Feature Importance Report (nMI + Permutation) — Jammer Classifier

Run folder: `run_20251217_122204`

This package documents *what the feature set measures*, *how importance was computed*, and *how to interpret the results*.
It is designed to be read top-to-bottom, but you can also jump to the feature-group sections.

---

1. Context and goals

You have a 4-class jammer classifier with classes:

- `NoJam`
- `Chirp`
- `NB` (narrowband)
- `WB` (wideband)

You want two complementary answers:

1. **Which features contain label-related structure in the data?** (data-centric view)
2. **Which features does the trained model actually rely on at inference time?** (model-centric view)

This report answers (1) using **normalized Mutual Information (nMI)** and (2) using **permutation importance** measured as **macro-F_1 drop** on the test set.

---

2. Snapshot of the evaluation run

2.1 Model + feature pipeline identifiers

- Model used for evaluation: `..\artifacts\finetuned\finetune_continue_20251216_160409\xgb_20251216_160409\xgb_finetuned_continue.joblib`
- Features directory used: `..\artifacts\finetuned\finetune_continue_20251216_160409\features`
- Number of engineered features: **78**

2.2 Test-set class distribution

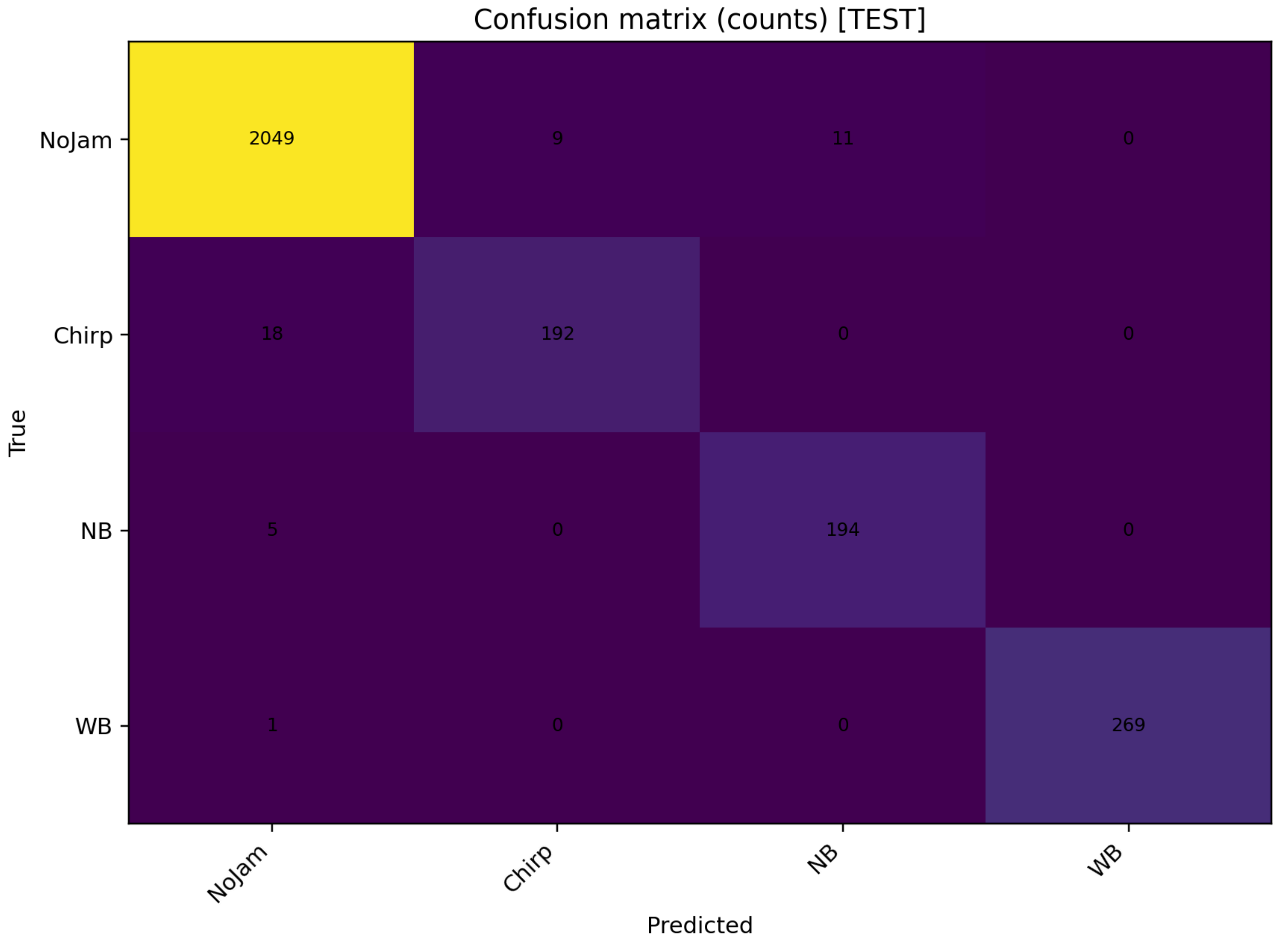
Total test samples: **2748**

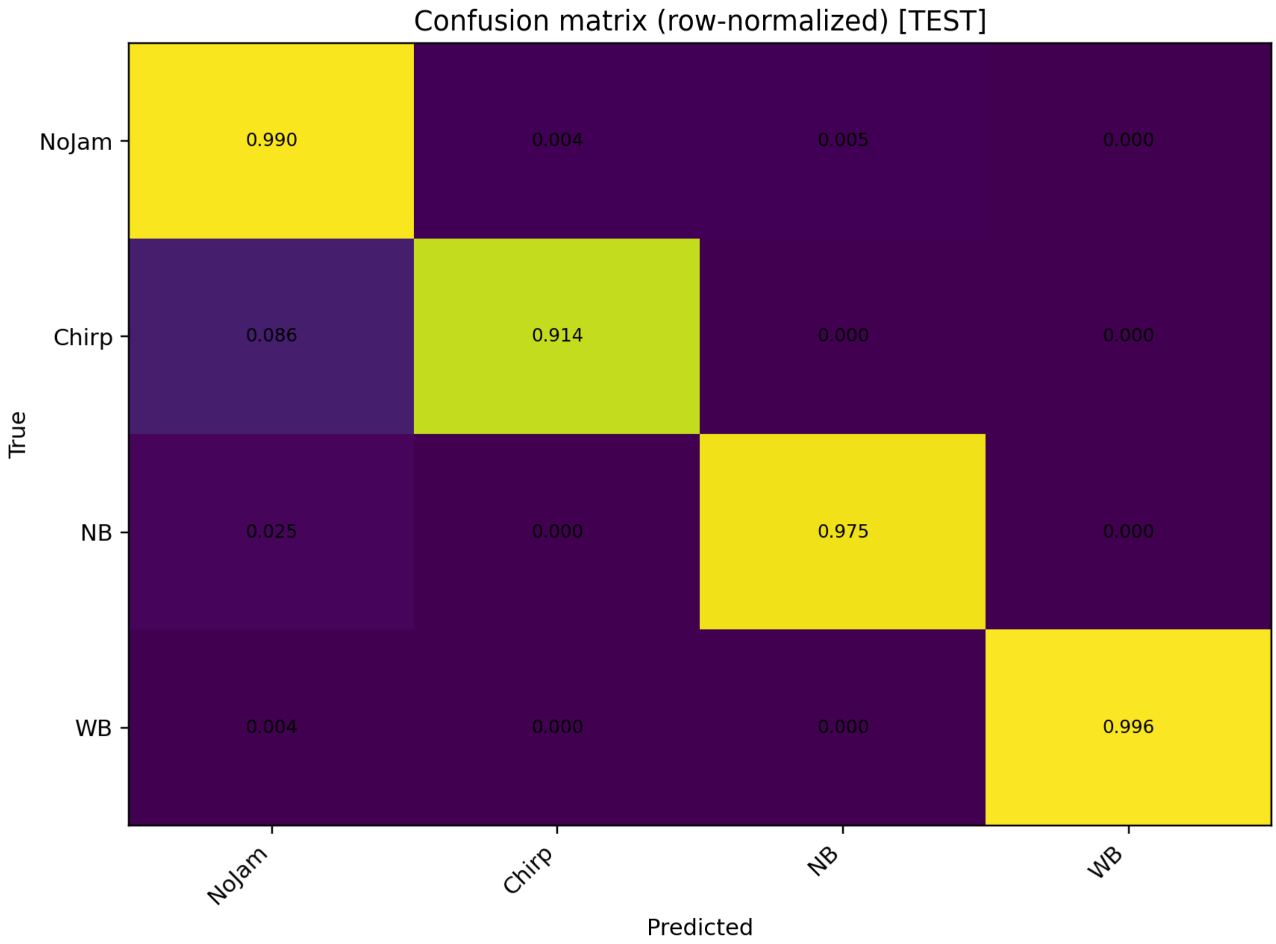
class	support	share
NoJam	2069	0.7529
Chirp	210	0.0764
NB	199	0.0724
WB	270	0.0983

2.3 Overall performance on the test set

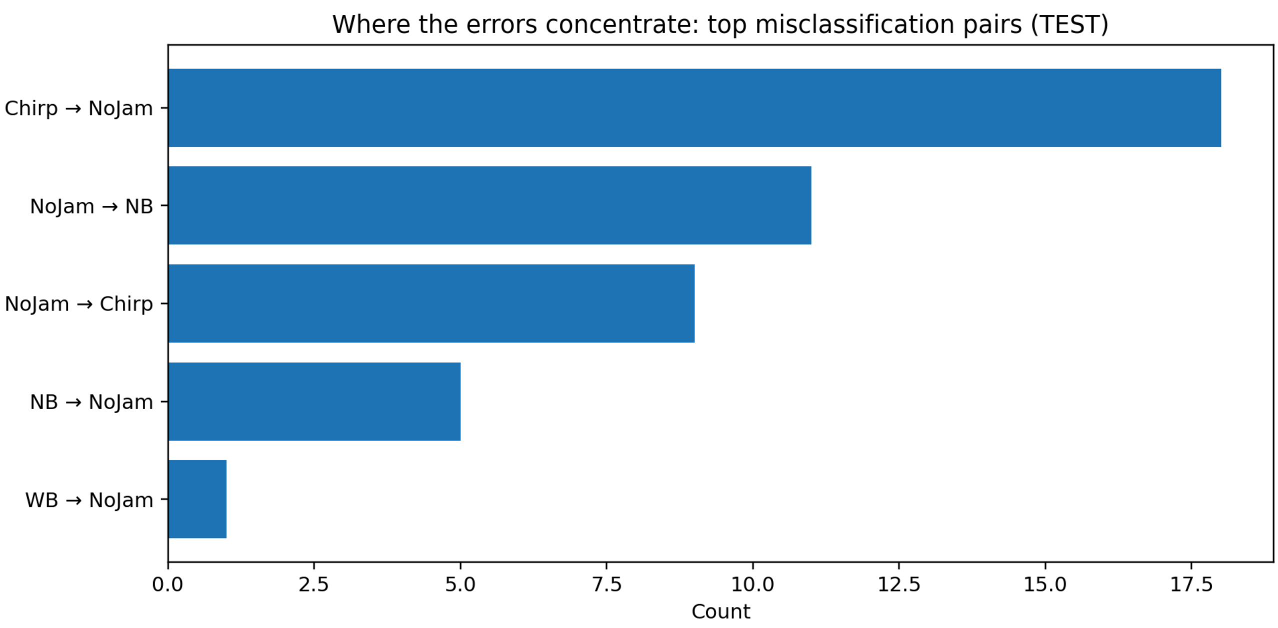
- Accuracy: **0.983988**
- Balanced accuracy: **0.968947**
- Macro F_1: **0.970556**
- Log loss: **0.049549**

Confusion matrices:





Where the errors concentrate:



High-confidence mistakes can be inspected in `assets/tables/high_confidence_errors_test.csv`.

---

3. Methods

3.1 Feature documentation source of truth

All feature definitions are taken from the local files shipped in this package:

- `source/features.md` (human-readable documentation, formulas, and intent)
- `source/feature_extractor.py` (actual implementation)

Whenever a feature is missing an explicit block in `features.md` (e.g. circularity or I/Q skew/kurtosis), this report documents it directly from `feature_extractor.py`.

3.2 Mutual information and normalized MI (nMI)

For each feature X and label Y, we estimate mutual information:

$$I(X;Y) = \int \sum_y p(x,y) \, \log \frac{p(x,y)}{p(x)p(y)} \, dx.$$

To make MI values comparable across label distributions, we normalize by the label entropy:

$$H(Y) = -\sum_y p(y) \log p(y), \qquad \mathrm{nMI}(X;Y) = \frac{I(X;Y)}{H(Y)}.$$

In this run, the label entropy was $H(Y) = 0.834435$ **nats** (low because the test distribution is imbalanced).

Key interpretation: nMI answers *"how much information about the class label is present in this feature alone, in the data distribution used for estimation."*

Important limitations (nMI):

- nMI is **marginal**: it does not account for feature interactions unless those interactions are already visible in the 1D distribution of X.
- nMI does not tell you whether the **trained model actually uses** the feature; it only describes label dependence in the data.
- Redundant/correlated features can *all* have high nMI even if only one is needed in a model.

3.3 Permutation importance (macro-F_1 drop)

Permutation importance is computed on the **held-out test set** as follows:

1. Compute the baseline score S_0 (here: macro-F_1).
2. For a feature column j, randomly permute that column across samples (breaking the feature–label association while keeping the marginal distribution of that feature).
3. Re-evaluate the score S_j with the permuted column.
4. Define importance as the score drop:

$$\Delta_j = S_0 - S_j.$$

We repeat the permutation multiple times per feature and report mean and standard deviation:

$$\mu_j = \mathbb{E}[\Delta_j], \qquad \sigma_j = \mathrm{Std}[\Delta_j].$$

Baseline test macro-F_1 was **0.970556** (this is the S_0 used for permutation drops).

Key interpretation: permutation answers *“if I destroy this feature’s alignment with the labels, how much does the model’s performance degrade?”*

Important limitations (permutation):

- If two features are highly correlated, permuting one may cause **little score drop** even if the feature is genuinely useful (the other feature can “cover” for it).
- Importance is **distribution-dependent**: a feature can look unimportant on one test distribution and crucial on another (e.g. different jammer SNRs).
- Small negative mean drops can happen due to Monte Carlo noise; interpret values near 0 as “no measurable effect.”

3.4 Combined score (normalized nMI + normalized permutation)

To get a single prioritization list, the script also provides a combined score:

$$\mathrm{score}_j = \mathrm{norm}(\mathrm{nMI}_j) + \mathrm{norm}(\mu_j),$$

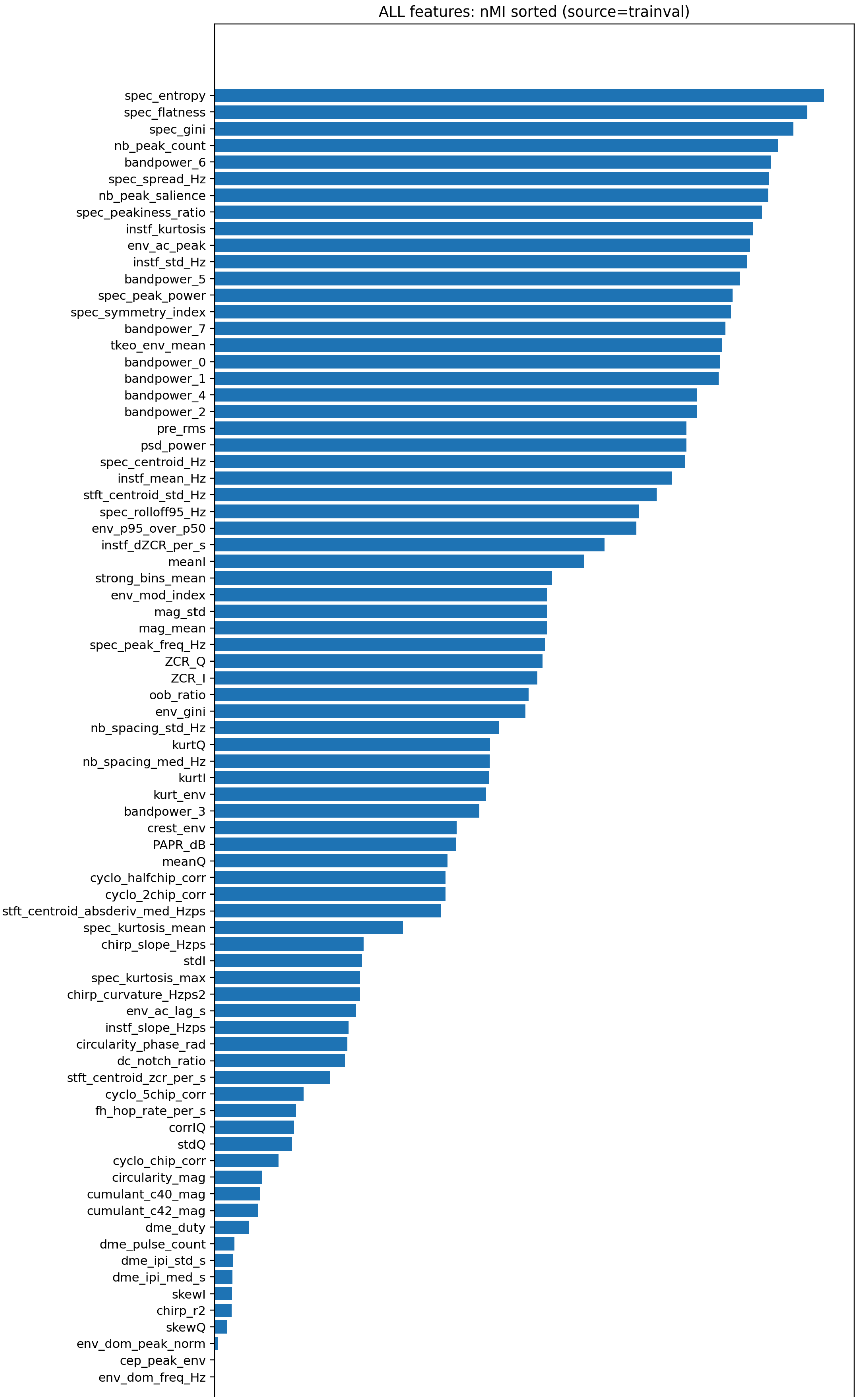
where each term is min–max normalized across the 78 features.

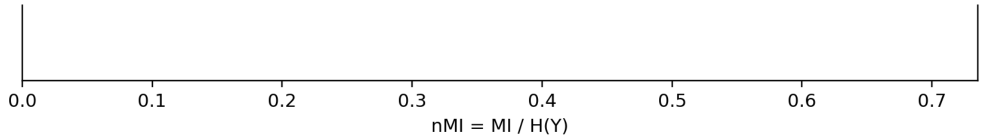
How to use it: as a *triage* list. For pruning decisions, always verify stability across retrains and across alternative test distributions.

