MSMA-PPG: CHANNEL-ALIGNED EXPERT ROUTING FOR SELECTED-WAVELENGTH PPG PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Selecting a photoplethysmography (PPG) wavelength by signal quality does not necessarily minimize physiological prediction error. We introduce MSMA-PPG, a framework that learns from synchronized multi-wavelength recordings and predicts from one selected PPG stream and acceleration. Wavelength-specific projections and residual adapters share an optical temporal backbone, full-to-subset alignment links paired training views, and a Channel-Aligned Mixture of Experts combines optical, motion, and fused representations through availability-masked top-2 routing. We evaluate single-stream representation transfer across heart rate (HR), blood pressure (BP), and atrial fibrillation (AF), alongside paired HR diagnostics of wavelength selection and expert routing. Across the seven-model PPG-only test results, with routing disabled, MSMA-PPG has the lowest reported participant-macro HR MAE on EarSet and PPG-DaLiA (5.020 and 9.246 bpm), the lowest systolic BP MAE (15.456 mmHg), and the second-highest AF AUPRC/AUROC (0.5177/0.8909). The paired diagnostics reveal a different ordering after adaptation: although frozen affine probes favor aligned initialization on EarSet and PPG-DaLiA, adapted aligned models have higher window-pooled MAE than reconstruction-only controls by 0.797, 1.485, and 0.373 bpm on WildPPG, EarSet, and PPG-DaLiA. Retrospective selection from recorded wavelengths also produces model-dependent responses: lower composite policy scores sometimes reduce HR error. Frozen expert interventions reveal strong dependence on the trained motion/shared pathways and mixture weights, despite the aligned model's higher absolute error. These findings distinguish competitive optical representation transfer from the separate effects of acquisition choices and fitted expert routing on downstream accuracy.

1 INTRODUCTION

Photoplethysmography (PPG) supports wearable physiological prediction, but its waveform varies with wavelength, placement, hardware, tissue, and motion (Allen, 2007; Allen et al., 2021). Adaptive sensing selects an optical path according to current signal conditions (Mäkinen et al., 2026). A high generic quality score, however, does not necessarily identify the wavelength with the lowest error for a particular target and fitted predictor. This mismatch connects two decisions: which waveform to acquire and how to use the resulting observation.

We study a downstream predictor constrained to one selected PPG wavelength and tri-axial acceleration (ACC). Synchronized multi-wavelength training supplies complementary views, while wavelength-specific processing retains acquisition identity. The Multi-Spectral Motion-Aware PPG model (MSMA-PPG) combines wavelength-specific projections and residual adapters with a shared optical temporal backbone, a motion encoder, and a Channel-Aligned Mixture of Experts (CA-MoE). A shared-weight full-view branch supervises subset representations. At inference, a sensing-conditioned router activates two pathways among the selected-wavelength, ACC, and shared experts; hard masks exclude experts for unrecorded wavelengths.

Two protocols examine different parts of this design. A seven-model PPG-only comparison disables routing and measures representation transfer to heart rate (HR), blood pressure (BP), and atrial fibrillation (AF). Paired HR experiments retain selected PPG, ACC, and routing to compare aligned and reconstruction-only training. Selection is simulated retrospectively: a frozen quality–energy–

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.

motion surrogate scores simultaneously recorded candidates before hiding unselected waveforms. These experiments measure the consequences of selecting recorded alternatives, without measuring online acquisition or device energy consumption.

Our contributions are:

- **A formulation separating acquisition and prediction.** We define HR selection regret relative to a frozen policy under a one-PPG-plus-ACC constraint, making disagreement between policy scores and predictor-specific wavelength errors measurable.
- **Channel-aligned representation learning and routing.** Wavelength-specific adapters, a shared backbone, full-to-subset alignment, and availability-masked top-2 experts connect paired training observations to prediction from fewer channels.
- **An analysis of transfer and routing behavior.** The reported PPG-only summaries place MSMA-PPG first on EarSet/PPG-DaLiA HR MAE and SBP MAE, and second on AF AUPRC/AUROC. Conversely, the adapted aligned HR model has higher MAE than reconstruction-only training across 129 matched policy conditions. Frozen interventions distinguish reliance on a learned mixture from improved absolute accuracy.

2 Related Work

Transferable physiological representations. PaPaGei, Pulse-PPG, AnyPPG, and SIGMA-PPG learn PPG representations for physiological prediction (Pillai et al., 2025; Saha et al., 2025; Nie et al., 2026; Guo et al., 2026); MantisV2 provides a general time-series comparator (Feofanov et al., 2026). Masked reconstruction and multimodal supervision learn from waveform structure and paired physiological signals (Thukral et al., 2026; Geenjaar et al., 2026). Our PPG-only comparison evaluates the optical backbone in this transfer setting, while the routed HR experiments examine its interaction with motion and fused evidence.

