ChronoProto: What a Single Positive Episode Can and Cannot Personalize

Anonymous authors
Paper under double-blind review

Abstract

When a seizure forecaster is personalized after a patient's first usable seizure, the only labeled target data are typically the preceding preictal EEG. This episode locates the patient's preictal windows, while ranking later windows also requires a reference for the interictal background. We formalize a positive-only protocol that releases this first preictal episode and keeps target background out of fitting. Under this protocol, ChronoProto contrasts a first-episode prototype with a fixed reference in a frozen source representation. We split the pairwise ranking error into prototype error and reference mismatch, and further split prototype error into within-episode sampling, governed by effective support, and episode-to-future shift. A one-dimensional construction admits opposite oracle rankings under the same observed information even when the positive distribution is known, establishing the need for an explicit reference assumption. On 28 later CHB–MIT seizures from 13 patients, changing only the reference moves patient-macro AP from 0.2656 to 0.3191. A source-selected linear SVM reaches 0.3175 AP, with paired AP and AUROC difference intervals against CP-0 spanning zero. These results motivate stating and justifying the reference as part of positive-only personalization.

1 Introduction

Seizure forecasting models adapt to each patient's preictal changes and background activity (Karoly et al., 2017; Kuhlmann et al., 2018). The first recorded seizure usually supplies a new patient's first labeled data: once its onset is annotated, the preceding EEG can be labeled under a prespecified preictal definition. This first usable episode is a natural starting point for personalization. It raises two questions for later ranking. First, temporal correlation determines how much independent evidence its windows provide about future positives. Second, the episode locates the patient's preictal activity, while ranking later windows requires a reference for the interictal background. Under positive-only supervision, this reference comes from outside the episode.

We study this problem under a *positive-only personalization* protocol. Source patients provide labeled training data, and the only target data released for adaptation is the first preictal episode. Target interictal EEG and unlabeled target recordings are kept out of fitting. Additional EEG is often available in practice. We withhold it to identify what can be learned from the first labeled episode alone, and the result gives a lower-information reference point for methods that also use target background. Under this protocol we introduce ChronoProto, a minimal method built to expose the two sources of uncertainty separately (Figure 1). A frozen source encoder and metric define the feature geometry, the first episode gives a positive prototype, and a distance contrast against a fixed reference ranks later windows. We compare two references, the source interictal center (CP-BG) and the origin of the source-standardized space (CP-0), and keep the positive support and the geometry fixed. This lets us ask how much independent evidence one correlated episode provides, how much prototype error comes from episode-to-future shift, and how strongly the missing background ties the ranking to the chosen reference.

The protocol comes with an information limit. We show that the same source records and target positives can be consistent with populations whose likelihood-ratio rankings are opposite (Proposition 2). A positive-only ranking rule therefore cannot infer its reference from the observed target positives, and the reference must come from an explicit cross-patient or structural assumption. Within ChronoProto, we can trace the consequences of that assumption. The contrast is linear in the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

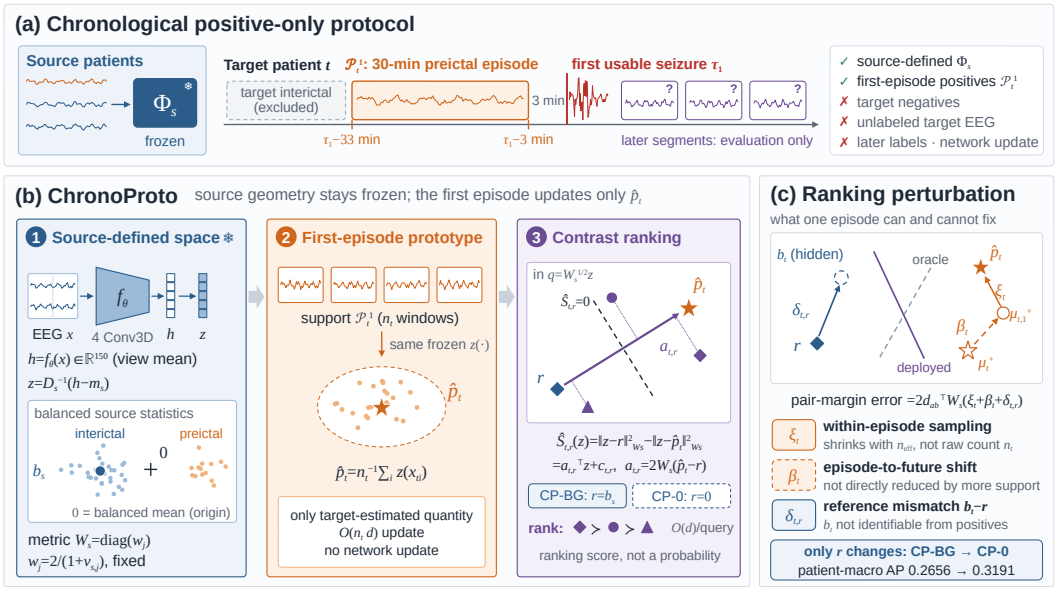

Figure 1: Overview of positive-only personalization with ChronoProto. (a) Source patients determine Φ_s, which holds the frozen encoder, feature geometry, and reference. For target patient t, personalization uses only the 30-min preictal episode $\mathcal{P}_t^1$, which ends 3 min before the first usable onset τ_1. Target negatives, unlabeled target EEG, and later labels are withheld, no network weights are updated, and later segments serve only for evaluation. (b) Windows are mapped into the source-defined space, the support is averaged into the prototype $\hat{p}_t$, and later windows are ranked by $\widehat{S}_{t,r}$ in Equation (5), with $r = b_s$ for CP-BG or $r = 0$ for CP-0. Both variants share the metric and the support. (c) The pairwise ranking error splits into within-episode sampling ξ_t, episode-to-future shift β_t, and reference mismatch $\delta_{t,r} = b_t - r$, where b_t cannot be identified from target positives. The AP values repeat Table 1. Coordinate weights come from source data and do not measure target reliability. Traces, embeddings, and rankings are schematic, and scores are not calibrated probabilities.

squared-distance features used by prototypical classifiers (Snell et al., 2017), so the error of each pairwise ranking splits into a prototype term and a reference term (Proposition 1). A correlated-sample analysis then shows how much independent evidence one episode can contribute.

We test these points on CHB–MIT with chronological patient-held-out folds. Across 28 later seizures from 13 patients, changing only the reference moves patient-macro AP from 0.2656 to 0.3191, while a source-selected linear SVM reaches 0.3175 AP. Both comparisons have paired AP and AUROC intervals spanning zero, so we interpret the observed differences as cohort-level sensitivity. Support ablations probe the first-episode representation, and source-only models establish performance under source supervision alone.

Our contributions are as follows.

1. **Protocol and identifiability.** We formalize positive-only personalization from the first usable seizure and characterize its information boundary: even exact knowledge of the target positive law does not determine the oracle ranking when the target background is unrestricted.

2. **Diagnostic decomposition.** We introduce ChronoProto, a minimal method whose pairwise ranking error separates into within-episode sampling error, episode-to-future shift, and reference mismatch, and we relate the sampling term to the effective support of a single correlated episode.

3. **Controlled evaluation.** Under chronological patient-held-out evaluation, we measure how sensitive the ranking is to the reference and to the representation of the first episode, and we keep these effects separate from comparisons with source-only models of other families.

2 Related Work

Seizure forecasting and patient-specific adaptation. Neural seizure forecasting methods learn patterns that separate preictal from interictal EEG (Truong et al., 2018; Tsiouris et al., 2018). These patterns vary considerably across patients, and later work has studied patient-independent learning and cross-patient transfer to improve generalization to unseen subjects (Dissanayake et al., 2021; 2022; Leguia et al., 2023; Peng et al., 2022; Hu et al., 2023). Other methods adapt a pretrained model with a small amount of target-patient data. Calibration from one or two seizures, for example, may use observations from both target classes (Shafiezadeh et al., 2024a). EEG foundation models such as CBraMod and NeuroLM learn representations intended to transfer across subjects and tasks (Wang et al., 2025; Jiang et al., 2025). We consider a more restricted form of personalization, in which the representation is fixed before adaptation and only the first target preictal episode is released. Evaluation is patient-held-out and chronological throughout, which avoids information leakage across patients or recordings (West et al., 2023; Shafiezadeh et al., 2024b).