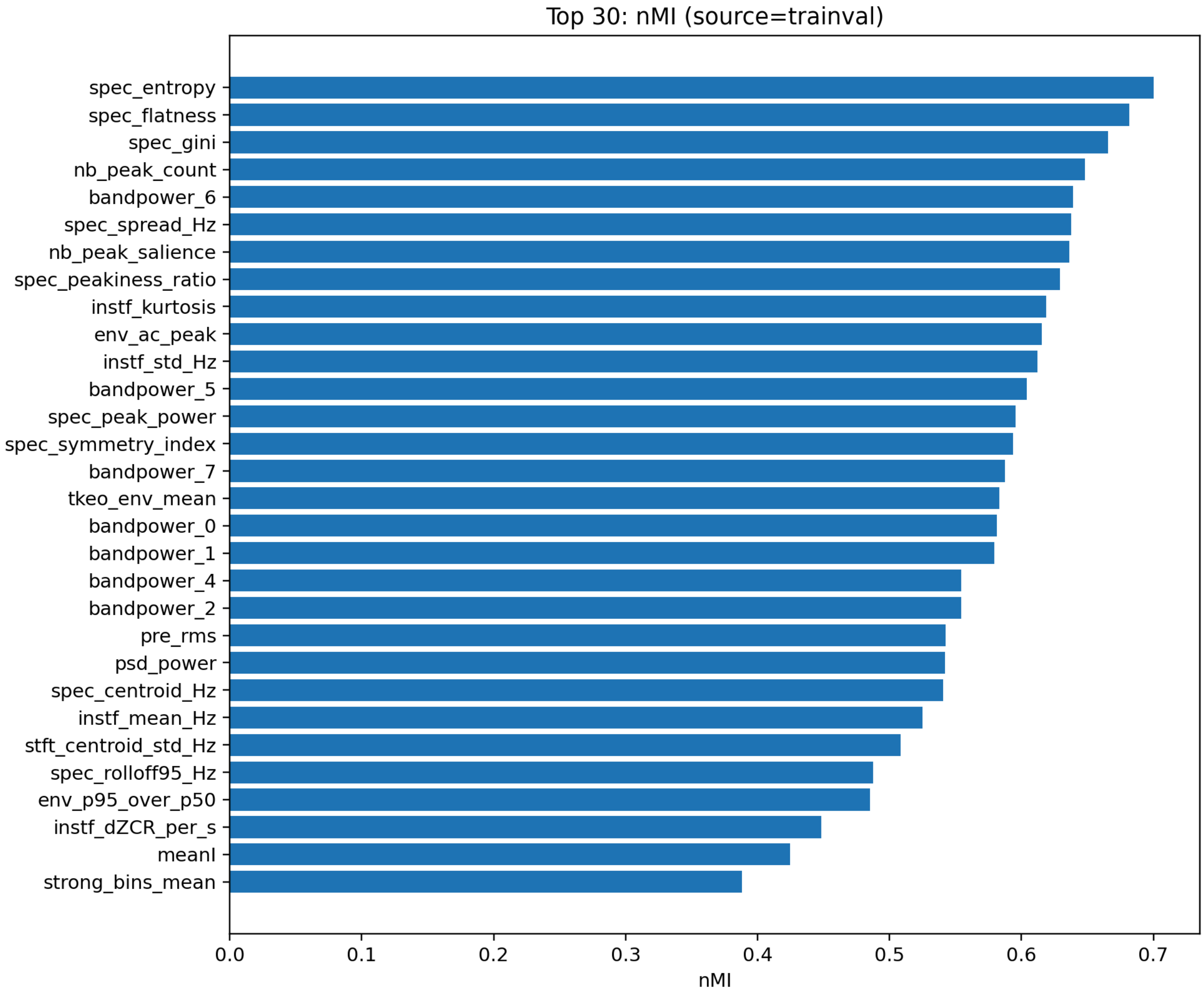
---

4. Visual overview of importance results

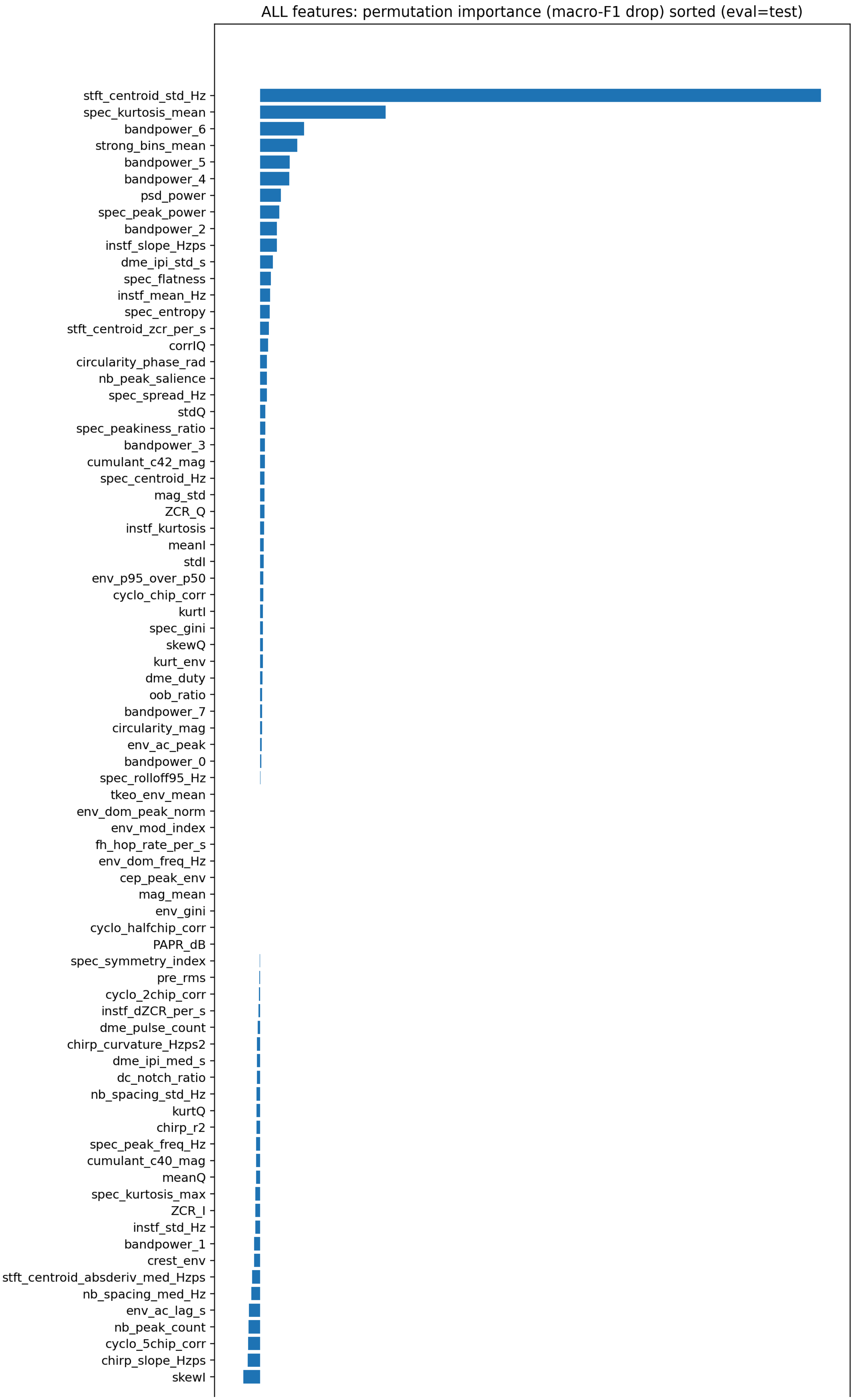
4.1 nMI ranking (train+val)

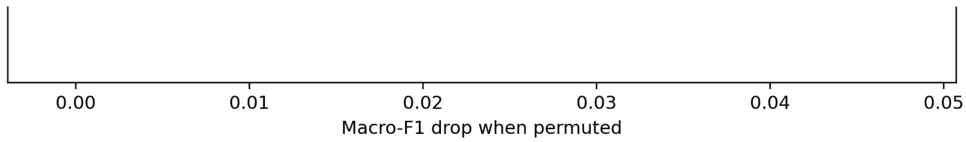

ALL features: nMI sorted (source=trainval)
spec_entropy
spec_flatness
spec_gini
nb_peak_count
bandpower_6
spec_spread_Hz
nb_peak_salience
spec_peakiness_ratio
instf_kurtosis
env_ac_peak
instf_std_Hz
bandpower_5
spec_peak_power
spec_symmetry_index
bandpower_7
tkeo_env_mean
bandpower_0
bandpower_1
bandpower_4
bandpower_2
pre_rms
psd_power
spec_centroid_Hz
instf_mean_Hz
stft_centroid_std_Hz
spec_rolloff95_Hz
env_p95_over_p50
instf_dZCR_per_s
meanI
strong_bins_mean
env_mod_index
mag_std
mag_mean
spec_peak_freq_Hz
ZCR_Q
ZCR_I
oob_ratio
env_gini
nb_spacing_std_Hz
kurtQ
nb_spacing_med_Hz
kurtI
kurt_env
bandpower_3
crest_env
PAPR_dB
meanQ
cyclo_halfchip_corr
cyclo_2chip_corr
stft_centroid_absderiv_med_Hzps
spec_kurtosis_mean
chirp_slope_Hzps
stdI
spec_kurtosis_max
chirp_curvature_Hzps2
env_ac_lag_s
instf_slope_Hzps
circularity_phase_rad
dc_notch_ratio
stft_centroid_zcr_per_s
cyclo_5chip_corr
fh_hop_rate_per_s
corrIQ
stdQ
cyclo_chip_corr
circularity_mag
cumulant_c40_mag
cumulant_c42_mag
dme_duty
dme_pulse_count
dme_ipi_std_s
dme_ipi_med_s
skewI
chirp_r2
skewQ
env_dom_peak_norm
cep_peak_env
env_dom_freq_Hz

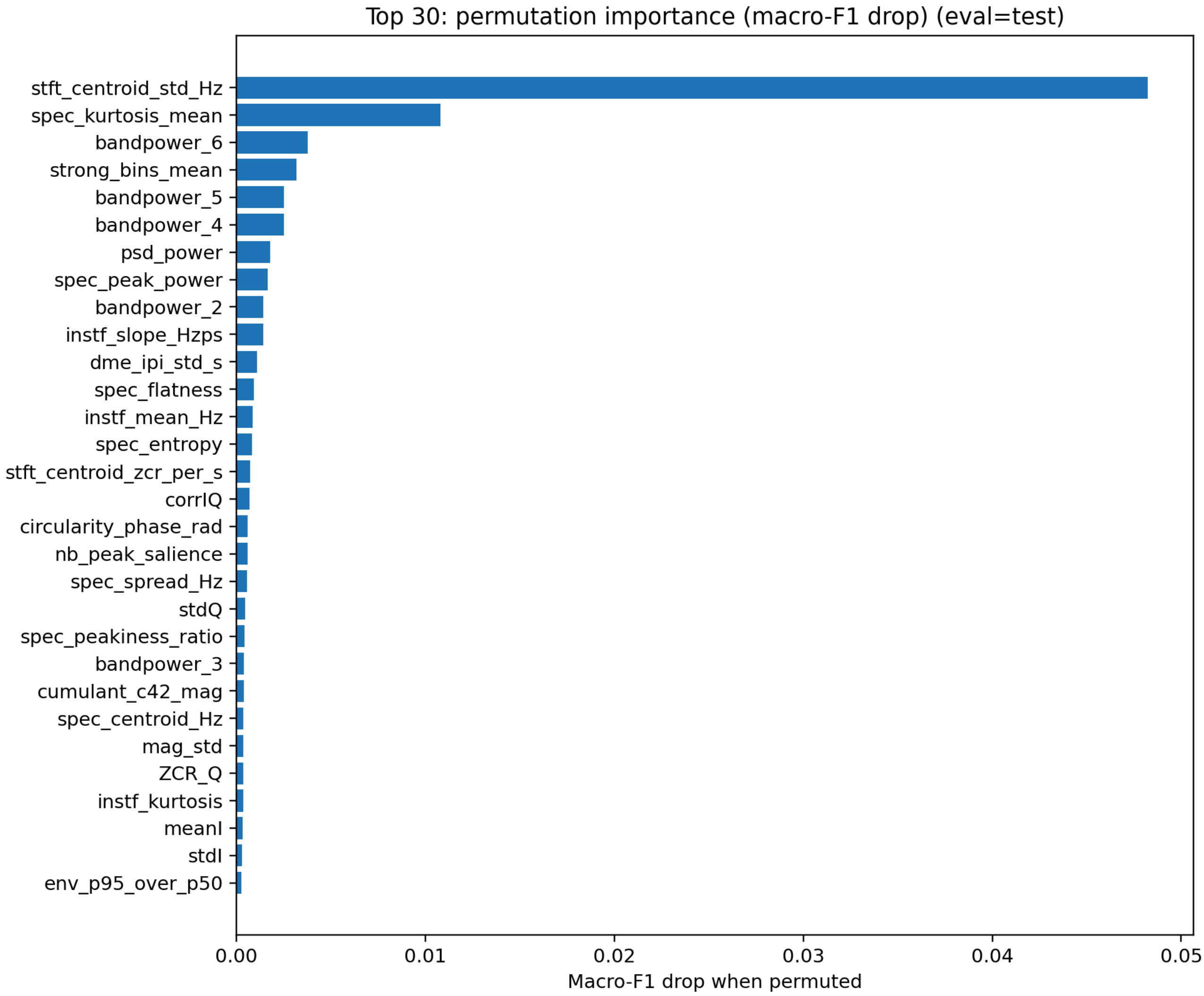




4.2 Permutation importance ranking (test macro-F_1 drop)

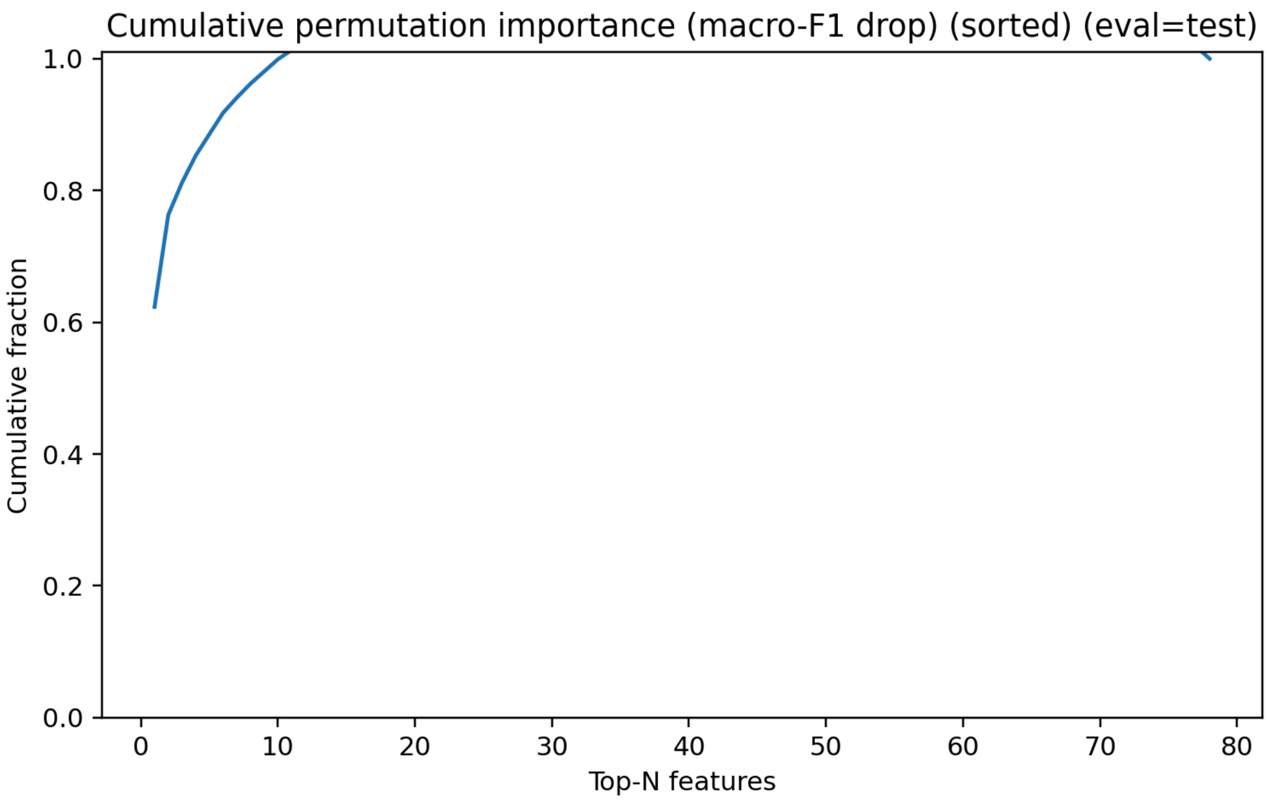

ALL features: permutation importance (macro-F1 drop) sorted (eval=test)
stft_centroid_std_Hz
spec_kurtosis_mean
bandpower_6
strong_bins_mean
bandpower_5
bandpower_4
psd_power
spec_peak_power
bandpower_2
instf_slope_Hzps
dme_ipi_std_s
spec_flatness
instf_mean_Hz
spec_entropy
stft_centroid_zcr_per_s
corrIQ
circularity_phase_rad
nb_peak_salience
spec_spread_Hz
stdQ
spec_peakiness_ratio
bandpower_3
cumulant_c42_mag
spec_centroid_Hz
mag_std
ZCR_Q
instf_kurtosis
meanI
stdI
env_p95_over_p50
cyclo_chip_corr
kurtI
spec_gini
skewQ
kurt_env
dme_duty
oob_ratio
bandpower_7
circularity_mag
env_ac_peak
bandpower_0
spec_rolloff95_Hz
tkeo_env_mean
env_dom_peak_norm
env_mod_index
fh_hop_rate_per_s
env_dom_freq_Hz
cep_peak_env
mag_mean
env_gini
cyclo_halfchip_corr
PAPR_dB
spec_symmetry_index
pre_rms
cyclo_2chip_corr
instf_dZCR_per_s
dme_pulse_count
chirp_curvature_Hzps2
dme_ipi_med_s
dc_notch_ratio
nb_spacing_std_Hz
kurtQ
chirp_r2
spec_peak_freq_Hz
cumulant_c40_mag
meanQ
spec_kurtosis_max
ZCR_I
instf_std_Hz
bandpower_1
crest_env
stft_centroid_absderiv_med_Hzps
nb_spacing_med_Hz
env_ac_lag_s
nb_peak_count
cyclo_5chip_corr
chirp_slope_Hzps
skewI



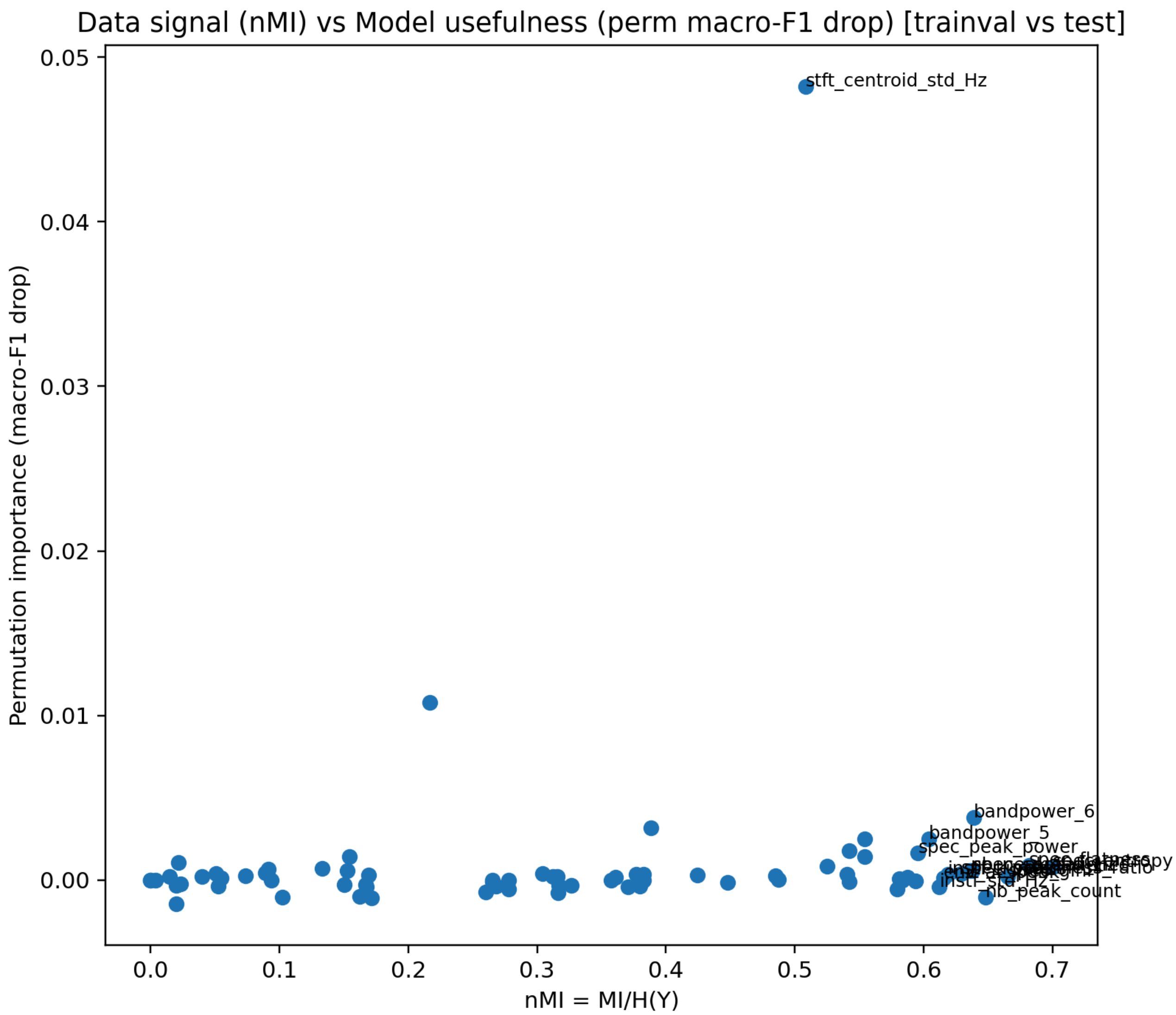


Permutation importance is extremely concentrated in this run:

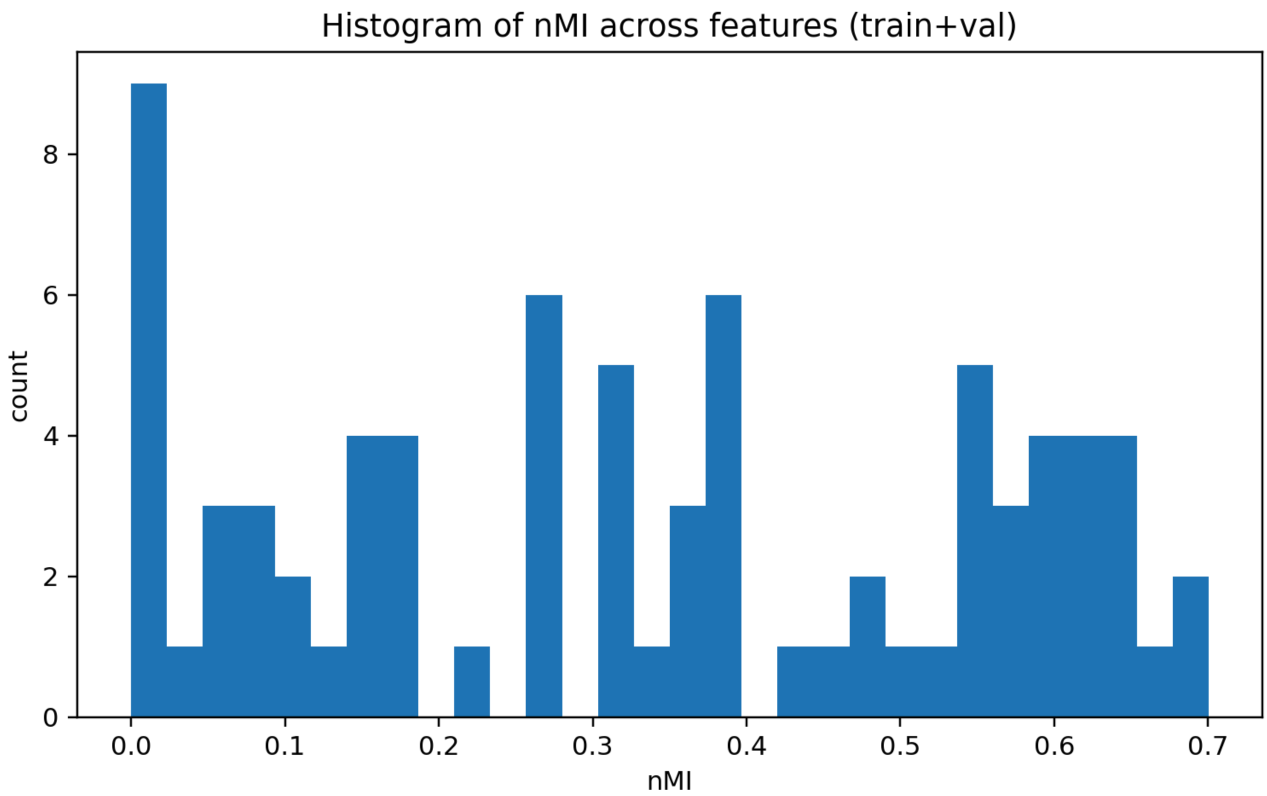
- The **top feature** (`stft_centroid_std_Hz`) accounts for about **53.5%** of the *total positive* permutation importance mass.
- The top **15** features explain about **90%** of the positive permutation importance mass (top **23** explain ~95%).

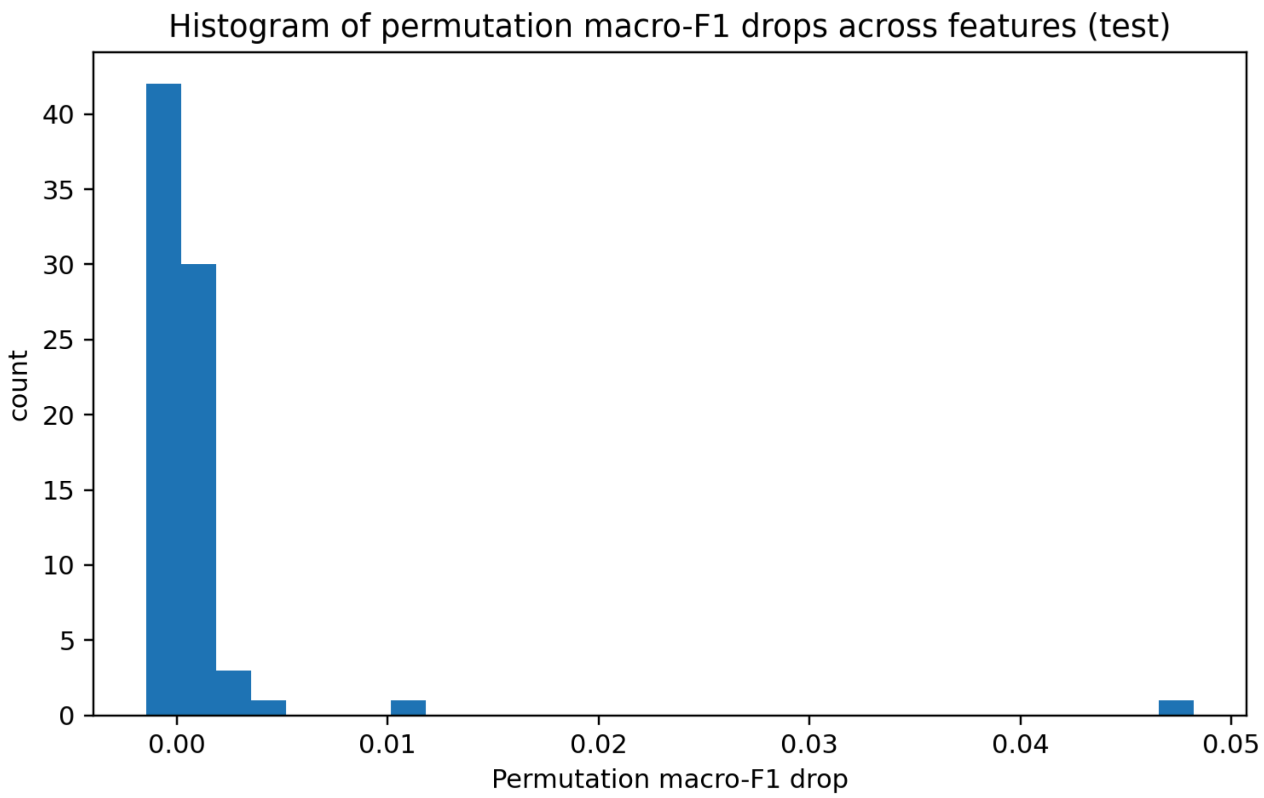


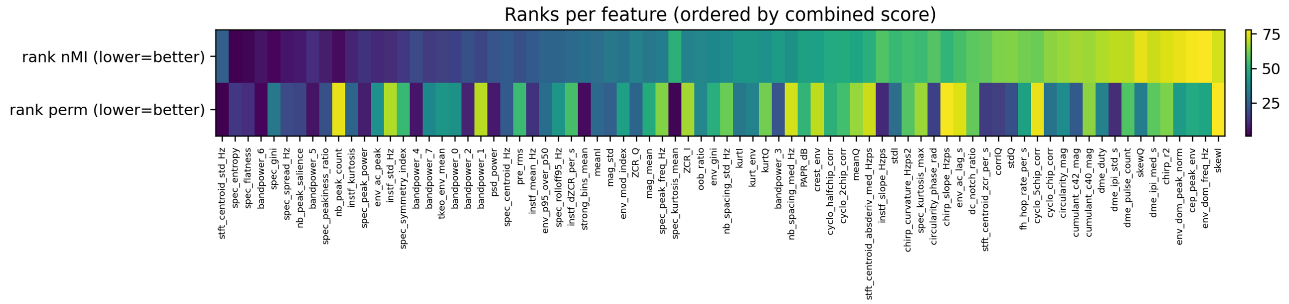
4.3 Cross-method alignment



Added diagnostics (generated for this package):







5. Top results and global interpretation

5.1 Top features by nMI (data signal)

feature	nMI	MI	rank_nMI
spec_entropy	0.70025	0.584314	1
spec_flatness	0.68167	0.56881	2
spec_gini	0.665434	0.555262	3
nb_peak_count	0.648161	0.540849	4
bandpower_6	0.639091	0.53328	5
spec_spread_Hz	0.637672	0.532096	6
nb_peak_salience	0.636317	0.530965	7
spec_peakiness_ratio	0.629294	0.525105	8
instf_kurtosis	0.618933	0.51646	9
env_ac_peak	0.61543	0.513537	10
instf_std_Hz	0.612333	0.510952	11
bandpower_5	0.604086	0.504071	12
spec_peak_power	0.595744	0.49711	13
spec_symmetry_index	0.593872	0.495547	14
bandpower_7	0.587283	0.49005	15
tkeo_env_mean	0.583284	0.486713	16
bandpower_0	0.581352	0.485101	17
bandpower_1	0.579438	0.483503	18
bandpower_4	0.554529	0.462719	19
bandpower_2	0.554255	0.46249	20
pre_rms	0.542506	0.452686	21
psd_power	0.542163	0.4524	22
spec_centroid_Hz	0.540587	0.451085	23

feature	nMI	MI	rank_nMI
instf_mean_Hz	0.525195	0.438241	24
stft_centroid_std_Hz	0.508264	0.424114	25
spec_rolloff95_Hz	0.487553	0.406831	26
env_p95_over_p50	0.485116	0.404798	27
instf_dZCR_per_s	0.448108	0.373917	28
meanI	0.424606	0.354306	29
strong_bins_mean	0.388255	0.323974	30

Interpretation:

- The top nMI features are dominated by **PSD-shape descriptors** (`spec_entropy`, `spec_flatness`, `spec_gini`) and by **coarse bandpower mass** in higher bands.
- This is exactly what you expect when the dataset contains strong contrasts between *tone-like* vs *spread-spectrum-like* interference: entropy/flatness/gini react sharply to that.

5.2 Top features by permutation (model usage)

feature	perm_macroF1_drop_mean	perm_macroF1_drop_std	ranl
stft_centroid_std_Hz	0.04823	0.004251	
spec_kurtosis_mean	0.010794	0.001272	
bandpower_6	0.003787	0.001829	
strong_bins_mean	0.003179	0.002924	
bandpower_5	0.002522	0.000961	
bandpower_4	0.002498	0.001232	
psd_power	0.00178	0.001258	
spec_peak_power	0.001648	0.002047	
bandpower_2	0.001433	0.000808	
instf_slope_Hzps	0.001429	0.000688	
dme_ipi_std_s	0.001093	0.000299	
spec_flatness	0.000914	0.000914	
instf_mean_Hz	0.000854	0.000963	
spec_entropy	0.000827	0.00072	
stft_centroid_zcr_per_s	0.00073	0.000794	
corrIQ	0.000685	0.000893	
circularity_phase_rad	0.000585	0.000805	
nb_peak_salience	0.00058	0.001243	

feature	perm_macroF1_drop_mean	perm_macroF1_drop_std	ranl
spec_spread_Hz	0.00056	0.000535	
stdQ	0.000443	0.000517	
spec_peakiness_ratio	0.000422	0.00064	
bandpower_3	0.000399	0.000892	
cumulant_c42_mag	0.000387	0.000619	
spec_centroid_Hz	0.000372	0.000559	
mag_std	0.000367	0.000765	
ZCR_Q	0.000353	0.000473	
instf_kurtosis	0.000349	0.000561	
meanI	0.000314	0.000629	
stdI	0.0003	0.001088	
env_p95_over_p50	0.000271	0.000938	

Interpretation:

- The model relies overwhelmingly on `stft_centroid_std_Hz` (STFT centroid standard deviation). This strongly suggests that **time-variation of spectral centroid** is a primary discriminator on this test set—classic for chirp-like interference.
- Several PSD-shape features (spectral kurtosis, peak power, entropy/flatness, some bandpowers) remain important, but their contributions are much smaller than the top STFT feature.

5.3 Top features by combined score (triage list)

feature	nMI	perm_macroF1_drop_mean	nMI_plus_perm_norn
stft_centroid_std_Hz	0.508264	0.04823	1.7258
spec_entropy	0.70025	0.000827	1.0171
spec_flatness	0.68167	0.000914	0.99242
bandpower_6	0.639091	0.003787	0.99118
spec_gini	0.665434	0.000241	0.95528
spec_spread_Hz	0.637672	0.00056	0.92223
nb_peak_salience	0.636317	0.00058	0.92072
bandpower_5	0.604086	0.002522	0.91496
spec_peakiness_ratio	0.629294	0.000422	0.90741
nb_peak_count	0.648161	-0.001013	0.90461
instf_kurtosis	0.618933	0.000349	0.891

feature	nMI	perm_macroF1_drop_mean	nMI_plus_perm_norn
spec_peak_power	0.595744	0.001648	0.884935
env_ac_peak	0.61543	0.000117	0.881291
instf_std_Hz	0.612333	-0.000415	0.865853
spec_symmetry_index	0.593872	-3.2e-05	0.847422
bandpower_4	0.554529	0.002498	0.843701
bandpower_7	0.587283	0.000165	0.8421
tkeo_env_mean	0.583284	4e-06	0.833043
bandpower_0	0.581352	7.7e-05	0.831807
bandpower_2	0.554255	0.001433	0.821226
bandpower_1	0.579438	-0.000525	0.816584
psd_power	0.542163	0.00178	0.811144
spec_centroid_Hz	0.540587	0.000372	0.779697
pre_rms	0.542506	-9.4e-05	0.772785
instf_mean_Hz	0.525195	0.000854	0.76772
env_p95_over_p50	0.485116	0.000271	0.698395
spec_rolloff95_Hz	0.487553	2.4e-05	0.696746
instf_dZCR_per_s	0.448108	-0.000146	0.63689
strong_bins_mean	0.388255	0.003179	0.620372
meanI	0.424606	0.000314	0.612881

How to read mismatches:

- High nMI + low permutation: the feature contains real class structure, but the trained model may not need it because similar information is already captured by other features.
- Low/medium nMI + high permutation: the feature may be used in a *nonlinear/interaction* way that marginal MI misses, or it may be exploiting a distribution quirk that is stable on this test set.

---

6. Feature-group level breakdown

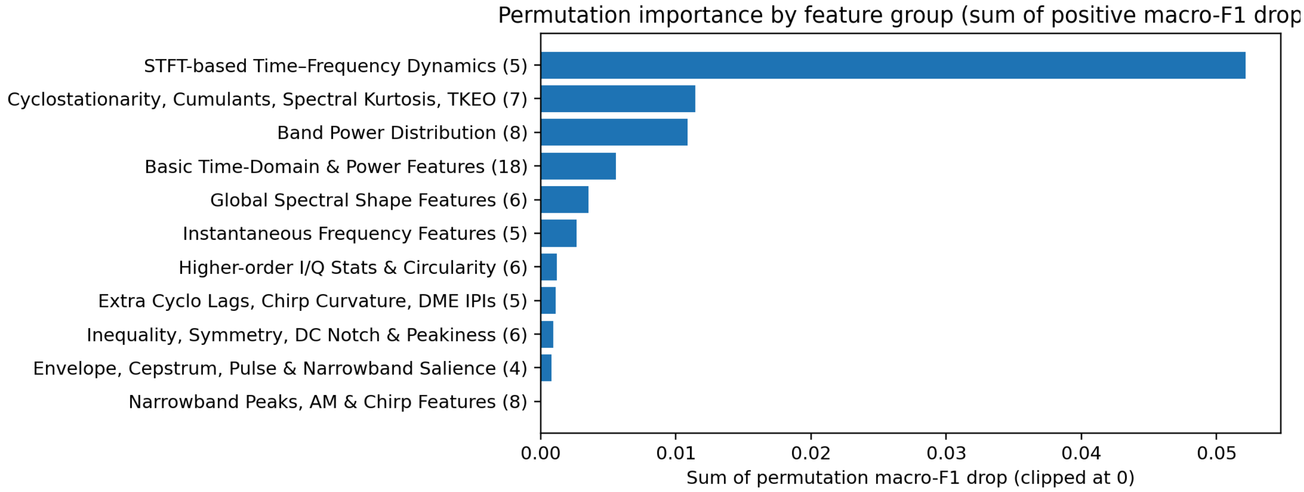
This section aggregates importance by the groups defined in `features.md`.