Multi-wavelength and optical–motion learning. Tri-spectral fusion uses simultaneously observed wavelengths (Meier et al., 2024a), and ProtoMM aligns PPG with acceleration (Mao et al., 2025). Our model uses paired optical views during training but receives only the selected PPG stream and ACC in the routed HR setting. WildPPG and EarSet supply optical alternatives; PPG-DaLiA supplies single-wavelength PPG–motion pairs (Meier et al., 2024b; Montanari et al., 2023; Reiss et al., 2019).

Missing inputs and expert routing. ModDrop, GMC, ADAPT, and LSM-2 handle incomplete modalities through dropout, alignment, or missingness-aware learning (Neverova et al., 2016; Poklukar et al., 2022; Mordacq et al., 2024; Xu et al., 2025). VCR combines shared/specific wearable representations with missing-aware routing (Weng et al., 2026); MMoE uses task-specific gates, and Flex-MoE supports variable observed modalities (Ma et al., 2018; Yun et al., 2024). We apply availability constraints after wavelength selection and examine how matched acquisition changes and frozen expert interventions alter the fitted predictor's accuracy.

3 Problem Formulation

Let $\mathcal{P} = \{\mathrm{R}, \mathrm{IR}, \mathrm{G}\}$ denote the red, infrared, and green optical channels, and let $\mathcal{P}_i \subseteq \mathcal{P}$ contain the wavelengths recorded for window i. A synchronized training example is $X_i^F = \{x_i^c : c \in \mathcal{P}_i\} \cup \{a_i\}$, where a_i is tri-axial acceleration. We study prediction from one selected optical channel and acceleration. An upstream policy π chooses $c_i^\pi \in \mathcal{P}_i$, giving the downstream observation

$$O_i^\pi = (x_i^{c_i^\pi}, a_i, m_i, q_i, s_i), \qquad S_i^\pi = \{c_i^\pi, \mathrm{ACC}\}, \tag{1}$$

where m_i encodes available inputs, q_i describes selected-signal quality, and s_i records the selection and policy state. The predictor and router receive only the selected PPG waveform, acceleration, and these covariates. Unselected waveforms, full candidate-quality scores, demographic metadata, and target labels are excluded from their inputs.

The evaluated policy is a retrospective quality–energy–motion surrogate: it scores all simultaneously recorded candidate waveforms within the current window and then exposes one to the predictor. This

construction makes alternative selections comparable on the same physiological interval. It evaluates the consequences of selection under a fixed downstream input budget; it does not implement acquisition with only one active wavelength.

For a fixed predictor f_t, target y_i^t, and evaluation loss ℓ_t^{eval}, we define selection regret relative to a frozen reference policy π_{ref} as

$$\mathcal{R}_t(\pi; \pi_{\mathrm{ref}}) = \mathbb{E}[\ell_t^{\mathrm{eval}}(f_t(O_i^{\pi}), y_i^t) - \ell_t^{\mathrm{eval}}(f_t(O_i^{\pi_{\mathrm{ref}}}), y_i^t)]. \tag{2}$$

For HR, $\ell_{\mathrm{HR}}^{\mathrm{eval}}(\hat{y}, y) = |\hat{y} - y|$, so the empirical window average is the pooled MAE difference in bpm. Positive regret indicates worse prediction than the reference and negative regret indicates improvement on the same examples. The surrogate ranking need not agree with prediction error (Appendix B.3).

The wavelength-selection analysis concerns HR on WildPPG and EarSet, which provide synchronized optical alternatives. Single-wavelength PPG-DaLiA supplies a no-change policy control. Separate PPG-only HR, BP, and AF comparisons assess representation transfer with task-specific heads and cohorts. These tasks are fitted separately to assess reuse of the optical representation, while the selection experiment tests how a fixed HR predictor uses the waveform actually selected.

4 METHOD

4.1 OVERVIEW

MSMA-PPG combines wavelength-specific optical processing, a shared temporal backbone, full-to-subset representation alignment, and availability-constrained expert routing (Figure 1). Paired training uses synchronized channels to learn representations for a student observing one PPG waveform and tri-axial acceleration. At inference, the selected optical representation, motion representation, and their fused representation form the inputs to CA-MoE. The separate PPG-only transfer comparison bypasses this router and fits task-specific heads to the optical representation.