Learning from positive-only target data. One-class classification describes an observed class from its examples. One-Class SVM estimates distributional support (Schölkopf et al., 2001), and SVDD and Deep SVDD learn enclosing descriptions in kernel and learned feature spaces (Tax & Duin, 2004; Ruff et al., 2018). Few-shot one-class meta-learning enables adaptation from sparse single-class support (Dahia & Pamplona Segundo, 2021; Frikha et al., 2021). Positive-unlabeled learning uses labeled positives and an unlabeled class mixture (Kiryo et al., 2017), and target-adaptation methods use unlabeled target observations (Jemal et al., 2024; Patel et al., 2025). Target-adaptation theory relies on additional target observations or source–target assumptions expressed through discrepancies, support conditions, or identifiable transformations (Ben-David et al., 2010; Johansson et al., 2019; Gulrajani & Hashimoto, 2022). Our protocol combines labeled source data with a single chronologically released, internally correlated positive episode. Target interictal samples and unlabeled mixtures remain outside fitting. Appendix B compares these information regimes.

Prototype-based scoring. Prototype-based classifiers represent each class by a point derived from its support and predict through distances in a learned feature space. Prototypical Networks adopt this formulation for few-shot classification and point out that, with squared Euclidean distance, the classifier is linear in the embedding (Snell et al., 2017). In seizure prediction, prototype methods build preictal and interictal prototypes for patient-level prediction (Gao et al., 2023; 2024). One-Way Prototypical Networks pair a single positive prototype with a zero reference for the unobserved class (Kruspe, 2019), the same construction as our CP-0 variant. We use these standard components as a controlled analytical tool. Since the representation, metric, and positive support stay fixed and only the reference varies, ChronoProto can separate the part of a positive-only ranking that the observed episode informs from the part supplied by an assumption about the missing background.

3 ChronoProto

ChronoProto personalizes a forecaster from the first usable preictal episode. Source patients determine the encoder, feature standardization, and distance metric (Figure 1(b)). The target episode is summarized by a positive prototype in this fixed space. A distance contrast between the prototype and a source-defined reference ranks later EEG windows.

The positive prototype is the only target-estimated quantity. The prefix *Chrono* denotes the chronological release of the prescribed preictal episode before the first usable seizure.

3.1 Positive-only personalization setting

For a held-out target patient t, the preictal windows associated with the first usable seizure form the support set

$$\mathcal{P}_t^1 = \{x_{t1}, \ldots, x_{tn_t}\}, \qquad \mathcal{A}_t = \mathcal{A}(\Phi_s, \mathcal{P}_t^1), \tag{1}$$

where x_{ti} is a support EEG window, $n_t = |\mathcal{P}_t^1|$ is the support size, Φ_s holds the source-defined encoder, feature geometry, and reference, and $\mathcal{A}$ is the personalization procedure that returns the scoring rule $\mathcal{A}_t$.

All patient-specific information comes from $\mathcal{P}_t^1$, and the rest of the model is inherited from Φ_s. Once personalization is done, $\mathcal{A}_t$ is frozen and applied to later EEG windows. We use *event* for a seizure onset and *episode* for its preictal support interval. An evaluation *segment* contains later preictal windows together with the retained interictal windows. Appendix B describes the data partitioning.

3.2 Source-defined representation space

To place the positive locations of different patients in a common coordinate system, we build a fixed representation space from the source patients. For an EEG window x,

$$h = f_\theta(x) \in \mathbb{R}^d, \qquad z(x) = D_s^{-1}(h - m_s), \tag{2}$$

where f_θ is the encoder trained on source patients, h is its d-dimensional representation, m_s is the balanced source mean, and D_s is a diagonal matrix of scale factors computed from source data. We call $z(x)$ the standardized representation.

In this space we use a fixed positive diagonal metric

$$W_s = \mathrm{diag}(w_1, \ldots, w_d) \succ 0, \qquad \|a\|_{W_s}^2 = a^\top W_s a, \tag{3}$$

with $w_j > 0$. Specifically, we set

$$w_j = \frac{2}{1 + v_{s,j}},$$

where $v_{s,j}$ is the balanced source interictal variance of coordinate j, so coordinates with high source interictal variance receive less weight. The analysis below only needs W_s to be a fixed positive diagonal matrix. Appendix C.1 gives the source statistics and the balancing details.

3.3 First-episode prototype and reference

We represent the positive location revealed by the first preictal episode by the mean of its standardized support features,

$$\hat{p}_t = \frac{1}{n_t} \sum_{x \in \mathcal{P}_t^1} z(x), \qquad n_t = |\mathcal{P}_t^1|, \tag{4}$$

where $\hat{p}_t \in \mathbb{R}^d$ is the first-episode prototype of patient t.

A later standardized window $z = z(x)$ is scored with the squared-distance contrast behind prototypical classifiers (Snell et al., 2017),

$$\widehat{S}_{t,r}(z) = \|z - r\|_{W_s}^2 - \|z - \hat{p}_t\|_{W_s}^2, \tag{5}$$

where $r \in \mathbb{R}^d$ is a fixed reference. A larger score means that the window lies closer to the patient's preictal prototype, relative to the reference.

We consider two references defined by the source data. CP-BG uses the balanced source interictal center, $r = b_s$. CP-0 uses the origin of the source-standardized space, $r = 0$, which parallels the zero reference of One-Way Prototypical Networks (Kruspe, 2019). The two variants share the encoder, standardization, metric, and first-episode prototype, so any difference between them is due to the reference alone.

3.4 Patient-specific linear ranking

Expanding Equation (5) gives

$$\begin{aligned} \widehat{S}_{t,r}(z) = a_{t,r}^\top z + c_{t,r} = 2\left[z - \frac{\hat{p}_t + r}{2}\right]^\top W_s(\hat{p}_t - r), \\ a_{t,r} = 2W_s(\hat{p}_t - r), \qquad c_{t,r} = r^\top W_s r - \hat{p}_t^\top W_s \hat{p}_t. \end{aligned} \tag{6}$$

Within a patient, $c_{t,r}$ cancels when two windows are compared, and the ranking depends only on the direction $a_{t,r} = 2W_s(\hat{p}_t - r)$. The first-episode prototype, the fixed reference, and the source metric together set this patient-specific direction.

After personalization we only cache $a_{t,r}$ and $c_{t,r}$. Leaving out the encoder, the update costs $O(n_t d)$ operations, each query costs $O(d)$, and the extra target storage is $O(d)$. Section 4 studies how errors in the prototype and the choice of reference perturb this ranking.

4 Analysis of Positive-Only Personalization

By Equation (6), the ChronoProto ranking depends on the first-episode prototype $\hat{p}_t$ and on the fixed reference r. We first split the ranking perturbation into prototype error and reference mismatch. We then ask whether the missing background can be identified at all, and finally split the first-episode prototype error into within-episode sampling and episode-to-future shift. The source geometry W_s is held fixed throughout.

4.1 Prototype error and reference mismatch

Let μ_t^+ and b_t be the population means of later target preictal and interictal embeddings. If both were known, we could use the target-center contrast

$$S_t^\star(z) = \|z - b_t\|_{W_s}^2 - \|z - \mu_t^+\|_{W_s}^2. \tag{7}$$

Under Gaussian class-conditionals with shared covariance W_s^{-1}, $S_t^\star$ induces the same ranking as the true log-likelihood ratio.

For two later windows z_a, z_b, define

$$\widehat{\Delta}_{ab,r} = \widehat{S}_{t,r}(z_a) - \widehat{S}_{t,r}(z_b), \qquad \Delta_{ab}^\star = S_t^\star(z_a) - S_t^\star(z_b), \tag{8}$$

and let $d_{ab} = z_a - z_b$, $e_t = \hat{p}_t - \mu_t^+$, and $\delta_{t,r} = b_t - r$.

Proposition 1 (Pairwise ranking perturbation)**.** *For any positive diagonal matrix W_s,*

$$\begin{aligned} \widehat{\Delta}_{ab,r} - \Delta_{ab}^\star &= 2d_{ab}^\top W_s(e_t + \delta_{t,r}), \\ \left|\widehat{\Delta}_{ab,r} - \Delta_{ab}^\star\right| &\le 2\|d_{ab}\|_{W_s}\left(\|e_t\|_{W_s} + \|\delta_{t,r}\|_{W_s}\right). \end{aligned} \tag{9}$$

The identity follows by substituting Equation (6) into the two pairwise margins, and the bound then follows from Cauchy–Schwarz and the triangle inequality.

