STFT Spectrogram

- 64-sample windows
- 32-sample overlap
- Z-score normalized
- 449 features

Wavelet Transform
Discrete WT (db4)

- 3-level decomposition
- Excellent time-freq. localization
- Captures discontinuities
- 58 features

Frequency Domain
FFT Statistics

- Mean, std, max, median
- Peak counting
- Magnitude sum
- 6 features

Integrated Feature Vector

Total: **513 dimensions** = 449 (Spectrogram) + 58 (Wavelet) + 6 (FFT Stats)

Dense Neural Network Architecture

Network Design

Progressive Dimensionality Reduction:

- Input: 513 features
- Hidden 1: 256 neurons
- Hidden 2: 128 neurons
- Output: 7 classes
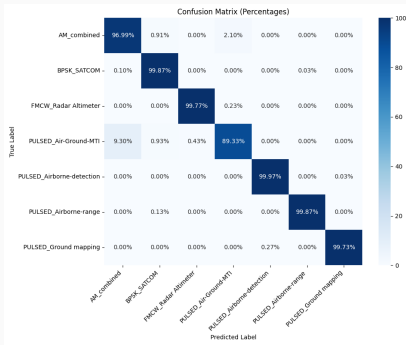
Total Parameters: 166,919

Hyperparameters

- **Optimizer:** Adam (lr=0.0005)
- **Loss:** Sparse categorical cross-entropy
- **Batch size:** 64
- **Early stopping:** 10 epochs patience
- **Class balancing:** Weighted loss

Results

Confusion Matrix (Percentages)

Key Results

- **Overall Accuracy:** 97.72%
- **Test Set:** 36,000 signals
- **SNR Range:** -20 dB to 18 dB
- **Model Complexity:** Low (167k params)

Main Confusions

AM_combined ↔ PULSED_Air-Ground-MTI

- 279 + 189 misclassifications
- Similar amplitude characteristics
- Consistent with technical similarities

Feature Domain Analysis

Table 2: Performance by Feature Domain

Feature Type	Accuracy (%)	Dimensions
Complete Feature Vector	97.72	513
Spectrogram Only	95.84	449
Wavelet Only	89.23	58
FFT Statistics Only	75.70	6
Spectrogram + Wavelet	96.41	507
Spectrogram + FFT	96.18	455
Wavelet + FFT	90.15	64

Key Insights

- Spectrograms contain **most discriminative information** (95.84%)
- Multi-domain fusion provides **1.88% improvement**
- Even simple FFT stats achieve **75.70% accuracy**

Processing Time Analysis

Table 3: Real-time Performance Analysis

Processing Stage	Time (ms)	Percentage	Signals/sec
Feature Extraction	2.147	84%	-
Model Inference	0.412	16%	-
Total Processing	**2.559**	**100%**	**390.8**

Hardware Performance

MacBook Pro M3:

- 8-core CPU, 10-core GPU
- 16GB unified memory
- 390 signals/second

Embedded Estimate

Raspberry Pi 4 / Jetson Nano:

- Estimated: 30-100 signals/sec
- Suitable for real-time apps
- Edge deployment ready

Comparison

State-of-the-Art Comparison

Method	Accuracy	Classes	Dataset	Complexity	SNR Range
Our Model	97.72%	7	Over-the-Air	167k params	-20 to 18 dB
Jagannath 2022	82.4% (OTA)	9 mod, 11 sig	OTA + Synth	253k params	0 dB
O'Shea 2018	95.6%	24	Synthetic	Medium (ResNet)	10 dB
Wang 2017	100%	7	Synthetic	Low (Manual)	0 dB
Schmidt 2017	95%	15	Synth/Real	Medium	-5 dB

Our Advantages

- **High accuracy** on real over-the-air signals across wide SNR range
- **Low complexity** model suitable for embedded deployment
- **Comprehensive evaluation** with detailed processing time analysis
- **Practical applicability** demonstrated through multi-domain features

Conclusions

Discussion and Key Findings

Strengths

- **Multi-domain approach** reduces classification errors by 50%
- **Efficient architecture** balances performance and complexity
- **Real-time viability** with 390+ signals/sec processing capability
- **Robust performance** across wide SNR range (-20 to 18 dB)

Challenges

- Confusion between AM_combined and PULSED_Air-Ground-MTI signals
- High intra-class similarity in amplitude-based signals
- Need for domain-specific descriptors for further improvement

Technical Improvements

- Signal-specific feature engineering
- Advanced architecture exploration
- Data augmentation strategies
- Ensemble methods integration

Implementation Focus

- Embedded system optimization
- Real-time processing enhancement
- Edge deployment strategies
- Power consumption optimization

Impact and Applications

This research enables practical deployment of ML-based signal classification in spectrum monitoring, electronic defense, and cognitive radio systems with demonstrated real-world performance.

Thank you for your attention!

Questions & Discussion

Multi-Domain Feature Extraction for ML-Based Over-the-Air RF Signal Classification

97.72% Accuracy — 167k Parameters — 390+ Signals/sec