4.2 CHANNEL-TYPED ENCODERS

Each observed optical waveform is divided into temporal patches. Wavelength-specific projections P_c, channel embeddings e_c, and residual adapters D_c surround a shared temporal Transformer T_θ. We include input normalization in P_c and the residual connection in D_c. With temporal-position embeddings e_{time}, the encoded sequence for $c \in S_i^{\pi} \cap \mathcal{P}$ is

$$U_i^c = D_c(T_\theta\left(P_c(\mathrm{Patch}(x_i^c)) + e_c + e_{\mathrm{time}}\right)). \tag{3}$$

The shared Transformer models temporal structure across optical channels, while the projections and adapters retain wavelength-dependent responses. A separate acceleration encoder preserves axis-specific patch embeddings and applies temporal and cross-axis processing. Pooling its output gives the motion embedding z_i^{ACC}. Cross-spectral attention and a fusion Transformer combine the available encoded sequences into a fused representation.

The reconstruction pathway additionally uses a two-way optical/motion gate to weight sequence tokens, with entropy and load-balance regularization during Stage II. This sequence gate and the downstream five-expert CA-MoE router operate at different points in the model. Selected-signal quality enters the latter through sensing features; the waveform encoder uses no additional quality or context token.

4.3 CROSS-CHANNEL ALIGNMENT

Let $z_i^c = \mathrm{Pool}_c(U_i^c)$ denote an optical channel embedding, and let $F_\theta(X_i, m_i, M_i)$ denote the fused representation under physical availability m_i and an artificial patch/channel mask M_i. The full-view and subset branches share parameters:

$$z_i^F = \mathrm{sg}\big(F_\theta(X_i^F, m_i^F, 0)\big), \qquad z_i^S = F_\theta(X_i^F, m_i^F, M_i^S), \tag{4}$$

Here m_i^F marks the channels present in the synchronized recording, M_i^S hides unselected waveforms, and sg stops gradients through the full-view target. The target is computed by a full-view

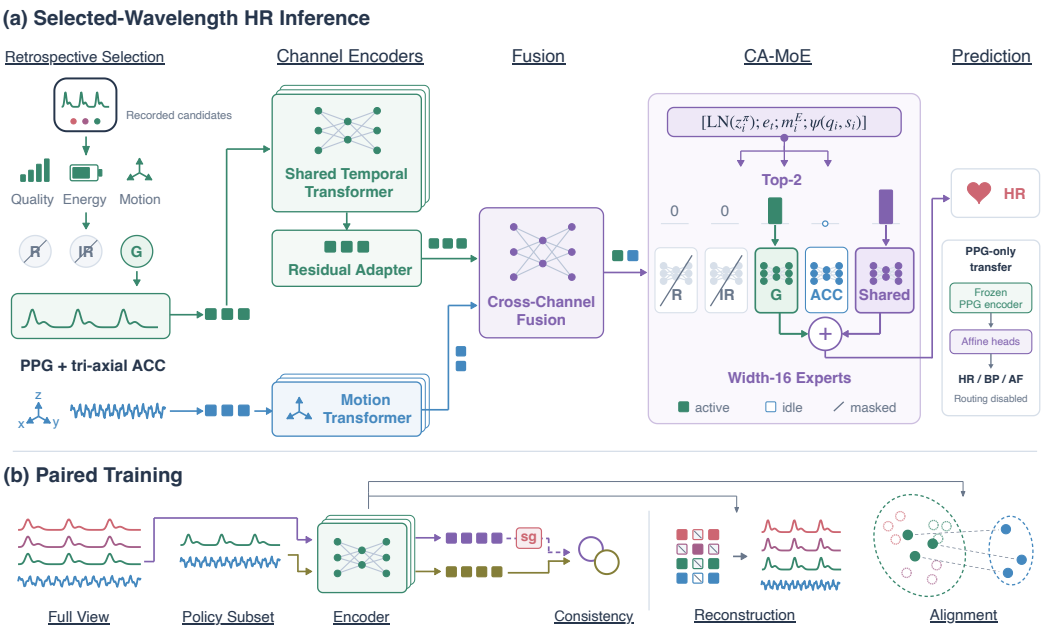


Figure 1: **MSMA-PPG. (a)** A selected PPG channel and ACC pass through typed encoders, fusion, and availability-masked top-2 routing. The illustrated decision activates the green and shared experts; the ACC expert is available but idle, while red/IR experts are masked. The paired HR policy and routing experiments use this pathway; the seven-model HR/BP/AF comparison uses PPG-only representations with routing disabled. Wavelength selection is simulated retrospectively from recorded alternatives. **(b)** Shared-weight full-view and subset branches learn through reconstruction, alignment, and embedding consistency. Artificially masked training channels retain mask tokens. sg denotes stop-gradient; dashed arrows are used only during training.

forward pass of the current encoder, with no separate teacher parameters. Masked positions retain learned mask tokens and participate in cross-channel processing. At inference, unrecorded channels instead have zero physical availability and are excluded from valid pooling. Artificial masking and physical absence therefore expose the model to different computational paths.