Here e_t measures prototype error relative to the future positive center, and $\delta_{t,r}$ measures reference mismatch with the target background. Both enter through d_{ab} and can add or partly cancel. More positive support can improve the prototype estimate. The fixed reference mismatch remains unchanged.

For two fixed references r_1, r_2 with the prototype and metric unchanged,

$$\widehat{\Delta}_{ab,r_2} - \widehat{\Delta}_{ab,r_1} = 2d_{ab}^\top W_s(r_1 - r_2). \tag{10}$$

The difference between CP-BG and CP-0 therefore comes entirely from the reference term, and its effect on a given pair depends on the pair direction d_{ab}.

4.2 An identifiability boundary for the missing background

By Equation (9), the reference mismatch depends on the unobserved target background b_t. A more basic question is whether target-positive information alone could determine a suitable reference.

Proposition 2 (Oracle ranking is not identifiable from target positives alone)**.** *Let a learner, possibly using independent randomness, construct a ranking score from $(\Phi_s, \mathcal{P}_t^1)$. Across target populations with the same source record and the same target-positive support law but unrestricted negative laws, no learner can recover the true likelihood-ratio ranking almost surely for every population.*

One dimension is enough to see this. Take two populations with the same source information and $Z^+ \sim \mathcal{N}(0,1)$, whose Gaussian negative distributions have equal variance and means on opposite sides of the positive center. The admitted positive-only information is the same for both, yet the two populations require opposite likelihood-ratio rankings. Appendix A gives the full proof.

Even exact knowledge of the target-positive law does not fix the background reference unless additional structure is assumed. CP-BG and CP-0 encode two such assumptions, both defined by the source data, and the first positive episode cannot tell which of them better matches the target background.

For example, if $b_t = b_s$ under the Gaussian comparator model, choosing $r = b_s$ gives $\delta_{t,r} = 0$, leaving only the first-episode prototype error.

4.3 Temporal effective support and episode-to-future shift

For a fixed reference, we split the prototype error as

$$e_t = \underbrace{\hat{p}_t - \mu^+_{t,1}}_{\xi_t} + \underbrace{\mu^+_{t,1} - \mu^+_t}_{\beta_t}, \tag{11}$$

where $\mu^+_{t,1}$ is the population mean of the first episode, ξ_t is the within-episode sampling error, and β_t is the episode-to-future shift.

Let $Z_i = z(x_{ti})$ be the support embeddings in temporal order, with $n = n_t$. We adapt the classical variance calculation for the mean of an autocorrelated sample (Bartlett, 1946). Under second-order stationarity and the separable lag-covariance model $\Gamma_k = \mathrm{Cov}(Z_i, Z_{i+k}) = \rho_k \Gamma_0$, the sampling covariance is

$$\mathrm{Cov}(\xi_t) = \frac{\Gamma_0}{n_{\text{eff}}}, \qquad n_{\text{eff}} = \frac{n^2}{n + 2\sum_{k=1}^{n-1}(n-k)\rho_k}. \tag{12}$$

Here n_{eff} is the number of independent windows with marginal covariance Γ_0 that would give the same prototype covariance as the observed correlated support, and the scalar denominator is assumed positive. When the weighted lag sum is positive, n_{eff} falls below the raw window count.

More effective support can reduce ξ_t, while β_t describes the separate episode-to-future shift. Combining Equations (9) and (11) gives

$$\widehat{\Delta}_{ab,r} - \Delta^\star_{ab} = 2d_{ab}^\top W_s \left(\xi_t + \beta_t + \delta_{t,r}\right).$$

In positive-only personalization, the pairwise ranking perturbation has three sources: the finite effective support of the first episode, the shift from that episode to future positives, and the mismatch between the fixed reference and the target background. In Section 5 we probe the corresponding sensitivities with support-coverage analyses and with the controlled comparison of CP-BG and CP-0.

5 Experiments

We address three questions: (i) how source-only and target-positive-only methods compare under the admitted information, (ii) how the ranking changes when only the reference varies, and (iii) how the representation of the first episode and patient heterogeneity affect the results. Transfer and richer-information analyses give secondary context.

5.1 Experimental setup

Data and chronological protocol. We evaluate on CHB–MIT scalp EEG (Guttag, 2010; Shoeb & Guttag, 2010) with chronological patient-held-out folds. Of 17 candidate patients, 13 have a first usable seizure and at least one later evaluable segment, which gives 28 later events. Target supervision comes only from the prescribed first preictal episode (33 to 3 minutes before onset), and target interictal EEG and later queries are excluded from fitting. In each outer fold, a four-block source CNN with a 150-dimensional hidden representation is trained on the remaining patients, and both ChronoProto variants share this representation, the source geometry, and the first-episode prototype. We use the FP1–F7 channel and nonoverlapping 60-s windows (details in Appendix C).

Comparators and model selection. The target-positive-only comparators are logistic regression, RBF SVM, and shrinkage LDA, each fitted on the first target episode together with source interictal examples. The source-only controls are logistic regression, linear SVM, a MobileNet-based architecture following Jang et al. (2026), and STFT-CNN-BiLSTM. For the linear controls, cross-fitting over source patients selects orientation and regularization jointly, with validation patients excluded from fitting. No target label selects a sign, hyperparameter, reference, or variant (Appendix C.3).

Metrics and uncertainty. AP and AUROC are the primary endpoints, and normalized partial AUROC up to a false-positive rate of 0.05 is secondary (McClish, 1989). Event-level metrics are averaged first within each patient and then with equal weight across patients:

$$\overline{M} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{E}_t|}\sum_{c\in\mathcal{E}_t} M_{tc}, \qquad \text{n-pAUC}_\alpha = \frac{1}{\alpha}\int_0^\alpha \mathrm{TPR}(u)\,du, \quad \alpha = 0.05. \tag{13}$$

Table 1: Patient-macro ranking performance on the primary CHB–MIT cohort. Target-positive-only methods use first-episode positives and, where stated, source interictal examples. Target negatives are excluded from fitting. Source-LR and Source-SVM jointly select score sign and regularization on source data.

Method	AP↑	AUROC↑	n-pAUC↑
No information	0.1809	0.5000	0.0250
Source-only (no target labels)			
Source-LR	0.3119	0.5799	0.1623
Source-SVM	0.3175	0.5843	0.1640
Jang26-MobileNet	0.2783	0.5813	0.1388
STFT-CNN-BiLSTM	0.2733	0.4971	0.1812
Target-positive-only (first target episode, no target negatives)			
Logistic regression	0.2338	0.4759	0.1368
RBF SVM	0.2494	0.5265	0.1339
Shrinkage LDA	0.2564	0.5423	0.1273
CP-BG	0.2656	0.5601	0.1367
CP-0	0.3191	0.5871	0.1861

Here T is the number of patients, $\mathcal{E}_t$ is the set of later events of patient t, and M_{tc} is an event metric. Paired 95% bootstrap intervals resample whole patients (Efron, 1979), and component intervals are unadjusted diagnostics (Appendix C.4).

5.2 Overall ranking performance

Among the target-positive-only methods, CP-0 has the highest cohort means: 0.3191 AP, 0.5871 AUROC, and 0.1861 n-pAUC (Table 1). Source-SVM reaches 0.3175 AP and 0.5843 AUROC. The differences are 0.0016 and 0.0029, with paired 95% intervals of $[-0.1401, 0.1322]$ and $[-0.1992, 0.1928]$. These intervals describe the uncertainty in the comparison across model families. Isolating the target-positive update requires a matched source-positive-prototype control.

Model selection on source data alone already has a large effect. Selecting orientation and regularization jointly raises linear SVM AP from 0.2056 to 0.3175 and AUROC from 0.3977 to 0.5843. For logistic regression, AP moves from 0.2056 to 0.3119 and AUROC from 0.3953 to 0.5799 (Figure 2(a)). Careful source-only selection is therefore necessary context for any personalization claim.