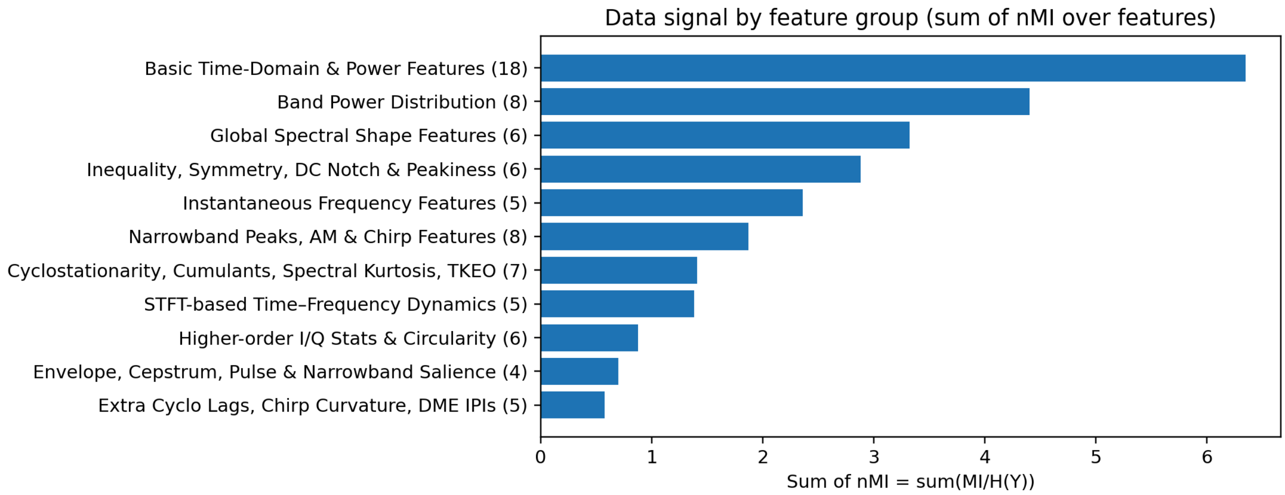
6.1 Group summary table

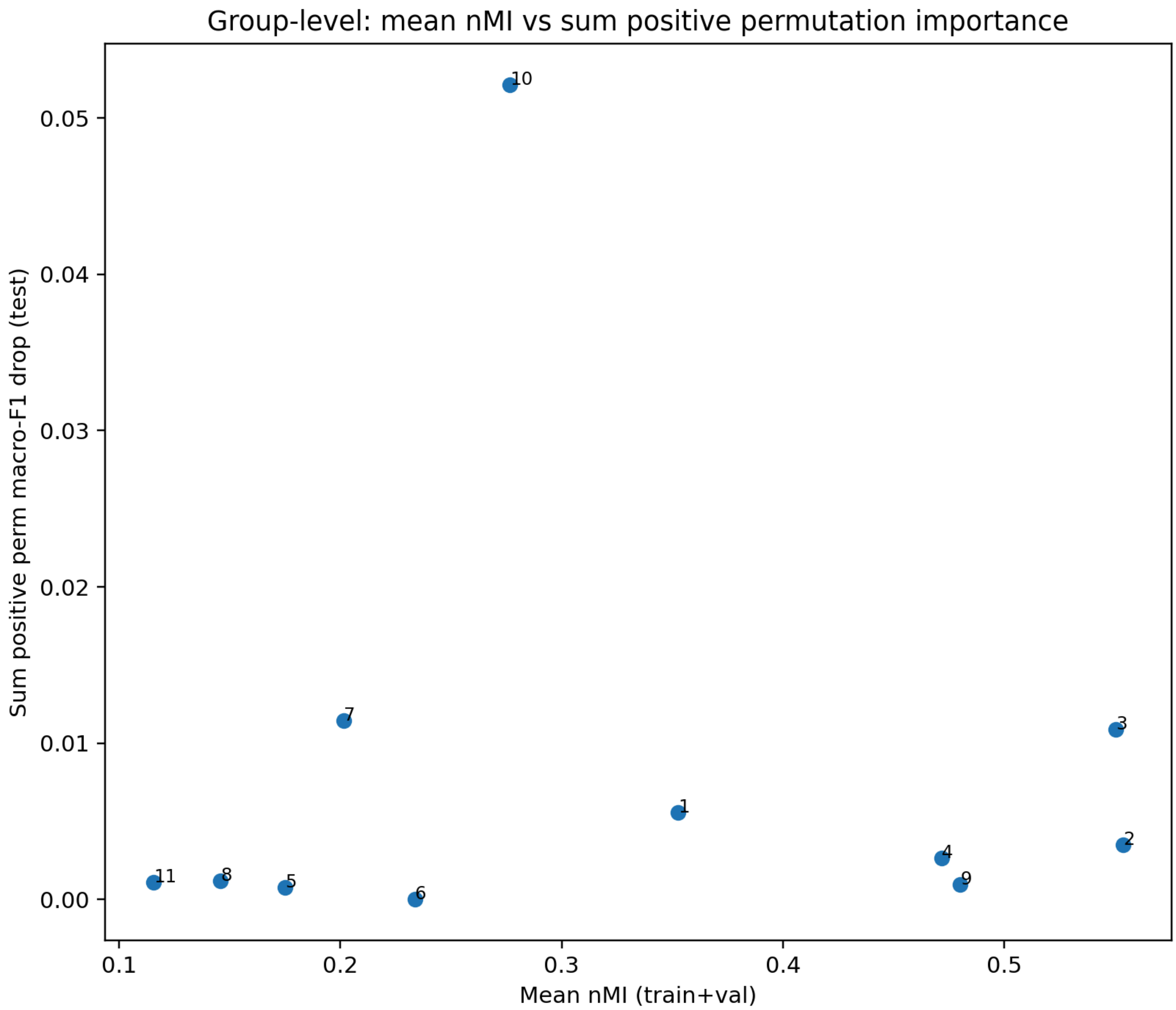
group_id	group_name	n_features	nMI_sum	nMI_share	perm_sum_pos
10	STFT-based Time–Frequency Dynamics (5)	5	1.38326	0.052939	0.052139
7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	7	1.40953	0.053945	0.011431
3	Band Power Distribution (8)	8	4.40439	0.168563	0.010882
1	Basic Time-Domain & Power Features (18)	18	6.34758	0.242931	0.005576
2	Global Spectral Shape Features (6)	6	3.32293	0.127173	0.003518
4	Instantaneous Frequency Features (5)	5	2.35873	0.090272	0.002632
8	Higher-order I/Q Stats & Circularity (6)	6	0.875254	0.033497	0.001211
11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	5	0.577675	0.022108	0.001093
9	Inequality, Symmetry, DC Notch & Peakiness (6)	6	2.88132	0.110273	0.000934
5	Envelope, Cepstrum, Pulse & Narrowband Salience (4)	4	0.699517	0.026772	0.000788
6	Narrowband Peaks, AM & Chirp Features (8)	8	1.86893	0.071527	0

Key takeaways:

- **Group 10 (STFT dynamics)** dominates permutation importance. If you only keep one group for chirp discrimination, it is this one.
- **Groups 2–3–7** contain the strongest *data-level* signal (nMI), consistent with global PSD shape and kurtosis capturing NB/WB differences.







---

7. Exhaustive per-group and per-feature analysis

This is the core of the report: for each group, we:

- explain what the group measures
- show the group ranking table
- review each feature with its definition + nMI + permutation impact + interpretation

If you are using this to decide what to prune, start with groups where both nMI and permutation are consistently low.

7.1 Group 1: Basic Time-Domain & Power Features (18)

Basic statistics computed directly on the complex IQ samples (means, variances, RMS, peakiness, envelope power). These are often sensitive to **overall interference strength**, AGC behavior, and burstiness.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macroF1_d
spec_entropy	0.70025	1	0.000827	
env_ac_peak	0.61543	10	0.000117	0
psd_power	0.542163	22	0.00178	0
pre_rms	0.542506	21	-9.4e-05	0
meanI	0.424606	29	0.000314	0
mag_std	0.382411	32	0.000367	0
ZCR_Q	0.377217	35	0.000353	0
mag_mean	0.382077	33	0	
ZCR_I	0.370882	36	-0.000409	0
oob_ratio	0.361104	37	0.000177	0
kurt_env	0.312344	43	0.000214	0
PAPR_dB	0.277864	46	-2.1e-05	0
crest_env	0.278135	45	-0.000543	0
meanQ	0.267961	47	-0.000366	0
stdI	0.169478	53	0.0003	0
env_ac_lag_s	0.162546	56	-0.000978	0
corrIQ	0.091598	63	0.000685	0
stdQ	0.089003	64	0.000443	0

Per-feature review:

`spec_entropy`

Definition & intent (from `features.md` / extractor):

18. `spec_entropy`

- **Intuition**: How “spread” vs “concentrated” the raw spectrum is.
  - White noise → high entropy.
  - One or a few strong tones → lower entropy.
- **Formula** using normalised PSD of $x[n]$:

$$P_{xx0}^{\text{prob}}[k] = \frac{\max(P_{xx0}[k], \varepsilon)}{\sum_j \max(P_{xx0}[j], \varepsilon)},$$

$$\text{spec} = -\sum_k P_{xx0}^{\text{prob}}[k] \, \log\left(P_{xx0}^{\text{prob}}[k]\right).$$

---

2. Global Spectral Shape Features (6)

Category intuition

These describe the **overall shape of the spectrum** of the normalized signal:

- Where it is centered,
- How wide it is,
- How flat vs peaky it is,
- Where the main peak is.

All use the normalized PSD of $z[n]$.

Let $(f_k, P_{xx}[k])$ be Welch PSD of $z[n]$ and

$$P_{xx}^{\text{norm}}[k] = \frac{\max(P_{xx}[k], \varepsilon)}{\sum_j \max(P_{xx}[j], \varepsilon)}.$$

Measured importance (this run):

- nMI (train+val): **0.700250** (MI = 0.584314 nats), rank **1/78** → *very high* data-signal.
- Permutation importance (test): **0.000827 ± 0.000720** macro-F_1 drop, rank **14/78** → *medium* model-usage, moderately stable (mean/std ≈ 1.15).
- Combined score (normalized nMI + normalized perm): **1.017156**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - These are global spectral-shape descriptors; they often separate narrowband vs wideband interference cleanly.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`env_ac_peak`

Definition & intent (from `features.md` / extractor):

11. `env_ac_peak`

- **Intuition**: Strength of the most pronounced *periodicity* in the envelope (excluding lag 0). Useful for repeatedly pulsed or AM signals.
- **Formula**:

$$\text{env ac} = \max_{1 \le k \le k_{\max}} r[k].$$

Measured importance (this run):

- nMI (train+val): **0.615430** (MI = 0.513537 nats), rank **10/78** → *very high* data-signal.
- Permutation importance (test): **0.000117 ± 0.000723** macro-F_1 drop, rank **40/78** → *very low* model-usage, noisy (mean/std ≈ 0.16).
- Combined score (normalized nMI + normalized perm): **0.881291**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`psd_power`

Definition & intent (from `features.md` / extractor):

14. `psd_power`

- **Intuition**: Total energy in the PSD estimate (basically the same information as `pre_rms` but in the frequency domain).
- **Formula**:

$$\text{psd} = \sum_k P_{xx0}[k],$$

```
where $P_{xx0}$ is the Welch PSD of $x[n]$.
```

Measured importance (this run):

- nMI (train+val): **0.542163** (MI = 0.452400 nats), rank **22/78** → *high* data-signal.
- Permutation importance (test): **0.001780 ± 0.001258** macro-F_1 drop, rank **7/78** → *high* model-usage, moderately stable (mean/std ≈ 1.41).
- Combined score (normalized nMI + normalized perm): **0.811144**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`pre_rms`

Definition & intent (from `features.md` / extractor):

13. `pre_rms`

- **Intuition**: Absolute power of the raw IQ chunk, as seen by the receiver.
- **Formula**:

$$\mathrm{pre} = \sqrt{\frac{1}{N}\sum_{n} |x[n]|^2}.$$

Measured importance (this run):

- nMI (train+val): **0.542506** (MI = 0.452686 nats), rank **21/78** → *high* data-signal.
- Permutation importance (test): **-0.000094 ± 0.000941** macro-F_1 drop, rank **54/78** → *very low* model-usage, noisy (mean/std ≈ -0.10).
- Combined score (normalized nMI + normalized perm): **0.772785**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`meanI`

Definition & intent (from `features.md` / extractor):

1. `meanI`
   - **Intuition**: Average value of I; if not ≈0, the I channel has a DC offset.
   - **Formula**:

$$\mathrm{meanI} = \frac{1}{N}\sum_{n=0}^{N-1} I[n].$$

Measured importance (this run):

- nMI (train+val): **0.424606** (MI = 0.354306 nats), rank **29/78** → *medium* data-signal.

- Permutation importance (test): **0.000314 ± 0.000629** macro-F_1 drop, rank **28/78** → *low* model-usage, noisy (mean/std ≈ 0.50).
- Combined score (normalized nMI + normalized perm): **0.612881**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`mag_std`

Definition & intent (from `features.md` / extractor):

7. `mag_std`
    - **Intuition**: How much the magnitude fluctuates.
        - Constant amplitude carrier → low.
        - Pulsed or heavily AM signal → higher.
    - **Formula**:

$$\mathrm{mag} = \sqrt{\frac{1}{N}\sum_n (|z[n]| - \mathrm{mag})^2}.$$

1.2 Zero crossings and PAPR

Measured importance (this run):

- nMI (train+val): **0.382411** (MI = 0.319097 nats), rank **32/78** → *medium* data-signal.
- Permutation importance (test): **0.000367 ± 0.000765** macro-F_1 drop, rank **25/78** → *medium* model-usage, noisy (mean/std ≈ 0.48).
- Combined score (normalized nMI + normalized perm): **0.553716**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`ZCR_Q`

Definition & intent (from `features.md` / extractor):

9. `ZCR_Q`
    - **Intuition**: Same for Q.
    - **Formula**:

$$\mathrm{ZCR} = \mathrm{ZCR}(Q[n]).$$

Measured importance (this run):

- nMI (train+val): **0.377217** (MI = 0.314763 nats), rank **35/78** → *medium* data-signal.
- Permutation importance (test): **0.000353 ± 0.000473** macro-F_1 drop, rank **26/78** → *low* model-usage, noisy (mean/std ≈ 0.74).
- Combined score (normalized nMI + normalized perm): **0.546000**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`mag_mean`

Definition & intent (from `features.md` / extractor):

6. `mag_mean`
    - **Intuition**: Average normalized magnitude; around 1 for sane signals because of the normalization.
    - **Formula**:

$$\mathrm{mag} = \frac{1}{N} \sum_n |z[n]|.$$

Measured importance (this run):

- nMI (train+val): **0.382077** (MI = 0.318818 nats), rank **33/78** → *medium* data-signal.
- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **49/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.545629**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`ZCR_I`

Definition & intent (from `features.md` / extractor):

8. `ZCR_I`
   - **Intuition**: How rapidly I changes sign → higher for high-frequency content.
   - **Formula**:

$$\mathrm{ZCR} = \mathrm{ZCR}(I[n]).$$

Measured importance (this run):

- nMI (train+val): **0.370882** (MI = 0.309477 nats), rank **36/78** → *medium* data-signal.
- Permutation importance (test): **-0.000409 ± 0.001043** macro-F_1 drop, rank **68/78** → *very low* model-usage, noisy (mean/std ≈ -0.39).
- Combined score (normalized nMI + normalized perm): **0.521152**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`oob_ratio`

Definition & intent (from `features.md` / extractor):

15. `oob_ratio`

- **Intuition**: How much power lives **outside** the "interesting" in-band GNSS window (e.g. wideband jammer).
- **Formula** (with in-band half-bandwidth $B = \mathrm{INB\,BW}$):

$$\mathcal{I} = \{k : |f_k| \le B\}, \quad \mathcal{O} = \{k : |f_k| > B\},$$

$$\mathrm{oob} = \frac{\sum_{k\in\mathcal{O}} P_{xx0}[k]}{\sum_{k\in\mathcal{I}} P_{xx0}[k] + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.361104** (MI = 0.301318 nats), rank **37/78** → *medium* data-signal.
- Permutation importance (test): **0.000177 ± 0.001069** macro-F_1 drop, rank **37/78** → *very low* model-usage, noisy (mean/std ≈ 0.17).
- Combined score (normalized nMI + normalized perm): **0.519357**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`kurt_env`

Definition & intent (from `features.md` / extractor):

17. `kurt_env`

- **Intuition**: How heavy-tailed the amplitude distribution is.
  - Gaussianish noise → kurtosis ≈ 3.
  - Occasional huge pulses → much larger.
- **Formula** (population kurtosis):

$$\text{kurt} = \frac{\mathbb{E}\left[(\text{env}_\text{raw} - \mu)^4\right]}{\left(\mathbb{E}[(\text{env}_\text{raw} - \mu)^2]\right)^2},$$

```
 where $\mu = \mathbb{E}[\text{env}_\text{raw}]$ (with safe defaults for
short / constant sequences).
```

Measured importance (this run):

- nMI (train+val): **0.312344** (MI = 0.260631 nats), rank **43/78** → *medium* data-signal.
- Permutation importance (test): **0.000214 ± 0.000798** macro-F_1 drop, rank **35/78** → *low* model-usage, noisy (mean/std ≈ 0.27).
- Combined score (normalized nMI + normalized perm): **0.450489**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`PAPR_dB`

Definition & intent (from `features.md` / extractor):

10. `PAPR_dB`
    - **Intuition**: Measures how “peaky” the amplitude is.
        - OFDM-like or pulsed signals → high PAPR.
        - Smooth constant-envelope signals → low PAPR.
    - **Formula**:

$$\mathrm{PAPR} = 20\log_{10}\left(\frac{\max_n |z[n]| + \varepsilon}{\frac{1}{N}\sum_n |z[n]| + \varepsilon}\right).$$

1.3 Envelope autocorrelation

Let $\mathrm{env}[n] = |z[n]|$ and $\tilde{\mathrm{env}}[n] = \mathrm{env}[n] - \overline{\mathrm{env}}$.

We compute its autocorrelation efficiently using FFT:

$$\mathrm{AC}[k] = \mathrm{IFFT}\big(|\mathrm{FFT}(\tilde{\mathrm{env}})|^2\big),$$

then normalise by $\mathrm{AC}[0]$:

$$r[k] = \frac{\mathrm{AC}[k]}{\mathrm{AC}[0]}.$$

We only look at lags up to $k_{\max} \approx \mathrm{MAX\ LAG} \cdot f_s$.

Measured importance (this run):

- nMI (train+val): **0.277864** (MI = 0.231860 nats), rank **46/78** → *low* data-signal.
- Permutation importance (test): **-0.000021 ± 0.000766** macro-F_1 drop, rank **52/78** → *very low* model-usage, noisy (mean/std ≈ -0.03).
- Combined score (normalized nMI + normalized perm): **0.396382**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
    - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`crest_env`

Definition & intent (from `features.md` / extractor):

16. `crest_env`

- **Intuition**: Another impulsiveness metric: how tall the biggest amplitude peak is compared to the average amplitude.
- **Formula**:

$$\text{crest} = \frac{\max_n \text{env}_{\text{raw}}[n]}{\frac{1}{N}\sum_n \text{env}_{\text{raw}}[n] + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.278135** (MI = 0.232086 nats), rank **45/78** → *medium* data-signal.
- Permutation importance (test): **-0.000543 ± 0.000503** macro-F_1 drop, rank **71/78** → *very low* model-usage, noisy (mean/std ≈ -1.08).
- Combined score (normalized nMI + normalized perm): **0.385940**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`meanQ`

Definition & intent (from `features.md` / extractor):

2. `meanQ`
    - **Intuition**: Same as above but for Q.
    - **Formula**:

$$\text{meanQ} = \frac{1}{N}\sum_{n=0}^{N-1} Q[n].$$

Measured importance (this run):

- nMI (train+val): **0.267961** (MI = 0.223596 nats), rank **47/78** → *low* data-signal.
- Permutation importance (test): **-0.000366 ± 0.000555** macro-F_1 drop, rank **66/78** → *very low* model-usage, noisy (mean/std ≈ -0.66).
- Combined score (normalized nMI + normalized perm): **0.375071**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`stdI`

Definition & intent (from `features.md` / extractor):

3. `stdI`
   - **Intuition**: How much I varies around its mean (spread / energy in I).
   - **Formula**:

$$\mathrm{stdI} = \sqrt{\frac{1}{N}\sum_{n}\left(I[n] - \mathrm{meanI}\right)^2}.$$

Measured importance (this run):

- nMI (train+val): **0.169478** (MI = 0.141418 nats), rank **53/78** → *low* data-signal.
- Permutation importance (test): **0.000300 ± 0.001088** macro-F_1 drop, rank **29/78** → *low* model-usage, noisy (mean/std ≈ 0.28).
- Combined score (normalized nMI + normalized perm): **0.248237**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`env_ac_lag_s`

Definition & intent (from `features.md` / extractor):

12. `env_ac_lag_s`

- **Intuition**: Time between those repeating patterns (period where the peak happens).
- **Formula**:

$$k^* = \arg\max_{1\le k\le k_{\max}} r[k], \quad \mathrm{env\ ac\ lag} = \frac{k^*}{f_s}.$$

1.4 Raw power, out-of-band, crest, kurtosis, entropy

Measured importance (this run):

- nMI (train+val): **0.162546** (MI = 0.135634 nats), rank **56/78** → *low* data-signal.
- Permutation importance (test): **-0.000978 ± 0.000916** macro-F_1 drop, rank **74/78** → *very low* model-usage, noisy (mean/std ≈ -1.07).

- Combined score (normalized nMI + normalized perm): **0.211848**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`corrIQ`

Definition & intent (from `features.md` / extractor):

5. `corrIQ`
   - **Intuition**: How linearly related I and Q are.
     - GNSS-like proper noise → low correlation.
     - Certain modulations → strong correlation.
   - **Formula**:

$$\mathrm{corrIQ} = \frac{\sum_n (I[n] - \mathrm{meanI})(Q[n] - \mathrm{meanQ})}{\sqrt{\sum_n (I[n] - \mathrm{meanI})^2}\,\sqrt{\sum_n (Q[n] - \mathrm{meanQ})^2}}.$$

Measured importance (this run):

- nMI (train+val): **0.091598** (MI = 0.076433 nats), rank **63/78** → *low* data-signal.
- Permutation importance (test): **0.000685 ± 0.000893** macro-F_1 drop, rank **16/78** → *medium* model-usage, noisy (mean/std ≈ 0.77).
- Combined score (normalized nMI + normalized perm): **0.145004**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Keep for now: the model is using it, even if marginal MI is modest (could be interaction-driven).