Spectral cross-correlation aligns eligible optical pairs, PPG–motion alignment links synchronized modalities, and cosine consistency brings the subset embedding toward the full-view target. The PPG–motion criterion supports cross-correlation and symmetrized InfoNCE variants, as detailed in Appendix C. Only recordings with verified wavelength identities contribute optical pair losses. These objectives encourage shared information across views while channel embeddings and adapters preserve wavelength identity. Their role is to transfer predictable structure from paired observations; information unique to an unavailable waveform remains unobserved.

4.4 Channel-Aligned Mixture of Experts

The expert set is $\mathcal{E} = \{E_{\mathrm{R}}, E_{\mathrm{IR}}, E_{\mathrm{G}}, E_{\mathrm{ACC}}, E_{\mathrm{shared}}\}$. With one PPG channel and ACC, the valid set is $\mathcal{V}_i = \{E_{c_i^\pi}, E_{\mathrm{ACC}}, E_{\mathrm{shared}}\}$. Optical and motion experts process their respective channel embeddings; the shared expert processes the fused embedding. Each expert is a two-layer residual bottleneck of width $r = 16$, where d is the embedding dimension, LN is layer normalization, and σ is the pointwise activation. Omitting biases and dropout,

$$E_e(z) = z + W^e_{\mathrm{up}}\,\sigma(W^e_{\mathrm{down}}\,\mathrm{LN}(z))\,, \quad W^e_{\mathrm{down}} \in \mathbb{R}^{16\times d},\ W^e_{\mathrm{up}} \in \mathbb{R}^{d\times 16}. \tag{5}$$

For downstream routing, define $X_i^\pi = \{x_i^{c_i^\pi}, a_i\}$ and let $z_i^\pi = F_\theta(X_i^\pi, m_i^{\mathrm{deploy}}, 0)$ be its fused representation. Expert availability is encoded by m_i^E, whose entry for E_e is one exactly when $E_e \in \mathcal{V}_i$. A router R_ϕ receives the fused representation, task embedding e_t, expert availability m_i^E,

and sensing features $\psi(q_i, s_i)$ describing selected-signal quality and the acquisition decision. Its logits are masked before sparse normalization:

$$v_i^e = \begin{cases} z_i^c, & E_e = E_c,\ c \in \{c_i^\pi, \mathrm{ACC}\}, \\ z_i^\pi, & E_e = E_{\mathrm{shared}}, \end{cases} \tag{6}$$

$$b_{i,t} = R_\phi([\mathrm{LN}(z_i^\pi); e_t; m_i^E; \psi(q_i, s_i)]), \tag{7}$$

$$\tilde{b}_{i,t,e} = \begin{cases} b_{i,t,e}, & E_e \in \mathcal{V}_i, \\ -\infty, & \text{otherwise}, \end{cases} \tag{8}$$

$$g_{i,t} = \mathrm{TopKSoftmax}(\tilde{b}_{i,t}, K = 2), \tag{9}$$

$$h_{i,t} = \sum_{e: E_e \in \mathcal{V}_i} g_{i,t,e} E_e(v_i^e). \tag{10}$$

Here TopKSoftmax retains the two largest valid logits, normalizes their exponentials to sum to one, and sets all remaining weights to zero. The weighted sum is evaluated only for active experts. The router therefore chooses two computational paths from the three available representations, after the sensing policy has fixed the observed wavelength. It cannot activate an expert for an unrecorded channel. Task head H_t produces $\hat{y}_i^t = H_t(h_{i,t})$.

The routed experiments fit HR separately, so e_t is fixed to the HR task identity. Conditioning and mixture interventions probe this predictor's use of inputs and expert weights (Appendices A.5 and A.6). The PPG-only comparison in Table 1 evaluates the encoder with routing disabled and therefore measures a different component of the framework.

5 Training and Optimization

Figure 2 summarizes initialization, paired alignment, task adaptation, and fixed-model evaluation.