5.3 Reference sensitivity

With the encoder, transform, metric, and prototype fixed, replacing CP-BG with CP-0 moves AP from 0.2656 to 0.3191, AUROC from 0.5601 to 0.5871, and n-pAUC from 0.1367 to 0.1861. Equation (10) attributes this change to the reference term. The paired AP and AUROC differences are 0.0536 and 0.0270, with 95% intervals of $[-0.0281, 0.1590]$ and $[-0.0303, 0.0986]$ (Figure 2(b)). Both intervals span zero. We interpret these changes as cohort-level reference sensitivity, complementing Section 4.1.

5.4 First-episode support and geometry

With CP-BG as the parent, full support exceeds eight distributed windows by 0.0310 AP, 0.0347 AUROC, and 0.0234 n-pAUC, with paired intervals above zero. Four distributed windows come closer to full support than eight do (Table 5), consistent with the role of temporal dependence in Section 4.3. This ablation jointly varies count and coverage and measures their combined sensitivity. Isotropic weighting changes AUROC by 0.0132 relative to CP-BG and slightly decreases n-pAUC, with the AP interval spanning zero. The geometry thus has endpoint-specific effects (Figure 4 and Table 5).

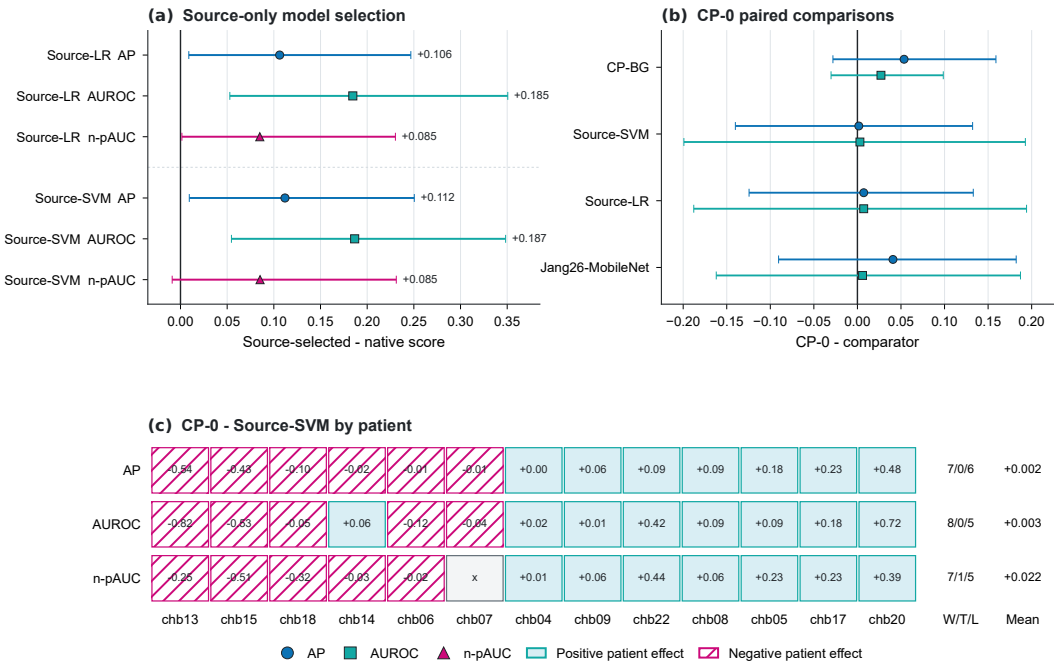

Figure 2: Primary effects on 13 patients. (a) Effect of jointly selecting sign and regularization on source data, relative to the native score. (b) CP-0 minus each comparator. (c) Patient-level differences from Source-SVM. Source-LR and Source-SVM are the source-selected linear controls. Error bars are 95% paired-patient bootstrap intervals from 100,000 resamples. Hatched cells are negative, filled cells are positive, and crosses mark ties.

5.5 Patient-level heterogeneity

Relative to Source-SVM, CP-0 has positive and negative AP differences for seven and six patients, respectively, with AUROC counts of eight and five (Figure 2(c)). For CP-0 versus CP-BG, the mean AP difference is 0.0536, with a component-bootstrap interval of $[-0.0283, 0.1580]$. Excluding the largest AP contributor gives a difference of 0.0134. Excluding the largest AUROC contributor shifts the AUROC difference from 0.0270 to -0.0019. Median prototypes show similar patterns (Appendix D.1). These diagnostics characterize patient influence within the observed cohort.

5.6 Transfer and richer target information

Transfer analysis. We transfer the source geometry and CP-BG rule fixed on the primary cohort to two held-out CHB–MIT patients with seven later events and three Siena patients with five later events. Relative to the native Source CNN, CP-BG raises AP from 0.0890 to 0.3477 on held-out CHB–MIT. On Siena, it raises AP from 0.4868 to 0.5099, AUROC from 0.6900 to 0.7181, and n-pAUC from 0.3172 to 0.3325. These results describe the two supplementary cohorts (Appendix D.2).

Richer target information. In a separate analysis, we release both classes of the first segment for nine patients with 18 events. Across a fixed logistic-regression sweep, AP ranges from 0.3028 to 0.4042, compared with 0.2711 for target-positive-only CP-BG. We report the complete sweep to describe performance under this richer, two-class supervision regime.

5.7 Limitations

The empirical scope is retrospective, single-channel ranking in 13 patients. A matched source-positive prototype would isolate the target-positive update, and larger multichannel cohorts with continuous monitoring would assess broader applicability. Episode shift, reference mismatch, and temporal effective support are characterized analytically and through sensitivity analyses. Direct

estimation requires window embeddings and labels. Transfer analysis evaluates CP-BG with the primary-cohort construction. Further work can compare references and source-only controls across cohorts and examine one-class methods, shrinkage, and selector robustness. The identifiability result applies independently of these implementation choices.

6 Conclusion

We studied seizure-forecaster personalization from the first target preictal episode under positive-only supervision. ChronoProto separates the episode-informed positive location from the reference required for ranking. Its pairwise ranking perturbation splits into finite effective support, episode-to-future shift, and reference mismatch.

With unrestricted target background, identifying the oracle ranking requires information beyond the target-positive law. On CHB–MIT, changing only the reference moves patient-macro AP from 0.2656 to 0.3191, with patient-specific variation. The first episode locates the patient's positives, and cross-patient information or an explicit structural assumption supplies the ranking reference. ChronoProto provides a controlled baseline for positive-only personalization and for comparisons with methods using richer target information.

Ethics statement

This retrospective methodological study uses publicly released, de-identified EEG databases. Evaluation keeps patients separated and reports the eligible sample size after chronological screening. The study concerns retrospective ranking. Clinical translation requires prospective validation under continuous monitoring, subgroup analysis, and assessment of operating points.

Reproducibility statement

Section 5 specifies the protocol, cohort, preprocessing, metrics, and patient-level resampling. Appendices A and B provide the proofs, chronological eligibility rules, and target-information regimes. Appendix C details the source statistics, preprocessing parameters, architecture, regularization candidates, model selection, evaluation, and bootstrap seeds. Appendix D.1 reports supplementary component analyses. The released artifact covers the numerical ChronoProto scoring rule, evaluation functions, and de-identified aggregate and patient-level inputs for reproducing the figures and summaries. Raw-EEG training procedures are documented in Appendix C.

AI use statement

We used generative AI tools for literature discovery, research ideation and execution, and drafting and polishing the paper, including checks of propositions, proofs, and result interpretation. We reviewed all AI-assisted content and take responsibility for the paper.

References

M. S. Bartlett. On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 8(1):27–41, 1946. doi: 10.2307/2983611.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79:151–175, 2010. doi: 10.1007/s10994-009-5152-4. URL https://link.springer.com/article/10.1007/s10994-009-5152-4.

Gabriel Dahia and Maurício Pamplona Segundo. Meta learning for few-shot one-class classification. *AI*, 2(2):195–208, 2021. doi: 10.3390/ai2020012. URL https://www.mdpi.com/2673-2688/2/2/12.

Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals. *IEEE Sensors Journal*, 21(7):9377–9388, April 2021. doi: 10.1109/JSEN.2021.3057076.

