

Supplementary Material 1

BioAdaptive Decoder: Complete Specification, Guard Conditions, and Validation-only Selection

Chen Xu

School of Information Science and Engineering, Lanzhou University

Corresponding author: xchen2024@lzu.edu.cn

ORCID: [0009-0009-8592-4336](https://orcid.org/0009-0009-8592-4336)

This Online Resource accompanies the manuscript “Clinical Evidence Morphology–Rhythm for ECG (CEMR-ECG): a traceable evidence framework for imbalanced heartbeat classification” submitted to Physical and Engineering Sciences in Medicine. It provides the full specification of the BioAdaptive decoder family, including all class-multiplier formulae, the validation-only selection algorithm, and the fixed guard conditions referenced in Methods Section 2.3.

1 Notation recap

For class $c \in \mathcal{C} = \{N, S, V, F, Q\}$, let n_c be the count of beats in the training split and define the N-to-class imbalance ratio

$$r_c = \frac{n_N}{\max(n_c, 1)}. \quad (1)$$

The fused backbone–evidence probability vector $\mathbf{q}_i^{(\alpha)}$ is defined in the main text (Eq. 3). The BioAdaptive decoder transforms $\mathbf{q}_i^{(\alpha)}$ to the final decision score $\hat{p}_{i,c}^{(\alpha,k)}$ using two clipped operations: a probability power γ_k and a class-specific multiplier $m_c^{(k)}$:

$$u_{i,c}^{(\alpha,k)} = \frac{(q_{i,c}^{(\alpha)})^{\gamma_k}}{\sum_{c' \in \mathcal{C}} (q_{i,c'}^{(\alpha)})^{\gamma_k}}, \quad s_{i,c}^{(\alpha,k)} = u_{i,c}^{(\alpha,k)} m_c^{(k)}, \quad \hat{p}_{i,c}^{(\alpha,k)} = \frac{s_{i,c}^{(\alpha,k)}}{\sum_{c' \in \mathcal{C}} s_{i,c'}^{(\alpha,k)}}. \quad (2)$$

2 Full candidate decoder family

The BioAdaptive decoder considers six fixed candidate policies. Each policy fixes γ_k and defines class multipliers $m_c^{(k)}$ as clipped powers of the imbalance ratio r_c . Table 1 lists the complete formulae.

3 Boundedness of the multiplier family

Each non-trivial multiplier has the form $m_c = \text{clip}(r_c^\beta, l, u)$ with fixed $\beta \in \mathbb{R}$ and $0 < l \leq u < \infty$. Two properties follow directly:

1. **Monotonicity in rarity.** Because $r_c = n_N / \max(n_c, 1) > 0$ and r_c increases as class c becomes rarer relative to N, the unclipped derivative $\partial(r_c^\beta) / \partial r_c = \beta r_c^{\beta-1}$ has the sign of β . Positive exponents therefore increase the multiplier as the class becomes rarer; the small negative V exponent in the conservative-default policy mildly damps V when its support is comparatively high.

Table 1: Complete BioAdaptive candidate decoder family. The candidates use bounded, prevalence-aware multipliers so that minority recovery is constrained and selected only through internal validation.

