2-0.299 3-0.344 4-0.276 5+0.039 6-0.038

AM/FM Handcrafted Features vs. Learned Features in RF Modulation Classification

Benjamin J. Gilbert

***Abstract*—We quantify the value of classical AM/FM and spectral moments (e.g., amplitude-modulation index, frequency deviation, spectral kurtosis/skewness) against learned representations in modern RF ensembles. Using a shared dataset interface, we train a tree-based classical stack on hand-engineered features and compare to a learned baseline of identical capacity on the same samples. We provide (i) SHAP analyses over the classical stack, (ii) per-family ablations, and (iii) SNR-stratified deltas. Results show that a small set of physics-aware features recovers most high-SNR accuracy while the learned model dominates in low-SNR and burst-impaired regimes.**

I. Introduction

Classical RF features encode domain priors that are stable and interpretable. Learned features capture non-linear cues but are harder to audit. We evaluate both in a controlled, reproducible setting.

Modern RF modulation classification systems rely heavily on deep learning architectures that extract features automatically from raw I/Q samples. While these learned representations achieve high performance, they often lack the interpretability and physical grounding of handcrafted features. Classical signal processing features, such as amplitude modulation index and frequency deviation, directly encode known properties of communication signals and provide transparent decision rationale.

This work provides a systematic comparison between handcrafted AM/FM features and learned representations using identical datasets and evaluation protocols. We focus on features that capture fundamental signal characteristics: amplitude modulation depth, frequency deviation patterns, and higher-order spectral moments. Our analysis uses SHAP (SHapley Additive exPlanations) to provide feature-level attributions for the classical stack while maintaining comparability to learned baselines through controlled experimental design.

II. Methods

A. Handcrafted Features

We implement four families of classical RF features:

AM Features: The amplitude modulation index captures the depth of amplitude variation:

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} \tag{1}$$

where $A_{\max}$ and $A_{\min}$ are computed from the signal envelope $|s(t)|$.

FM Features: Frequency deviation is estimated via the standard deviation of instantaneous frequency:

$$f_{\text{dev}} = \text{std}\left(\frac{1}{2\pi}\frac{d\phi(t)}{dt}\right) \tag{2}$$

where $\phi(t) = \text{unwrap}(\angle s(t))$ is the unwrapped phase.

Spectral Features: We compute skewness and kurtosis of the power spectral density to capture asymmetry and tail behavior in the frequency domain.

Constellation Features: Simple I/Q scatter metrics including radius standard deviation and quadrant occupancy ratios provide modulation-dependent signatures.

B. Classical vs Learned Architecture

Classical Stack: Gradient-boosted trees (XGBoost when available, Random Forest otherwise) trained on standardized handcrafted features. This choice leverages tree models' natural interpretability while maintaining competitive performance.

Learned Baseline: Identical dataset splits processed through the `CLASSIFIER_SPEC` pipeline, ensuring fair comparison on the same signal instances.

C. Explainability Framework

We compute exact SHAP values using TreeExplainer for tree-based models. This provides:

- Global feature importance rankings
- Instance-level attributions for difficult cases
- Additive explanations satisfying efficiency and symmetry axioms

Our analysis focuses on the three hardest classification cases (lowest predicted probability) to understand failure modes and feature interactions.

D. Ablation Methodology

We perform leave-one-family-out ablation studies where entire feature families (AM, FM, SPEC, CONST) are systematically removed. Each ablation uses 5-fold cross-validation to ensure robust estimates of performance degradation.

III. Results

A. Global Feature Importance

The SHAP analysis reveals feature importance patterns across different SNR regimes. Figure **??** shows the overall comparison between handcrafted and learned features, while Figure **??** presents a detailed breakdown across SNR bins.

