PHYSIOSPLIT: DELAY-AWARE CONTEXT-SELECTIVE FUSION UNDER COHORT SHIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

ECG and PPG provide complementary views of cardiovascular dynamics, but their timing, beat-level events, and slowly varying subject context can shift differently across patients, sensors, and acquisition protocols. Treating these factors as one undifferentiated representation makes it difficult to preserve transferable dynamics while adapting to a new subject. We introduce PhysioSplit, a delay-aware fusion framework that estimates a bounded fractional ECG–PPG arrival-delay proxy with a continuous-delay state-space encoder and factorizes the resulting representation into context and event codes. Context encoder adaptation updates only the context pathway during target calibration, while the event pathway and shared timing representation remain fixed. Across five cohorts, and over the three waveform-transfer settings, PhysioSplit reduces the external waveform degradation ratio from 2.23 for SiamQuality to 1.55 and lowers RMSE on the most challenging waveform setting from 13.21 to 6.74 mmHg. It achieves Macro-F1 values of 0.847 on the source cohort and 0.724 on the zero-shot ICD-derived phenotype task. Progressive ablations, delay-agreement diagnostics, temporal reliability measures, and frozen probes connect these gains to sub-sample alignment and structured information routing. The resulting model provides a compact interface for frozen transfer and small-support context calibration without treating learned codes as direct vascular measurements.

1 INTRODUCTION

Physiological recognition increasingly combines ECG and PPG. ECG provides an electrical timing reference, whereas PPG records a peripheral waveform after a subject- and sensor-dependent delay. Their joint morphology supports arterial-pressure reconstruction and phenotype recognition, but sensor placement, calibration, reference devices, and evaluation populations remain part of the measurement problem (Sastimoglu et al., 2025; Stergiou et al., 2022; Kachuee et al., 2015; Wu et al., 2022a). Patient-level resources such as VitalDB and PulseDB now support cross-dataset evaluation and multimodal physiological modeling (Lee et al., 2022; Wang et al., 2023; John et al., 2025).

Recent ECG–PPG estimators improve waveform fidelity with PPG-to-ABP translation, timing cues, reconstruction objectives, and quality-aware encoders (Ibtehaz et al., 2022; Long & Wang, 2023; Zhang et al., 2024; Ding et al., 2024). Their performance can still deteriorate when the reference device, sensor placement, or patient population changes. Most transfer methods expose one fused representation to both prediction and adaptation, without specifying which information should remain stable (Moulaeifard et al., 2025; Kim & Baek, 2026).

This is a structural gap. ECG–PPG arrival delay is an alignment variable, beat events change rapidly, and vascular or sensor context changes more slowly. Domain-alignment and robust-optimization methods generally transform a shared feature space, while continuous-time and state-space models capture long dependencies without defining the deployment adaptation subspace (Ganin et al., 2016; Sun & Saenko, 2016; Arjovsky et al., 2019; Zhou et al., 2022; Lin et al., 2022; Gulrajani & Lopez-Paz, 2020; Chen et al., 2018b; Gu et al., 2021; Gu & Dao, 2023).

We propose PhysioSplit, which combines a continuous-delay state-space encoder (CDSE), a factorized context/event representation, and context encoder adaptation (CEA). CDSE estimates a bounded ECG-to-PPG arrival-delay proxy and applies fractional interpolation before the shared state update. The encoder routes pooled, slowly varying information to a context code and sequence-level beat information to an event code; CEA updates only the context branch during target calibration, leaving the event encoder, timing representation, decoder dynamics, and task heads fixed. These codes are computational variables, not direct vascular measurements (Chen et al., 2018a; Li

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.

& Mandt, 2018; Yao et al., 2022). We evaluate patient-disjoint VitalDB training, frozen transfer to MIMIC-III and UCI, 32-fold leave-one-subject-out wearable transfer, and zero-shot phenotype transfer to MC-MED. Where calibration is available, we use four support and four query windows and keep ABP-derived and ICD-derived phenotype readouts separate. Waveform error, delay agreement, phenotype recognition, temporal reliability, and frozen semantic probes test the proposed routing and adaptation choices. The contributions are:

- CDSE provides delay-aware multimodal encoding with sub-sample ECG–PPG alignment and linear sequence complexity;
- The factorized context/event representation exposes an explicit context-only calibration interface while retaining a fixed event pathway;
- The five-cohort protocol separates waveform transfer, phenotype transfer, and subject-level calibration under fixed support/query budgets; and
- Mechanistic analyses connect external performance to fractional alignment, temporal reliability, decoder routing, and code-specific probes.

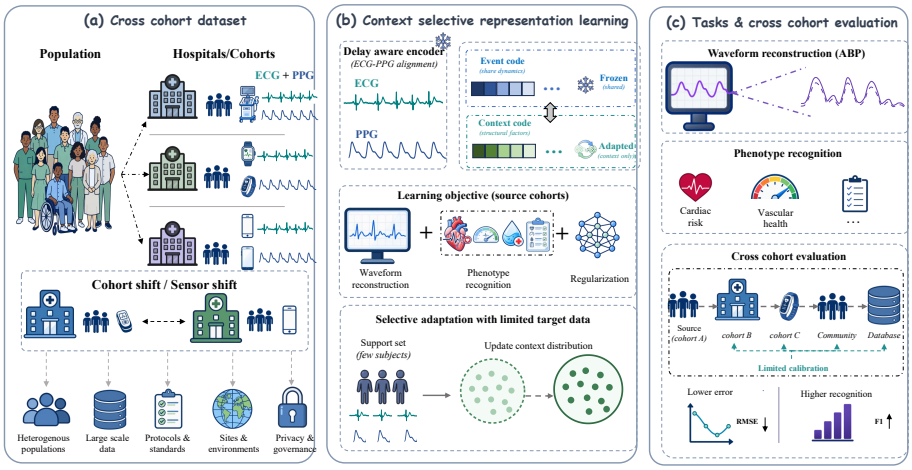


Figure 1: Study overview. (a) Cohort and sensor sources of shift; (b) delay-aware fusion, context/event factorization, and context-only calibration; (c) signal- and subject-level evaluation.

2 Related Work

Cross-cohort physiological transfer. Cross-dataset physiological estimation has exposed a persistent gap between in-distribution accuracy and external reliability. Transferable forecasting under causal conditional shift provides a related perspective: predictive structure can remain useful while the observed cohort changes (Li et al., 2024). PPG-to-ABP translation provides a strong waveform baseline, while ECG–PPG fusion adds electrical timing and complementary morphology (Ibtehaz et al., 2022; Long & Wang, 2023). Reconstruction networks and quality-aware encoders address feature fidelity and decoder stability (Zhang et al., 2024; Ding et al., 2024). Multi-level fusion has also been used to combine heterogeneous sensor summaries for pressure prediction, and related wearable work shows that fusion choices affect transfer across sensing conditions (Simjanoska et al., 2020; John et al., 2025). Public patient-level resources such as VitalDB and PulseDB make these comparisons possible, but the resulting benchmarks still differ in reference device, label construction, and cohort composition (Lee et al., 2022; Wang et al., 2023). Recent generalization and transfer studies quantify this external degradation (Moulaeifard et al., 2025; Kim & Baek, 2026); they motivate our separation of waveform transfer, phenotype transfer, and target calibration rather than a pooled score.

Domain alignment and personalization. Domain-adversarial alignment and covariance alignment reduce marginal mismatch by transforming a shared feature space (Ganin et al., 2016; Sun & Saenko, 2016). Invariant-risk formulations and Bayesian variants instead seek predictors that remain stable across source environments (Arjovsky et al., 2019; Lin et al., 2022). Distributionally robust optimization and broader domain-generalization studies emphasize worst-group behavior and source-only training protocols (Sagawa et al., 2019; Gulrajani & Lopez-Paz, 2020; Zhou et al., 2022). Meta-

learning supplies a complementary route to rapid subject adaptation, while clinical personalization studies show the value of small target support sets (Finn et al., 2017; Liu et al., 2023; Zhu et al., 2022). Transfer-learning evaluations in physiological sensing further demonstrate that a global update can improve one cohort while degrading another (Kim & Baek, 2026). These methods generally do not specify which information should remain stable during deployment. In contrast, PhysioSplit defines an explicit context branch that may change and evaluates the retained event branch through waveform, recognition, and representation-level readouts.

Continuous-time fusion and factorized representations. Neural ordinary differential equations and structured state-space models provide efficient ways to represent long temporal dependencies (Chen et al., 2018b; Gu et al., 2021). Selective state-space layers and ECG-oriented state-space encoders extend this line to high-rate physiological sequences (Gu & Dao, 2023; Qiang et al., 2024). Transformer and time-series image formulations offer alternative long-context constructions, including patch-based, two-dimensional, and irregular-sampling views (Nie et al., 2022; Wu et al., 2022b; Li et al., 2023). On the representation side, beta-VAE and sequential disentanglement methods use factorized objectives to expose statistically distinct sources of variation (Chen et al., 2018a; Burgess et al., 2018; Li & Mandt, 2018; Yao et al., 2022). The identifiability literature cautions that unsupervised factors are not automatically physiological and motivates auxiliary-variable or causal interpretations (Locatello et al., 2019; Lee, 1998; Schölkopf et al., 2021; Hyvarinen et al., 2019). Our contribution is to use timing-aware fusion and factorized routing as a measurable adaptation interface, rather than treating continuous-time modeling or latent separation as an end in itself.

3 METHOD

3.1 PROBLEM FORMULATION AND FUSION ARCHITECTURE

For a synchronized ECG–PPG window $\mathbf{X} \in \mathbb{R}^{2\times L}$, Track 1 reconstructs an ABP waveform $\hat{\mathbf{Y}}$ and Track 2 predicts a phenotype distribution. We model cohort variation through a slowly varying context code while treating short-time event dynamics as shared:

$$d\mathbf{s}(t) = \mathscr{F}(\mathbf{s}(t), \mathbf{z}_{\text{ctx}}^{(i)})dt + \mathscr{G}(\mathbf{s}(t))d\mathbf{W}(t), \tag{1}$$

where $\mathbf{z}_{\text{ctx}}$ is learned rather than directly measured. CDSE first extracts continuous-time morphology from each modality and couples the streams through a bounded delay:

$$\frac{d\mathbf{h}(t)}{dt} = \mathbf{A}\mathbf{h}(t) + \mathbf{B}_E\mathbf{h}_E(t) + \mathbf{B}_P\mathbf{h}_P(t-\tau(t)), \qquad \tau_k = \tau_{\min} + (\tau_{\max}-\tau_{\min})\sigma(\mathbf{w}_\tau^\top \mathbf{h}_{E,k} + b_\tau). \tag{2}$$

The delay is an ECG-to-PPG arrival-delay proxy. We split it into an integer offset d_k and fractional remainder γ_k, and interpolate the PPG feature before the state update:

$$\tilde{\mathbf{p}}_k = (1-\gamma_k)\mathbf{h}_{k-d_k}^{(P)} + \gamma_k \mathbf{h}_{k-d_k-1}^{(P)}. \tag{3}$$

The resulting recurrence is linear in window length. The estimator and its interpolation/hold bound are given in Appendix B.

More explicitly, each modality is first transformed by a spectrally normalized continuous-time front end,

$$\frac{d\mathbf{h}^{(m)}(t)}{dt} = f_{\theta_{\text{ode}}^{(m)}}(\mathbf{h}^{(m)}(t), t), \qquad m \in \{E, P\}, \tag{4}$$

where the terminal states provide local ECG and PPG morphology. The gate in Eq. equation 2 is predicted from the ECG state and is constrained to the 100–400 ms arrival-delay range used by the experiments. At sample interval δ_t, we write $d_k = \lfloor \tau_k/\delta_t \rfloor$ and $\gamma_k = (\tau_k - d_k\delta_t)/\delta_t$. The delayed PPG feature in Eq. equation 3 therefore retains its position inside one sample interval instead of rounding the delay to the nearest recurrent step. For a twice continuously differentiable feature trajectory with $\sup_t \|\mathbf{g}''(t)\|_2 \le G_2$, the interpolation component satisfies

$$\|\tilde{\mathbf{p}}_k - \mathbf{g}(t_k - \tau_k)\|_2 \le G_2\delta_t^2/8. \tag{5}$$

This bound concerns interpolation at the supplied delay and does not equate the learned gate with a direct vascular measurement. The delayed feature is then propagated by the diagonal zero-order-hold state update

$$\mathbf{h}_k = \bar{\mathbf{a}}_k \odot \mathbf{h}_{k-1} + \bar{\mathbf{b}}_{E,k}\mathbf{h}_k^{(E)} + \bar{\mathbf{b}}_{P,k}\tilde{\mathbf{p}}_k, \tag{6}$$

where $\bar{\mathbf{a}}_k = \exp(\boldsymbol{\Delta}_k \odot \mathbf{a})$ and $\bar{\mathbf{b}}_{\cdot,k}$ are the ZOH coefficients. The diagonal structure avoids dense matrix exponentials and preserves linear complexity in L; the complete buffered implementation is given in Appendix D.