`stdQ`

Definition & intent (from `features.md` / extractor):

4. `stdQ`

- **Intuition**: Same idea, but for Q.
- **Formula**:

$$\mathrm{stdQ} = \sqrt{\frac{1}{N}\sum_{n}\left(Q[n] - \mathrm{meanQ}\right)^2}.$$

Measured importance (this run):

- nMI (train+val): **0.089003** (MI = 0.074268 nats), rank **64/78** → *low* data-signal.
- Permutation importance (test): **0.000443 ± 0.000517** macro-F_1 drop, rank **20/78** → *medium* model-usage, noisy (mean/std ≈ 0.86).
- Combined score (normalized nMI + normalized perm): **0.136279**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

7.2 Group 2: Global Spectral Shape Features (6)

Global descriptors of the **power spectral density (PSD) shape**. These usually separate narrowband vs wideband patterns (e.g. entropy/flatness) and capture how "tone-like" vs "noise-like" the spectrum is.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macr
spec_flatness	0.68167	2	0.000914	
spec_spread_Hz	0.637672	6	0.00056	
spec_peak_power	0.595744	13	0.001648	
spec_centroid_Hz	0.540587	23	0.000372	
spec_rolloff95_Hz	0.487553	26	2.4e-05	
spec_peak_freq_Hz	0.379703	34	-0.000347	

Per-feature review:

`spec_flatness`

Definition & intent (from `features.md` / extractor):

21. `spec_flatness`

- **Intuition**: 1 for perfectly flat spectrum, near 0 for spectra with strong peaks. Good to distinguish narrowband tones from wideband noise.
- **Formula** (Wiener flatness):

$$\text{spec} = \frac{\exp\left(\frac{1}{K}\sum_k \ln(P_{xx}^{\text{norm}}[k])\right)}{\frac{1}{K}\sum_k P_{xx}^{\text{norm}}[k]}.$$

Measured importance (this run):

- nMI (train+val): **0.681670** (MI = 0.568810 nats), rank **2/78** → *very high* data-signal.
- Permutation importance (test): **0.000914 ± 0.000914** macro-F_1 drop, rank **12/78** → *medium* model-usage, moderately stable (mean/std ≈ 1.00).
- Combined score (normalized nMI + normalized perm): **0.992426**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - These are global spectral-shape descriptors; they often separate narrowband vs wideband interference cleanly.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`spec_spread_Hz`

Definition & intent (from `features.md` / extractor):

20. `spec_spread_Hz`

- **Intuition**: Effective bandwidth: how far the energy spreads around the centroid.
- **Formula**:

$$\text{spec spread} = \sqrt{\sum_k (f_k - \text{spec centroid})^2\, P_{xx}^{\text{norm}}[k]}.$$

Measured importance (this run):

- nMI (train+val): **0.637672** (MI = 0.532096 nats), rank **6/78** → *very high* data-signal.
- Permutation importance (test): **0.000560 ± 0.000535** macro-F_1 drop, rank **19/78** → *medium* model-usage, moderately stable (mean/std ≈ 1.05).
- Combined score (normalized nMI + normalized perm): **0.922237**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`spec_peak_power`

Definition & intent (from `features.md` / extractor):

24. `spec_peak_power`

- **Intuition**: How strong that main peak is relative to total power (because PSD is normalized).
- **Formula**:

$$\text{spec peak} = P_{xx}^{\text{norm}}[k^*].$$

---

3. Band Power Distribution (8)

Category intuition

Here we split the whole band into **8 equal slices** from $-f_s/2$ to $+f_s/2$ and measure how much normalized power lies in each. This is a coarse "spectral histogram".

Edges:

$$e_i = -\frac{f_s}{2} + i \cdot \frac{f_s}{8}, \quad i = 0, \ldots, 8.$$

For band i:

$$\mathcal{B}_i = \{k : e_i \leq f_k < e_{i+1}\},$$

$$B_i = \sum_{k \in \mathcal{B}_i} P_{xx}^{\text{norm}}[k], \quad \text{bandpower}_i = \frac{B_i}{\sum_{j=0}^{7} B_j + \varepsilon}.$$

Features:

25–32. `bandpower_0` … `bandpower_7`

- **Intuition**: Fraction of energy in each of the 8 sub-bands.
  Together they roughly sum to 1 and describe where the spectrum lives.
- **Formula**: as above.

---

4. Instantaneous Frequency Features (5)

Category intuition

These look at the **instantaneous frequency over time** (from the IQ phase). They are good for detecting:

- Frequency drift and chirps,
- Jitter in the carrier,
- Rapid frequency fluctuations (possibly FH-like behaviour).

We first compute:

- Unwrapped phase:

$$\phi[n] = \mathrm{unwrap}(\arg z[n]).$$

- Instantaneous frequency:

$$f_{\mathrm{inst}}[k] = \frac{f_s}{2\pi}\,(\phi[k+1] - \phi[k]), \quad k = 0, \ldots, N-2,$$

then clip out extreme percentiles for robustness.

Measured importance (this run):

- nMI (train+val): **0.595744** (MI = 0.497110 nats), rank **13/78** → *high* data-signal.
- Permutation importance (test): **0.001648 ± 0.002047** macro-F_1 drop, rank **8/78** → *high* model-usage, noisy (mean/std ≈ 0.81).
- Combined score (normalized nMI + normalized perm): **0.884935**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`spec_centroid_Hz`

Definition & intent (from `features.md` / extractor):

19. `spec_centroid_Hz`

- **Intuition**: “Center of mass” of the spectrum; indicates which side (positive/negative) holds more energy.

- **Formula**:

$$\text{spec centroid} = \sum_k f_k \, P_{xx}^{\text{norm}}[k].$$

Measured importance (this run):

- nMI (train+val): **0.540587** (MI = 0.451085 nats), rank **23/78** → *high* data-signal.
- Permutation importance (test): **0.000372 ± 0.000559** macro-F_1 drop, rank **24/78** → *medium* model-usage, noisy (mean/std ≈ 0.67).
- Combined score (normalized nMI + normalized perm): **0.779697**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`spec_rolloff95_Hz`

Definition & intent (from `features.md` / extractor):

22. `spec_rolloff95_Hz`

- **Intuition**: Frequency such that 95% of total spectral energy lies below it → "edge" of effective band.
- **Formula**:

$$C[m] = \sum_{k \le m} P_{xx}^{\text{norm}}[k],$$

```
find smallest $m$ with $C[m] \ge 0.95$, then
```

$$\text{spec rolloff95} = f_m.$$

Measured importance (this run):

- nMI (train+val): **0.487553** (MI = 0.406831 nats), rank **26/78** → *medium* data-signal.
- Permutation importance (test): **0.000024 ± 0.000301** macro-F_1 drop, rank **42/78** → *very low* model-usage, noisy (mean/std ≈ 0.08).
- Combined score (normalized nMI + normalized perm): **0.696746**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`spec_peak_freq_Hz`

Definition & intent (from `features.md` / extractor):

23. `spec_peak_freq_Hz`

- **Intuition**: Frequency of the strongest spectral component (e.g. a CW jammer).
- **Formula**:

$$k^* = \arg\max_k P_{xx}^{\text{norm}}[k], \quad \text{spec peak freq} = f_{k^*}.$$

Measured importance (this run):

- nMI (train+val): **0.379703** (MI = 0.316838 nats), rank **34/78** → *medium* data-signal.
- Permutation importance (test): **-0.000347 ± 0.000558** macro-F_1 drop, rank **64/78** → *very low* model-usage, noisy (mean/std ≈ -0.62).
- Combined score (normalized nMI + normalized perm): **0.535036**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

7.3 Group 3: Band Power Distribution (8)

Eight relative PSD integrals over equally spaced frequency bands across $[-f_s/2, f_s/2]$. These act like a coarse histogram of where spectral energy sits.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macroF1_d
bandpower_6	0.639091	5	0.003787	0
bandpower_5	0.604086	12	0.002522	0
bandpower_4	0.554529	19	0.002498	0
bandpower_7	0.587283	15	0.000165	0
bandpower_0	0.581352	17	7.7e-05	0
bandpower_2	0.554255	20	0.001433	0

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macroF1_dı
bandpower_1	0.579438	18	-0.000525	0
bandpower_3	0.304355	44	0.000399	0

Per-feature review:

`bandpower_6`

Definition & intent (from `features.md` / extractor):

`bandpower_6` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

  $$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

  Then:

  $$\mathrm{bandpower}_6 = \int_{B_6} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.639091** (MI = 0.533280 nats), rank **5/78** → *very high* data-signal.
- Permutation importance (test): **0.003787 ± 0.001829** macro-F_1 drop, rank **3/78** → *very high* model-usage, stable (mean/std ≈ 2.07).
- Combined score (normalized nMI + normalized perm): **0.991183**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_5`

Definition & intent (from `features.md` / extractor):

`bandpower_5` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

  Then:

$$\mathrm{bandpower}_5 = \int_{B_5} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.604086** (MI = 0.504071 nats), rank **12/78** → *high* data-signal.
- Permutation importance (test): **0.002522 ± 0.000961** macro-F_1 drop, rank **5/78** → *high* model-usage, stable (mean/std ≈ 2.62).
- Combined score (normalized nMI + normalized perm): **0.914964**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_4`

Definition & intent (from `features.md` / extractor):

`bandpower_4` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

  Then:

$$\mathrm{bandpower}_4 = \int_{B_4} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.554529** (MI = 0.462719 nats), rank **19/78** → *high* data-signal.
- Permutation importance (test): **0.002498 ± 0.001232** macro-F_1 drop, rank **6/78** → *high* model-usage, stable (mean/std ≈ 2.03).
- Combined score (normalized nMI + normalized perm): **0.843701**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_7`

Definition & intent (from `features.md` / extractor):

`bandpower_7` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

  Then:

$$\text{bandpower}_7 = \int_{B_7} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \text{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.587283** (MI = 0.490050 nats), rank **15/78** → *high* data-signal.
- Permutation importance (test): **0.000165 ± 0.001494** macro-F_1 drop, rank **38/78** → *very low* model-usage, noisy (mean/std ≈ 0.11).
- Combined score (normalized nMI + normalized perm): **0.842100**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`bandpower_0`

Definition & intent (from `features.md` / extractor):

`bandpower_0` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).

The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \dots, 8.$$

Then:

$$\mathrm{bandpower}_0 = \int_{B_0} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.581352** (MI = 0.485101 nats), rank **17/78** → *high* data-signal.
- Permutation importance (test): **0.000077 ± 0.001088** macro-F_1 drop, rank **41/78** → *very low* model-usage, noisy (mean/std ≈ 0.07).
- Combined score (normalized nMI + normalized perm): **0.831807**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_2`

Definition & intent (from `features.md` / extractor):

`bandpower_2` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

Then:

$$\text{bandpower}_2 = \int_{B_2} P(f)\,df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \text{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.554255** (MI = 0.462490 nats), rank **20/78** → *high* data-signal.
- Permutation importance (test): **0.001433 ± 0.000808** macro-F_1 drop, rank **9/78** → *high* model-usage, moderately stable (mean/std ≈ 1.77).
- Combined score (normalized nMI + normalized perm): **0.821226**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_1`

Definition & intent (from `features.md` / extractor):

`bandpower_1` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\,df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

$$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

Then:

$$\mathrm{bandpower}_1 = \int_{B_1} P(f)\, df.$$

- **Range:** $[0, 1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.579438** (MI = 0.483503 nats), rank **18/78** → *high* data-signal.
- Permutation importance (test): **-0.000525 ± 0.000999** macro-F_1 drop, rank **70/78** → *very low* model-usage, noisy (mean/std ≈ -0.53).
- Combined score (normalized nMI + normalized perm): **0.816584**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`bandpower_3`

Definition & intent (from `features.md` / extractor):

`bandpower_3` (Band Power Distribution)

- **Definition:** Let $P(f)$ be the normalized PSD of the complex IQ signal on the two-sided frequency axis $f \in [-f_s/2, f_s/2]$ (so that $\int P(f)\, df = 1$).
  The bandpowers split the Nyquist band into 8 equal-width intervals:

  $$B_i = [f_i, f_{i+1}), \quad f_i = -\frac{f_s}{2} + i\,\frac{f_s}{8}, \quad i = 0, 1, \ldots, 8.$$

  Then:

  $$\mathrm{bandpower}_3 = \int_{B_3} P(f)\, df.$$

- **Range:** $[0,1]$ and $\sum_{i=0}^{7} \mathrm{bandpower}_i \approx 1$ (up to numerical error).
- **Intuition:** Captures *where* spectral energy lives. Narrowband interference concentrates power in few bands; wideband interference spreads it; chirps move energy over time but still change the long-term distribution depending on sweep width.
- **Caveats:** If the PSD is normalized before integration (as in your extractor), these are *relative* powers; they ignore absolute power unless coupled with total-power features (e.g. `mag_mean`, `PAPR_dB`).

Measured importance (this run):

- nMI (train+val): **0.304355** (MI = 0.253965 nats), rank **44/78** → *medium* data-signal.
- Permutation importance (test): **0.000399 ± 0.000892** macro-F_1 drop, rank **22/78** → *medium* model-usage, noisy (mean/std ≈ 0.45).
- Combined score (normalized nMI + normalized perm): **0.442921**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Bandpower features are *relative* PSD mass; interpret jointly (energy redistribution across bands).

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

7.4 Group 4: Instantaneous Frequency Features (5)

Instantaneous frequency (IF) features derived from phase differences. Designed to expose **frequency ramps** and variability typical of chirps and some swept interference.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macroF
instf_kurtosis	0.618933	9	0.000349	
instf_std_Hz	0.612333	11	-0.000415	
instf_mean_Hz	0.525195	24	0.000854	
instf_dZCR_per_s	0.448108	28	-0.000146	
instf_slope_Hzps	0.154167	57	0.001429	

Per-feature review:

`instf_kurtosis`

Definition & intent (from `features.md` / extractor):

36. `instf_kurtosis`

- **Intuition**: Whether the inst. frequency has occasional big jumps (heavy tails) vs more Gaussian noise.
- **Formula** (population kurtosis):

$$\text{instf} = \frac{\mathbb{E}\big[(f_{\text{inst}} - \mu_f)^4\big]}{\big(\mathbb{E}[(f_{\text{inst}} - \mu_f)^2]\big)^2}.$$

Measured importance (this run):

- nMI (train+val): **0.618933** (MI = 0.516460 nats), rank **9/78** → *very high* data-signal.
- Permutation importance (test): **0.000349 ± 0.000561** macro-F_1 drop, rank **27/78** → *low* model-usage, noisy (mean/std ≈ 0.62).
- Combined score (normalized nMI + normalized perm): **0.891100**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`instf_std_Hz`

Definition & intent (from `features.md` / extractor):

34. `instf_std_Hz`

- **Intuition**: How much the instantaneous frequency wiggles around its mean (frequency jitter).
- **Formula**:

$$\text{instf std} = \sqrt{\frac{1}{M}\sum_k (f_{\text{inst}}[k] - \text{instf mean})^2}.$$

Measured importance (this run):

- nMI (train+val): **0.612333** (MI = 0.510952 nats), rank **11/78** → *high* data-signal.
- Permutation importance (test): **-0.000415 ± 0.000433** macro-F_1 drop, rank **69/78** → *very low* model-usage, noisy (mean/std ≈ -0.96).
- Combined score (normalized nMI + normalized perm): **0.865853**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`instf_mean_Hz`

Definition & intent (from `features.md` / extractor):

33. `instf_mean_Hz`

- **Intuition**: Average carrier offset of the chunk. Non-zero means the centre frequency is shifted.
- **Formula** (with $M = N - 1$):

$$\text{instf mean} = \frac{1}{M} \sum_{k=0}^{M-1} f_{\text{inst}}[k].$$

Measured importance (this run):

- nMI (train+val): **0.525195** (MI = 0.438241 nats), rank **24/78** → *high* data-signal.
- Permutation importance (test): **0.000854 ± 0.000963** macro-F_1 drop, rank **13/78** → *medium* model-usage, noisy (mean/std ≈ 0.89).
- Combined score (normalized nMI + normalized perm): **0.767720**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`instf_dZCR_per_s`

Definition & intent (from `features.md` / extractor):

37. `instf_dZCR_per_s`

- **Intuition**: How often the *change* in inst. frequency flips sign per second → how "zig-zaggy" the frequency evolution is.
- **Formula**:

$$d_f[k] = f_{\text{inst}}[k+1] - f_{\text{inst}}[k],$$

$$\text{instf dZCR per} = \text{ZCR}(d_f) \cdot f_s.$$

---

5. Envelope, Cepstrum, Pulse & Narrowband Salience (4)

Category intuition

These features look at the **amplitude envelope** and at how much the spectrum is dominated by a few peaks. They catch things like:

- Periodic amplitude modulation (AM),
- DME-style pulses,
- Tones that dominate the spectrum.

Measured importance (this run):

- nMI (train+val): **0.448108** (MI = 0.373917 nats), rank **28/78** → *medium* data-signal.
- Permutation importance (test): **-0.000146 ± 0.000423** macro-F_1 drop, rank **56/78** → *very low* model-usage, noisy (mean/std ≈ -0.35).
- Combined score (normalized nMI + normalized perm): **0.636890**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`instf_slope_Hzps`

Definition & intent (from `features.md` / extractor):

35. `instf_slope_Hzps`

- **Intuition**: Linear trend of frequency vs time, i.e. chirp slope.
  - Positive → frequency ramps up.
  - Negative → ramps down.
  - Near zero → stationary carrier.
- **Formula**: least-squares fit $f_{\text{inst}}[k] \approx a t_k + b$ with $t_k = k/f_s$:

$$\text{instf slope} = a.$$

Measured importance (this run):

- nMI (train+val): **0.154167** (MI = 0.128642 nats), rank **57/78** → *low* data-signal.
- Permutation importance (test): **0.001429 ± 0.000688** macro-F_1 drop, rank **10/78** → *high* model-usage, stable (mean/std ≈ 2.08).
- Combined score (normalized nMI + normalized perm): **0.249789**.

Interpretation:

- Cross-method read: surprising: low nMI but high permutation impact (possible interaction/nonlinear usage, or reliance on a distribution quirk).
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Keep for now: the model is using it, even if marginal MI is modest (could be interaction-driven).

7.5 Group 5: Envelope, Cepstrum, Pulse & Narrowband Salience (4)

Features targeting **envelope modulation**, cepstral periodicity and pulsing. Particularly relevant when interference has repetitive bursts or structured AM.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macroF
nb_peak_salience	0.636317	7	0.00058	
dme_duty	0.03996	69	0.000208	
dme_pulse_count	0.02324	70	-0.000212	
cep_peak_env	0	77	0	

Per-feature review:

`nb_peak_salience`

Definition & intent (from `features.md` / extractor):

41. `nb_peak_salience`

- **Intuition**: How much of the spectral energy is concentrated in the **top 5 peaks** vs the rest.
  - Large value → strong tones.
  - Small value → more spread / noise-like.
- **Formula**: using normalized PSD $P_{xx}^{\mathrm{norm}}[k]$,
  - Let $\mathcal{T}$ be indices of 5 largest bins.
  - Top power: $P_{\mathrm{top}} = \sum_{k\in\mathcal{T}} P_{xx}^{\mathrm{norm}}[k]$.
  - Remaining: $P_{\mathrm{rest}} = 1 - P_{\mathrm{top}}$.

$$\mathrm{nb\ peak} = \frac{P_{\mathrm{top}}}{P_{\mathrm{rest}} + \varepsilon}.$$

---

6. Narrowband Peaks, AM & Chirp Features (8)

Category intuition

This group refines the view on **narrowband tones**, **AM behaviour**, and **chirp-like sweeps**:

- How many peaks,
- How regularly spaced they are,
- How strongly the amplitude is modulated,
- Whether the signal behaves like a clean chirp.