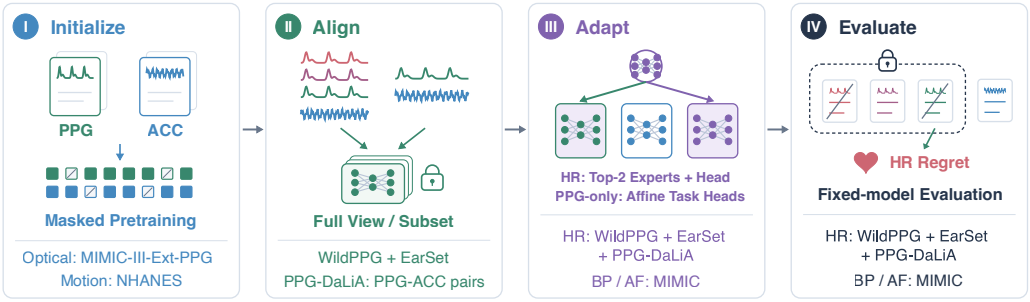


Figure 2: **Staged learning.** The Stage-II lock denotes frozen temporal backbones with trainable wavelength calibration. HR policy and routing experiments use top-2 experts; the separate PPG-only HR/BP/AF comparison uses frozen encoders and affine heads, except for the end-to-end supervised baseline. Stage IV fixes the evaluated checkpoints.

Stage I: channel initialization. Masked reconstruction on MIMIC-III-Ext-PPG initializes shared optical temporal processing (Moulaeifard et al., 2026c). Its single physical PLETH stream has unspecified wavelength and supplies no cross-wavelength supervision. NHANES supplies motion-initialization data.

Stage II: paired alignment. WildPPG and EarSet provide synchronized multi-wavelength PPG and motion; PPG-DaLiA contributes single-wavelength optical–motion pairs. Wavelength-specific projectors, normalization, residual adapters, and cross-channel calibration blocks are trained while the shared optical Transformer and motion encoder remain fixed. Reconstruction, paired alignment, subset consistency, and reconstruction-gate regularization form the objective (Appendix C).

Stage III: task-specific adaptation. Aligned and reconstruction-only HR models use one selected PPG channel plus ACC and the same adaptation schedule; frozen probes instead fit a 193-parameter

affine head. Adaptation updates the HR predictor using task and auxiliary losses. The separate seven-model PPG-only comparison freezes pretrained encoders, disables routing, and fits affine heads for HR, BP, and AF independently. BP/AF use MIMIC PPG. ResNet1D is trained end to end.

Stage IV: frozen evaluation. HR interventions use validation-selected checkpoints on matched test windows, retaining one PPG plus ACC except in sensor-removal controls. The PPG-only panel evaluates separate task heads on held-out test cohorts. Test data are excluded from all initialization, alignment, and task-fitting stages.

Dataset provenance and evaluation protocols are detailed in Appendices B.1 and B.4.

6 EXPERIMENTS

6.1 DATASETS, EVALUATION SETTINGS, AND METRICS

We evaluate PPG-only transfer to HR, systolic/diastolic BP (SBP/DBP), and AF with routing disabled. Separate paired HR experiments compare aligned (A) and reconstruction-only (R) recipes using selected PPG, ACC, and routing on identical windows and acquisition choices. These settings have different inputs and adaptation procedures, so their errors are reported separately (Appendix Table 12).

WildPPG, EarSet, and PPG-DaLiA contain 16, 30, and 15 participants, respectively (Meier et al., 2024b; Montanari et al., 2023; Reiss et al., 2019). The first two provide optical alternatives; DaLiA provides green PPG and motion. BP and AF use MIMIC PPG. Appendix B.1 separates public totals from the evaluated cohorts.

HR participant-macro MAE weights participants equally; pooled MAE and RMSE weight windows equally. BP reports SBP/DBP MAE, RMSE, and prediction-minus-target bias. AF uses record-level area under the precision–recall curve (AUPRC) and receiver operating characteristic curve (AUROC). All scores are held-out test results. Test data are excluded from every model's pretraining and downstream fitting; validation data select checkpoints. Metric definitions and comparison protocols appear in Appendices B.2 and B.4.

6.2 FOUNDATION-MODEL COMPARISON ACROSS HR, BP, AND AF

Table 1 compares PaPaGei-S, Pulse-PPG, AnyPPG, SIGMA-PPG Pro, MantisV2, MSMA-PPG, and supervised ResNet1D (Pillai et al., 2025; Saha et al., 2025; Nie et al., 2026; Guo et al., 2026; Feofanov et al., 2026). All foundation-model encoders are pretrained anew under our splits, then frozen for separate affine-head fitting on single-PPG input. ResNet1D trains end to end and matches MSMA-PPG's stored encoder parameters within 2%. The protocol specifies matched labelled-data and validation-search budgets within each task; adapters and pretraining provenance are detailed in Appendix B.4.

For AF, the source interval contains 1,024 samples at 128 Hz. The SIGMA-PPG Pro and Pulse-PPG adapters both resample this 8-second interval to 400 samples at 50 Hz. Pulse-PPG's input sampling rate thus matches its published evaluation. These preprocessing choices define the implementations compared here.