Candidate	γ_k	Multiplier definition and intended role
Evidence-argmax	1.00	$m_N = m_S = m_V = m_F = m_Q = 1$; uses the fused backbone-evidence probabilities without prevalence rescaling. Serves as the safety baseline activated by guard conditions.
Conservative-default	1.00	$m_N = 0.85$; $m_S = \text{clip}(r_S^{0.23}, 1.00, 2.60)$; $m_V = \text{clip}(r_V^{-0.02}, 0.90, 1.05)$; $m_F = \text{clip}(r_F^{0.15}, 1.00, 2.20)$; $m_Q = 1$. Moderate S/F recovery with normal-class damping. Used as the fixed policy for MIT-BIH DS1→DS2 transfer.
Stable-prior	0.90	$m_N = 1$; $m_S = \text{clip}(r_S^{0.25}, 1.00, 4.00)$; $m_V = \text{clip}(r_V^{0.03}, 1.00, 1.50)$; $m_F = \text{clip}(r_F^{0.20}, 1.00, 4.00)$; $m_Q = 1$. Long-tail prior support, broader upper bounds than conservative-default.
S-recovery	0.95	$m_N = 0.90$; $m_S = \text{clip}(r_S^{0.28}, 1.00, 4.50)$; $m_V = 1$; $m_F = \text{clip}(r_F^{0.08}, 1.00, 1.60)$; $m_Q = 1$. S-focused recovery; preferred when validation S support is large but F support is limited.
F-recovery	0.95	$m_N = 0.90$; $m_S = \text{clip}(r_S^{0.08}, 1.00, 1.60)$; $m_V = 0.98$; $m_F = \text{clip}(r_F^{0.22}, 1.00, 4.00)$; $m_Q = 1$. F-focused recovery; activated only when validation F support is sufficient (see Table 2).
Mild-prior	0.95	$m_N = 0.95$; $m_S = \text{clip}(r_S^{0.18}, 1.00, 2.50)$; $m_V = 1$; $m_F = \text{clip}(r_F^{0.12}, 1.00, 2.20)$; $m_Q = 1$. Conservative same-dataset prevalence correction with mild N damping.

2. **Hard boundedness.** The clip operator restricts m_c to $[l, u]$ regardless of r_c . This prevents very-low-support classes from producing unbounded recovery factors and is the structural reason the BioAdaptive outputs are reported as decision scores rather than calibrated probabilities.

For the conservative-default policy, Eq. (3) writes out the complete multiplier vector used for MIT-BIH DS1→DS2:

$$\mathbf{m}^{\text{clin}} = [0.85, \text{clip}(r_S^{0.23}, 1.00, 2.60), \text{clip}(r_V^{-0.02}, 0.90, 1.05), \text{clip}(r_F^{0.15}, 1.00, 2.20), 1.00]. \quad (3)$$

4 Validation-only selection algorithm

Given validation backbone probabilities $\{\mathbf{p}_i^B\}_{i \in \mathcal{V}}$, validation evidence probabilities $\{\mathbf{p}_i^E\}_{i \in \mathcal{V}}$, validation labels $\{y_i\}_{i \in \mathcal{V}}$, and training counts $\{n_c\}$, the BioAdaptive adapter sweeps the joint grid of (α, k) where

$$\alpha \in \{0.20, 0.35, 0.50, 0.65, 0.80\} \quad \text{and} \quad k \in \{\text{the six candidates in Table 1}\}.$$

For every (α, k) pair the adapter computes a validation score

$$\text{score}(\alpha, k) = \text{M-F1}(4)(\hat{\mathbf{p}}_{\mathcal{V}}^{(\alpha, k)}, y_{\mathcal{V}}) - 0.25 \cdot \text{frac_absent_NSVF}(\hat{\mathbf{p}}_{\mathcal{V}}^{(\alpha, k)}),$$

where the penalty `frac_absent_NSVF` is the fraction of validation samples assigned to an N/S/V/F class that is absent from the validation fold. The adapter selects the pair maximizing $\text{score}(\alpha, k)$, breaking ties first by combined S/F F1 and then by accuracy.

Table 2: Fixed BioAdaptive guard conditions. If any condition is met, the selected recovery candidate is replaced by evidence-argmax.

Guard	Threshold	Rationale
Validation M-F1(4) gain over evidence-argmax	< 0.015	Requires a minimal validation gain before adding recovery; smaller validation deltas are interpreted as noise rather than signal.
Validation accuracy drop	> 0.004	Limits aggregate degradation; recovery is not allowed to trade more than a small slice of accuracy for macro-F1.
Validation S-class F1 drop	> 0.08	Prevents a large loss on the most recurrent minority endpoint; S recovery is treated as the most reliable minority signal.
F-class gain with < 10 validation F beats	< 10 beats	Prevents a sparse validation fold from authorising F-specific recovery; under this threshold the selection reverts to evidence-argmax for F.