6.1. Narrowband peaks and spacing

Measured importance (this run):

- nMI (train+val): **0.636317** (MI = 0.530965 nats), rank **7/78** → *very high* data-signal.
- Permutation importance (test): **0.000580 ± 0.001243** macro-F_1 drop, rank **18/78** → *medium* model-usage, noisy (mean/std ≈ 0.47).
- Combined score (normalized nMI + normalized perm): **0.920721**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Narrowband helper features are designed to trigger on tonal/line interference.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`dme_duty`

Definition & intent (from `features.md` / extractor):

40. `dme_duty`

- **Intuition**: Fraction of time where the smoothed envelope is “high” → how busy the pulsed interference is.
- **Formula**:

$$\mathrm{dme} = \frac{1}{N}\sum_n a[n].$$

Measured importance (this run):

- nMI (train+val): **0.039960** (MI = 0.033344 nats), rank **69/78** → *very low* data-signal.
- Permutation importance (test): **0.000208 ± 0.000456** macro-F_1 drop, rank **36/78** → *low* model-usage, noisy (mean/std ≈ 0.46).
- Combined score (normalized nMI + normalized perm): **0.061377**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`dme_pulse_count`

Definition & intent (from `features.md` / extractor):

39. `dme_pulse_count`

- **Intuition**: Rough count of strong pulses in the envelope (designed with DME-like bursts in mind).
- **Formula idea**:
  - Smooth $\mathrm{env}_{\mathrm{raw}}$ with moving average of length $\approx 0.5\,\mu$s to get $\mathrm{env}_s[n]$.
  - Threshold $T = \mathbb{E}[\mathrm{env}_s] + 3 \cdot \mathrm{std}(\mathrm{env}_s)$.
  - Boolean above threshold: $a[n] = 1$ if $\mathrm{env}_s[n] > T$, else 0.
  - Rising edges $r[n] = 1$ when $a[n] = 1$ and $a[n-1] = 0$.

$$\mathrm{dme\ pulse} = \sum_n r[n].$$

Measured importance (this run):

- nMI (train+val): **0.023240** (MI = 0.019392 nats), rank **70/78** → *very low* data-signal.
- Permutation importance (test): **-0.000212 ± 0.000324** macro-F_1 drop, rank **57/78** → *very low* model-usage, noisy (mean/std ≈ -0.65).
- Combined score (normalized nMI + normalized perm): **0.028794**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`cep_peak_env`

Definition & intent (from `features.md` / extractor):

38. `cep_peak_env`

- **Intuition**: Strength of a **periodic pattern** in the envelope (e.g. regularly spaced pulses) in a given quefrency range (here ≈0.2–5 ms).
- **Formula** (simplified):
  - Envelope $e[n] = |z[n]| - \overline{|z|}$, window $w[n]$.
  - Spectrum:

$$S[k] = \mathrm{FFT}(e[n]w[n]).$$

- Log magnitude:

$$L[k] = \log(|S[k]| + \varepsilon).$$

- Real cepstrum:

$$c[q] = \mathrm{IFFT}(L[k]).$$

- With quefrency q/f_s in $[q_{\min}, q_{\max}]$ (e.g. $[2 \cdot 10^{-4}, 5 \cdot 10^{-3}]$ s):

$$\text{cep peak} = \max_{q \in [q_{\min}, q_{\max}]} c[q].$$

Measured importance (this run):

- nMI (train+val): **0.000000** (MI = 0.000000 nats), rank **77/78** → *very low* data-signal.
- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **48/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.000000**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.6 Group 6: Narrowband Peaks, AM & Chirp Features (8)

Hand-crafted detectors for **narrowband lines/peaks**, amplitude modulation indicators, and chirp proxies. These are often complementary to the more generic PSD metrics.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_mac
nb_peak_count	0.648161	4	-0.001013	
env_mod_index	0.382492	31	0	
nb_spacing_std_Hz	0.326801	39	-0.000324	
nb_spacing_med_Hz	0.316325	41	-0.000769	
chirp_slope_Hzps	0.171327	52	-0.001067	
chirp_r2	0.019719	74	-0.00033	
env_dom_peak_norm	0.004101	76	0	
env_dom_freq_Hz	0	78	0	

Per-feature review:

`nb_peak_count`

Definition & intent (from `features.md` / extractor):

42. `nb_peak_count`

- **Intuition**: Number of significant spectral peaks above a prominence threshold. Multi-tone jammers → more peaks; single CW → 1 strong peak.
- **Formula**:
  - $P_{xx}[k]$: PSD (unnormalised) on z.
  - $P_{\max} = \max_k P_{xx}[k]$, threshold $\mathrm{prom} = 0.03 P_{\max}$.
  - Use `find_peaks` (SciPy) with this prominence to get peak index set $\mathcal{P}$.

$$\mathrm{nb\ peak} = |\mathcal{P}|.$$

Measured importance (this run):

- nMI (train+val): **0.648161** (MI = 0.540849 nats), rank **4/78** → *very high* data-signal.
- Permutation importance (test): **-0.001013 ± 0.000941** macro-F_1 drop, rank **75/78** → *very low* model-usage, noisy (mean/std ≈ -1.08).
- Combined score (normalized nMI + normalized perm): **0.904618**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Narrowband helper features are designed to trigger on tonal/line interference.

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`env_mod_index`

Definition & intent (from `features.md` / extractor):

45. `env_mod_index`

- **Intuition**: How strongly the amplitude is modulated.
  - Constant envelope → near 0.
  - Strong AM → larger.
- **Formula**:

$$\mathrm{env\ mod} = \frac{\mathbb{E}[(\mathrm{env} - \mu)^2]}{\mu^2 + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.382492** (MI = 0.319165 nats), rank **31/78** → *medium* data-signal.

- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **45/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.546221**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`nb_spacing_std_Hz`

Definition & intent (from `features.md` / extractor):

44. `nb_spacing_std_Hz`

- **Intuition**: If there are several peaks, these metrics tell you how **regularly spaced** they are in frequency.
  - E.g. a comb of tones → fairly constant spacing.
- **Formula** (if $|\mathcal{P}| \geq 2$):
  - Peak frequencies f_i (sorted),
  - Spacings $s_j = f_{j+1} - f_j$.

$$\text{nb spacing med} = \text{median}(s_j), \quad \text{nb spacing std} = \text{std}(s_j).$$

If fewer than 2 peaks, both are 0.

6.2. AM envelope features

Let $\text{env}[n] = |z[n]|$, mean μ, zero-mean $e[n] = \text{env}[n] - \mu$.

Measured importance (this run):

- nMI (train+val): **0.326801** (MI = 0.272694 nats), rank **39/78** → *medium* data-signal.
- Permutation importance (test): **-0.000324 ± 0.000767** macro-F_1 drop, rank **61/78** → *very low* model-usage, noisy (mean/std ≈ -0.42).
- Combined score (normalized nMI + normalized perm): **0.459964**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:

    - Narrowband helper features are designed to trigger on tonal/line interference.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`nb_spacing_med_Hz`

Definition & intent (from `features.md` / extractor):

43. `nb_spacing_med_Hz`

Measured importance (this run):

- nMI (train+val): **0.316325** (MI = 0.263953 nats), rank **41/78** → *medium* data-signal.
- Permutation importance (test): **-0.000769 ± 0.000189** macro-F_1 drop, rank **73/78** → *very low* model-usage, noisy (mean/std ≈ -4.07).
- Combined score (normalized nMI + normalized perm): **0.435778**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
    - Narrowband helper features are designed to trigger on tonal/line interference.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`chirp_slope_Hzps`

Definition & intent (from `features.md` / extractor):

48. `chirp_slope_Hzps`

- **Intuition**: Average frequency change rate across the whole chunk, estimated from these centroids.
  Basically another chirp slope (complementing `instf_slope_Hzps`).
- **Formula**: Fit $c_s \approx a t_s + b$ by least squares and take

$$\text{chirp slope} = a.$$

Measured importance (this run):

- nMI (train+val): **0.171327** (MI = 0.142962 nats), rank **52/78** → *low* data-signal.
- Permutation importance (test): **-0.001067 ± 0.000468** macro-F_1 drop, rank **77/78** → *very low* model-usage, noisy (mean/std ≈ -2.28).

- Combined score (normalized nMI + normalized perm): **0.222546**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`chirp_r2`

Definition & intent (from `features.md` / extractor):

49. `chirp_r2`

- **Intuition**: How well a **linear** model explains the centroid evolution.
  - Close to 1 → clean linear chirp.
  - Small → behaviour is messy or non-chirp.
- **Formula**:

$$SS_{\mathrm{res}} = \sum_s (c_s - \hat{c}_s)^2, \quad \hat{c}_s = a t_s + b,$$

$$SS_{\mathrm{tot}} = \sum_s (c_s - \bar{c})^2 + \varepsilon,$$

$$\mathrm{chirp} = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}}.$$

---

7. Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)

Category intuition

GNSS signals are **cyclostationary** at chip rate and have specific higher-order statistics. This category tries to capture:

- GNSS-like chip periodicity,
- Modulation type / non-Gaussianity,
- Time-frequency “burstiness”,
- Rapid energy changes.

7.1. Cyclostationary proxies

We define:

$$\text{cyclo lag}(z, L) = \frac{\left|\sum_{n=0}^{N-L-1} z[n+L]\,\overline{z[n]}\right|}{\sqrt{\left(\sum_n |z[n+L]|^2\right)\left(\sum_n |z[n]|^2\right)} + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.019719** (MI = 0.016454 nats), rank **74/78** → *very low* data-signal.
- Permutation importance (test): **-0.000330 ± 0.000723** macro-F_1 drop, rank **63/78** → *very low* model-usage, noisy (mean/std ≈ -0.46).
- Combined score (normalized nMI + normalized perm): **0.021307**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`env_dom_peak_norm`

Definition & intent (from `features.md` / extractor):

47. `env_dom_peak_norm`

- **Intuition**:
  - `env_dom_freq_Hz` → dominant **modulation frequency** of the envelope (in 30–7000 Hz band).
  - `env_dom_peak_norm` → how dominant that modulation is relative to all others.
- **Formula**:
  - FFT: $E[k] = \mathrm{FFT}(e[n]w[n])$, envelope power $P_e[k] = |E[k]|^2$.
  - Frequencies $f_k^{(e)}$.
  - Band mask $\mathcal{B}_e = \{k : f_{\min} \le f_k^{(e)} \le f_{\max}\}$.

$$k^* = \arg\max_{k \in \mathcal{B}_e} P_e[k], \quad \text{env dom freq} = f_{k^*}^{(e)},$$

$$\text{env dom peak} = \frac{P_e[k^*]}{\sum_{k \in \mathcal{B}_e} P_e[k] + \varepsilon}.$$

6.3. Chirp slope and linearity

We split $z[n]$ into S equal segments and per segment compute the spectral centroid c_s and its center time t_s.

Measured importance (this run):

- nMI (train+val): **0.004101** (MI = 0.003422 nats), rank **76/78** → *very low* data-signal.

- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **44/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.005857**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`env_dom_freq_Hz`

Definition & intent (from `features.md` / extractor):

46. `env_dom_freq_Hz`

Measured importance (this run):

- nMI (train+val): **0.000000** (MI = 0.000000 nats), rank **78/78** → *very low* data-signal.
- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **47/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.000000**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.7 Group 7: Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)

Cyclostationarity and higher-order spectral statistics (e.g. spectral kurtosis). These aim to detect **non-Gaussianity in frequency**, periodic structure, and distinctive jammer fingerprints.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_macr
tkeo_env_mean	0.583284	16	4e-06	
spec_kurtosis_mean	0.216656	51	0.010794	
cyclo_2chip_corr	0.265585	49	-0.000112	
spec_kurtosis_max	0.167275	54	-0.000407	
cyclo_chip_corr	0.073565	65	0.000247	
cumulant_c42_mag	0.050652	68	0.000387	
cumulant_c40_mag	0.052513	67	-0.000362	

Per-feature review:

`tkeo_env_mean`

Definition & intent (from `features.md` / extractor):

56. `tkeo_env_mean`

- **Intuition**: Sensitive to **instantaneous energy changes** in the envelope (like a second derivative adapted for energy).
  Higher for signals with rapid local changes.
- **Formula**:

  Let $e[n] = |z[n]|$ and for $n = 1, \ldots, N-2$:

$$\psi[n] = e[n]^2 - e[n-1]e[n+1],$$

then clamp $\psi[n] \geq 0$ and

$$\text{tkeo env} = \frac{\mathbb{E}[\psi[n]]}{\mathbb{E}[e[n]]^2 + \varepsilon}.$$

---

8. Higher-order I/Q Stats & Circularity (6)

Category intuition

Proper complex Gaussian noise has:

- zero skewness,
- kurtosis ≈ 3,
- and is **circular** (no preferred axis in I/Q plane).

These features measure how far we are from that ideal, giving clues about modulation and interference structure.

---

57. `skewI`, 58. `skewQ`

- **Intuition**: Asymmetry of I and Q histograms.
  Heavy skew can indicate offset or one-sided modulation.
- **Formula**:

$$\text{skewI} = \frac{\mathbb{E}[(I-\mu_I)^3]}{\sigma_I^3+\varepsilon}, \quad \text{skewQ} = \frac{\mathbb{E}[(Q-\mu_Q)^3]}{\sigma_Q^3+\varepsilon}.$$

59. `kurtI`, 60. `kurtQ`

- **Intuition**: Tail heaviness of I and Q distributions.
  Pulses or outliers increase these.
- **Formula**:

$$\text{kurtI} = \frac{\mathbb{E}[(I-\mu_I)^4]}{\left(\mathbb{E}[(I-\mu_I)^2]\right)^2},$$

$$\text{kurtQ} = \frac{\mathbb{E}[(Q-\mu_Q)^4]}{\left(\mathbb{E}[(Q-\mu_Q)^2]\right)^2}.$$

61. `circularity_mag`, 62. `circularity_phase_rad`

- **Intuition**:
  - Proper circular complex noise → $E[z^2] \approx 0$, so circularity magnitude ≈ 0.
  - Strong modulation confined to I or Q → large magnitude, phase capturing orientation.
- **Formula**:

$$d = \mathbb{E}[|z|^2] + \varepsilon, \quad \rho = \frac{\mathbb{E}[z^2]}{d},$$

$$\text{circularity} = |\rho|, \quad \text{circularity phase} = \arg(\rho).$$

---

9. Inequality, Symmetry, DC Notch & Peakiness (6)

Category intuition

These features summarise:

- How unevenly power is distributed across frequencies and amplitudes (Gini, peakiness),

- How symmetric the spectrum is around DC,
- Whether we have a notch or spike near DC.

Measured importance (this run):

- nMI (train+val): **0.583284** (MI = 0.486713 nats), rank **16/78** → *high* data-signal.
- Permutation importance (test): **0.000004 ± 0.001250** macro-F_1 drop, rank **43/78** → *very low* model-usage, noisy (mean/std ≈ 0.00).
- Combined score (normalized nMI + normalized perm): **0.833043**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`spec_kurtosis_mean`

Definition & intent (from `features.md` / extractor):

54. `spec_kurtosis_mean`

- **Intuition**: Average "burstiness" across all frequencies.
  If many frequencies are sometimes very loud and sometimes quiet, this rises.
- **Formula**:

$$\text{spec kurtosis} = \frac{1}{I}\sum_{i=1}^{I}\text{kurt}_i.$$

Measured importance (this run):

- nMI (train+val): **0.216656** (MI = 0.180786 nats), rank **51/78** → *low* data-signal.
- Permutation importance (test): **0.010794 ± 0.001272** macro-F_1 drop, rank **2/78** → *very high* model-usage, stable (mean/std ≈ 8.48).
- Combined score (normalized nMI + normalized perm): **0.533208**.

Interpretation:

- Cross-method read: surprising: low nMI but high permutation impact (possible interaction/nonlinear usage, or reliance on a distribution quirk).
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Keep for now: the model is using it, even if marginal MI is modest (could be interaction-driven).

`cyclo_2chip_corr`

Definition & intent (from `features.md` / extractor):

51. `cyclo_2chip_corr`

- **Intuition**: Same idea but at 2 chip periods.
- **Formula**:

$$L_2 = \mathrm{round}\big(f_s/2.046\,\mathrm{MHz}\big),$$

$$\mathrm{cyclo\ 2chip} = \mathrm{cyclo\ lag}(z, L_2).$$

7.2. Higher-order cumulants

We normalise z to unit average power and compute 4th-order cumulants.

Let

$$z_c[n] = z[n] - \overline{z}, \quad p = \mathbb{E}[|z_c|^2] + \varepsilon, \quad z_n[n] = \frac{z_c[n]}{\sqrt{p}}.$$

Moments:

$$m_{20} = \mathbb{E}[z_n^2], \quad m_{40} = \mathbb{E}[z_n^4], \quad m_{42} = \mathbb{E}[|z_n|^2 z_n^2].$$

Cumulants:

$$c_{40} = m_{40} - 3m_{20}^2, \quad c_{42} = m_{42} - |m_{20}|^2 - 2.$$

Measured importance (this run):

- nMI (train+val): **0.265585** (MI = 0.221613 nats), rank **49/78** → *low* data-signal.
- Permutation importance (test): **-0.000112 ± 0.000743** macro-F_1 drop, rank **55/78** → *very low* model-usage, noisy (mean/std ≈ -0.15).
- Combined score (normalized nMI + normalized perm): **0.376949**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`spec_kurtosis_max`

Definition & intent (from `features.md` / extractor):

55. `spec_kurtosis_max`

- **Intuition**: Maximal burstiness at any frequency.
  Good for detecting a single frequency that occasionally spikes.
- **Formula**:

$$\text{spec kurtosis} = \max_i \text{kurt}_i.$$

7.4. Teager–Kaiser on envelope

Measured importance (this run):

- nMI (train+val): **0.167275** (MI = 0.139580 nats), rank **54/78** → *low* data-signal.
- Permutation importance (test): **-0.000407 ± 0.000976** macro-F_1 drop, rank **67/78** → *very low* model-usage, noisy (mean/std ≈ -0.42).
- Combined score (normalized nMI + normalized perm): **0.230433**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`cyclo_chip_corr`

Definition & intent (from `features.md` / extractor):

50. `cyclo_chip_corr`

- **Intuition**: Correlation between the signal and a copy shifted by **1 chip period** (approx). GNSS-like signals should have non-zero structure here; pure noise or generic jammers less so.
- **Formula**: lag

$$L_1 = \mathrm{round}\big(f_s/1.023\,\mathrm{MHz}\big),$$

$$\mathrm{cyclo\ chip} = \mathrm{cyclo\ lag}(z, L_1).$$

Measured importance (this run):

- nMI (train+val): **0.073565** (MI = 0.061385 nats), rank **65/78** → *low* data-signal.
- Permutation importance (test): **0.000247 ± 0.000598** macro-F_1 drop, rank **31/78** → *low* model-usage, noisy (mean/std ≈ 0.41).
- Combined score (normalized nMI + normalized perm): **0.110171**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`cumulant_c42_mag`

Definition & intent (from `features.md` / extractor):

53. `cumulant_c42_mag`

- **Intuition**: Similar to above for C_{42}; together, C_{40} and C_{42} help discriminate between modulations (BPSK, QPSK, etc.) and noise.
- **Formula**:

$$\mathrm{cumulant\ c42} = |c_{42}|.$$

7.3. Spectral kurtosis

We compute a spectrogram $S_{xx}[i, j]$ (frequency i, time j) of $z[n]$ (PSD mode). For each frequency bin i we look at how its power varies across time.