HR transfer across wearable cohorts. MSMA-PPG has the lowest reported participant-macro MAE on EarSet and PPG-DaLiA and the second-lowest on WildPPG. Relative to the ResNet1D summaries, its MAE is lower by 2.160 bpm on WildPPG, 3.602 bpm on EarSet, and 2.898 bpm on DaLiA, corresponding to reductions of 18.2%, 41.8%, and 23.9%. The same direction across the three datasets shows that the favorable HR results are not restricted to one dataset. This comparison concerns the complete pretrained representation and fitting procedure; it does not assign the reductions to an individual loss term.

The differences from the strongest pretrained alternatives are smaller. On EarSet, MSMA-PPG is below MantisV2 by 0.209 bpm (4.0%); on DaLiA, it is below AnyPPG by 0.045 bpm (0.5%). MantisV2 has the lowest WildPPG MAE, at 9.576 versus 9.693 bpm. MSMA-PPG's RMSE rank matches its MAE rank on EarSet and WildPPG, but differs on DaLiA: MSMA-PPG has an RMSE

Table 1: PPG-only test results. Bold/underlined entries denote the best/second-best reported values: lower errors and absolute bias, higher AF scores. HR uses participant-macro MAE and pooled RMSE (bpm); BP uses pooled errors and signed bias (mmHg); AF uses record-level metrics. MSMA-PPG uses aligned Stage-II weights.

(a) Heart rate and atrial fibrillation

	WildPPG		EarSet		PPG-DaLiA		AF	
Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	AUPRC	AUROC
ResNet1D	11.853	17.391	8.622	11.984	12.144	18.050	0.3624	0.6352
MantisV2	**9.576**	**13.991**	<u>5.229</u>	<u>9.119</u>	9.599	<u>12.929</u>	0.4922	0.8699
PaPaGei-S	12.098	16.221	9.596	15.212	14.742	17.942	0.3060	0.7999
Pulse-PPG	11.216	15.972	7.641	11.118	11.842	17.381	0.3905	0.6423
AnyPPG	11.507	15.818	6.462	10.732	<u>9.291</u>	**12.797**	**0.6348**	**0.9236**
SIGMA-PPG Pro	10.112	14.668	7.211	10.013	11.563	16.767	0.4252	0.6621
MSMA-PPG (ours)	<u>9.693</u>	<u>14.062</u>	**5.020**	**9.041**	**9.246**	13.116	<u>0.5177</u>	<u>0.8909</u>

(b) Blood pressure

	SBP			DBP		
Model	MAE	RMSE	Bias	MAE	RMSE	Bias
ResNet1D	18.241	22.686	-4.324	12.915	16.392	-2.051
MantisV2	16.293	20.377	-3.980	8.738	11.446	-1.272
PaPaGei-S	17.140	21.336	**-3.101**	9.165	11.926	<u>-1.253</u>
Pulse-PPG	16.939	21.402	-3.767	12.242	15.681	-1.620
AnyPPG	<u>15.820</u>	<u>19.761</u>	-3.235	**8.376**	**10.913**	**-0.789**
SIGMA-PPG Pro	16.220	20.093	<u>-3.188</u>	11.804	15.065	-1.352
MSMA-PPG (ours)	**15.456**	**19.697**	-3.596	<u>8.549</u>	<u>11.326</u>	-1.443

of 13.116 bpm, above AnyPPG's 12.797 and MantisV2's 12.929. Participant-macro MAE weights participants equally, whereas pooled RMSE weights windows equally and penalizes large errors more strongly. The two metrics therefore describe different aspects of performance, and the DaLiA result supports a lower average participant MAE rather than uniformly lower error under both aggregations.

BP transfer: target-specific accuracy and bias. MSMA-PPG has the lowest reported SBP MAE and RMSE, at 15.456 and 19.697 mmHg. Its SBP MAE is 0.364 mmHg (2.3%) below AnyPPG and 2.785 mmHg (15.3%) below ResNet1D. For DBP, AnyPPG has the lowest MAE, at 8.376 mmHg, followed by MSMA-PPG at 8.549 mmHg; the latter is 4.366 mmHg (33.8%) below ResNet1D. The SBP and DBP orderings differ, so the results distinguish the two pressure targets rather than assigning a single overall BP rank.

Absolute error and signed bias give different assessments. Every model has negative mean bias for both targets, indicating average underprediction in each reported evaluation. Despite its lowest SBP MAE, MSMA-PPG has an SBP bias of -3.596 mmHg, compared with -3.101 for PaPaGei-S. Its DBP bias is -1.443 mmHg, compared with AnyPPG's -0.789. Thus the lower SBP error does not coincide with the smallest systematic offset, and the DBP summaries favor AnyPPG in both error and absolute bias.