Guard substitution. The selected recovery candidate is replaced by evidence-argmax when any of four fixed guard conditions is met (Table 2).

MIT-BIH DS1→DS2 exception. For MIT-BIH the decoder k is fixed to the conservative-default policy because the internal validation fold is sampled from DS1, whereas the final test set is the independent DS2 record group. Allowing DS1 validation labels to choose class-specific multipliers could favor minority patterns that are useful within DS1 but less transferable to DS2. The fixed policy keeps the DS1→DS2 test as a patient-independent transfer check rather than a decoder-selection exercise. The mixture coefficient α remains method-specific and validation-selected because it controls the continuous backbone–evidence balance rather than the class-specific recovery policy.

5 Reproducibility settings of the evidence encoder

The CEMR-ECG evidence encoder uses the following fixed settings, set before the framework comparison and not tuned on the test set:

- Estimator: extremely randomized trees [ExtraTrees, [1]], with 700 trees.
- Feature sampling ratio: 0.35.
- Minimum leaf size: 2.
- Bootstrap sampling: disabled.
- Fixed class-cost vector: $w_N = 1.0$, $w_S = 3.0$, $w_V = 1.5$, $w_F = 12.0$, $w_Q = 1.0$.

The minority-sensitive class-cost profile was a deliberate design choice: F received the largest cost because it combines low support with the difficult fusion morphology endpoint; S received a moderate cost because it is the most recurrent minority recovery target; V received a milder weighting because it had higher support; and Q was retained for reporting rather than as a primary endpoint. Representative sensitivity analyses for no-cost, balanced, moderate-tail, and strong-tail alternative profiles are summarized in the main text Section 3.5.

6 Class-cost sensitivity (representative-interface check)

A representative-interface check (75 runs across five backbones $\times$ three datasets $\times$ five seeds) varied the class-cost profile around the default setting. The default profile reached 59.76% mean M-F1(4); no-cost, moderate-tail, and strong-tail profiles ranged from 59.74% to 60.07%. The balanced-profile value of 57.92%, computed with class costs equalised, indicates that fully removing prevalence awareness can reduce evidence-encoder performance in this imbalanced setting.

7 BioAdaptive constant sensitivity

For the ExtraTrees representative interface, S/F exponents, S/F upper bounds, score power, and N damping were varied from 0.5 to 1.5 times their fixed values where applicable. Mean changes in M-F1(4) ranged from -1.76 to $+0.03$ percentage points relative to the default setting, with the largest decreases occurring when the minority upper bounds were halved. Thus, the selected constants were not a single-point optimum that collapsed under nearby perturbation, although aggressive reduction of S/F recovery predictably weakened the minority-sensitive endpoint.

8 Interface perturbation and F-support guard checks

Interface perturbation check. Stored validation and test probability vectors from ExtraTrees, CAT-Net, and TimeMixer were perturbed by adding zero-mean Gaussian noise to the backbone probability entries and renormalizing the rows. The mild and moderate perturbation levels used standard deviations of 0.03 and 0.07, respectively. These values were fixed stress-test levels on the probability scale: 0.03 perturbs confident class probabilities without usually changing the dominant class, whereas 0.07 is a stronger interface distortion intended to test whether the adapter-decoder behavior degrades gracefully under noisier upstream probabilities. They are sensitivity levels rather than estimates of a real device-noise distribution. The adapter and decoder were rerun using the same validation-only selection rule; positive gains were retained in all tested settings.

F-support guard check. F beats were removed from the validation subset used for adapter and decoder selection, then selection was rerun. The conservative guard substitution from Table 2 fired in 45/45 evaluated runs, reverting any tentative F-specific recovery to the evidence-argmax baseline. This check supports the intended role of the F-support guard: preventing the policy machinery from authorising unsupported F promotion.

References

- [1] Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. Mach Learn 63(1):3–42. <https://doi.org/10.1007/s10994-006-6226-1>