Theekshana Dissanayake, Tharindu Fernando, Simon Denman, Sridha Sridharan, and Clinton Fookes. Geometric deep learning for subject independent epileptic seizure prediction using scalp EEG signals. *IEEE Journal of Biomedical and Health Informatics*, 26(2):527–538, February 2022. doi: 10.1109/JBHI.2021.3100297.

Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, January 1979. doi: 10.1214/aos/1176344552.

Ahmed Frikha, Denis Krompaß, Hans-Georg Köpken, and Volker Tresp. Few-shot one-class classification via meta-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7448–7456, 2021. doi: 10.1609/aaai.v35i8.16913. URL https://ojs.aaai.org/index.php/AAAI/article/view/16913.

Yikai Gao, Aiping Liu, Lanlan Wang, Ruobing Qian, and Xun Chen. A self-interpretable deep learning model for seizure prediction using a multi-scale prototypical part network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1847–1856, March 2023. doi: 10.1109/TNSRE.2023.3260845.

Yikai Gao, Aiping Liu, Heng Cui, Ruobing Qian, and Xun Chen. An interpretable and generalizable deep learning model for iEEG-based seizure prediction using prototype learning and contrastive learning. *Computers in Biology and Medicine*, 183:109257, December 2024. doi: 10.1016/j.compbiomed.2024.109257.

Ishaan Gulrajani and Tatsunori Hashimoto. Identifiability conditions for domain adaptation. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 7982–7997. PMLR, 2022. URL https://proceedings.mlr.press/v162/gulrajani22a.html.

John Guttag. CHB-MIT Scalp EEG Database. PhysioNet, version 1.0.0, June 2010. RRID: SCR_007345.

Shuaicong Hu, Jian Liu, Rui Yang, Ya'Nan Wang, Aiguo Wang, Kuanzheng Li, Wenxin Liu, and Cuiwei Yang. Exploring the applicability of transfer learning and feature engineering in epilepsy prediction using hybrid transformer model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1321–1332, February 2023. doi: 10.1109/TNSRE.2023.3244045.

Dabin Jang, Ki-Young Jung, Young-Gyu Jeon, Tae-Joon Kim, Sang Kun Lee, and Kyeong-Yuk Min. Single-channel EEG-based seizure prediction using deep learning. *Scientific Reports*, 16(1):15888, April 2026. doi: 10.1038/s41598-026-44670-7.

Imene Jemal, Lina Abou-Abbas, Khadidja Henni, Amar Mitiche, and Neila Mezghani. Domain adaptation for EEG-based, cross-subject epileptic seizure prediction. *Frontiers in Neuroinformatics*, 18:1303380, February 2024. doi: 10.3389/fninf.2024.1303380.

Wei-Bang Jiang, Yansen Wang, Bao-Liang Lu, and Dongsheng Li. NeuroLM: A universal multi-task foundation model for bridging the gap between language and EEG signals. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/8b4add8b0aa8749d80a34ca5d941c355-Abstract-Conference.html.

Fredrik D. Johansson, David Sontag, and Rajesh Ranganath. Support and invertibility in domain-invariant representations. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 527–536. PMLR, 2019. URL https://proceedings.mlr.press/v89/johansson19a.html.

Philippa J. Karoly, Hoameng Ung, David B. Grayden, Levin Kuhlmann, Kent Leyde, Mark J. Cook, and Dean R. Freestone. The circadian profile of epilepsy improves seizure forecasting. *Brain*, 140 (8):2169–2182, August 2017. doi: 10.1093/brain/awx173.

Ryuichi Kiryo, Gang Niu, Marthinus C. du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://papers.nips.cc/paper_files/paper/2017/hash/7cce53cf90577442771720a370c3c723-Abstract.html.

Anna Kruspe. One-way prototypical networks. *arXiv preprint arXiv:1906.00820*, 2019. doi: 10.48550/arXiv.1906.00820. URL https://arxiv.org/abs/1906.00820.

Levin Kuhlmann, Klaus Lehnertz, Mark P. Richardson, Björn Schelter, and Hitten P. Zaveri. Seizure prediction—ready for a new era. *Nature Reviews Neurology*, 14(10):618–630, October 2018. doi: 10.1038/s41582-018-0055-2.

Marc G. Leguia, Vikram R. Rao, Thomas K. Tcheng, Jonas Duun-Henriksen, Troels W. Kjær, Timothée Proix, and Maxime O. Baud. Learning to generalize seizure forecasts. *Epilepsia*, 64(S4): S99–S113, December 2023. doi: 10.1111/epi.17406.

Donna Katzman McClish. Analyzing a portion of the ROC curve. *Medical Decision Making*, 9(3): 190–195, August 1989. doi: 10.1177/0272989X8900900307.

Gaurav Patel, Christopher M. Sandino, Behrooz Mahasseni, Ellen Zippi, Erdrin Azemi, Ali Moin, and Juri Minxha. Efficient source-free time-series adaptation via parameter subspace disentanglement. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e1fcd183ab33714a8464e4e9a20ac710-Abstract-Conference.html.

Peizhen Peng, Yang Song, Lu Yang, and Haikun Wei. Seizure prediction in EEG signals using STFT and domain adaptation. *Frontiers in Neuroscience*, 15:825434, January 2022. doi: 10.3389/fnins.2021.825434.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4393–4402. PMLR, 2018. URL https://proceedings.mlr.press/v80/ruff18a.html.

Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7): 1443–1471, 2001. doi: 10.1162/089976601750264965. URL https://research.google/pubs/estimating-the-support-of-a-high-dimensional-distribution/.

Sina Shafiezadeh, Gian Marco Duma, Giovanni Mento, Alberto Danieli, Lisa Antoniazzi, Fiorella Del Popolo Cristaldi, Paolo Bonanni, and Alberto Testolin. Calibrating deep learning classifiers for patient-independent electroencephalogram seizure forecasting. *Sensors*, 24(9):2863, April 2024a. doi: 10.3390/s24092863.

Sina Shafiezadeh, Gian Marco Duma, Marco Pozza, and Alberto Testolin. A systematic review of cross-patient approaches for EEG epileptic seizure prediction. *Journal of Neural Engineering*, 21 (6):061004, December 2024b. doi: 10.1088/1741-2552/ad9682.

Ali H. Shoeb and John V. Guttag. Application of machine learning to epileptic seizure detection. In *Proc. 27th Int. Conf. Machine Learning*, pp. 975–982, June 2010.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4077–4087, December 2017.

David M. J. Tax and Robert P. W. Duin. Support vector data description. *Machine Learning*, 54(1): 45–66, 2004. doi: 10.1023/B:MACH.0000008084.60811.49. URL https://research.tudelft.nl/en/publications/support-vector-data-description/.

Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Samuel Ippolito, and Omid Kavehei. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, September 2018. doi: 10.1016/j.neunet.2018.04.018.

Kostas M. Tsiouris, Vasileios C. Pezoulas, Michalis Zervakis, Spiros Konitsiotis, Dimitrios D. Koutsouris, and Dimitrios I. Fotiadis. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Computers in Biology and Medicine*, 99:24–37, August 2018. doi: 10.1016/j.compbiomed.2018.05.019.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/bbbd6d915cb90be21c1254a82d45cedd-Abstract-Conference.html.

Joseph West, Zahra Dasht Bozorgi, Jeffrey Herron, Howard J. Chizeck, Jordan D. Chambers, and Lyra Li. Machine learning seizure prediction: One problematic but accepted practice. *Journal of Neural Engineering*, 20(1):016008, February 2023. doi: 10.1088/1741-2552/acae09.

A Theoretical Derivations and Further Analysis

A.1 Linearization and geometric interpretation

For symmetric W_s, expanding Equation (5) gives

$$\widehat{S}_{t,r}(z) = (z-r)^\top W_s(z-r) - (z-\hat{p}_t)^\top W_s(z-\hat{p}_t) \tag{14}$$

$$= 2z^\top W_s(\hat{p}_t - r) + r^\top W_s r - \hat{p}_t^\top W_s \hat{p}_t \tag{15}$$

$$= 2\left[z - \frac{\hat{p}_t + r}{2}\right]^\top W_s(\hat{p}_t - r). \tag{16}$$

Here z is a query embedding, $\hat{p}_t$ is the first-episode positive prototype, r is the fixed reference, and W_s is the source-defined metric. The quadratic terms $z^\top W_s z$ cancel, which gives the linear form used in Section 3.4.