AF transfer: score discrimination. MSMA-PPG has the second-highest reported AUPRC (0.5177) and AUROC (0.8909). Its AUPRC exceeds MantisV2's by 0.0255, while AnyPPG leads at 0.6348, a 0.1171 gap over MSMA-PPG. AUROC gives the same leading trio: AnyPPG, MSMA-PPG, and MantisV2 score 0.9236, 0.8909, and 0.8699. These metrics evaluate score discrimination across thresholds (Appendix B.2); they do not establish performance at a diagnostic operating point or probability calibration.

Cross-task synthesis. MSMA-PPG ranks first or second on every reported primary endpoint, with the leading model varying by target. These separate-head, PPG-only results characterize the optical representation. The following experiments assess the complete routed HR predictor.

6.3 HR SELECTION AND ROUTING ANALYSIS

We compare aligned Stage-II (A) and reconstruction-only (R) training recipes on held-out test cohorts of two WildPPG, four EarSet, and two PPG-DaLiA participants. Each model is selected by minimum validation loss and fixed before intervention. Validation and test cohorts are separate. All model parameters remain fixed during the interventions. Unless specified otherwise, errors are pooled MAE in bpm. Between-arm differences are A minus R; intervention effects are relative to each arm's reference.

Table 2: Frozen affine-probe and adapted HR test errors (bpm; lower is better). MAE pools windows, whereas macro MAE weights participants equally. Each model is selected by validation loss and fixed for testing. The regimes differ in both adaptation and checkpoint selection. A: aligned; R: reconstruction-only.

		Pooled MAE (bpm)			Macro MAE (bpm)	
Dataset	Regime	A	R	A−R	A	R
WildPPG	Frozen affine	15.605	12.926	+2.679	15.603	12.913
EarSet	Frozen affine	10.291	10.913	-0.622	10.740	11.498
PPG-DaLiA	Frozen affine	10.478	12.710	-2.232	10.486	12.720
WildPPG	Adapted	9.748	8.951	+0.797	9.766	8.970
EarSet	Adapted	9.415	7.930	+1.485	10.350	9.055
PPG-DaLiA	Adapted	8.291	7.918	+0.373	8.305	7.933

Reference accuracy and frozen transfer. A's adapted MAE is higher by 0.797/1.485/0.373 bpm on WildPPG/EarSet/PPG-DaLiA; participant-macro MAE preserves this ordering. Frozen encoders with a 193-parameter affine head instead favor A on EarSet and PPG-DaLiA by 0.622 and 2.232 bpm (Table 2). Useful frozen features therefore need not yield better adapted predictors. The regimes differ in adaptation and checkpoint selection, so this reversal does not isolate the effect of alignment.

Response to wavelength selection. Seven policies at six attempted rates yield 129 matched conditions per arm, including references and single-wavelength controls. Each policy selects one recorded PPG stream with ACC unchanged; the downstream router processes that evidence. A has higher pooled and macro MAE in every pair, despite smaller excess-error areas in three of fourteen dataset–policy combinations (Figure 3). Smaller reference-relative regret thus need not imply lower absolute error.

On EarSet, worst-score selection changes MAE by $+0.299$ in A and -1.042 in R. Fixed-wavelength MAE favors green over infrared in A (8.566 versus 11.119 bpm) and infrared over green in R (6.772 versus 6.909; Table 6). The policy switches 1,036 of 2,187 windows from green to infrared; replaying their fixed-wavelength predictions attributes $+0.634$ bpm of A's error change and -0.644 of R's to this transition. Their 1.278-bpm contrast accounts for most of the 1.341-bpm difference between total policy effects. The opposite responses therefore reflect predictor-specific wavelength errors. Attempted corruption can leave the selected wavelength unchanged, and all 43 single-wavelength PPG-DaLiA conditions coincide within each arm. Appendices A.3 and A.4 detail the dominant transition and effective change rates.

Expert assignments and the role of motion. With acquisition fixed, swapping ACC/shared expert modules penalizes A more than cycling optical modules on every cohort (Appendix A.1). Deleting the ACC path and renormalizing surviving weights raises A's MAE by 6.210/8.287/10.339 bpm, versus $+0.101/+0.229/-0.499$ for R. Conditional ACC-input removal raises A's MAE by 1.476/0.601/0.502 bpm (Table 5). Path deletion preserves motion information in fused features, unlike ACC-input removal; these interventions measure different dependencies, and their error changes are not additive sensor contributions. R nevertheless leads in 18/21 paired path conditions