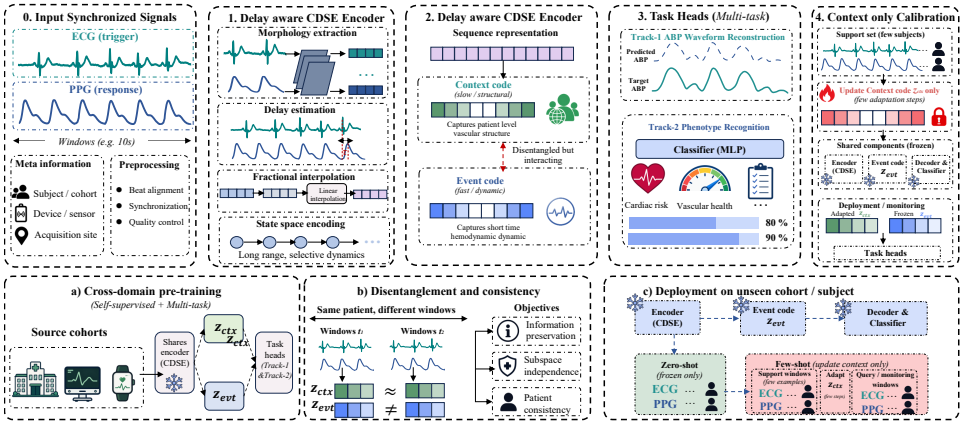


Figure 2: PhysioSplit architecture. CDSE aligns ECG and PPG; the factorized encoder produces context and event codes; AdaIN conditions the waveform decoder on context; CEA updates only the context branch for a new subject.

3.2 Context/Event Representation and Restricted Calibration

The factorized posterior is

$$q(\mathbf{z}_{\text{ctx}}, t \mid \mathbf{H}) = q_{\phi_{\text{ctx}}}(\mathbf{z}_{\text{ctx}} \mid \bar{\mathbf{H}}) q_{\phi_{\text{evt}}}(t \mid \mathbf{H}), \qquad \bar{\mathbf{H}} = L^{-1} \sum_k \mathbf{H}_k. \tag{7}$$

Temporal pooling restricts the context input, while the event branch retains the sequence needed for beat-level reconstruction. Context conditions decoder features through asymmetric AdaIN,

$$\text{AdaIN}(\mathbf{f}^{(\ell)}, \mathbf{z}_{\text{ctx}}) = \boldsymbol{\gamma}^{(\ell)}(\mathbf{z}_{\text{ctx}}) \odot \frac{\mathbf{f}^{(\ell)} - \mu_c(\mathbf{f}^{(\ell)})}{\sigma_c(\mathbf{f}^{(\ell)}) + \varepsilon} + \boldsymbol{\beta}^{(\ell)}(\mathbf{z}_{\text{ctx}}). \tag{8}$$

The training objective is

$$\mathscr{L}_{\text{total}} = \lambda_w \mathscr{L}_{\text{wave}} + \lambda_p \mathscr{L}_{\text{pheno}} + \lambda_d \mathscr{L}_{\text{dis}}, \tag{9}$$

where $\mathscr{L}_{\text{wave}}$ is the ABP Huber reconstruction loss and

$$\mathscr{L}_{\text{dis}} = \mathscr{L}_{\text{feat}} + \alpha \mathscr{L}_{\text{MI}} + \beta \mathscr{L}_{\text{TC}} + \gamma \mathscr{L}_{\text{MKL}} + \lambda_{\text{con}} \mathscr{L}_{\text{consist}}. \tag{10}$$

Here $\mathscr{L}_{\text{feat}}$ reconstructs the encoder feature summary and is distinct from the waveform term; the remaining terms control rate, cross-subspace dependence, marginal priors, and same-patient context stability. Full estimators are in Appendix B.

The factorization is motivated by a two-timescale pressure model rather than by a claim that either code is a measured physiological parameter. In a Windkessel-style decomposition,

$$P_{\text{art}}(t) = RQ(t) + C^{-1} \int_0^t Q(\xi)\, d\xi, \tag{11}$$

R and C vary more slowly than the beat-level flow $Q(t)$. We use this relationship to motivate the routing: the context encoder receives the temporal mean $\bar{\mathbf{H}}$, whereas the event encoder receives the full sequence. Temporal pooling suppresses rapid fluctuations in the context input, and the consistency term encourages nearby windows from one patient to produce similar context codes. The event path is not forced to be invariant; it keeps the variation needed to reconstruct individual beats.

The total-correlation term measures dependence between the two whole latent blocks,

$$\mathscr{L}_{\text{TC}} = D_{\text{KL}}\big(q(\mathbf{z}_{\text{ctx}}, \mathbf{z}_{\text{evt}}) \,\|\, q(\mathbf{z}_{\text{ctx}}) q(\mathbf{z}_{\text{evt}})\big). \tag{12}$$

The feature decoder reconstructs $\bar{\mathbf{H}}$ from both codes, so $\mathscr{L}_{\text{feat}}$ preserves information that may not be visible in the waveform loss. The mutual-information, total-correlation, and marginal-KL terms are the three terms in the standard ELBO decomposition over the two blocks; their minibatch-weighted estimator and sign convention are listed in Appendix B.4. This separation makes clear that waveform reconstruction is counted once, while feature reconstruction and representation regularization enter through $\mathscr{L}_{\text{dis}}$.

The two branches are parameterized as distributions rather than deterministic feature slices. The context posterior is obtained from the pooled state and the event posterior from the ordered state sequence; sampling during training therefore exposes the decoder to variation in both summaries. The phenotype head receives their concatenation, whereas the waveform decoder receives the event-conditioned feature map through context-dependent scale and shift. This asymmetry is important for calibration: changing ϕ_{ctx} changes the modulation parameters without changing the temporal event code or the shared delay representation.

The regularizers address different failure modes. The rate term limits uncontrolled information growth in the posterior; total correlation discourages information that is redundantly shared by both blocks; and the marginal KL terms keep each block near its prior. The same-patient consistency term is applied to paired windows and to the context branch only, so beat-level changes are not penalized as instability. Feature reconstruction provides the complementary information-preservation path: it asks the two codes jointly to retain the pooled state, while the waveform and phenotype terms determine which retained information is useful for the two tasks. The objective is thus a joint fusion objective rather than a post-hoc disentanglement penalty.

For a target subject, CEA freezes CDSE, the event encoder, decoder dynamics, and task heads, and updates only ϕ_{ctx} from a learned initializer. With support set S_i, query set Q_i, and M inner steps,

$$\mathscr{L}_{\text{calib}}^{(m)} = \mathscr{L}_{\text{wave}}(S_i; \phi_{\text{ctx}}^{(m-1)}) + \lambda_{\text{reg}} \|\phi_{\text{ctx}}^{(m-1)} - \phi_{\text{ctx}}^{*}\|_2^2, \quad \phi_{\text{ctx}}^{(m)} = \phi_{\text{ctx}}^{(m-1)} - \alpha_{\text{in}} \nabla \mathscr{L}_{\text{calib}}^{(m)}. \tag{13}$$

Zero-shot inference uses no target support and freezes all learned parameters. This distinction separates training-time episodic learning from test-time context calibration.

During meta-training, the query objective is evaluated after the support update and used to update the shared parameters and the context initializer,

$$\min_{\mathbf{\Phi}, \phi_{\text{ctx}}^{*}} \mathbb{E}_{\mathscr{T}_i} \left[\mathscr{L}_{\text{total}}(Q_i; \mathbf{\Phi}, \phi_{\text{ctx},i}^{(M)}) \right]. \tag{14}$$

Thus the support loss defines the deployment update, whereas the disjoint query loss determines whether that update improves subsequent windows. At test time, the same copy-and-update operation is applied only when target support is available; A–B remain frozen zero-shot evaluations. The complete training and streaming procedures are provided as pseudocode in Appendix D.

This protocol creates two distinct comparisons. Frozen external testing asks whether episodic training produces a representation that transfers without seeing a target window. RWW calibration asks whether a small target support set can update the designated branch and improve disjoint query windows. The support windows are never used for reporting, and the query objective is not used to choose the inner-step count. External waveform transfer and wearable calibration are therefore reported as separate evaluation regimes rather than averaged into one transfer score.

4 Experimental setup

VitalDB is the sole training source and is split by patient into train, validation, and held-out test partitions (6:2:2). All cohorts use quality-controlled, non-overlapping 10-s windows at 125 Hz. We organize the evaluation as a transfer ladder: Settings A–B are frozen tests on MIMIC-III and UCI, Setting C is 32-fold RWW leave-one-subject-out transfer with context-only calibration, and Setting D is zero-shot phenotype recognition on MC-MED. MC-MED uses the same semantic classes but derives labels from ICD-10 visits rather than ABP, so its scores are reported separately. CEA uses four support and four query windows with three inner updates. Stochastic results are mean$\pm$SD over three independently initialized runs; the release configuration records seed 42 as the default and a fixed three-seed reproduction protocol. Deterministic descriptive analyses are reported once. The cohort inventory, exclusion rules, pseudo-domain construction, and hyperparameters are given in Appendix E.

We report complementary evidence rather than a single aggregate score. The signal track uses waveform RMSE, MAE, Pearson correlation, and dynamic time-warping distance; the delay track compares the learned ECG-to-PPG arrival-delay proxy with the R-peak-to-PPG-foot reference using correlation, absolute error, within-patient ICC, and notch-timing error. Recognition is reported as setting-specific Macro-F1 for ABP-derived and ICD-derived labels, without pooling the two evidence families. Reliability uses ICC, within-patient ICC, LOSO transfer, and frozen probes for age, BMI, and SBP. RWW support windows are used only for the prescribed context update, while disjoint query windows are reserved for reporting.

Table 1: Cross-cohort results. Waveform RMSE (mean±SD over three runs) and setting-specific phenotype Macro-F1. Set C (with CEA) uses the prescribed RWW context calibration; external mean/ρ are derived from displayed means.

Waveform RMSE (mmHg), lower is better					
Method	Source	Set A	Set B	Set C	Ext. mean/ρ
SiamQuality	4.61 ± .09	8.03 ± .27	9.67 ± .34	13.21 ± .59	10.30/2.23
IRM	4.81 ± .11	7.93 ± .26	9.38 ± .33	12.64 ± .56	9.98/2.08
w/o CEA	4.29 ± .09	7.48 ± .22	8.97 ± .31	11.64 ± .52	9.36/2.18
PhysioSplit	**3.87 ± .06**	**5.43 ± .14**	**5.89 ± .17**	**6.74 ± .24**	**6.02/1.55**
Phenotype Macro-F1, setting-specific					
Variant	Source	Set A	Set B	Set C	Set D
No disentanglement	.610	.548	.573	.428	–
w/o CEA	.741	.684	.699	.581	–
PhysioSplit	**.847**	**.791**	**.763**	**.641**	**.724**

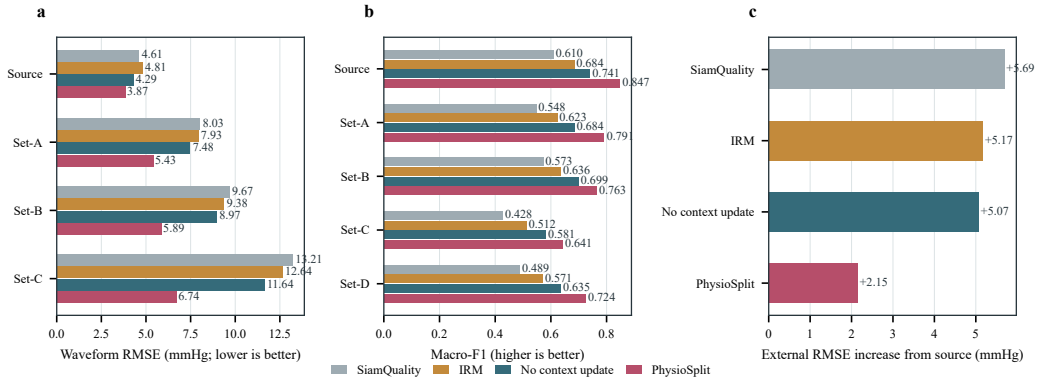


Figure 3: Cross-cohort benchmark. (a) Waveform RMSE; (b) setting-specific phenotype Macro-F1; (c) external RMSE increase from the source cohort.

5 Results

5.1 Cross-cohort transfer

Table 1 shows the central cross-cohort comparison. PhysioSplit reduces external mean waveform RMSE from 10.30 to 6.02 mmHg and the degradation ratio from 2.23 to 1.55 relative to SiamQuality. The source RMSE is 0.74 mmHg lower, while the reductions on Sets A–C are 2.60, 3.78, and 6.47 mmHg. Phenotype scores remain setting-specific because ABP-derived and ICD-derived labels use different construction mechanisms.

Figure 3 summarizes waveform error, phenotype transfer, and external degradation. Frozen A–B tests receive no target support, whereas RWW is a separate context update. The *w/o* CEA row isolates that update within PhysioSplit; external baselines receive no target update, so Set C is an end-to-end calibrated deployment evaluation rather than a pure source-only domain-generalization comparison. The fixed source head reaches Macro-F1 0.724 on the ICD-derived Setting D labels.

Two aspects of the transfer ladder are important for interpretation. First, the source-to-external increase is reported relative to the same source reference for every method, so the degradation ratio does not depend on a different normalization for each cohort. Second, the phenotype rows retain the semantic class names but keep ABP-derived and ICD-derived labels separate. Thus the external waveform comparison measures transfer of the signal task, whereas Setting D measures whether the fixed source head remains useful under an alternate label-construction process.

This separation also identifies the source of the apparent transfer gain. Sets A–B test whether the representation survives changed cohort and sensing conditions without target support, whereas Set C tests whether the same learned representation can be calibrated with a small, disjoint support set. The *w/o* CEA row compares those two Set-C conditions within the proposed model; it is therefore the relevant control for the value of context calibration. The table is not interpreted as evidence that an unreported global or event-branch update would perform worse.