Per-bin kurtosis:

$$\mathrm{kurt}_i = \frac{\mathbb{E}_j[(S_{xx}[i,j] - \mu_i)^4]}{(\mathbb{E}_j[(S_{xx}[i,j] - \mu_i)^2])^2}, \quad \mu_i = \mathbb{E}_j[S_{xx}[i,j]].$$

Measured importance (this run):

- nMI (train+val): **0.050652** (MI = 0.042266 nats), rank **68/78** → *very low* data-signal.
- Permutation importance (test): **0.000387 ± 0.000619** macro-F_1 drop, rank **23/78** → *medium* model-usage, noisy (mean/std ≈ 0.62).
- Combined score (normalized nMI + normalized perm): **0.080349**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`cumulant_c40_mag`

Definition & intent (from `features.md` / extractor):

52. `cumulant_c40_mag`

- **Intuition**: Magnitude of 4th-order cumulant C_{40}; sensitive to modulation format and non-Gaussianity.
- **Formula**:

$$\text{cumulant c40} = |c_{40}|.$$

Measured importance (this run):

- nMI (train+val): **0.052513** (MI = 0.043819 nats), rank **67/78** → *very low* data-signal.
- Permutation importance (test): **-0.000362 ± 0.000688** macro-F_1 drop, rank **65/78** → *very low* model-usage, noisy (mean/std ≈ -0.53).
- Combined score (normalized nMI + normalized perm): **0.067486**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.8 Group 8: Higher-order I/Q Stats & Circularity (6)

Higher-order moments of I/Q and complex circularity. Useful for distinguishing proper noise-like signals from improper or deterministic components, and for detecting I/Q imbalance artifacts.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_mac
kurtI	0.315569	42	0.000245	
kurtQ	0.316901	40	-0.000325	
circularity_phase_rad	0.152867	58	0.000585	

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_mac
circularity_mag	0.054925	66	0.000153	
skewQ	0.014931	75	0.000228	
skewI	0.020062	73	-0.00145	

Per-feature review:

`kurtI`

Definition & intent (from `features.md` / extractor):

`kurtI` (Higher-order moments of I)

- **Definition:** Population kurtosis (non-excess) of I:

$$\mathrm{kurtI} = \frac{\mathbb{E}\left[(I - \mu_I)^4\right]}{\left(\mathbb{E}\left[(I - \mu_I)^2\right]\right)^2}.$$

  For a Gaussian distribution, kurtosis is 3 (when using the non-excess convention).
- **Intuition:** Detects heavy tails / impulsiveness. Impulsive interference, bursts, or clipping can increase kurtosis.
- **Extractor detail:** Returns 3.0 for short or constant arrays (by design).

Measured importance (this run):

- nMI (train+val): **0.315569** (MI = 0.263322 nats), rank **42/78** → *medium* data-signal.
- Permutation importance (test): **0.000245 ± 0.000531** macro-F_1 drop, rank **32/78** → *low* model-usage, noisy (mean/std ≈ 0.46).
- Combined score (normalized nMI + normalized perm): **0.455736**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`kurtQ`

Definition & intent (from `features.md` / extractor):

`kurtQ` (Higher-order moments of Q)

- **Definition:** Population kurtosis (non-excess) of Q:

$$\mathrm{kurtQ} = \frac{\mathbb{E}\left[(Q-\mu_Q)^4\right]}{\left(\mathbb{E}\left[(Q-\mu_Q)^2\right]\right)^2}.$$

  For a Gaussian distribution, kurtosis is 3 (when using the non-excess convention).
- **Qntuition:** Detects heavy tails / impulsiveness. Qmpulsive interference, bursts, or clipping can increase kurtosis.
- **Extractor detail:** Returns 3.0 for short or constant arrays (by design).

Measured importance (this run):

- nMI (train+val): **0.316901** (MI = 0.264433 nats), rank **40/78** → *medium* data-signal.
- Permutation importance (test): **-0.000325 ± 0.000714** macro-F_1 drop, rank **62/78** → *very low* model-usage, noisy (mean/std ≈ -0.45).
- Combined score (normalized nMI + normalized perm): **0.445819**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`circularity_phase_rad`

Definition & intent (from `features.md` / extractor):

`circularity_phase_rad` (Complex circularity phase)

- **Definition:** Using $\rho = \mathbb{E}[z^2]/\mathbb{E}[|z|^2]$ as above,

$$\mathrm{circularity\_phase\_rad} = \arg(\rho).$$

- **Range:** $(-\pi, \pi]$.
  - **Intuition:** If the IQ cloud is elongated (improper), the phase of ρ encodes the *orientation* of that elongation in the IQ plane (related to I/Q imbalance / axis rotation).
- **Caveat:** When $|\rho|$ is tiny (near 0), the phase is numerically unstable and should be interpreted cautiously.

Measured importance (this run):

- nMI (train+val): **0.152867** (MI = 0.127558 nats), rank **58/78** → *low* data-signal.
- Permutation importance (test): **0.000585 ± 0.000805** macro-F_1 drop, rank **17/78** → *medium* model-usage, noisy (mean/std ≈ 0.73).
- Combined score (normalized nMI + normalized perm): **0.230425**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Circularity features can expose I/Q imbalance or strong deterministic components; treat phase cautiously if magnitude is tiny.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`circularity_mag`

Definition & intent (from `features.md` / extractor):

`circularity_mag` (Complex circularity / impropriety)

- **Definition:** Let $z = I + jQ$. Define the (normalized) circularity coefficient

$$\rho = \frac{\mathbb{E}[z^2]}{\mathbb{E}[|z|^2]}.$$

  Then:

$$\text{circularity\_mag} = |\rho|.$$

- **Range:** $[0, 1]$ in typical cases.
- **Intuition:** Measures how "proper" (circular) the complex distribution is. Proper complex Gaussian noise has $\rho \approx 0$. Strong deterministic tones, imbalance, or real-valued leakage can make the distribution improper (larger $|\rho|$).

Measured importance (this run):

- nMI (train+val): **0.054925** (MI = 0.045831 nats), rank **66/78** → *very low* data-signal.
- Permutation importance (test): **0.000153 ± 0.000611** macro-F_1 drop, rank **39/78** → *very low* model-usage, noisy (mean/std ≈ 0.25).
- Combined score (normalized nMI + normalized perm): **0.081610**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:

- Circularity features can expose I/Q imbalance or strong deterministic components; treat phase cautiously if magnitude is tiny.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`skewQ`

Definition & intent (from `features.md` / extractor):

`skewQ` (Higher-order moments of Q)

- **Definition:** Sample skewness of the quadrature component Q:

$$\mathrm{skewQ} = \mathbb{E}\left[\left(\frac{Q-\mu_Q}{\sigma_Q}\right)^3\right].$$

- **Qntuition:** Measures asymmetry of the amplitude distribution. Strong non-Gaussian components, clipping, or asymmetric interference can shift skewness away from 0.
- **Extractor detail:** Uses a “safe” implementation returning 0 for very short or constant arrays.

Measured importance (this run):

- nMI (train+val): **0.014931** (MI = 0.012459 nats), rank **75/78** → *very low* data-signal.
- Permutation importance (test): **0.000228 ± 0.000599** macro-F_1 drop, rank **34/78** → *low* model-usage, noisy (mean/std ≈ 0.38).
- Combined score (normalized nMI + normalized perm): **0.026048**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`skewI`

Definition & intent (from `features.md` / extractor):

`skewI` (Higher-order moments of I)

- **Definition:** Sample skewness of the in-phase component I:

$$\text{skewI} = \mathbb{E}\left[\left(\frac{I-\mu_I}{\sigma_I}\right)^3\right].$$

- **Intuition:** Measures asymmetry of the amplitude distribution. Strong non-Gaussian components, clipping, or asymmetric interference can shift skewness away from 0.
- **Extractor detail:** Uses a “safe” implementation returning 0 for very short or constant arrays.

Measured importance (this run):

- nMI (train+val): **0.020062** (MI = 0.016741 nats), rank **73/78** → *very low* data-signal.
- Permutation importance (test): **-0.001450 ± 0.000773** macro-F_1 drop, rank **78/78** → *very low* model-usage, noisy (mean/std ≈ -1.88).
- Combined score (normalized nMI + normalized perm): **-0.001415**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Higher-order moments are fragile under outliers; they can be informative for impulsive or clipped signals.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.9 Group 9: Inequality, Symmetry, DC Notch & Peakiness (6)

Inequality / symmetry / DC-notch and peakiness metrics. These tend to track *how concentrated* the spectrum is and whether there is a strong DC component or sharp peaks.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_ma
spec_gini	0.665434	3	0.000241	
spec_peakiness_ratio	0.629294	8	0.000422	
spec_symmetry_index	0.593872	14	-3.2e-05	
env_p95_over_p50	0.485116	27	0.000271	
env_gini	0.357265	38	-1e-06	
dc_notch_ratio	0.150343	59	-0.000284	

Per-feature review:

`spec_gini`

Definition & intent (from `features.md` / extractor):

63. `spec_gini`

- **Intuition**: Gini coefficient of the normalized PSD.
  - 0 → perfectly equal power per bin.
  - 1 → all power in a single bin.
    Another "peakiness" measure.
- **Formula**:

  Let $x_k = P_{xx}^{\mathrm{norm}}[k]$ sorted ascending, $S = \sum_k x_k$:

$$G = \frac{2\sum_{k=1}^{K} k x_k}{KS} - \frac{K+1}{K},$$

clipped to $[0, 1]$.

Measured importance (this run):

- nMI (train+val): **0.665434** (MI = 0.555262 nats), rank **3/78** → *very high* data-signal.
- Permutation importance (test): **0.000241 ± 0.000778** macro-F_1 drop, rank **33/78** → *low* model-usage, noisy (mean/std ≈ 0.31).
- Combined score (normalized nMI + normalized perm): **0.955285**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).
- Notes:
  - These are global spectral-shape descriptors; they often separate narrowband vs wideband interference cleanly.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`spec_peakiness_ratio`

Definition & intent (from `features.md` / extractor):

68. `spec_peakiness_ratio`

- **Intuition**: Simple ratio of max PSD to median PSD.
  - A few tones over noise → very large.

- Flat noise → closer to 1.
- **Formula**:

$$\text{spec peakiness} = \frac{\max_k P_{xx}[k]}{\text{median}_k(P_{xx}[k]) + \varepsilon}.$$

---

10. STFT-based Time–Frequency Dynamics (5)

Category intuition

These use an STFT (short-time Fourier transform) to follow the spectrum over time. They capture:

- How the **spectral centroid** moves,
- Whether it jumps (FH-like),
- Whether at each time we see a broad or narrow ridge of power.

We compute a spectrogram of $z[n]$:

$$(f_i, t_j, S_{xx}[i,j]) = \text{spectrogram}(z, f_s)$$

(Hann window, `STFT_NPERSEG`, `STFT_NOVERLAP`, `STFT_NFFT`, PSD mode).

Time hop:

$$\Delta t = \frac{\text{STFT} - \text{STFT}}{f_s}.$$

Normalise each time column:

$$S_{xx}^{\text{norm}}[i,j] = \frac{S_{xx}[i,j]}{\sum_l S_{xx}[l,j] + \varepsilon}.$$

Define spectral centroid per frame:

$$c_j = \sum_i f_i S_{xx}^{\text{norm}}[i,j].$$

Measured importance (this run):

- nMI (train+val): **0.629294** (MI = 0.525105 nats), rank **8/78** → *very high* data-signal.
- Permutation importance (test): **0.000422 ± 0.000640** macro-F_1 drop, rank **21/78** → *medium* model-usage, noisy (mean/std ≈ 0.66).
- Combined score (normalized nMI + normalized perm): **0.907414**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated "cluster"; prune only after checking stability across folds and retrains.

`spec_symmetry_index`

Definition & intent (from `features.md` / extractor):

66. `spec_symmetry_index`

- **Intuition**: Whether positive and negative frequencies carry similar power.
  LO offsets or IF design can make things asymmetric.
- **Formula**:

$$P_+ = \sum_{k:f_k>0} P_{xx}[k], \quad P_- = \sum_{k:f_k<0} P_{xx}[k],$$

$$\text{spec symmetry} = \frac{P_+ - P_-}{P_+ + P_- + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.593872** (MI = 0.495547 nats), rank **14/78** → *high* data-signal.
- Permutation importance (test): **-0.000032 ± 0.001013** macro-F_1 drop, rank **53/78** → *very low* model-usage, noisy (mean/std ≈ -0.03).
- Combined score (normalized nMI + normalized perm): **0.847422**.

Interpretation:

- Cross-method read: misaligned: strong signal but low permutation impact (likely redundancy/correlation or the model prefers an alternative).

Pruning / engineering notes:

- Do **not** prune blindly: high nMI suggests real structure, but the model may already capture it via correlated features.

`env_p95_over_p50`

Definition & intent (from `features.md` / extractor):

65. `env_p95_over_p50`

- **Intuition**: "How much bigger are the large amplitudes than the median?".
  Pulsed signals → 95th percentile much larger than median.
- **Formula**:

$$p_{95} = \text{percentile}_{95}(\text{env}_{\text{raw}}), \quad p_{50} = \text{percentile}_{50}(\text{env}_{\text{raw}}),$$

$$\text{env p95 over} = \frac{p_{95}}{p_{50} + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.485116** (MI = 0.404798 nats), rank **27/78** → *medium* data-signal.
- Permutation importance (test): **0.000271 ± 0.000938** macro-F_1 drop, rank **30/78** → *low* model-usage, noisy (mean/std ≈ 0.29).
- Combined score (normalized nMI + normalized perm): **0.698395**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`env_gini`

Definition & intent (from `features.md` / extractor):

64. `env_gini`

- **Intuition**: Same concept but applied to the envelope samples.
  High value → only a few samples carry most amplitude (strong pulses).
- **Formula**:

  Let

$$x_n = \frac{\max(\text{env}_{\text{raw}}[n], 0)}{\sum_n \max(\text{env}_{\text{raw}}[n], 0) + \varepsilon},$$

sort and apply the same Gini formula.

Measured importance (this run):

- nMI (train+val): **0.357265** (MI = 0.298115 nats), rank **38/78** → *medium* data-signal.
- Permutation importance (test): **-0.000001 ± 0.000316** macro-F_1 drop, rank **50/78** → *very low* model-usage, noisy (mean/std ≈ -0.00).
- Combined score (normalized nMI + normalized perm): **0.510182**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Amplitude-envelope features can reflect burstiness, pulsing, or clipping, but are also affected by AGC.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`dc_notch_ratio`

Definition & intent (from `features.md` / extractor):

67. `dc_notch_ratio`

- **Intuition**: Power near DC compared to power in a wider central band.
  - Notch filter at DC → low ratio.
  - Strong DC spike → high ratio.
- **Formula**:

$$\mathcal{D} = \{k : |f_k| \le 0.5\,\mathrm{MHz}\}, \quad \mathcal{R} = \{k : |f_k| \le 5\,\mathrm{MHz}\},$$

$$\text{dc notch} = \frac{\sum_{k\in\mathcal{D}} P_{xx}[k]}{\sum_{k\in\mathcal{R}} P_{xx}[k] + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.150343** (MI = 0.125452 nats), rank **59/78** → *low* data-signal.
- Permutation importance (test): **-0.000284 ± 0.001028** macro-F_1 drop, rank **60/78** → *very low* model-usage, noisy (mean/std ≈ -0.28).
- Combined score (normalized nMI + normalized perm): **0.208803**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.10 Group 10: STFT-based Time–Frequency Dynamics (5)

STFT-based time–frequency dynamics: how the spectral centroid and energy distribution **move over time**. In practice, these can dominate chirp detection because they see time variation that a single PSD cannot.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean
stft_centroid_std_Hz	0.508264	25	0.04823
strong_bins_mean	0.388255	30	0.003179
stft_centroid_absderiv_med_Hzps	0.260058	50	-0.000698
stft_centroid_zcr_per_s	0.133065	60	0.00073
fh_hop_rate_per_s	0.093619	62	0

Per-feature review:

`stft_centroid_std_Hz`

Definition & intent (from `features.md` / extractor):

69. `stft_centroid_std_Hz`

- **Intuition**: How far the centroid moves around its average position.
  Useful for spotting very mobile interference (e.g. FH).
- **Formula**:

$$\bar{c} = \frac{1}{J}\sum_j c_j,$$

$$\text{stft centroid std} = \sqrt{\frac{1}{J}\sum_j (c_j - \bar{c})^2}.$$

Measured importance (this run):

- nMI (train+val): **0.508264** (MI = 0.424114 nats), rank **25/78** → *high* data-signal.
- Permutation importance (test): **0.048230 ± 0.004251** macro-F_1 drop, rank **1/78** → *very high* model-usage, stable (mean/std ≈ 11.35).
- Combined score (normalized nMI + normalized perm): **1.725833**.

Interpretation:

- Cross-method read: aligned: strong signal *and* the model uses it heavily.
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`strong_bins_mean`

Definition & intent (from `features.md` / extractor):

73. `strong_bins_mean`

- **Intuition**: On average, how many time-frequency bins are "strong" (above half of the max at each time).
  - Wideband jammers → many strong bins.
  - Narrow tones → very few.
- **Formula**:

$$M_j = \max_i S_{xx}^{\mathrm{norm}}[i,j],$$

$$\text{strong mask}_{i,j} = \begin{cases} 1, & S_{xx}^{\mathrm{norm}}[i,j] > 0.5M_j \\ 0, & \text{otherwise} \end{cases},$$

$$\text{strong bins} = \frac{1}{IJ}\sum_{i,j} \text{strong mask}_{i,j}.$$

---

11. Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)

Category intuition

Finally we have:

- Extra cyclostationary lags around the chip period,
- A curvature term to detect non-linear chirps,
- Detailed interpulse-interval stats for DME-like interference.

11.1. Extra cyclostationarity

Measured importance (this run):

- nMI (train+val): **0.388255** (MI = 0.323974 nats), rank **30/78** → *medium* data-signal.
- Permutation importance (test): **0.003179 ± 0.002924** macro-F_1 drop, rank **4/78** → *high* model-usage, moderately stable (mean/std ≈ 1.09).
- Combined score (normalized nMI + normalized perm): **0.620372**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Treat as part of a correlated “cluster”; prune only after checking stability across folds and retrains.

`stft_centroid_absderiv_med_Hzps`

Definition & intent (from `features.md` / extractor):

70. `stft_centroid_absderiv_med_Hzps`

- **Intuition**: Typical speed (in Hz/s) at which the centroid moves.
- **Formula**:

$$d_c[j] = c_{j+1} - c_j,$$

$$\text{stft centroid absderiv med} = \text{median}_j \left(\left| \frac{d_c[j]}{\Delta t} \right| \right).$$

Measured importance (this run):

- nMI (train+val): **0.260058** (MI = 0.217002 nats), rank **50/78** → *low* data-signal.
- Permutation importance (test): **-0.000698 ± 0.000745** macro-F_1 drop, rank **72/78** → *very low* model-usage, noisy (mean/std ≈ -0.94).
- Combined score (normalized nMI + normalized perm): **0.356910**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`stft_centroid_zcr_per_s`

Definition & intent (from `features.md` / extractor):

71. `stft_centroid_zcr_per_s`

- **Intuition**: How often the centroid’s velocity changes sign per second (back-and-forth movement).
- **Formula**:

$$\text{ZCR}(d_c) = \frac{|\{j : d_c[j] \cdot d_c[j+1] < 0\}|}{J - 1},$$

$$\text{stft centroid zcr per} = \frac{\mathrm{ZCR}(d_c)}{\Delta t}.$$

Measured importance (this run):

- nMI (train+val): **0.133065** (MI = 0.111034 nats), rank **60/78** → *low* data-signal.
- Permutation importance (test): **0.000730 ± 0.000794** macro-F_1 drop, rank **15/78** → *medium* model-usage, noisy (mean/std ≈ 0.92).
- Combined score (normalized nMI + normalized perm): **0.205166**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.
- Notes:
  - Time-frequency dynamics features are typically the strongest indicators for chirp-like jammers.

Pruning / engineering notes:

- Keep for now: the model is using it, even if marginal MI is modest (could be interaction-driven).