Consider the metric-transformed coordinates $q = W_s^{1/2} z$. The final expression above is twice the inner product between the query displacement from the transformed midpoint and the direction from the transformed reference to the transformed prototype. For $\hat{p}_t \neq r$, division by $2\|\hat{p}_t - r\|_{W_s}$ gives the signed distance to the bisecting hyperplane between the prototype and the reference. Because this denominator is positive and independent of the query, the transformation preserves within-patient ranking. In the degenerate case $\hat{p}_t = r$, the prototype and the reference coincide and the contrast in Equation (5) is identically zero.

A.2 Target reference model and proof of Proposition 1

Assume Gaussian class-conditional distributions with target positive and negative means μ_t^+ and b_t, respectively, and shared covariance W_s^{-1}. The target log-likelihood ratio is

$$g_t^\star(z) = z^\top W_s(\mu_t^+ - b_t) - \frac{1}{2}(\mu_t^+ + b_t)^\top W_s(\mu_t^+ - b_t). \tag{17}$$

Expanding Equation (7) gives $S_t^\star(z) = 2g_t^\star(z)$, so under this comparator model $S_t^\star$ and the target log-likelihood ratio induce the same within-patient ranking. Class priors only add a query-independent constant to the log posterior odds and leave the ranking unchanged.

For $d_{ab} = z_a - z_b$, Equation (8) together with the linear form in Equation (6) gives

$$\widehat{\Delta}_{ab,r} = 2d_{ab}^\top W_s(\hat{p}_t - r), \tag{18}$$

$$\Delta_{ab}^\star = 2d_{ab}^\top W_s(\mu_t^+ - b_t). \tag{19}$$

Substituting $\hat{p}_t = \mu_t^+ + e_t$ and $b_t = r + \delta_{t,r}$ yields

$$\widehat{\Delta}_{ab,r} - \Delta_{ab}^\star = 2d_{ab}^\top W_s(e_t + \delta_{t,r}),$$

which is the identity in Equation (9). Applying the Cauchy–Schwarz inequality under the W_s norm gives

$$\left|d_{ab}^\top W_s(e_t + \delta_{t,r})\right| \leq \|d_{ab}\|_{W_s}\|e_t + \delta_{t,r}\|_{W_s}.$$

The triangle inequality then yields

$$\left|\widehat{\Delta}_{ab,r} - \Delta_{ab}^\star\right| \leq 2\|d_{ab}\|_{W_s}\left(\|e_t\|_{W_s} + \|\delta_{t,r}\|_{W_s}\right).$$

If a positive target-center margin exceeds this perturbation bound, the deployed pairwise margin remains positive. Reversing the pair order gives the corresponding statement for a negative target-center margin.

Finally, for two references r_1 and r_2,

$$\widehat{\Delta}_{ab,r_2} - \widehat{\Delta}_{ab,r_1} = 2d_{ab}^\top W_s(r_1 - r_2),$$

which gives Equation (10).

A.3 Proof of Proposition 2

Fix an identical source record and source geometry with $W_s = 1$. In both target populations, let the first-episode support consist of draws from $\mathcal{N}(0, 1)$, and let later positives follow the same distribution. Define the two target-negative distributions as $\mathcal{N}(a, 1)$ and $\mathcal{N}(-a, 1)$, where $a > 0$. Both populations then share the source information and the target-positive law. Their admitted adaptation inputs have the same distribution, and coupling the first-episode draws makes the realized inputs identical as well.

For the population with negative mean a, the log-likelihood ratio is $-az + a^2/2$, whereas for the population with negative mean $-a$, it is $az + a^2/2$. The two functions are strictly monotone in opposite directions, so for any two distinct query values the oracle pair order is reversed between the populations.

A deterministic learner receives the same admitted information in both populations and therefore produces the same scoring rule, which cannot realize both opposite orderings. For a randomized learner, fix any distinct query pair and let q_+ and q_- denote the probabilities of producing the two possible strict orders under the common admitted-input distribution and independent learner randomness. Since the correct orders in the two populations are opposite and $q_+ + q_- \leq 1$, at least one population has probability at most $1/2$ of receiving the correct strict order. Hence, no learner can recover the oracle ranking almost surely for both populations.

Because any unrestricted family containing these two populations inherits this indistinguishability, uniform recovery of the oracle ranking is impossible from target-positive information alone. Exact knowledge of the target-positive distribution does not distinguish the two populations either, which establishes the claimed non-identifiability.

A.4 Temporal covariance and episode shift

Condition on the source geometry and on fixed first-episode and future-positive population means. Let $Z_1, \ldots, Z_n$ denote the temporally ordered support representations, with $\mathbb{E}[Z_i] = \mu^+_{t,1}$, and assume second-order stationarity with lag-k cross-covariance $\Gamma_k = \mathrm{Cov}(Z_i, Z_{i+k})$. Because $\hat{p}_t = n^{-1} \sum_{i=1}^n Z_i$, the covariance of the within-episode sampling term is

$$\begin{aligned} C_\xi &= \mathrm{Cov}(\xi_t) = \mathrm{Cov}(\hat{p}_t) \\ &= \frac{1}{n^2} \left[n\Gamma_0 + \sum_{k=1}^{n-1} (n-k)(\Gamma_k + \Gamma_k^\top) \right]. \end{aligned} \tag{20}$$

This follows by expanding $n^{-2} \sum_{i,j} \mathrm{Cov}(Z_i, Z_j)$ and collecting terms with the same temporal lag.

Under the separable covariance model $\Gamma_k = \rho_k \Gamma_0$, the expression becomes $C_\xi = \Gamma_0 / n_{\text{eff}}$, with

$$n_{\text{eff}} = \frac{n^2}{n + 2\sum_{k=1}^{n-1} (n-k)\rho_k},$$

as defined in Equation (12). A positive weighted lag sum reduces n_{eff} below the raw support count n, while zero nonzero-lag correlations give $n_{\text{eff}} = n$. For general multivariate temporal dependence, the sampling covariance above remains the more general representation.

Using $e_t = \xi_t + \beta_t$ and $\mathbb{E}[\xi_t] = 0$, we obtain

$$\mathbb{E}\|e_t\|^2_{W_s} = \mathrm{tr}(W_s C_\xi) + \|\beta_t\|^2_{W_s}, \tag{21}$$

where $\mathrm{tr}(\cdot)$ is the matrix trace. The first term is the uncertainty due to finite, temporally correlated support within the first episode, and the second is the mean shift from the first episode to the future positive population. More effective support reduces the sampling term, whereas the episode-to-future term is set by how the observed episode relates to future positives.

A.5 Conditional ranking reversal and AUROC disagreement

For a fixed later pair and reference, further assume

$$\mathbb{E}[\xi_t \mid d_{ab}, \delta_{t,r}, \beta_t] = 0, \qquad \mathrm{Cov}(\xi_t \mid d_{ab}, \delta_{t,r}, \beta_t) = C_{\xi|ab}.$$

Under these conditions, Equation (11) gives

$$\mathbb{E}[\widehat{\Delta}_{ab,r} - \Delta^{\star}_{ab} \mid d_{ab}, \delta_{t,r}, \beta_t] = 2d_{ab}^{\top} W_s(\beta_t + \delta_{t,r}), \tag{22}$$

$$\mathrm{Var}(\widehat{\Delta}_{ab,r} - \Delta^{\star}_{ab} \mid d_{ab}, \delta_{t,r}, \beta_t) = 4d_{ab}^{\top} W_s C_{\xi|ab} W_s d_{ab} = \sigma_R^2. \tag{23}$$

For a positive target-center margin $m = \Delta^{\star}_{ab} > 0$, define

$$q = m + 2d_{ab}^{\top} W_s(\beta_t + \delta_{t,r}),$$

which is the conditional mean of the deployed margin. When $q > 0$, Cantelli's inequality gives

$$\Pr(\widehat{\Delta}_{ab,r} \le 0 \mid d_{ab}, \delta_{t,r}, \beta_t) \le \frac{\sigma_R^2}{\sigma_R^2 + q^2}. \tag{24}$$

If the conditional fluctuation is Gaussian and $\sigma_R > 0$, the reversal probability becomes $\Phi(-q/\sigma_R)$, where Φ denotes the standard normal cumulative distribution function.