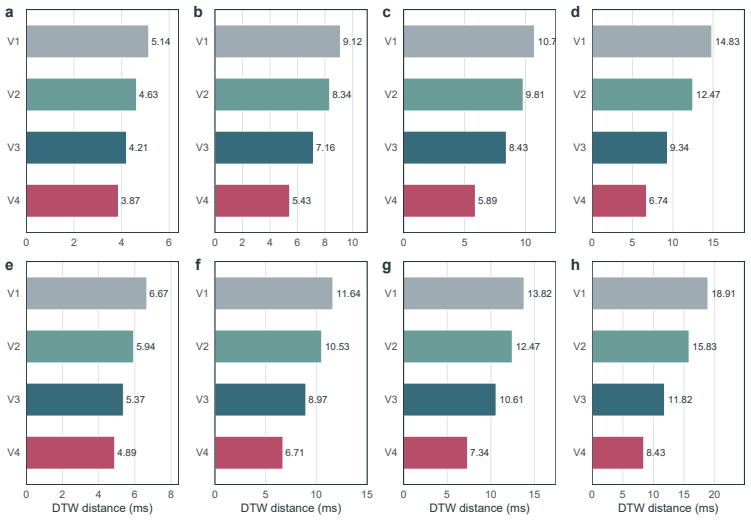


Figure 4: Progressive encoder ablation. (a–d) waveform RMSE and (e–h) DTW distance for the source cohort and Sets A–C.

5.2 Signal-level transfer and pressure agreement

The cross-cohort error is not uniform across waveforms. The representative residual atlas and cohort-level agreement plots are provided in Appendix S2–S3. They separate amplitude error from waveform-shape agreement: a high Pearson correlation can coexist with a large pressure error when the reconstructed waveform has a systematic offset. The atlas contrasts best, median, and high-error windows, whereas the cohort plots summarize bias and dispersion for SBP and DBP. These views make the error mode explicit and, together with the RMSE table, show that the improvement is not driven by a single favorable segment.

The signal-level diagnostics also determine which error mode is being reduced. RMSE captures absolute pressure deviation, correlation captures waveform shape, and DTW measures residual temporal displacement. Reading them together avoids interpreting a high correlation as accurate pressure calibration when a nearly constant offset remains. The same separation is used for the notch-timing diagnostic, which is sensitive to local morphology and is not absorbed by an aggregate waveform score.

5.3 Mechanistic evidence for delay-aware fusion

The cumulative encoder ablation keeps the decoder, representation objective, and calibration protocol fixed, so each row isolates an encoder addition. Continuous morphology, integer delay, and fractional delay reduce Set-C RMSE from 14.83 to 12.47, 9.34, and 6.74 mmHg, respectively; Set-C DTW falls from 18.91 to 8.43 ms and notch-timing error from 21.4 to 4.2 ms. The fractional-delay estimator tracks the ECG-to-PPG arrival-delay proxy with $r = 0.934$ on VitalDB and $r = 0.848$ on MIMIC-III; its MAE is 4.7 and 6.1 ms, respectively (Appendix B.2). Thus the ablation links the largest external improvement to sub-sample alignment rather than to a generic increase in decoder capacity.

The agreement diagnostics provide an independent check on this interpretation. The fractional-delay variant improves the learned-delay correlation and reduces the absolute delay error on both cohorts, while waveform and notch metrics improve at the same time. Alignment is therefore reflected both in the internal delay estimate and in reconstructed pressure morphology, making the ablation more informative than a change in RMSE alone.

The delay result is consistent with the intended fusion mechanism. The reference is the R-peak-to-PPG-foot interval averaged within a 10-s window, used as an arrival-delay proxy rather than a direct vascular measurement. Fixed-delay and attention-derived controls have lower correlation and larger absolute error, while the fractional model retains sub-sample variation without adding recurrent states. The agreement and error results consequently provide two checks on the timing module: one compares the learned gate with a signal landmark, and the other measures its downstream effect on reconstructed pressure. The full fixed-delay and attention-derived controls are reported in Appendix S5.

Table 2: Progressive CDSE ablation. The left block reports waveform error and degradation; the right block reports alignment and recognition diagnostics. Entries are mean±SD over three runs where applicable.

		Waveform RMSE and domain degradation					Alignment and recognition diagnostics						
V	Components added	Source	Set-A	Set-B	Set-C	$\rho\downarrow$	Source DTW	Set-A DTW	Set-B DTW	Set-C DTW	Δt_{notch} (ms)	F1	Sen_{Hypo}
V1	Discrete SSM	5.14 ± .15	9.12 ± .38	10.73 ± .47	14.83 ± .74	2.25	6.67 ± .21	11.64 ± .49	13.82 ± .61	18.91 ± .96	21.4 ± 3.7	.683	.594
V2	+ Neural ODE	4.63 ± .11	8.34 ± .31	9.81 ± .39	12.47 ± .61	2.19	5.94 ± .17	10.53 ± .41	12.47 ± .53	15.83 ± .78	16.8 ± 2.9	.721	.638
V3	+ Integer delay	4.21 ± .09	7.16 ± .24	8.43 ± .29	9.34 ± .43	1.97	5.37 ± .13	8.97 ± .31	10.61 ± .38	11.82 ± .54	10.3 ± 2.1	.791	.749
V4	**+ Fractional delay**	**3.87 ± .06**	**5.43 ± .14**	**5.89 ± .17**	**6.74 ± .24**	**1.55**	**4.89 ± .09**	**6.71 ± .18**	**7.34 ± .22**	**8.43 ± .31**	**4.2 ± .8**	**.847**	**.814**
V1→V4 change		−24.7%	−40.5%	−45.1%	−54.6%	−31.1%	−26.7%	−42.4%	−46.9%	−55.4%	−80.4%	+24.0%	+37.1%

Table 3: Delay plausibility across cohorts. Entries are mean±SD where applicable; "–" denotes an unavailable within-patient ICC.

	VitalDB (in-distribution)				MIMIC-III (OOD)			
Method	$r\uparrow$	MAE (ms)$\downarrow$	ICC$(2,1)\uparrow$	$\text{ICC}_{\text{wp}}\uparrow$	$r\uparrow$	MAE (ms)$\downarrow$	ICC$(2,1)\uparrow$	$\text{ICC}_{\text{wp}}\uparrow$
Heuristic arrival-delay baselines								
Fixed ($\tau = 250$ ms)†	0.000	47.3 ± 18.6	0.000	0.000 ± .000	0.000	51.2 ± 20.1	0.000	0.000 ± .000
PAT (R→PPG-Peak)	.741	18.4 ± 7.2	.683	0.694 ± .107	.718	20.1 ± 8.3	.661	0.673 ± .118
Adaptive PAT	.831	11.9 ± 4.8	.803	0.812 ± .083	.806	13.7 ± 5.4	.779	0.784 ± .094
Attention surrogates (specificity control)								
PatchTST $\bar{d}_{\text{attn}}$	.412	31.7 ± 12.4	.374	–	.389	34.2 ± 13.8	.351	–
CrossAttn $\bar{d}_{\text{attn}}$	.531	24.6 ± 9.7	.497	–	.507	27.1 ± 10.9	.471	–
CDSE ablation								
V3 (integer delay)	.871	8.3 ± 3.4	.849	0.843 ± .064	.848	9.6 ± 3.9	.824	0.831 ± .073
V4 (fractional delay)	**.934**	**4.7 ± 1.9**	**.918**	**0.921 ± .038**	**.908**	**6.1 ± 2.4**	**.893**	**0.903 ± .047**

5.4 Representation reliability and semantic selectivity

The complete representation reaches ICC$(2,1) = 0.812$, within-patient ICC $= 0.834$, and RWW LOSO ICC $= 0.634 \pm .041$. Removing total correlation reduces source ICC to 0.531; removing temporal consistency reduces it to 0.614. Frozen probes show that $\mathbf{z}_{\text{ctx}}$ predicts age and BMI with $R^2 = 0.531$ and 0.489, versus 0.094 and 0.078 from $\mathbf{z}_{\text{evt}}$; the zero-shot RWW SBP probe reaches AUROC 0.712 versus 0.597. These results support code-specific information routing without assigning a unique physiological meaning to individual latent coordinates. The reliability analysis tests both persistence and selectivity: ICC and within-patient ICC assess stability across nearby windows, whereas LOSO transfer tests whether that stability survives a new subject. Frozen probes provide a controlled readout of information content, not an identifiability claim about individual coordinates. Extended reliability, DCI/MIG, decoder, and application analyses are in Appendix F.

The phenotype-transfer view is complementary to reliability because it tests whether the same factorization remains useful when labels are derived from a different clinical process. Removing the delay module, context/event factorization, or CEA produces larger reductions on the more distant settings, while Setting D remains a separate ICD-derived readout. The complete setting-specific heat maps are provided in Appendix S4; the main table reports the corresponding Macro-F1 values without pooling label-construction mechanisms. The reliability measures provide a second constraint on this interpretation: ICC and within-patient ICC quantify persistence across nearby windows, LOSO evaluates transfer to an unseen subject, and frozen probes test code selectivity. Their agreement links the observed phenotype transfer to the same routing structure rather than to a single metric, while the separation of label mechanisms prevents reliability evidence from being conflated with generalization across clinical labels.

5.5 Decoder conditioning and information routing

The factorized representation does not imply that the two codes are independent of the prediction head. Decoder conditioning determines how the branches interact after encoding. Among concatenation, cross-attention, FiLM, and AdaIN, the asymmetric AdaIN decoder gives the lowest residual Set-C error in the controlled comparison (Appendix F). Its channel-wise scale and shift allow context to modulate waveform envelope and reflection morphology while event features carry beat-level content. This routing provides a concrete computational explanation for why context calibration can change pressure shape without updating the event branch.

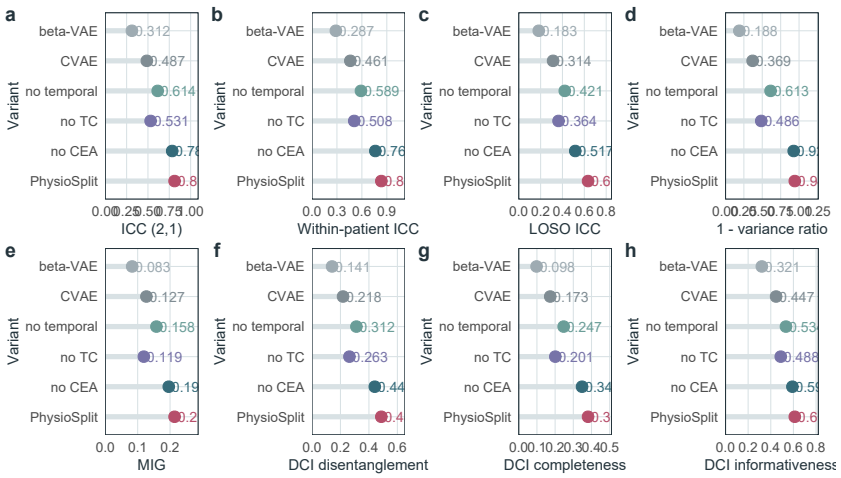


Figure 5: Representation reliability and disentanglement across variants.

Table 4: Temporal invariance and disentanglement of context and event codes. F1–F3 are prespecified reliability criteria; LOSO is 32-fold RWW leave-one-subject-out evaluation.

Method	ICC$(2,1)$ (95% CI)	F1	F2	ICC_{wp}	F3	ρ_{var}	LOSO ICC (RWW)	MIG	DCI-$\mathscr{D}$	DCI-$\mathscr{C}$	DCI-$\mathscr{I}$
β-VAE (w/o TC)	.312 [.271,.358]	×	×	.287	×	.812	.183± .031	.083	.141	.098	.321
Standard CVAE	.487 [.443,.531]	×	×	.461	×	.631	.314± .038	.127	.218	.173	.447
w/o temporal loss	.614 [.578,.649]	×	×	.589	×	.387	.421± .043	.158	.312	.247	.534
w/o total correlation	.531 [.496,.568]	×	×	.508	×	.514	.364± .041	.119	.263	.201	.488
w/o MAML	.784 [.751,.814]	✓	✓	.762	×	.073	.517± .046	.196	.441	.348	.591
PhysioSplit (ours)	**.812 [.781,.843]**	✓	✓	**.834**	✓	**.058**	**.634 ± .041**	**.214**	**.487**	**.381**	**.612**

The complete variant and layer-wise comparison is provided in Appendix S6. The layer-wise pattern is consistent with the intended division of labor: context-sensitive gains appear in the decoder stages that control waveform envelope and morphology, whereas the event pathway remains tied to beat-level structure. This analysis therefore interprets the routing mechanism and its calibration interface, rather than introducing another benchmark.

6 Discussion

Across the evaluated settings, alignment, preservation, and adaptation play distinct roles. CDSE corrects ECG–PPG timing before fusion, factorization separates persistent context from event dynamics, and CEA updates only the context branch. Progressive ablations and delay-agreement diagnostics link transfer gains to delay modeling, while temporal consistency, total correlation, reliability, and frozen probes support the intended routing. At the decoder, asymmetric AdaIN lets context updates modulate waveform envelope and morphology while event features preserve beat-level structure, clarifying why calibration need not rewrite the shared temporal representation. The evidence supports selective adaptation under the tested protocol, within the evaluated sensor and acquisition range, rather than complete removal of cohort shift or a direct vascular interpretation of the learned delay and latent codes.

7 Conclusion

PhysioSplit addresses cross-cohort ECG–PPG transfer by separating arrival delay, persistent context, and beat-level event dynamics. CDSE corrects fractional timing before fusion, while context/event factorization exposes only the context pathway to calibration and keeps the event pathway and shared timing fixed. The method improves external waveform transfer while preserving phenotype and representation evidence; ablations, delay agreement, reliability, and frozen probes connect these gains to alignment and structured information routing. Deployment supports frozen zero-shot transfer without target data and small-support context updates when calibration is available. These conclusions are bounded by the evaluated sensors and acquisition conditions, and the learned delay and latent codes remain computational representations rather than direct vascular measurements.