`fh_hop_rate_per_s`

Definition & intent (from `features.md` / extractor):

72. `fh_hop_rate_per_s`

- **Intuition**: Approximate **hop rate** of strong frequency jumps → aimed at FH jammers.
- **Formula**:

$$\mathrm{mad} = \mathrm{median}_j(|d_c[j] - \mathrm{median}(d_c)|) + 10^{-6},$$

$$T_{\text{hop}} = \max(5 \cdot 10^5, 6 \cdot \mathrm{mad}),$$

$$N_{\text{hops}} = |\{j : |d_c[j]| > T_{\text{hop}}\}|, \quad T_{\text{dur}} = (J-1)\Delta t,$$

$$\text{fh hop rate per} = \frac{N_{\text{hops}}}{T_{\text{dur}} + \varepsilon}.$$

Measured importance (this run):

- nMI (train+val): **0.093619** (MI = 0.078119 nats), rank **62/78** → *low* data-signal.
- Permutation importance (test): **0.000000 ± 0.000000** macro-F_1 drop, rank **46/78** → *very low* model-usage, unknown.
- Combined score (normalized nMI + normalized perm): **0.133693**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

7.11 Group 11: Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)

Extra cyclostationary lags, chirp curvature, and DME-like inter-pulse interval metrics. This group is more specialized; it can be very informative on datasets where these specific structures exist.

Group table (sorted by combined score):

feature	nMI	rank_nMI	perm_macroF1_drop_mean	perm_ma
cyclo_halfchip_corr	0.265595	48	-1.9e-05	
chirp_curvature_Hzps2	0.167081	55	-0.000272	
cyclo_5chip_corr	0.102548	61	-0.001026	
dme_ipi_std_s	0.021522	71	0.001093	
dme_ipi_med_s	0.020929	72	-0.000283	

Per-feature review:

`cyclo_halfchip_corr`

Definition & intent (from `features.md` / extractor):

74. `cyclo_halfchip_corr`

- **Intuition**: Cyclo correlation at **half a chip**. Gives extra granularity on how chip-like the structure is.
- **Formula**:

$$L_{\frac{1}{2}} = \mathrm{round}\left(\frac{f_s}{2 \cdot 1.023\,\mathrm{MHz}}\right),$$

$$\mathrm{cyclo\ halfchip} = \mathrm{cyclo\ lag}(z, L_{\frac{1}{2}}).$$

Measured importance (this run):

- nMI (train+val): **0.265595** (MI = 0.221622 nats), rank **48/78** → *low* data-signal.
- Permutation importance (test): **-0.000019 ± 0.000363** macro-F_1 drop, rank **51/78** → *very low* model-usage, noisy (mean/std ≈ -0.05).
- Combined score (normalized nMI + normalized perm): **0.378898**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`chirp_curvature_Hzps2`

Definition & intent (from `features.md` / extractor):

76. `chirp_curvature_Hzps2`

- **Intuition**: Measures **curvature** of the frequency trajectory.
  - Linear chirp → curvature ≈ 0.
  - Non-linear sweep → non-zero curvature.
- **Formula**: Fit

$$c_j \approx at_j^2 + bt_j + c$$

and take

$$\text{chirp curvature} = 2a.$$

11.3. DME interpulse intervals (IPIs)

For DME-like pulsed interference we look at the time between pulses.

- Smooth env_{raw} with a shorter window ($\approx 0.3\,\mu$s) to get $\text{env}_s[n]$.
- Threshold as before: $T = \mathbb{E}[\text{env}_s] + 3 \cdot \text{std}(\text{env}_s)$.
- Detect peaks $p_0, \ldots, p_{K-1}$ (at least $\approx 0.2\,\mu$s apart).

Interpulse intervals:

$$\text{IPI}_k = \frac{p_{k+1} - p_k}{f_s}, \quad k = 0, \ldots, K-2.$$

Measured importance (this run):

- nMI (train+val): **0.167081** (MI = 0.139418 nats), rank **55/78** → *low* data-signal.
- Permutation importance (test): **-0.000272 ± 0.001043** macro-F_1 drop, rank **58/78** → *very low* model-usage, noisy (mean/std ≈ -0.26).
- Combined score (normalized nMI + normalized perm): **0.232963**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`cyclo_5chip_corr`

Definition & intent (from `features.md` / extractor):

75. `cyclo_5chip_corr`

- **Intuition**: Cyclo correlation at **5 chips**. Checks for longer-range chip periodicity.
- **Formula**:

$$L_5 = \text{round}\left(5 \cdot \frac{f_s}{1.023\,\text{MHz}}\right),$$

$$\text{cyclo 5chip} = \text{cyclo lag}(z, L_5).$$

11.2. Chirp curvature from STFT

We again use the STFT centroid sequence c_j and times $t_j = j\Delta t$.

Measured importance (this run):

- nMI (train+val): **0.102548** (MI = 0.085570 nats), rank **61/78** → *low* data-signal.
- Permutation importance (test): **-0.001026 ± 0.000343** macro-F_1 drop, rank **76/78** → *very low* model-usage, noisy (mean/std ≈ -2.99).
- Combined score (normalized nMI + normalized perm): **0.125169**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

`dme_ipi_std_s`

Definition & intent (from `features.md` / extractor):

78. `dme_ipi_std_s`

- **Intuition**: How regular those IPIs are.
  - Very periodic pulses → low std.
  - Irregular bursts → higher std.
- **Formula**:

$$\text{dme ipi std} = \text{std}_k(\text{IPI}_k).$$

12. Big-picture summary

- The feature vector combines:
  - **Time-domain shape** (DC, variance, ZCR, PAPR, envelope stats),
  - **Global and local spectral shape** (centroid, spread, band powers, peaks),
  - **Instantaneous frequency dynamics** (drift, slope, jitter, hops),
  - **Envelope modulation and pulses** (AM, DME-style, cepstrum),
  - **Cyclostationarity & higher-order structure** (GNSS chip periodicity, cumulants),
  - **Non-Gaussian / non-circular behaviour** (skew, kurtosis, circularity),
  - **Time–frequency evolution** via STFT features.
- Together they form a **strict, interpretable fingerprint** of a GNSS+jammer IQ chunk, suitable for supervised learning and for qualitative inspection by humans.

Measured importance (this run):

- nMI (train+val): **0.021522** (MI = 0.017959 nats), rank **71/78** → *very low* data-signal.
- Permutation importance (test): **0.001093 ± 0.000299** macro-F_1 drop, rank **11/78** → *medium* model-usage, stable (mean/std ≈ 3.65).
- Combined score (normalized nMI + normalized perm): **0.053404**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Keep for now: the model is using it, even if marginal MI is modest (could be interaction-driven).

`dme_ipi_med_s`

Definition & intent (from `features.md` / extractor):

77. `dme_ipi_med_s`

- **Intuition**: Typical spacing between pulses (seconds).
  Useful to recognise specific pulsed systems like DME.
- **Formula**:

$$\mathrm{dme\ ipi\ med} = \mathrm{median}_k(\mathrm{IPI}_k).$$

Measured importance (this run):

- nMI (train+val): **0.020929** (MI = 0.017464 nats), rank **72/78** → *very low* data-signal.

- Permutation importance (test): **-0.000283 ± 0.000346** macro-F_1 drop, rank **59/78** → *very low* model-usage, noisy (mean/std ≈ -0.82).
- Combined score (normalized nMI + normalized perm): **0.024022**.

Interpretation:

- Cross-method read: moderate/weak: neither clearly dominant in both views.

Pruning / engineering notes:

- Candidate for *early* pruning tests: low nMI and negligible permutation impact on this test set.

---

8. Cross-method patterns worth acting on

8.1 Features that are strong in both nMI and permutation

These are your *most defensible* "core" features: they show strong label structure and the trained model measurably depends on them.

feature	group_id	group_name	nMI	perm_macroF1_drop_mean
spec_entropy	1	Basic Time-Domain & Power Features (18)	0.70025	0.000827
spec_flatness	2	Global Spectral Shape Features (6)	0.68167	0.000914
bandpower_6	3	Band Power Distribution (8)	0.639091	0.003787
bandpower_5	3	Band Power Distribution (8)	0.604086	0.002522
spec_peak_power	2	Global Spectral Shape Features (6)	0.595744	0.001648

8.2 High nMI but negligible permutation impact

Common explanation: redundancy/correlation. In pruning experiments, you typically drop these *only after* validating that your strongest correlated alternative stays in the set.

feature	group_id	group_name	nMI	perm_macroF1_drop
nb_peak_count	6	Narrowband Peaks, AM & Chirp Features (8)	0.648161	-0.0
env_ac_peak	1	Basic Time-Domain & Power Features (18)	0.61543	0.0
instf_std_Hz	4	Instantaneous Frequency Features (5)	0.612333	-0.0
spec_symmetry_index	9	Inequality, Symmetry, DC Notch & Peakiness (6)	0.593872	-3
tkeo_env_mean	7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	0.583284	

8.3 High permutation but not especially high nMI

These features can matter because the model uses them in interactions, or because they are stable proxies for a phenomenon not cleanly visible in marginal MI.

feature	group_id	group_name	nMI	perm_macroF1_drop
spec_kurtosis_mean	7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	0.216656	0
instf_slope_Hzps	4	Instantaneous Frequency Features (5)	0.154167	0
dme_ipi_std_s	11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	0.021522	0

feature	group_id	group_name	nMI	perm_macroF1_dro
stft_centroid_zcr_per_s	10	STFT-based Time–Frequency Dynamics (5)	0.133065	
corrIQ	1	Basic Time-Domain & Power Features (18)	0.091598	0

8.4 Low in both views (initial pruning candidates)

These are the safest features to try removing first. The correct pruning workflow is still empirical: retrain and measure drift.

feature	group_id	group_name	nMI	perm_mac
kurtQ	8	Higher-order I/Q Stats & Circularity (6)	0.316901	
nb_spacing_med_Hz	6	Narrowband Peaks, AM & Chirp Features (8)	0.316325	
PAPR_dB	1	Basic Time-Domain & Power Features (18)	0.277864	
crest_env	1	Basic Time-Domain & Power Features (18)	0.278135	
cyclo_halfchip_corr	11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	0.265595	
cyclo_2chip_corr	7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	0.265585	
meanQ	1	Basic Time-Domain & Power Features (18)	0.267961	
stft_centroid_absderiv_med_Hzps	10	STFT-based Time–Frequency Dynamics (5)	0.260058	

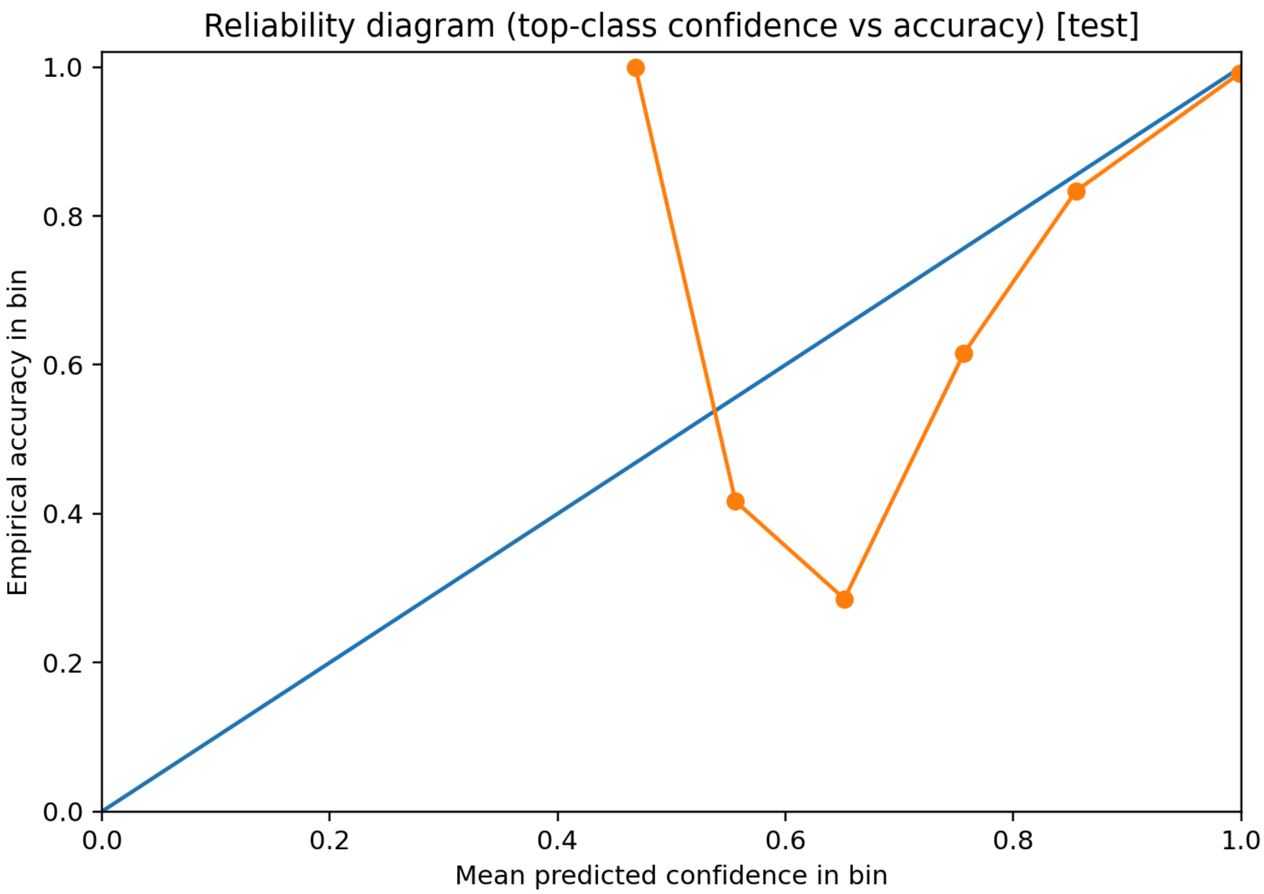
feature	group_id	group_name	nMI	perm_mac
chirp_curvature_Hzps2	11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	0.167081	
spec_kurtosis_max	7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	0.167275	
chirp_slope_Hzps	6	Narrowband Peaks, AM & Chirp Features (8)	0.171327	
env_ac_lag_s	1	Basic Time-Domain & Power Features (18)	0.162546	
dc_notch_ratio	9	Inequality, Symmetry, DC Notch & Peakiness (6)	0.150343	
fh_hop_rate_per_s	10	STFT-based Time–Frequency Dynamics (5)	0.093619	
cyclo_5chip_corr	11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	0.102548	
cumulant_c40_mag	7	Cyclostationarity, Cumulants, Spectral Kurtosis, TKEO (7)	0.052513	
dme_pulse_count	5	Envelope, Cepstrum, Pulse & Narrowband Salience (4)	0.02324	
dme_ipi_med_s	11	Extra Cyclo Lags, Chirp Curvature, DME IPIs (5)	0.020929	
chirp_r2	6	Narrowband Peaks, AM & Chirp Features (8)	0.019719	

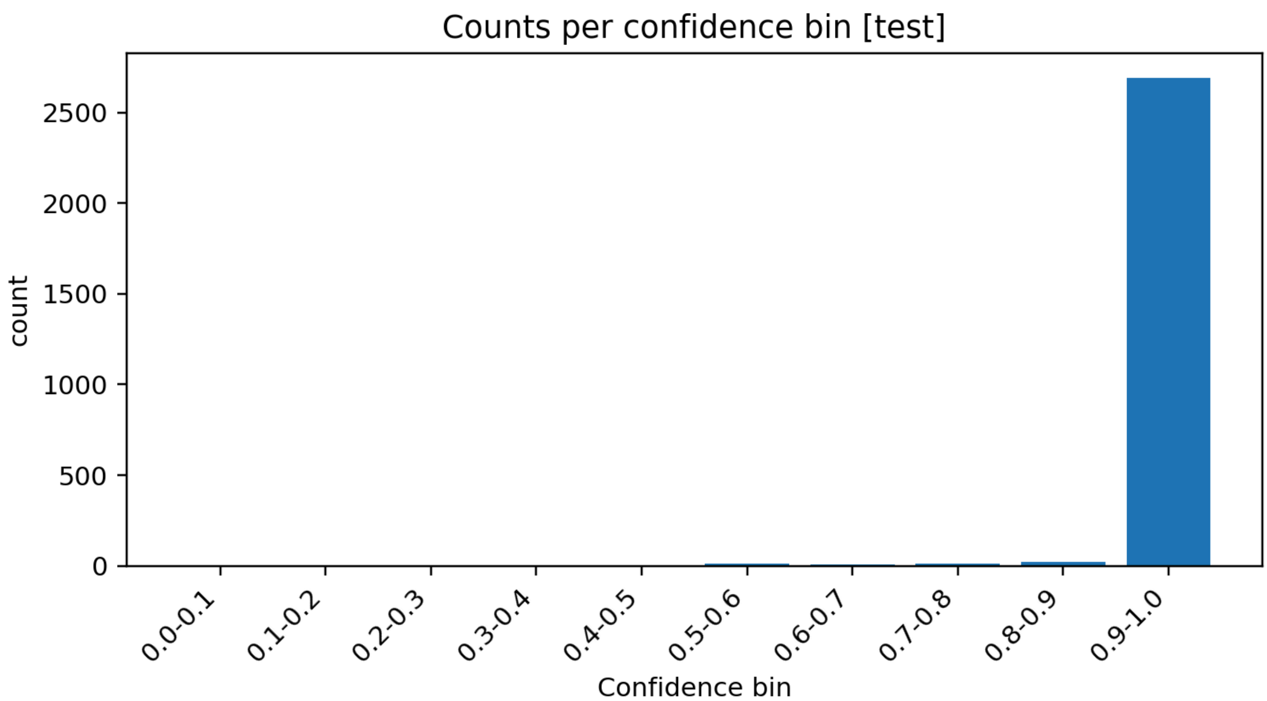
feature	group_id	group_name	nMI	perm_mac
env_dom_peak_norm	6	Narrowband Peaks, AM & Chirp Features (8)	0.004101	
cep_peak_env	5	Envelope, Cepstrum, Pulse & Narrowband Salience (4)	0	
env_dom_freq_Hz	6	Narrowband Peaks, AM & Chirp Features (8)	0	
skewI	8	Higher-order I/Q Stats & Circularity (6)	0.020062	

---

9. Model confidence and calibration diagnostics (test)

These plots are not feature importance per se, but they help you interpret why permutation drops are dominated by a small subset of features (a highly confident model is often less sensitive to many small cues).





---

10. Recommendations and next experiments

10.1 If your goal is *robust classification* across environments

- Keep the **STFT-dynamics group** (Group 10) intact until you validate robustness across different sweep rates, SNRs, and bandwidths.
- Treat the PSD-shape cluster (`spec_entropy` , `spec_flatness` , `spec_gini` , bandpowers) as a *redundant block*. You usually do not need all of them, but you should not prune without checking correlated replacements.
- For the narrowband class, retain both global PSD shape and the dedicated NB detectors (`nb_peak_count` , `nb_peak_prom_mean` , etc.).

10.2 If your goal is *feature pruning / model compression*

A disciplined pruning plan:

1. Remove only **low–low** features first (Section 8.4).
2. Retrain with identical hyperparameters and compare macro-F_1 and per-class F_1.
3. If stable, iterate by removing the next batch.
4. Only then consider pruning in the *high nMI / low permutation* cluster (Section 8.2), one at a time.

10.3 What would make this report even stronger

- **Per-class permutation importance:** compute permutation drops for each class's F_1 (one-vs-rest). This often reveals class-specific features that macro-F_1 hides.
- **Cross-validation stability:** rerun permutation on multiple folds to see whether the same features stay dominant.
- **Cluster analysis of feature correlations:** permutation can understate importance in correlated groups; correlation clustering helps interpret that.

Appendix A — Files included in this package

- `feature_importance_report.md` — this report
- `full_ranking_with_groups.csv` — all 78 features with nMI, permutation, ranks, combined score, and group mapping
- `group_stats_nMI_perm.csv` — group aggregates
- `top30_features_by_nMI.csv` / `top30_features_by_perm_macroF1_drop.csv` — convenience slices
- `per_group_tables/` — one CSV per feature group
- `assets/plots/` — all plots (original run + extra diagnostics)
- `assets/tables/` — confusion matrices, prediction logs, and error summaries
- `source/` — the exact run outputs + `features.md` + `feature_extractor.py`