To connect pairwise disagreement with AUROC, define

$$g(\Delta) = \mathbf{1}\{\Delta > 0\} + \frac{1}{2}\mathbf{1}\{\Delta = 0\}.$$

For a random later positive-negative pair, let

$$\mathcal{D} = \{g(\widehat{\Delta}_{ab,r}) \neq g(\Delta^{\star}_{ab})\}$$

denote the disagreement event. Then

$$\left|\mathrm{AUROC}(\widehat{S}_{t,r}) - \mathrm{AUROC}(S^{\star}_t)\right| \le \mathbb{E}\left|g(\widehat{\Delta}_{ab,r}) - g(\Delta^{\star}_{ab})\right| \le \Pr(\mathcal{D}). \tag{25}$$

The probability of pairwise disagreement therefore bounds the AUROC difference between the deployed score and the target-center comparator. AP also depends on prevalence and on the positions of all examples in the ranked list, so it would need a full-list analysis that goes beyond this pairwise relation.

B Dataset and Protocol Details

Table 2 lists typical information regimes for target adaptation. Source labels are available for training in every setting, while labels from later target segments are reserved for evaluation.

Table 2: Examples of information regimes during target adaptation.

Regime	Target +	Target −	Unlabeled target	Target update
Source-only	No	No	No	None
Target-positive-only (ours)	First episode	No	No	Positive center
Positive-unlabeled	Yes	No	Mixture	Mixture-based fitting
Unlabeled adaptation	No	No	Mixture	Unlabeled-data update
Two-class calibration	Yes	Yes	No	Fitted target boundary

ChronoProto belongs to the target-positive-only regime. The first target episode provides positive support, and target interictal observations and unlabeled target mixtures stay out of fitting.

Chronological eligibility is determined at the patient level. The first usable event must contain a complete preictal interval under the fixed labeling rule. Each patient must then have at least one later seizure-centered segment containing retained preictal and interictal windows for evaluation. Interictal windows from the first segment are excluded from all positive-only estimators, and later segments are used for evaluation after personalization.

Events are first averaged within each patient before cohort-level aggregation. The available study inventory supports the reported 17-to-13 chronological eligibility screening for the primary cohort.

Figure 3 summarizes the fixed EEG input and chronological protocol. Gray scalp locations provide anatomical orientation, while FP1–F7 is the only derivation used by the models.

Table 3: Cohort composition after chronological eligibility screening. Available denotes the study inventory.

Cohort	Available	Eligible	Later events
CHB–MIT primary	17	13	28
CHB–MIT held-out	4	2	7
Siena held-out	14	3	5

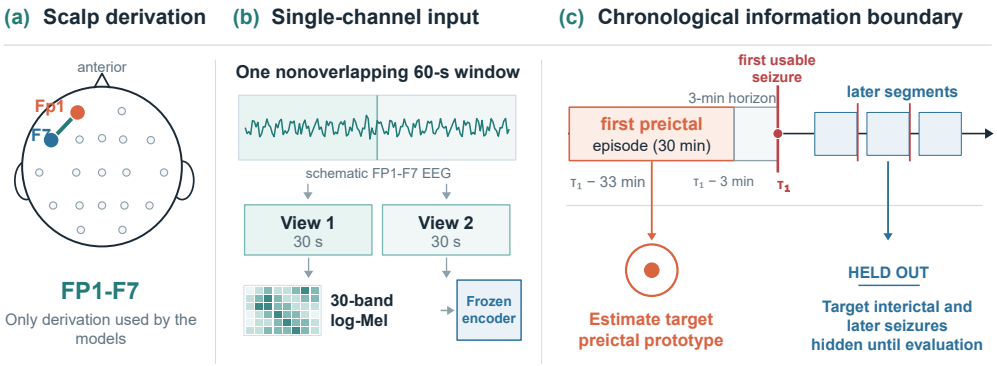

Figure 3: Input and chronological protocol. Only FP1–F7 is used. The first preictal episode estimates the target prototype; target interictal data and later seizures remain hidden until evaluation.

C Implementation and Evaluation Details

C.1 Source geometry and balancing

Here we specify the source map and metric used in Section 3.2. All source statistics are estimated separately within each outer source fold and remain fixed during target personalization.

Let $h_{ucyi} = f_\theta(x_{ucyi}) \in \mathbb{R}^d$, where u indexes source patient, c indexes retained segment, $y \in \{-,+\}$ denotes interictal or preictal class, and i indexes windows within a stratum.

To prevent patients with longer recordings from dominating the source geometry, we balance patients, retained segments, and the two classes through

$$\begin{aligned} \omega_{ucyi} &= \frac{1}{|\mathcal{U}_s|\,|\mathcal{C}_u|\,2n_{ucy}}, \\ m_s &= \sum_{u,c,y,i} \omega_{ucyi} h_{ucyi}, \\ r_{s,j}^2 &= \sum_{u,c,y,i} \omega_{ucyi}(h_{ucyi,j} - m_{s,j})^2. \end{aligned} \tag{26}$$

Here, $\mathcal{U}_s$ is the source-patient set, $\mathcal{C}_u$ is the patient-specific retained-segment set, and $n_{ucy} > 0$ is the number of distinct windows in stratum (u, c, y). The weights sum to one. The quantities m_s and $r_{s,j}$ are the weighted source mean and standard deviation of coordinate j.

The common standardized transform is

$$z_j(x) = \frac{f_\theta(x)_j - m_{s,j}}{\widetilde{r}_{s,j}}, \qquad \widetilde{r}_{s,j} = \begin{cases} r_{s,j}, & r_{s,j} \geq \epsilon, \\ 1, & r_{s,j} < \epsilon, \end{cases} \qquad \epsilon = 10^{-8}. \tag{27}$$

Equivalently, $D_s = \mathrm{diag}(\widetilde{r}_{s,1}, \ldots, \widetilde{r}_{s,d})$, which is the scale matrix in Equation (2). Under this standardization, the origin represents the source-weighted mean.

Source interictal embeddings are then used to define the fixed background reference and coordinate dispersion:

$$b_s = \sum_{u,c,i} \widetilde{\omega}^{-}_{uci} z(x_{uc(-)i}),$$
$$v_{s,j} = \sum_{u,c,i} \widetilde{\omega}^{-}_{uci} \left[z_j(x_{uc(-)i}) - b_{s,j} \right]^2, \tag{28}$$

where $\widetilde{\omega}^{-}_{uci} = 2\omega_{uc(-)i}$ renormalizes the class-balanced interictal weights to sum to one.

The source metric is

$$W_s = \mathrm{diag}(w_1, \ldots, w_d), \qquad w_j = \frac{2}{1 + v_{s,j}}, \qquad \|a\|^2_{W_s} = a^\top W_s a. \tag{29}$$

The weighting is a bounded inverse dispersion. It gives unit weight at unit variance, amplifies no coordinate by more than a factor of 2, and can be read as inverse dispersion after equal shrinkage toward unit variance.

C.2 Replication invariance of source statistics

Consider a fixed encoder, fixed embeddings, and fixed patient-segment-class assignments. Uniformly replicate every window in one retained stratum (u, c, y) by an integer factor k. Its size changes from n_{ucy} to kn_{ucy}. According to Equation (26), each replicated copy receives weight ω_{ucyi}/k. The k copies therefore preserve the original total weight and weighted feature contribution of the stratum. Applying the same argument to the first and second moments preserves m_s, r_s, the deterministic scale fallback in Equation (27), and the resulting standardized embeddings.

The renormalized interictal moments b_s and v_s are consequently unchanged, as is the source metric W_s. The released implementation additionally removes repeated physical waveform identifiers before computing these source statistics. The common factor 2 in Equation (29) scales all ChronoProto scores by the same positive constant and therefore leaves within-patient ranking unchanged.

C.3 Preprocessing, training, and comparator configuration

We use the bipolar FP1–F7 derivation sampled at 256 Hz. Nonoverlapping 60-s intervals contain two 30-s views. Preprocessing applies a third-order Butterworth bandpass from 0.1 to 100 Hz, a 50–60 Hz bandstop, and standardization within each view. The short-time Fourier transform uses a 256-sample window and a 38-sample hop. The resulting spectra are mapped to 30 log-Mel bands spanning 0.1–70 Hz.