AI USE STATEMENT

Generative AI tools were used for language polishing of the manuscript. The authors reviewed the assisted text, verified the data processing and analyses, and take responsibility for the final manuscript.

ETHICS STATEMENT

Public datasets were accessed under their repository terms. The in-house cohort was collected with institutional ethics approval and written informed consent; all records were coded before analysis. The study evaluates PhysioSplit as a representation-learning method and does not claim clinical validation.

REPRODUCIBILITY STATEMENT

Section 4 and Supplementary Sections A–F describe the data, training procedure, evaluation protocol, and additional results. The anonymized implementation and configuration files are available at `https://anonymous.4open.science/r/PhysioSplit/`.

REFERENCES

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. Understanding disentangling in β-VAE. *arXiv preprint arXiv:1804.03599*, 2018.

Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. In *Advances in Neural Information Processing Systems*, volume 31, 2018a.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018b.

Cheng Ding, Zhicheng Guo, Zhaoliang Chen, Randall J Lee, Cynthia Rudin, and Xiao Hu. Siamquality: a convnet-based foundation model for photoplethysmography signals. *Physiological Measurement*, 45(8):085004, 2024.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. *arXiv preprint arXiv:2007.01434*, 2020.

Aapo Hyvarinen, Hiroaki Sasaki, and Richard Turner. Nonlinear ICA using auxiliary variables and generalized contrastive learning. In *The 22nd international conference on artificial intelligence and statistics*, pp. 859–868. Pmlr, 2019.

Nabil Ibtehaz, Sakib Mahmud, Muhammad EH Chowdhury, Amith Khandakar, Muhammad Salman Khan, Mohamed Arselene Ayari, Anas M Tahir, and M Sohel Rahman. PPG2ABP: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms. *Bioengineering*, 9(11):692, 2022.

Arlene John, He Wang, Barry Cardiff, Keshab K. Parhi, and Deepu John. Multimodal fusion for robust respiratory rate estimation in wearable sensing. *Information Fusion*, 123:103253, 2025.

Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE international symposium on circuits and systems (ISCAS)*, pp. 1006–1009. IEEE, 2015.

Aman Kansal, Emma Chen, Tom Jin, Pranav Rajpurkar, and David Kim. Multimodal clinical monitoring in the emergency department (MC-MED). *PhysioNet*, 2025. Version 1.0.0.

Yoon Chul Kim and Hyun Jae Baek. Evaluating cross-dataset transfer learning for photoplethysmography-based blood pressure estimation. *Scientific Reports*, 2026.

Hyung-Chul Lee, Yoonsang Park, Soo Bin Yoon, Seong Mi Yang, Dongnyeok Park, and Chul-Woo Jung. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data*, 9(1):279, 2022.

Te-Won Lee. Independent component analysis. In *Independent component analysis: Theory and applications*, pp. 27–66. Springer, 1998.

Yingzhen Li and Stephan Mandt. Disentangled sequential autoencoder. *arXiv preprint arXiv:1803.02991*, 2018.

Zekun Li, Shiyang Li, and Xifeng Yan. Time series as images: Vision transformer for irregularly sampled time series. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49187–49204, 2023.

Zijian Li, Ruichu Cai, Tom ZJ Fu, Zhifeng Hao, and Kun Zhang. Transferable time-series forecasting under causal conditional shift. *IEEE transactions on pattern analysis and machine intelligence*, 46(4):1932–1949, 2024.

Yong Lin, Hanze Dong, Hao Wang, and Tong Zhang. Bayesian invariant risk minimization. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16000–16009. IEEE, 2022.

Zhenxing Liu, Yujie Chen, Yong Zhang, Shaolin Ran, Cheng Cheng, and Guili Yang. Diagnosis of arrhythmias with few abnormal ECG samples using metric-based meta learning. *Computers in biology and medicine*, 153:106465, 2023.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pp. 4114–4124. PMLR, 2019.

Weicai Long and Xingjun Wang. BPNet: A multi-modal fusion neural network for blood pressure estimation using ECG and PPG. *Biomedical Signal Processing and Control*, 86:105287, 2023.

Mohammad Moulaeifard, Peter H Charlton, and Nils Strodthoff. Generalizable deep learning for photoplethysmography-based blood pressure estimation–a benchmarking study. *Machine Learning: Health*, 1(1):010501, 2025.

Mohammad Moulaeifard, Peter H Charlton, and Nils Strodthoff. MIMIC-III-Ext-PPG: A PPG benchmark dataset for cardiorespiratory analysis. *PhysioNet*, 2026. Version 1.0.0.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Yupeng Qiang, Xunde Dong, Xiuling Liu, Yang Yang, Fei Hu, and Rongjia Wang. ECGMamba: Towards ECG classification with state space models. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 6498–6505. IEEE, 2024.

Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731*, 2019.

Ziya Sastimoglu, Sophini Subramaniam, Abu Ilius Faisal, Wei Jiang, Andrew Ye, and M Jamal Deen. Wearable PPG based BP estimation methods: a systematic review and meta-analysis. *IEEE Journal of Biomedical and Health Informatics*, 29(4):2439–2452, 2025.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Monika Simjanoska, Stefan Kochev, Jovan Tanevski, Ana Madevska Bogdanova, Gregor Papa, and Tome Eftimov. Multi-level information fusion for learning a blood pressure predictive model using sensor data. *Information Fusion*, 58:24–39, 2020.

George S Stergiou, Ramakrishna Mukkamala, Alberto Avolio, Konstantinos G Kyriakoulis, Stephan Mieke, Alan Murray, Gianfranco Parati, Aletta E Schutte, James E Sharman, Roland Asmar, et al. Cuffless blood pressure measuring devices: review and statement by the european society of hypertension working group on blood pressure monitoring and cardiovascular variability. *Journal of hypertension*, 40(8):1449–1460, 2022.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pp. 443–450. Springer, 2016.

Weinan Wang, Pedram Mohseni, Kevin L Kilgore, and Laleh Najafizadeh. PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health*, 4:1090854, 2023.

Bing-Jhang Wu, Bing-Fei Wu, and Chi-Po Hsu. Camera-based blood pressure estimation via windkessel model and waveform features. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2022a.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022b.

Weiran Yao, Guangyi Chen, and Kun Zhang. Temporally disentangled representation learning. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26492–26503, 2022.

Gengjia Zhang, Daegil Choi, and Jaehyo Jung. Reconstruction of arterial blood pressure waveforms based on squeeze-and-excitation network models using electrocardiography and photoplethysmography signals. *Knowledge-Based Systems*, 295:111798, 2024.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4396–4415, 2022.

Taiyu Zhu, Kezhi Li, Pau Herrero, and Pantelis Georgiou. Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Transactions on Biomedical Engineering*, 70(1):193–204, 2022.

Supplementary Material

A Additional Background and Notation

This section fixes the notation used in the derivations and lists abbreviations that recur in the main text and the supplementary analyses.

A.1 Abbreviations

Table S1: Abbreviations used in the manuscript and Appendix.

Term	Definition	Term	Definition
AAMI	Association for the Advancement of Medical Instrumentation	ABP	arterial blood pressure
AdaIN	adaptive instance normalization	AUROC	area under the receiver operating characteristic curve
CEA	context encoder adaptation	CDSE	continuous-delay state-space encoder
CNAP	continuous non-invasive arterial pressure	CVAE	conditional variational autoencoder
DBP	diastolic blood pressure	DCI	disentanglement, completeness, and informativeness
DG	domain generalization	DTW	dynamic time warping
ECG	electrocardiography	F-CVAE	factorized conditional variational autoencoder
FO-MAML	first-order model-agnostic meta-learning	ICC	intraclass correlation coefficient
LOSO	leave-one-subject-out	MAE	mean absolute error
MAML	model-agnostic meta-learning	MC-MED	Multimodal Clinical Monitoring in the Emergency Department
MIG	mutual information gap	ODE	ordinary differential equation
PAT	pulse arrival time	PPG	photoplethysmography
PTT	pulse transit time; not directly measured in this study	RMSE	root mean square error
RWW	real-world wearable cohort	SBP	systolic blood pressure
SQI	signal quality index	TC	total correlation
ZOH	zero-order hold		

B Mathematical Properties and Estimator Details

This section records the regularity, interpolation, and representation properties needed to interpret the estimator and its factorized objective.

B.1 Neural ODE Well-Posedness

Proof. Let $f_\theta(\mathbf{h},t) = \mathbf{W}_2\,\sigma(\mathbf{W}_1\mathbf{h}+\mathbf{b}_1)+\mathbf{b}_2$ with σ denoting the element-wise Swish function. For any $\mathbf{h},\mathbf{h}' \in \mathbb{R}^d$,

$$\begin{aligned}\|f_\theta(\mathbf{h},t) - f_\theta(\mathbf{h}',t)\|_2 &\le \|\mathbf{W}_2\|_2 L_\sigma \|\mathbf{W}_1\|_2 \|\mathbf{h}-\mathbf{h}'\|_2 \\ &\le \kappa^2 L_\sigma \|\mathbf{h}-\mathbf{h}'\|_2. \end{aligned} \tag{15}$$

Since f_θ is globally Lipschitz in $\mathbf{h}$ and continuous in t, Picard–Lindelöf guarantees existence and uniqueness. □

B.2 Fractional-Delay Interpolation and Hold Error

Lemma S1 (Fractional ZOH truncation error). Let $\mathbf{g}(t) = \mathbf{h}_{\text{loc}}^{(P)}(t)$ be twice continuously differentiable, with $G_2 = \sup_t \|\mathbf{g}''(t)\|_2$. At the sampled delay τ_k, the interpolant satisfies

$$\|\tilde{\mathbf{p}}_k - \mathbf{g}(t_k - \tau_k)\|_2 \le \frac{G_2}{8}\,\delta_t^2. \tag{16}$$

Proof. The interpolation remainder is bounded by $G_2\gamma_k(1-\gamma_k)\delta_t^2/2$. Since $0 \le \gamma_k \le 1$, the product $\gamma_k(1-\gamma_k)$ is at most $1/4$, giving the stated bound. □

If $\tau_k = \tau(t_k)$ is held over an adjacent interval and $|\dot{\tau}(t)| \le \varepsilon$, its displacement from $\tau(t)$ is at most $\varepsilon\delta_t$. The corresponding feature displacement is bounded by $\|\dot{\mathbf{g}}\|_\infty \varepsilon\delta_t$. Combining interpolation and

delay holding therefore gives an error bound of $G_2\delta_t^2/8 + \|\dot{\mathbf{g}}\|_\infty \varepsilon \delta_t$. The interpolation component is second order in the sampling interval; the combined bound is second order when $\varepsilon = \mathcal{O}(\delta_t)$. These bounds concern discretization at the supplied delay and do not bound delay-estimation error.

B.3 Conditional Separation Under Nonlinear-ICA Assumptions

Remark S2 (Sufficient-condition separation interpretation). Assume that (C1) the aggregate posterior factorizes across $\mathbf{z}_{\text{ctx}}$ and $\mathbf{z}_{\text{evt}}$, (C2) the context code is stable across paired windows from the same patient, and (C3) the decoder uses asymmetric AdaIN modulation from $\mathbf{z}_{\text{ctx}}$. These conditions motivate distinct functional roles for the two branches. Recovering individual nonlinear latent sources additionally requires identifiability assumptions on the generative model and auxiliary information. Independence and non-Gaussianity alone do not establish that recovery. The argument below concerns the architectural and temporal constraints on the partition.