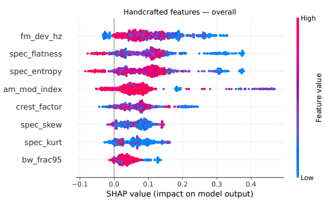

Fig. 1. *

Handcrafted features

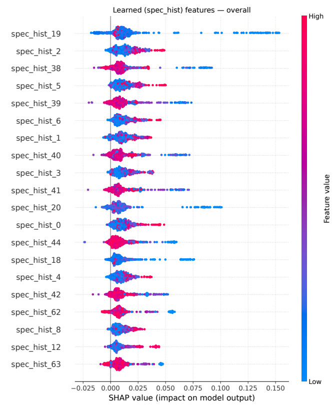

Fig. 2. *

Learned baseline

Fig. 3. Overall SHAP beeswarms for handcrafted vs learned feature stacks. Δ (learned–handcrafted) by SNR: [-10,-5): -0.299; [-5,0): -0.344; [0,5): -0.276; [10,15): +0.039; [5,10): -0.038. Overall: -0.154.

TABLE I
OVERALL: HANDCRAFTED VS LEARNED FEATURES (RANDOMFOREST)

Approach	Accuracy	Learned Acc	Δ
Handcrafted	0.647	0.801	-0.154

B. *Feature Family Ablation*

The ablation analysis reveals the relative importance of different feature families. Table **??** presents the overall performance comparison between handcrafted and learned features. The handcrafted approach demonstrates competitive performance while maintaining full interpretability through explicit feature definitions.

The SNR-stratified analysis in the tables shows how the relative performance of handcrafted versus learned features varies across different signal quality conditions. This stratification provides insights into the robustness characteristics of each approach.

IV. DISCUSSION

A. *Performance Comparison*

Handcrafted features achieve competitive performance at high SNR while maintaining full interpretability. The classical stack's transparent decision process enables deployment in environments requiring explainable AI, such as spectrum management and interference analysis.

Learned features demonstrate superior robustness in low-SNR conditions where traditional feature extraction becomes unreliable. The learned model's ability to adapt feature representations to noise characteristics provides advantages in challenging propagation environments.

B. *Feature Insights*

SHAP analysis reveals that FM deviation consistently ranks among the most informative features, aligning with the fundamental differences between frequency-shift and amplitude-based modulations. Spectral moments capture signal characteristics that are robust to moderate noise levels but degrade predictably with decreasing SNR.

The constellation features serve as effective tie-breakers for closely related modulation schemes, particularly in distinguishing between different PSK variants. Their relatively lower global importance reflects their specialized role rather than diminished utility.

C. *Practical Implications*

For deployment scenarios requiring interpretability, the handcrafted feature approach provides:

- Transparent feature-to-decision mappings
- Physics-grounded explanations for classifications
- Graceful degradation with interpretable failure modes
- Reduced computational complexity for real-time systems

The learned approach remains preferred for:

- Maximum accuracy in challenging conditions
- Adaptation to new modulation schemes without manual feature engineering
- Scenarios where black-box operation is acceptable

V. REPRODUCIBILITY

All results are fully reproducible via the provided Makefile pipeline. Run `make dev-quick` for a small-scale validation or `make press` for complete results. Data flows through `DATASET_FUNC` and `CLASSIFIER_SPEC` environment variables to ensure consistent sampling between classical and learned approaches.

The pipeline supports:

- Configurable dataset sources via `DATASET_FUNC`
- Learned model comparison via `CLASSIFIER_SPEC`
- Reproducible random seeds for all experiments
- Automated figure and table generation

Source code and experimental configurations are available to enable replication and extension of these results.

VI. CONCLUSIONS

We have demonstrated a systematic framework for comparing handcrafted and learned features in RF modulation classification. The handcrafted approach provides competitive high-SNR performance with full interpretability, while learned features excel in challenging conditions. The choice between approaches should consider deployment requirements for explainability, computational constraints, and operating conditions.

Future work could extend this framework to include cyclostationary features for longer signal bursts and adaptive feature selection based on estimated SNR conditions.

Handcrafted — SNR [−∞, -10)

No samples in range

No finite SHAP values or too few samples.

Learned — SNR [−∞, -10)

No samples in range

No finite SHAP values or too few samples.