The single-channel source CNN contains four convolutional blocks with 64 $3\times3\times3$ kernels, rectified linear units, dropout of 0.5, and three $1 \times 2 \times 2$ max-pooling operations. A 150-unit rectified hidden layer precedes the two-class output. The native source score is the preictal-minus-interictal logit margin averaged across the two 30-s views. ChronoProto instead averages the corresponding hidden activations and applies the frozen source transform defined in Section C.1.

For source-only logistic regression and linear SVM, $C \in \{0.01, 0.1, 1, 10\}$ and the score sign are selected through four source cross-fitting blocks. Each validation patient is excluded from both encoder and classifier fitting, and its later events are used to evaluate candidate configurations. Candidate AP, AUROC, and n-pAUC are first averaged by event and then by patient. The selector prioritizes n-pAUC while requiring nondecreasing AP, AUROC noninferiority, and a paired one-standard-error margin. The selected configuration is then refit using all outer-source patients and applied once to the held-out target patient.

Target-positive-only logistic regression, RBF SVM, and shrinkage LDA use the first target preictal episode as positive support and source interictal embeddings as negative examples. All hyperparameters are selected using source patients. CP-BG and CP-0 use the same encoder, source transform, metric, and first-episode support. Their only difference is $r = b_s$ for CP-BG and $r = 0$ for CP-0.

C.4 Evaluation and uncertainty details

Events are ordered by annotated seizure onset. Eligibility requires a first usable event for personalization and at least one later evaluable segment. The 13 eligible primary-cohort patients contribute 328 first-episode preictal windows used for personalization, 617 later preictal windows used for evaluation, and 11,173 later interictal windows used for evaluation. The preictal interval spans 33 to 3 minutes before seizure onset. Labels from later windows are used only during evaluation.

Chance references for AP, AUROC, and n-pAUC are prevalence, 0.5, and $\alpha/2$, respectively, with $\alpha = 0.05$. The AP reference of 0.1809 in Table 1 is the event-then-patient mean prevalence, following the same equal-patient aggregation used for the reported model performance.

Primary paired comparisons use 100,000 whole-patient bootstrap resamples with percentile 95% intervals and seed 71071. Component analyses use 20,000 whole-patient resamples with seed 70001 and report unadjusted diagnostic intervals. All later events from a resampled patient remain grouped together.

For the patient-influence diagnostic, the patient with the largest endpoint-specific benefit is removed, and the paired mean difference is recomputed over the remaining 12 patients. The diagnostic shows how much high-influence patients contribute to the cohort mean. Absolute differences at most 10^{-12} are counted as ties. Paired differences are computed before cohort means are rounded for display.

D Additional Empirical Analyses

D.1 Source-only selection and component sensitivity

Table 4 compares native source scores with scores obtained after joint source-only selection of orientation and regularization.

Table 4: Source-only score selection diagnostic.

Classifier	Score	AP	AUROC	n-pAUC
Logistic	Native	0.2056	0.3953	0.0773
Logistic	Source-selected	0.3119	0.5799	0.1623
Linear SVM	Native	0.2056	0.3977	0.0788
Linear SVM	Source-selected	0.3175	0.5843	0.1640

Joint source-only selection changes all three ranking metrics substantially, and the selected models serve as the source-only controls in the main comparison.

Table 5 reports the absolute performance of the target-positive-only component variants. Each perturbation starts from CP-BG and changes one component at a time.

Table 5: Target-positive-only component analysis.

Variant	AP	AUROC	n-pAUC
CP-BG	0.2656	0.5601	0.1367
CP-0	0.3191	0.5871	0.1861
Geometric median	0.2906	0.5629	0.1676
Coordinate median	0.3011	0.5681	0.1784
Four distributed windows	0.2577	0.5571	0.1258
Eight distributed windows	0.2345	0.5254	0.1133
No source standardization	0.2829	0.5823	0.1483
Isotropic metric	0.2707	0.5733	0.1358

The median prototypes, support-coverage variants, and source-geometry variants each probe a different kind of sensitivity in ChronoProto. Both median variants give positive full-cohort AP differences relative to CP-BG, but the patient-influence analyses show that part of these changes is concentrated among high-impact patients. The reduced-support variants show that the behavior of

the first episode depends on how the support is spread across it, which is consistent with the temporal effective-support analysis in Section 4.3.

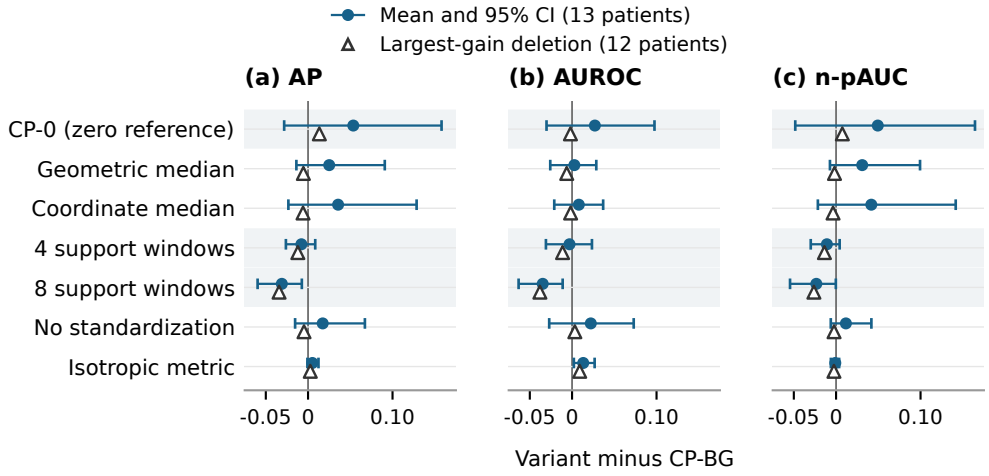

Figure 4: Component sensitivity relative to CP-BG on 13 patients. Circles and horizontal bars show paired means and 95% whole-patient bootstrap intervals based on 20,000 resamples with seed 70001. Triangles show the paired mean after removing the endpoint-specific largest-benefit patient. One component is changed per row.

D.2 Transfer analysis

Table 6 reports supplementary transfer results on additional patient inventories. Applying the same chronological eligibility rule leaves two eligible CHB–MIT patients with seven later events and three eligible Siena patients with five later events. The source geometry is fixed from the primary-cohort construction, and CP-BG is compared with the native Source CNN. Because only aggregate transfer results are retained, this section reports descriptive cohort means.

Table 6: Descriptive transfer results.

Cohort	Score	AP	AUROC	n-pAUC
CHB–MIT ($n = 2$)	Native Source CNN	0.0890	0.4142	0.0620
CHB–MIT ($n = 2$)	CP-BG	0.3477	0.7871	0.3564
Siena ($n = 3$)	Native Source CNN	0.4868	0.6900	0.3172
Siena ($n = 3$)	CP-BG	0.5099	0.7181	0.3325

On the held-out CHB–MIT patients, CP-BG increases AP from 0.0890 to 0.3477 and AUROC from 0.4142 to 0.7871. On Siena, AP increases from 0.4868 to 0.5099 and AUROC from 0.6900 to 0.7181. With two and three patients, these numbers give only a descriptive view of the same ChronoProto construction on additional patient groups.

D.3 Richer target supervision

Table 7 examines a richer target-information setting in which both classes from the first target segment are available. This analysis includes nine primary-cohort patients and 18 later evaluation events. CP-BG continues to use only the positive support on the same patient subset, while target logistic regression and the full-network update use both first-segment classes.

Table 7: Richer-supervision sensitivity on nine patients and 18 later events.

Method	AP	AUROC	n-pAUC
CP-BG (target-positive-only)	0.2711	0.5361	0.1564
Target logistic, $C = 0.01$	0.3028	0.5824	0.1814
Target logistic, $C = 0.1$	0.3377	0.6088	0.2109
Target logistic, $C = 1$	0.3835	0.6335	0.2579
Target logistic, $C = 10$	0.4042	0.6607	0.2657
Full-network update	0.2045	0.4188	0.0339

As target-background labels become available, logistic regression moves from 0.3028 AP at $C = 0.01$ to 0.4042 AP at $C = 10$, compared with 0.2711 AP for positive-only CP-BG on the same patient subset. The sweep thus contrasts the positive-only protocol with a richer two-class personalization regime on the same patients.