Supporting argument. Under (C1), $\mathrm{TC}(q)=0$ implies that the aggregate posterior factorizes: $q(\mathbf{z}) = q(\mathbf{z}_{\text{ctx}})\, q(\mathbf{z}_{\text{evt}})$. This factorization constrains dependence between the blocks without assigning semantic meaning to their coordinates. (C2) reduces the between-subspace permutation ambiguity: changes that lower the cross-window cosine similarity of $\mathbf{z}_{\text{ctx}}$ incur a temporal-consistency penalty, biasing the solution toward slowly varying context representations. (C3) further reduces exchangeability between the subspaces: the AdaIN decoder assigns asymmetric generative roles $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ exclusively to $\mathbf{z}_{\text{ctx}}$, which cannot be exchanged with the content role of $\mathbf{z}_{\text{evt}}$ without altering the decoder's functional form. Transformations within either block may still be absorbed by the encoder and decoder. The constraints therefore motivate the context/event assignment; temporal reliability and frozen probes evaluate its empirical meaning. □

B.4 F-CVAE Objective and Minibatch-Weighted TC Estimator

The main manuscript uses TC as an independence constraint. We next give its full decomposition and the minibatch estimator used in implementation. Define the aggregate posterior

$$q(\mathbf{z}) = \frac{1}{N}\sum_{n=1}^{N} q_\phi(\mathbf{z} \mid \mathbf{H}^{(n)}), \qquad \mathbf{z} = (\mathbf{z}_{\text{ctx}}, \mathbf{z}_{\text{evt}}), \tag{17}$$

where the sum is over all training windows. The expected KL term in the ELBO can be decomposed as follows. Throughout this decomposition, $j \in \{\text{ctx}, \text{evt}\}$ indexes a whole latent block, and the prior factorizes over the two blocks. TC therefore measures between-block dependence, while the marginal KL terms act on the block distributions.

$$\begin{aligned}
&\mathbb{E}_{p_{\text{data}}(\mathbf{H})}\left[D_{\text{KL}}\left(q_\phi(\mathbf{z} \mid \mathbf{H}) \,\|\, p(\mathbf{z})\right)\right] \\
&\quad = \underbrace{I_q(\mathbf{H};\mathbf{z})}_{\mathscr{L}_{\text{MI}}} + \underbrace{D_{\text{KL}}\left(q(\mathbf{z}) \,\middle\|\, \prod_j q(\mathbf{z}_j)\right)}_{\mathscr{L}_{\text{TC}}} + \underbrace{\sum_j D_{\text{KL}}(q(\mathbf{z}_j) \,\|\, p(\mathbf{z}_j))}_{\mathscr{L}_{\text{MKL}}}.
\end{aligned} \tag{18}$$

The three terms respectively control code information, dependence between latent blocks, and marginal prior matching. The F-CVAE contribution is

$$\begin{aligned}
\mathscr{L}_{\text{dis}} = &\underbrace{-\mathbb{E}_q\left[\log p_{\omega_{\text{feat}}}\left(\tilde{\mathbf{H}} \mid \mathbf{z}_{\text{ctx}}, \mathbf{z}_{\text{evt}}\right)\right]}_{\mathscr{L}_{\text{feat}}} \\
&+ \alpha \underbrace{I_q(\mathbf{H};\mathbf{z}_{\text{ctx}},\mathbf{z}_{\text{evt}})}_{\mathscr{L}_{\text{MI}}} + \beta \underbrace{D_{\text{KL}}(q(\mathbf{z}_{\text{ctx}},\mathbf{z}_{\text{evt}}) \,\|\, q(\mathbf{z}_{\text{ctx}})q(\mathbf{z}_{\text{evt}}))}_{\mathscr{L}_{\text{TC}}} \\
&+ \gamma \underbrace{\sum_{j\in\{\text{ctx},\text{evt}\}} D_{\text{KL}}\left(q(\mathbf{z}_j) \,\|\, p(\mathbf{z}_j)\right)}_{\mathscr{L}_{\text{MKL}}} + \lambda_{\text{con}}\mathscr{L}_{\text{consist}}.
\end{aligned} \tag{19}$$

Here $p_{\omega_{\text{feat}}}$ reconstructs the pooled encoder feature $\tilde{\mathbf{H}}$ from the two latent codes. This term is distinct from $\mathscr{L}_{\text{wave}}$, which reconstructs the ABP waveform; hence the feature-space and waveform reconstruction enter the objective through separate terms. The aggregate marginals are intractable and

are approximated from a minibatch $\mathscr{B}$. For each subspace, let w_{il} denote the normalized sampling weight assigned to example l when evaluating sample i, with $\sum_{l\in\mathscr{B}} w_{il} = 1$. Then

$$\log \widehat{q}(\mathbf{z}_j^{(i)}) \approx \log \sum_{l\in\mathscr{B}} w_{il} q_{\phi_j}(\mathbf{z}_j^{(i)} \mid \mathbf{H}^{(l)}), \tag{20}$$

with the context encoder conditioned on $\bar{\mathbf{H}}^{(l)}$. The joint aggregate posterior is estimated using the factored posterior:

$$\log \widehat{q}(\mathbf{z}^{(i)}) \approx \log \sum_{l\in\mathscr{B}} w_{il} q_{\phi_{\text{ctx}}}(\mathbf{z}_{\text{ctx}}^{(i)} \mid \bar{\mathbf{H}}^{(l)}) q_{\phi_{\text{evt}}}(\mathbf{z}_{\text{evt}}^{(i)} \mid \mathbf{H}^{(l)}). \tag{21}$$

These minibatch estimates give the TC estimator

$$\widehat{\mathscr{L}_{\text{TC}}} = \frac{1}{|\mathscr{B}|} \sum_{i=1}^{|\mathscr{B}|} \left[\log \widehat{q}(\mathbf{z}^{(i)}) - \sum_{j\in\{\text{ctx,evt}\}} \log \widehat{q}(\mathbf{z}_j^{(i)}) \right]. \tag{22}$$

The weights are evaluated in log space and account for the dataset-to-minibatch sampling ratio. To limit dependence induced by within-patient temporal autocorrelation, each minibatch contains at most one segment per patient when patient identifiers are available.

Figure S1 compares the raw disentanglement metrics with their rank- and column-normalized counterparts, separating absolute metric values from consistency across criteria.

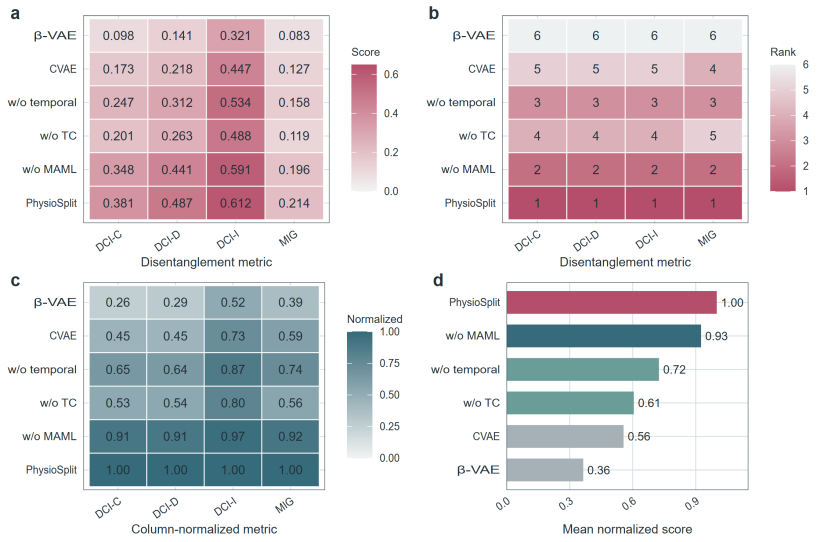


Figure S1: Disentanglement across generative baselines and ablations. (a) Raw DCI completeness, DCI disentanglement, DCI informativeness, and MIG; (b) within-metric ranks; (c) column-normalized scores; and (d) their mean. Higher normalized scores and lower ranks indicate more consistent separation of context and event information.

C Extended Main-Paper Figures

The following figures contain signal-, pressure-, phenotype-, delay-, and decoder-level views referenced from the main text. They are retained here to keep the main paper focused on the five central visual summaries.

D Supplementary Algorithms

The following algorithms make the implementation path explicit, from buffered fractional-delay fusion to loss evaluation and context-restricted inference.

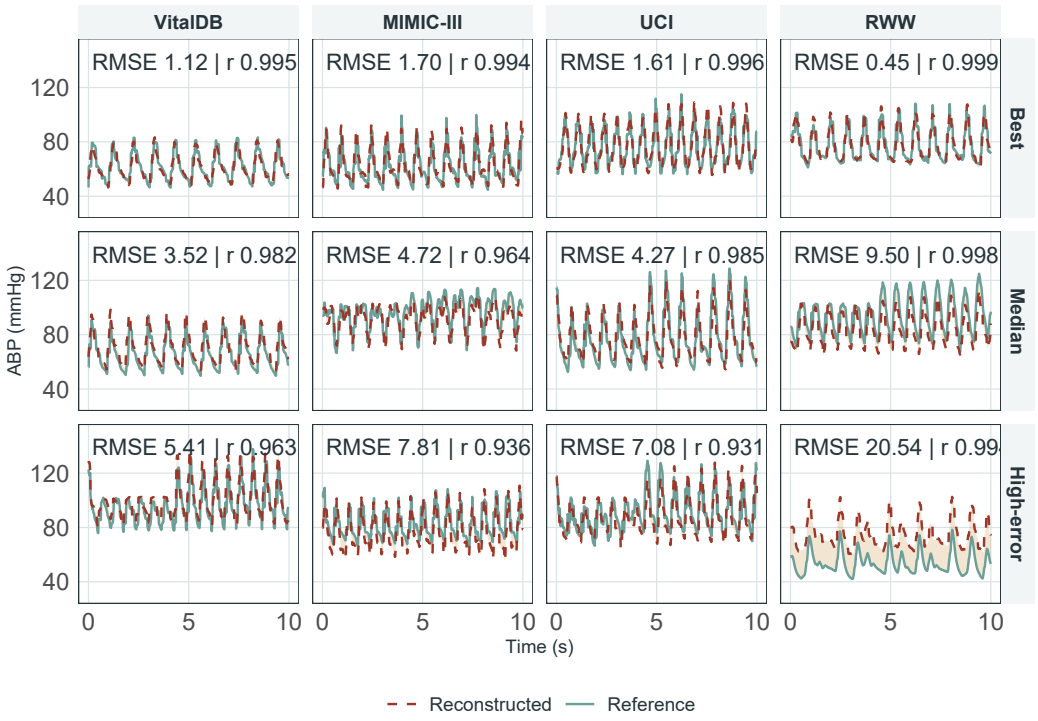


Figure S2: Representative waveform residuals across four cohorts. Rows show best, median, and high-error windows with RMSE and Pearson correlation.

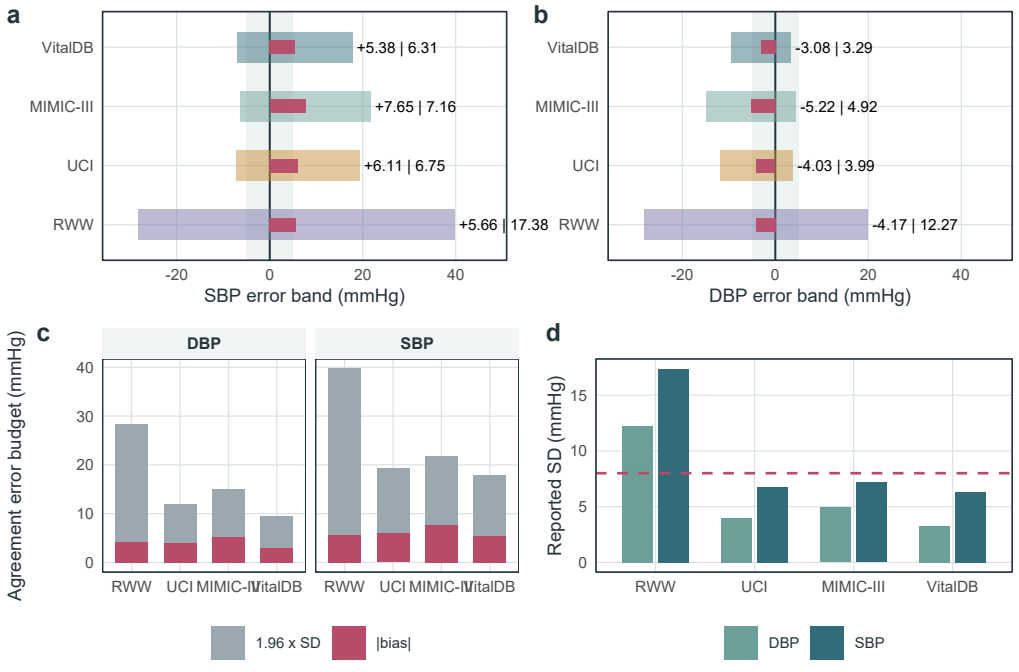


Figure S3: Cohort-level pressure agreement and error budgets. Panels report SBP/DBP bias, dispersion, and associated error decomposition.

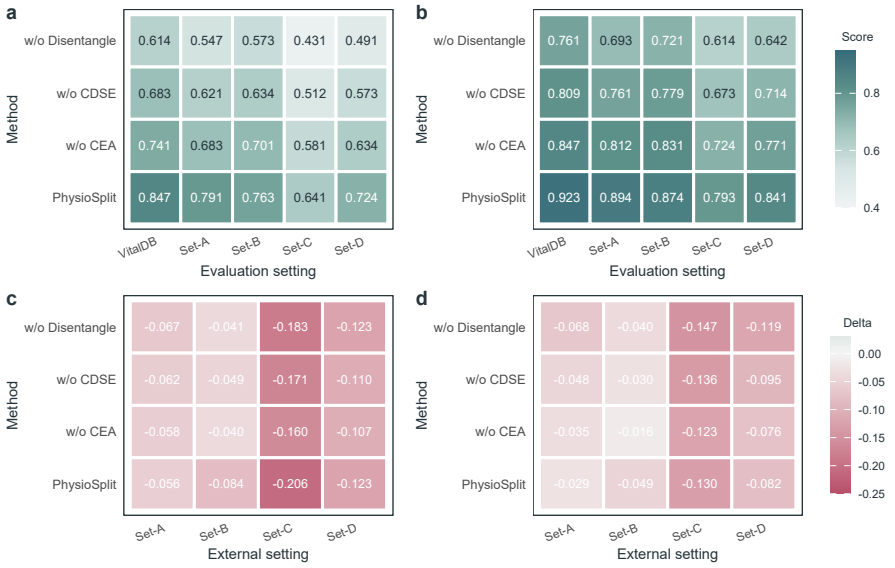


Figure S4: Phenotype transfer across source and external settings. (a,b) Macro-F1 and AUROC; (c,d) changes from the source cohort. Setting D uses shared semantic classes with an ICD-derived label-construction mechanism.

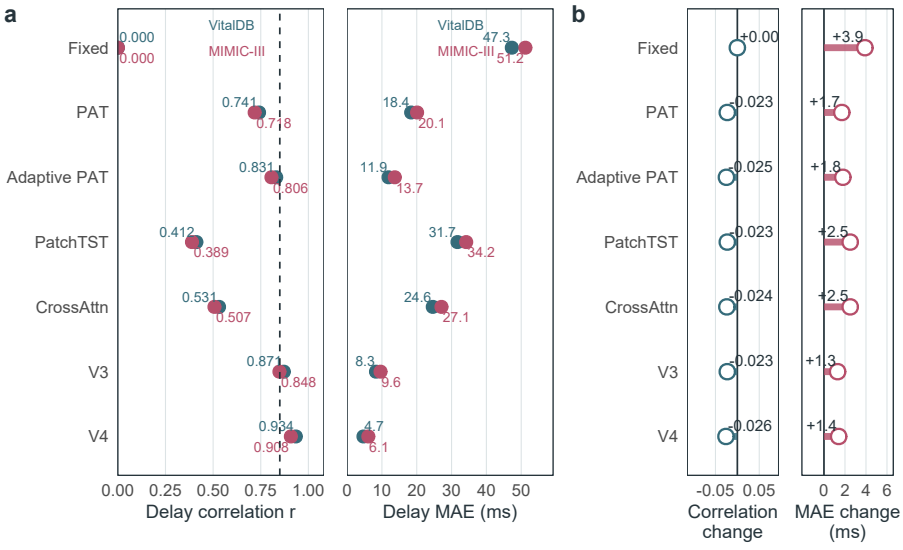


Figure S5: Delay diagnostics for VitalDB and MIMIC-III. (a) Learned-delay correlation and MAE; (b) changes relative to the fixed-delay reference.

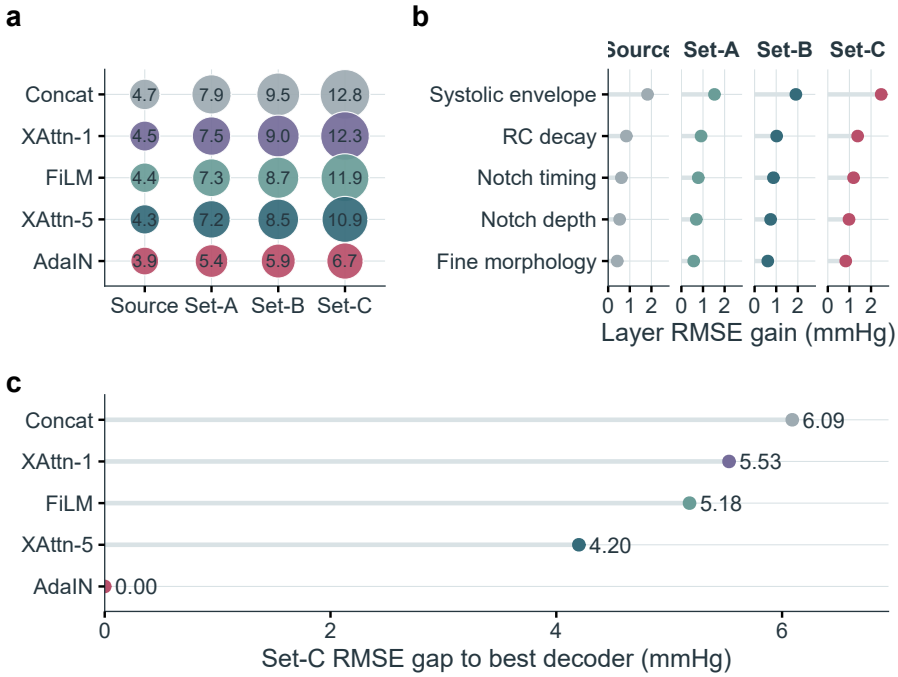


Figure S6: Decoder conditioning analysis. (a) Variant performance; (b) layer-specific RMSE gains; (c) residual Set-C RMSE relative to AdaIN.

D.1 CDSE Recurrence and Fractional-Delay Implementation

CDSE implements the fractional-delay recurrence in three steps. First, the ECG branch estimates the bounded delay τ_k and decomposes it into integer and fractional offsets (d_k, γ_k). Second, a circular buffer stores PPG local features and returns the interpolated delayed feature $\tilde{\mathbf{p}}_k$. Third, the diagonal selective-state recurrence applies element-wise exponentials and Hadamard products to update $\mathbf{h}_k$. The diagonal state matrix avoids dense matrix exponentials, so the implementation remains linear in sequence length. At the start of each window, delayed PPG features without sufficient history are zero-padded. The number of affected samples is determined by $\lceil \tau_k/\delta_t \rceil$ rather than fixed a priori.

Under the diagonal structure of $\mathbf{A}$ and a zero-order-hold assumption over $[t_k, t_{k+1})$, the n-th hidden dimension has the closed-form update

$$h_n(t_{k+1}) = e^{a_n \Delta_{k,n}} h_n(t_k) + \frac{e^{a_n \Delta_{k,n}} - 1}{a_n} \left([\mathbf{B}_E]_{n,:} \mathbf{g}_k^{(E)} + [\mathbf{B}_P]_{n,:} \tilde{\mathbf{g}}_k^{(P)} \right). \tag{23}$$

Vectorizing this expression gives the recurrence used by the encoder:

$$\bar{\mathbf{a}}_k = \exp(\mathbf{\Delta}_k \odot \mathbf{a}), \tag{24}$$

$$\bar{\mathbf{b}}_{\cdot,k} = (\mathbf{\Delta}_k \odot \mathbf{a})^{-1} \odot (\bar{\mathbf{a}}_k - \mathbf{1}) \odot (\mathbf{\Delta}_k \odot \mathbf{b}_{\cdot,k}), \tag{25}$$

$$\mathbf{h}_k = \bar{\mathbf{a}}_k \odot \mathbf{h}_{k-1} + \bar{\mathbf{b}}_{E,k} \mathbf{g}_k^{(E)} + \bar{\mathbf{b}}_{P,k} \tilde{\mathbf{g}}_k^{(P)}, \tag{26}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{h}_k + \mathbf{D}\left(\mathbf{g}_k^{(E)} \,\|\, \tilde{\mathbf{g}}_k^{(P)}\right). \tag{27}$$

The scalar derivation follows from variation of constants:

$$\begin{aligned} h_n(t_{k+1}) &= e^{a_n \Delta_{k,n}} h_n(t_k) + \int_0^{\Delta_{k,n}} e^{a_n(\Delta_{k,n} - s)} u_{n,k}\, ds \\ &= e^{a_n \Delta_{k,n}} h_n(t_k) + u_{n,k} \frac{e^{a_n \Delta_{k,n}} - 1}{a_n}, \end{aligned} \tag{28}$$

where $u_{n,k} = [\mathbf{B}_E]_{n,:} \mathbf{g}_k^{(E)} + [\mathbf{B}_P]_{n,:} \tilde{\mathbf{g}}_k^{(P)}$. Because $\mathbf{A}$ is diagonal, all exponentials and inversions are element-wise operations rather than dense matrix exponentials.

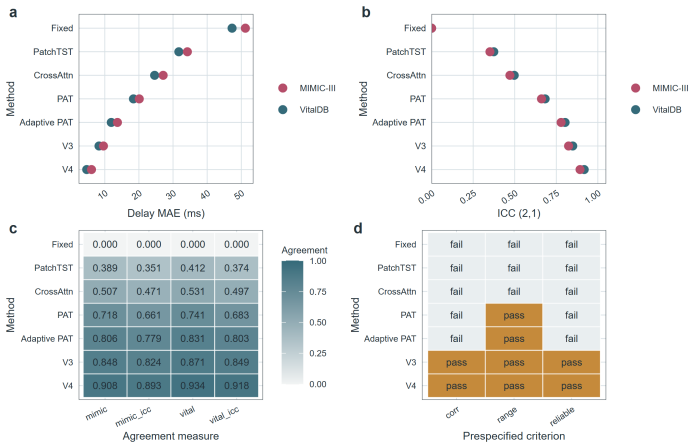


Figure S7: Arrival-delay agreement on VitalDB and MIMIC-III. Panels show (a) delay MAE, (b) ICC$(2,1)$, (c) correlation-based agreement measures, and (d) whether each method satisfies the prespecified correlation, range, and reliability criteria.

Algorithm S1 CDSE buffered phase operator and fused-feature emission

Require: ECG and PPG windows $\mathbf{X}^{(E)}, \mathbf{X}^{(P)} \in \mathbb{R}^{1\times L}$; encoder parameters Θ_{enc}; sampling interval δ_t.

Ensure: Fused feature sequence $\mathbf{H} \in \mathbb{R}^{L\times d}$ and delay sequence $\{\tau_k\}_{k=1}^{L}$.

1: Extract local ECG and PPG morphology features $\mathbf{g}_k^{(E)}, \mathbf{g}_k^{(P)}$ with the continuous front-end.
2: Initialize recurrent state $\mathbf{h}_0 \leftarrow \mathbf{0}$, a delay buffer, and an output accumulator.
3: **for** $k = 1$ **to** L **do**
4: Predict τ_k from $\mathbf{g}_k^{(E)}$ and clip it to $[\tau_{\min}, \tau_{\max}]$.
5: Convert τ_k/δ_t into a queue index d_k and a remainder γ_k.
6: Fetch the two queue entries bracketing d_k and create the delayed PPG feature by linear interpolation.
7: Apply the diagonal transition to the previous state and the aligned ECG/PPG pair; append the emitted feature to the accumulator.
8: Push $\mathbf{g}_k^{(P)}$ into the delay buffer and discard entries that cannot be queried by a future window.
9: **end for**
10: Stack the emitted features in temporal order to obtain $\mathbf{H}$.
11: **return** $\mathbf{H}$ and $\{\tau_k\}_{k=1}^{L}$.

Figures S7 and S8 report cohort-wise agreement and transfer changes for the learned ECG–PPG arrival delay.

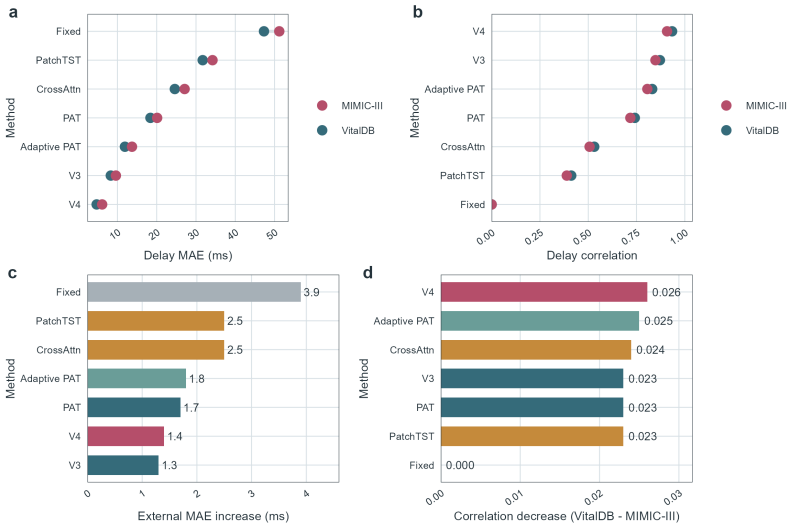


Figure S8: Cohort-level behavior of learned ECG–PPG arrival delays. Panels compare (a) delay MAE, (b) delay correlation, (c) external MAE increase, and (d) the VitalDB-to-MIMIC-III correlation decrease across heuristic, attention-based, integer-delay, and fractional-delay methods.

D.2 F-CVAE Forward Pass and Disentanglement-Loss Computation

Algorithm S2 F-CVAE forward pass and disentanglement-loss computation

Require: Paired window feature sequences from the same patient: $\mathscr{B}=\{(\mathbf{H}^{(t_1)},\mathbf{H}^{(t_2)})_i\}_{i=1}^{|\mathscr{B}|}$; weights $\alpha,\beta,\gamma,\lambda_{\text{con}}$; dataset size N.

Ensure: Latent codes $\{\mathbf{z}_{\text{ctx},i},\mathbf{z}_{\text{evt},i}\}$; disentanglement loss $\mathscr{L}_{\text{dis}}$.

for $i=1$ **to** $|\mathscr{B}|$ **do**
  for $w\in\{t_1,t_2\}$ **do**
    $\bar{\mathbf{H}}_i^{(w)} \leftarrow \frac{1}{L}\sum_{k=1}^{L}\mathbf{H}_{i,k}^{(w)}$
    $(\boldsymbol{\mu}_{\text{ctx}}^{(w)},\log(\boldsymbol{\sigma}_{\text{ctx}}^2)^{(w)}) \leftarrow \text{Enc}_{\phi_{\text{ctx}}}(\bar{\mathbf{H}}_i^{(w)})$
    $(\boldsymbol{\mu}_{\text{evt}}^{(w)},\log(\boldsymbol{\sigma}_{\text{evt}}^2)^{(w)}) \leftarrow \text{Enc}_{\phi_{\text{evt}}}(\mathbf{H}_i^{(w)})$
    for $j\in\{\text{ctx},\text{evt}\}$ **do**
      $\boldsymbol{\sigma}_j^{(w)} \leftarrow \exp\left(\frac{1}{2}\log(\boldsymbol{\sigma}_j^2)^{(w)}\right)$
      $\mathbf{z}_j^{(w)} \leftarrow \boldsymbol{\mu}_j^{(w)} + \boldsymbol{\sigma}_j^{(w)}\odot\boldsymbol{\varepsilon}_{j,i}^{(w)}, \quad \boldsymbol{\varepsilon}_{j,i}^{(w)}\sim\mathscr{N}(\mathbf{0},\mathbf{I})$
    end for
    $\hat{\bar{\mathbf{H}}}_i^{(w)} \leftarrow \text{Dec}_{\omega_{\text{feat}}}(\mathbf{z}_{\text{ctx}}^{(w)}\|\mathbf{z}_{\text{evt}}^{(w)})$
  end for
  Estimate $\mathscr{L}_{\text{feat},i}$, $\mathscr{L}_{\text{MI},i}$, $\mathscr{L}_{\text{TC},i}$, $\mathscr{L}_{\text{MKL},i}$ via the TC decomposition defined above
  $\mathscr{L}_{\text{consist},i} \leftarrow \max(0,\,\tau_m-\cos(\mathbf{z}_{\text{ctx}}^{(t_1)},\mathbf{z}_{\text{ctx}}^{(t_2)}))$
end for
$\mathscr{L}_{\text{dis}} \leftarrow \frac{1}{|\mathscr{B}|}\sum_i(\mathscr{L}_{\text{feat},i}+\alpha\mathscr{L}_{\text{MI},i}+\beta\mathscr{L}_{\text{TC},i}+\gamma\mathscr{L}_{\text{MKL},i}+\lambda_{\text{con}}\mathscr{L}_{\text{consist},i})$
return $\{\mathbf{z}_{\text{ctx},i},\mathbf{z}_{\text{evt},i}\}$; $\mathscr{L}_{\text{dis}}$

D.3 CEA Meta-Training and Streaming Inference

The following formulation expands the inner and outer objectives in the main manuscript. For task $\mathscr{T}_i$, initialize the context branch from the meta-prior and perform M support-set updates:

$$\mathscr{L}_{\text{calib}}^{(m)} = \mathscr{L}_{\text{wave}}\left(\mathscr{D}_i^{\text{sup}};\boldsymbol{\Phi},\phi_{\text{ctx},i}^{(m-1)}\right) + \lambda_{\text{reg}}\left\|\phi_{\text{ctx},i}^{(m-1)}-\phi_{\text{ctx}}^{*}\right\|_2^2. \tag{29}$$

$$\begin{aligned}\phi_{\text{ctx},i}^{(0)} &= \phi_{\text{ctx}}^{*},\\ \phi_{\text{ctx},i}^{(m)} &= \phi_{\text{ctx},i}^{(m-1)} - \alpha_{\text{in}}\nabla_{\phi_{\text{ctx}}}\mathscr{L}_{\text{calib}}\left(\mathscr{D}_i^{\text{sup}};\boldsymbol{\Phi},\phi_{\text{ctx},i}^{(m-1)}\right).\end{aligned} \tag{30}$$

After adaptation, the task-level query objective and the exact meta-gradient with respect to the context initializer are

$$\mathscr{J}_i(\mathbf{\Phi}, \phi_{\text{ctx}}^*) = \mathscr{L}_{\text{total}}\left(\mathscr{D}_i^{\text{qry}}; \mathbf{\Phi}, \phi_{\text{ctx},i}^{(M)}\right). \tag{31}$$

$$\nabla_{\phi_{\text{ctx}}^*} \mathscr{J}_i = \left[\prod_{m=1}^{M}\left(\mathbf{I} - \alpha_{\text{in}} \nabla^2_{\phi\phi} \mathscr{L}_{\text{calib}}^{(m)}\right)\right]^\top \nabla_{\phi_{\text{ctx},i}^{(M)}} \mathscr{J}_i. \tag{32}$$

FO-MAML drops the Hessian-vector products in the exact-gradient expression above, giving

$$\nabla_{\phi_{\text{ctx}}^*} \mathscr{J}_i \approx \nabla_{\phi_{\text{ctx},i}^{(M)}} \mathscr{J}_i. \tag{33}$$

Proposition 1 (FO-MAML gradient approximation)**.** *Suppose the calibration loss is twice differentiable and its Hessian along the M inner updates satisfies* $\|\nabla^2_{\phi\phi}\mathscr{L}_{\text{calib}}^{(m)}\|_2 \le L_c$*. The difference between the exact second-order meta-gradient and the first-order approximation satisfies*

$$\left\|\nabla_{\phi^*}^{\text{SO}} - \nabla_{\phi^*}^{\text{FO}}\right\|_2 \le \left[(1+\alpha_{\text{in}}L_c)^M - 1\right] \left\|\nabla_{\phi_{\text{ctx},i}^{(M)}} \mathscr{J}_i\right\|_2. \tag{34}$$

Proof. The exact second-order gradient contains the inner-loop Jacobian $\prod_{m=1}^{M}(\mathbf{I} - \alpha_{\text{in}}\nabla^2_{\phi\phi}\mathscr{L}_{\text{calib}}^{(m)})$. FO-MAML replaces this product by the identity. Each factor has norm at most $1 + \alpha_{\text{in}}L_c$; applying the product perturbation bound and multiplying by the query-gradient norm gives the bound in equation 34. □

This approximation preserves the subspace restriction: the support-set updates change only ϕ_{ctx}, whereas $\mathbf{\Phi}$ is updated by the outer query loss and is frozen at deployment. The deployment procedure is listed in Supplementary Algorithm S3.

Algorithm S3 Context-restricted calibration and streaming inference

Require: Frozen $\mathbf{\Phi}^*$; meta-prior ϕ_{ctx}^*; $\alpha_{\text{in}}, M, \lambda_{\text{reg}}$.
Ensure: Continuous waveform estimate $\hat{\mathbf{Y}}$ and phenotype probabilities $\hat{\mathbf{p}}$ per segment.

Phase 1: Bedside calibration
$\phi_{\text{ctx}} \leftarrow \phi_{\text{ctx}}^*$
for $m = 1$ **to** M **do**
  $\mathbf{H} \leftarrow \text{CDSE}(\mathbf{X}_{\text{cal}}; \Theta_{\text{enc}}^*)$
  $\mathbf{z}_{\text{ctx}} \leftarrow \boldsymbol{\mu}_{\text{ctx}}(\text{Enc}_{\phi_{\text{ctx}}}(\mathbf{H}))$
  $\mathbf{z}_{\text{evt}} \leftarrow \boldsymbol{\mu}_{\text{evt}}(\text{Enc}_{\phi_{\text{evt}}^*}(\mathbf{H}))$
  $\hat{\mathbf{Y}} \leftarrow \text{AdaIN-Dec}(\mathbf{z}_{\text{ctx}}, \mathbf{z}_{\text{evt}}, \mathbf{H}; \omega_{\text{wave}}^*)$
  $\phi_{\text{ctx}} \leftarrow \phi_{\text{ctx}} - \alpha_{\text{in}}\nabla_{\phi_{\text{ctx}}}(\mathscr{L}_{\text{wave}} + \lambda_{\text{reg}}\|\phi_{\text{ctx}} - \phi_{\text{ctx}}^*\|_2^2)$
end for
Phase 2: Continuous monitoring
while sensor stream active **do**
  $\mathbf{H} \leftarrow \text{CDSE}(\mathbf{X}; \Theta_{\text{enc}}^*)$
  $\hat{\mathbf{Y}} \leftarrow \text{AdaIN-Dec}(\mathbf{z}_{\text{ctx}}, \mathbf{z}_{\text{evt}}, \mathbf{H}; \omega_{\text{wave}}^*)$
  $\hat{\mathbf{p}} \leftarrow \text{Softmax}(\mathscr{F}_{\psi^*}(\mathbf{z}_{\text{ctx}} \| \mathbf{z}_{\text{evt}}))$
  Emit $\hat{\mathbf{Y}}$ and $\hat{\mathbf{p}}$ to the monitoring interface.
end while

E Dataset Construction and Evaluation Protocols

This section specifies cohort construction, filtering, split logic, metric aggregation, comparator training, and the fixed optimization configuration.

E.1 MC-MED ICD-Based Label Construction

MC-MED supports zero-shot Setting D phenotype evaluation using synchronized ECG and PPG without a continuous ABP reference. Discharge ICD-10 codes provide weak visit-level labels rather than waveform-level pressure measurements. Hypertension-positive visits are identified from the ICD-10 hypertension families I10–I16; hypotension-positive visits are identified from I95.* codes; visits without either family are assigned to the normotension reference group for the three-class

Track-2 evaluation. Visits containing contradictory hypertension and hypotension code families are excluded before aggregation. The source-trained phenotype head remains fixed, and MC-MED contributes no samples to encoder training, model selection, calibration, or domain-shift-ratio computation. Visit labels follow the order normotension–hypertension–hypotension. Scores remain setting-specific because source and target share class semantics but use different label-construction mechanisms.

E.2 Preprocessing, Signal-Quality Checks, and Rejection Criteria

All datasets are processed into non-overlapping 10-s windows at 125 Hz after bidirectional sixth-order Chebyshev anti-aliasing. A window is rejected if any required modality is missing, contains non-finite values, or shows flat-line saturation for more than 1 s. For ABP-equipped cohorts, windows are additionally rejected when the reference waveform has baseline wander, systolic/diastolic ordering violations, or physiologically implausible pressure values (SBP outside 60–250 mmHg, DBP outside 30–150 mmHg, or MAP outside 40–180 mmHg). For MIMIC-III-Ext-PPG, Track-1 waveform reconstruction uses only windows with valid ABP and non-negative signal-quality index; lower-quality windows contribute to the designated noise-stratification analysis but not waveform reconstruction. Filtering precedes patient-level partitioning, preventing rejected windows from influencing model selection.

E.3 Patient-Level Splits, LOSO, Zero-Shot, and Few-Shot Calibration

VitalDB is the sole training source and is partitioned 6:2:2 by patient into training, validation, and held-out in-domain test splits. External cohorts are not used during training or hyperparameter selection. Setting A and Setting B are evaluated as zero-shot source-to-target transfer. Setting C uses 32-fold leave-one-subject-out evaluation on the RWW cohort; when few-shot calibration is reported, support windows from the held-out subject are used only for the CEA inner update and are kept disjoint from query windows used for reporting. Setting D is strictly zero-shot on MC-MED and is evaluated at visit level by aggregating per-window softmax probabilities within each visit.

Table S2: Cohort inventory and experimental role. Counts describe the analysis inventory used by the reported experiments.

Cohort	Analysis units	10-s segments	f_s (Hz)	Role
VitalDB (Lee et al., 2022)	1,200 subjects	480,000	$500 \rightarrow 125$	Training, validation, and patient-disjoint source testing
MIMIC-III-Ext-PPG (Moulaeifard et al., 2026)	2,371 subjects[†]	$\sim$5,670,000[‡]	125	Zero-shot ICU/ward transfer (Setting A)
UCI cuff-less BP (Kachuee et al., 2015)	942 subjects	12,000	125	Zero-shot ward-style transfer (Setting B)
MC-MED (Kansal et al., 2025)	70,545 visits[§]	–	125	Zero-shot visit-level phenotype recognition (Setting D)
RWW (private in-house cohort)	32 subjects	$\sim$3,840	$500 \rightarrow 125$	Wearable LOSO and context calibration (Setting C)

[†]Subjects retained after quality filtering requiring valid ABP. [‡]Underlying 30-s records divided into three non-overlapping 10-s windows. [§]Visit-level ICD-10 analysis units; MC-MED provides no continuous ABP waveform. VitalDB uses a 6:2:2 patient-level split.

The RWW cohort comprises 32 adults aged 25–78 years (mean 52.4, SD 14.7; 18 male and 14 female). Each participant contributed at least 20 min of synchronized wearable ECG–PPG and CNAP-referenced recording. The inventory reports subjects, visits, or segments according to the unit used by each analysis rather than treating these counts as interchangeable samples. Patient-level partitioning is applied before window aggregation; consequently, the segment totals describe evaluation workload, whereas the subject and visit counts define the independence structure for transfer and phenotype analyses. VitalDB, MIMIC-III-Ext-PPG, and the UCI Cuff-Less Blood Pressure Estimation dataset are accessed under their respective repository or dataset terms. MC-MED is a PhysioNet resource used only for visit-level ICD-10 phenotype evaluation. RWW is a private, consented in-house cohort and is not redistributed with the code. Class membership is constructed after the quality filters and label rules below; because ABP-derived and ICD-derived labels use different evidence sources, we report their class-wise performance by setting rather than pooling class counts across cohorts.

E.4 METRIC DEFINITIONS AND AGGREGATION RULES

Waveform reconstruction is evaluated on the same non-overlapping windows used for model selection and transfer. For a reference waveform $\mathbf{y}$ and estimate $\hat{\mathbf{y}}$ with N samples, we report

$$\mathrm{RMSE} = \sqrt{N^{-1}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}, \qquad \mathrm{MAE} = N^{-1}\sum_{n=1}^{N}|y_n - \hat{y}_n|. \tag{35}$$

Pearson correlation is computed within each evaluation window and then aggregated over the evaluation units. The external degradation ratio is the mean external RMSE divided by the source-test RMSE; lower values indicate smaller relative degradation. Dynamic time-warping (DTW) distance and notch timing error are computed after the same waveform alignment used for the reported pressure metrics.

For phenotype recognition, Macro-F1 is the unweighted mean of the per-class F1 scores. ABP-derived phenotype scores and the ICD-derived MC-MED scores are reported by setting because their label-construction mechanisms differ, even though the semantic classes are shared. Representation reliability is measured with $\mathrm{ICC}(2,1)$ and within-patient ICC; the RWW value aggregates the 32 held-out subjects after leave-one-subject-out evaluation. Frozen-probe results use the source-trained probe without target calibration. Values involving stochastic training are reported as mean $\pm$ SD over three independent runs; ICC confidence intervals are retained where shown in the main or supplementary tables, while deterministic descriptive analyses are reported once.

Table S3: Evaluation metrics and aggregation rules used in the manuscript.

Evidence	**Metric**	**Aggregation and interpretation**
Waveform	RMSE, MAE	Window-level error in mmHg; lower is better.
Waveform	Pearson r	Within-window agreement between reference and estimate.
Transfer	External degradation ratio ρ	Mean external RMSE divided by source-test RMSE; lower is better.
Alignment	DTW, notch timing	Shape distance and landmark timing error after common preprocessing.
Phenotype	Macro-F1	Unweighted mean of class-wise F1; reported separately by label mechanism.
Representation	$\mathrm{ICC}(2,1)$, $\mathrm{ICC}_{\mathrm{wp}}$	Reliability across repeated windows and within-patient windows.
Probe	R^2, AUROC	Frozen source-trained probes evaluated on the specified target cohort.

E.5 DOMAIN-GENERALIZATION PSEUDO-DOMAIN CONSTRUCTION

Domain-generalization comparators use four VitalDB pseudo-domains defined by training-set age quartiles after the patient split is fixed. The 25th, 50th, and 75th percentiles delimit Q1–Q4; target-cohort metadata and unlabeled target windows are excluded. This construction preserves the single-source transfer protocol used by PhysioSplit. The main manuscript reports SiamQuality and IRM as task-specific and invariant-risk references, while Table S4 provides the extended single-source DG comparison.

Table S4: Single-source domain-generalization results. Set C for PhysioSplit is reported with CEA; the other rows are frozen source-only evaluations.

Method	**Source**	**Set A**	**Set B**	**Set C**	**External mean**	$\boldsymbol{\rho}$
DANN	4.94±.13	8.12±.28	9.53±.34	12.87±.58	10.17	2.06
CORAL	5.07±.14	8.28±.30	9.71±.36	13.14±.61	10.38	2.05
IRM	4.81±.11	7.93±.26	9.38±.33	12.64±.56	9.98	2.08
GroupDRO	4.88±.12	8.07±.27	9.49±.34	12.91±.59	10.16	2.08
PhysioSplit	**3.87±.06**	**5.43±.14**	**5.89±.17**	**6.74±.24**	**6.02**	**1.55**

The comparison uses one source cohort throughout. DANN, CORAL, IRM, and GroupDRO receive the same patient-level split, input windows, optimizer schedule, and validation rule as the main model; their only additional signal is the age-quartile pseudo-domain assigned within the VitalDB training split. No target samples or target metadata enter either training or model selection. The external mean and ρ therefore summarize a common source-only generalization test, while the PhysioSplit Set-C value additionally reflects the context-calibration protocol used for Setting C. The *w/o* CEA row in the main table is the corresponding no-update control. We do not interpret this comparison as a four-policy ranking: event-only and global fine-tuning were not part of the reported protocol. This table therefore separates source-domain regularization from the documented deployment-time context update without assigning unmeasured values to alternative policies.

E.6 Baseline Implementation Details

All baselines use the same 10-s, 125 Hz windows, patient-level VitalDB partitions, and external target splits as PhysioSplit. Public implementations are adapted only at the ECG–PPG input and ABP output interfaces; unavailable implementations are reproduced in the shared PyTorch pipeline. Model selection uses the VitalDB validation split, the AdamW/cosine schedule, and the early-stopping rules in Table S5. The protocol excludes target samples and metadata and supplies only the age-quartile pseudo-domain labels defined in Sec. A3.4.

E.7 Training and Loss Hyperparameter Summary

Table S5 consolidates the fixed training configuration and identifies quantities selected on the VitalDB validation split. The table is organized by the point at which each setting enters the pipeline: input preprocessing, source training, representation learning, task balancing, calibration, and evaluation. Fixed values are used for every cohort and run; only quantities marked as validation-selected are tuned on the source validation partition. The loss rows distinguish waveform reconstruction from feature-space reconstruction and representation regularization, while the episodic-sampling row specifies the support/query budget used for each calibration task. This separation makes the training configuration and the deployment update rule reproducible without requiring target-cohort tuning.

Table S5: Selected training, loss, and protocol hyperparameters.

Group	Quantity	Setting	Role
Input protocol	K, f_s	10-s windows; 125 Hz after resampling	Common segment length and sampling grid.
Benchmark training	optimizer, learning rate, weight decay	AdamW; 10^{-3}; 10^{-4}	Main benchmark optimizer setting.
Benchmark schedule	epochs, batch size, patience	50 epochs; 256; early-stop patience 7	Training budget and early stopping.
Learning-rate schedule	scheduler, minimum learning rate	cosine annealing; 10^{-6}	Smooth decay used in benchmark runs.
Meta-pretraining	$\eta_{\text{base}}, \eta_{\phi_{\text{ctx}}}$, gradient clipping	3×10^{-4}, 3×10^{-4}, 1.0	Base and context-branch updates.
Meta-pretraining schedule	epochs, batch size, patience	20 epochs; 256; early-stop patience 5	Pretraining budget before transfer.
Latent size	$d_{\text{ctx}}, d_{\text{evt}}$	64, 64	Main benchmark latent widths.
Delay range	$\tau_{\min}, \tau_{\max}$	100–400 ms physiological ECG-to-PPG arrival-delay range	Bounds the learned arrival-delay gate.
F-CVAE	$\alpha, \beta, \gamma, T_{\text{anneal}}$	1.0, 4.0, 1.0; 10000 steps	MI, TC, and marginal-KL weights.
Temporal consistency	$\tau_m, \lambda_{\text{con}}$	$\tau_m = 0.9$; $\lambda_{\text{con}} = 0.5$	Within-patient code stability.
Task losses	$\lambda_w(t), \lambda_p(t), \lambda_d(t)$	$\lambda_w = 1$, $\lambda_p = 0.3$, $\lambda_d(t) = 0.1\min\{1,(t+1)/10000\}$	Linear disentanglement warm-up over 10,000 steps.
Waveform robustness	δ_H	5.0	Huber threshold for ABP errors.
CEA	$\alpha_{\text{in}}, M, \lambda_{\text{reg}}$	0.01; 3 inner steps; 0.1	Context-branch calibration strength.
Episodic sampling	support/query size, update order, frequency	4/4 segments; first-order; every 20 steps	Patient-level support/query adaptation.
Evaluation configuration	random seed	42 (default); repeat protocol 42, 43, 44	The release configuration records the default seed and the three-seed reproduction list. Reported stochastic entries are mean±SD over three independently initialized runs.

E.8 Model Architecture Parameters

Table S6 lists the PhysioSplit architecture. The cumulative variants change only the encoder mechanisms shown in the table; the latent partition, input grid, and adaptation interface remain shared.

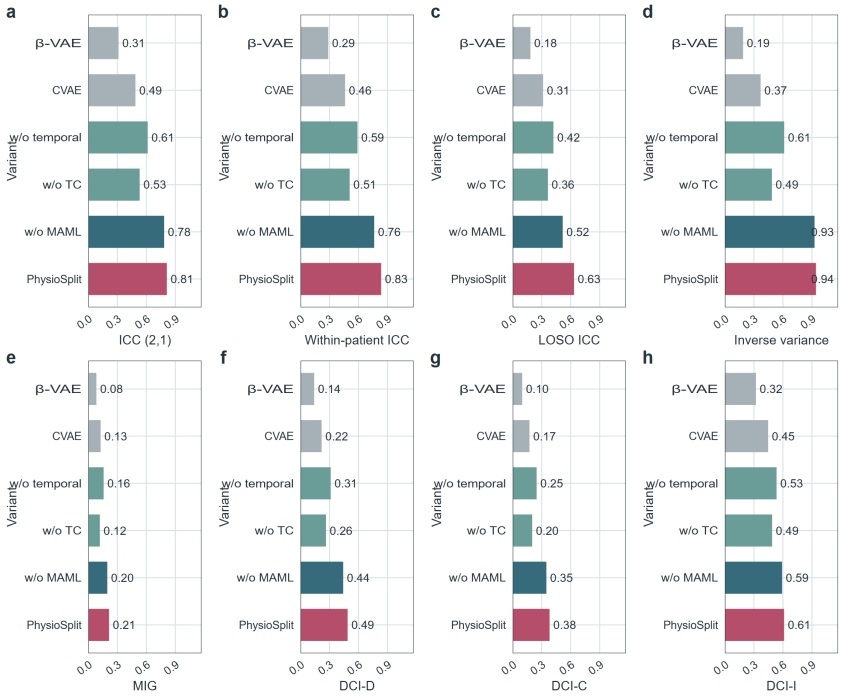


Figure S9: Dimension-level representation reliability across generative baselines and ablations. Panels (a–d) report ICC$(2,1)$, within-patient ICC, RWW LOSO ICC, and inverse variance; panels (e–h) report MIG and the three DCI components. Higher values indicate stronger reliability or separation; for inverse variance, a larger value corresponds to lower dispersion.

Table S6: Architecture configuration shared by the PhysioSplit experiments.

Component	**Configuration**
Input signals	Synchronized ECG and PPG; 10-s windows at 125 Hz
Latent partition	$\mathbf{z}_{\text{ctx}}$: 64 dimensions; $\mathbf{z}_{\text{evt}}$: 64 dimensions
Delay module	Learnable fractional delay; bounded to 100–400 ms
Encoder variants	V1 discrete SSM; V2 continuous morphology; V3 integer delay; V4 fractional delay
Context adaptation	Context branch updated; event branch and shared decoder dynamics retained
Decoder conditioning	Context code applied through asymmetric AdaIN modulation
Training objective	Waveform and feature reconstruction, phenotype, mutual-information, total-correlation, and temporal-consistency terms

F Extended Latent-Space and Vascular Semantics Analyses

These analyses test whether the factorized codes retain stable and selective information across repeated windows, held-out subjects, and probe targets.

Reliability and test–retest window definition. ICC(2,1) is computed from non-overlapping windows drawn from the same subject under the held-out evaluation split. For VitalDB analyses, the paired windows are sampled from the held-out validation or test partition without sharing windows with model selection. For RWW LOSO, the paired windows are drawn from the held-out subject's recording after signal-quality screening and before any target-subject fine-tuning. The resulting ICC quantifies representation-level temporal reliability rather than device-level repeatability or patient identification.

Figure S9 provides dimension-level ICC, MIG, and DCI summaries under the same held-out evaluation protocol.

F.1 VASCULAR SEMANTIC LINEAR PROBE OF $\mathbf{z}_{\text{ctx}}$

Table S7: Frozen linear-probe results across source, external clinical, and wearable-transfer levels ($\mathbf{z}_{\text{ctx}}$ vs. $\mathbf{z}_{\text{evt}}$). Values are mean ± SD.

Level	Target	Metric	$\mathbf{z}_{\text{ctx}}\uparrow$	$\mathbf{z}_{\text{evt}}$	Δ
L1 VitalDB	Age	R^2	**0.531**±.028	0.094±.019	**+.437**
	BMI	R^2	**0.489**±.033	0.078±.022	**+.411**
	Sex	AUROC	0.552±.036	0.537±.043	+.015
L2 MIMIC-III	HTN	AUROC	**0.763**±.024	0.691±.031	**+.072**
	AF	AUROC	**0.741**±.027	0.708±.032	**+.033**
	PN†	AUROC	0.514±.041	0.507±.044	+.007
L3 RWW	SBP‡	AUROC	**0.712**±.049	0.597±.063	**+.115**

Δ=$\mathbf{z}_{\text{ctx}}-\mathbf{z}_{\text{evt}}$; bold Δ values satisfy $|\Delta|>0.03$. †Pneumonia (ICD-9 486); both at chance ($p>0.35$, binomial test). ‡SBP>130 vs. <120 mmHg; trained on VitalDB, applied zero-shot to the RWW cohort.

At L1, $\mathbf{z}_{\text{ctx}}$ improves age and BMI prediction by 0.437 and 0.411 in R^2 relative to $\mathbf{z}_{\text{evt}}$, consistent with preferential encoding of persistent subject attributes. At L2, the context advantage is larger for hypertension (+0.072 AUROC) than for atrial fibrillation (+0.033), whereas the non-cardiovascular pneumonia control remains near chance in both codes. The zero-shot RWW SBP probe retains a +0.115 AUROC advantage, showing that its predictive information is retained in the wearable setting. Figure S10 shows the corresponding target-wise probe scores and context–event differences.

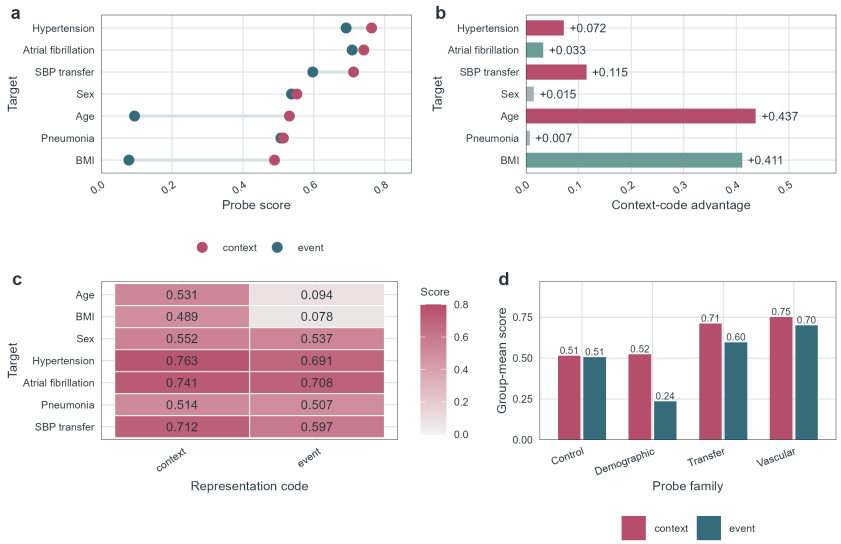


Figure S10: Frozen-probe readouts of context and event information. (a) Target-wise probe scores; (b) context-code advantages; (c) the full target-by-code score matrix; and (d) mean performance by probe family. Positive differences indicate stronger predictive information in $\mathbf{z}_{\text{ctx}}$.