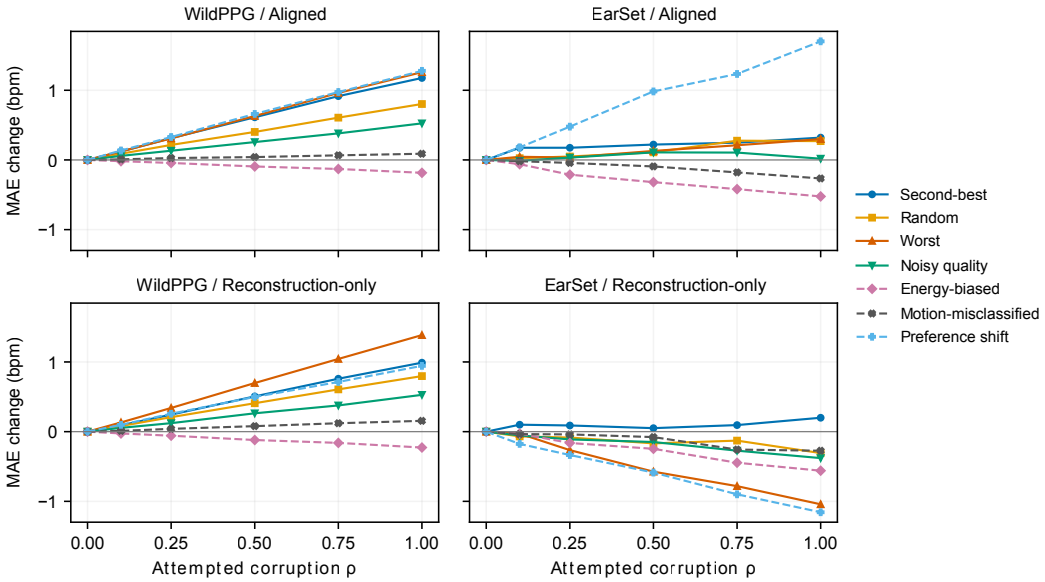


Figure 3: Pooled MAE changes from each arm's reference under matched wavelength-selection policies. Panels share axis scales; attempted corruption need not change the wavelength.

(Appendix A.7). A's stronger dependence on motion/fused pathways therefore accompanies worse reference accuracy.

Learned weighting and router conditioning. Equalizing A's original top-2 weights raises MAE by 0.043/0.964/0.184 bpm; top-1 raises it by 6.372/5.425/9.194 bpm. The learned mixture helps this fitted predictor, although R remains better in all fifteen paired routing conditions (Appendix A.6). Zeroing sensing covariates changes A's highest-weight expert in 43.60–48.47% of windows, yet raises MAE by 0.691 bpm on WildPPG and lowers it by 1.691/0.541 on EarSet/PPG-DaLiA (Appendix A.5). Removing availability covariates improves both arms' MAE on all cohorts while retaining the physical hard mask. Routing sensitivity thus does not consistently improve prediction. These frozen interventions characterize fitted-model dependence; alternative architectures would require retraining.

7 Conclusion

MSMA-PPG combines wavelength-specific calibration, a shared optical backbone, full-to-subset alignment, and availability-constrained routing. The reported PPG-only summaries give its representation the lowest HR participant-macro MAE on EarSet and PPG-DaLiA and the lowest SBP MAE, with second-ranked DBP MAE and AF AUPRC/AUROC. Rankings remain target-specific: MantisV2 leads WildPPG HR, while AnyPPG leads DBP and AF.

The routed HR experiments reveal a different outcome. Despite lower frozen-probe MAE on two cohorts, the adapted aligned model has higher MAE than reconstruction-only training throughout the paired policy grid. Its greater sensitivity to motion/shared pathways and mixture substitutions reflects reliance on the fitted configuration. Opposite EarSet responses to worst-score selection further show that policy scores and task error can disagree. Representation transfer, acquisition-policy effects, and routing dependence therefore require separate assessment.

AI USE STATEMENT

In this work, we used generative AI tools (OpenAI Codex) to assist with methodology and evaluation-protocol discussions, implementation of model adapters and data-preparation and evaluation code, interpretation of experiment summaries, and translation of manuscript sections for author review. We have not used generative AI tools to generate the physiological recordings, labels, or experimental measurements reported in this paper. Additionally, we used generative AI tools for code debugging and testing, literature searches, figure and table preparation, manuscript drafting and language editing, reference verification, and PDF layout checks.

We have reviewed the AI-assisted manuscript content. This review included checking manuscript claims against cited sources and available experiment summaries, verifying the transcription and arithmetic of reported results, and inspecting figures, tables, and rendered manuscript pages. We take responsibility for the final content of this work, including text, claims, code, figures, and other artifacts produced with the aid of generative AI.

REFERENCES

John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1–R39, 2007. doi: 10.1088/0967-3334/28/3/R01.

John Allen, Dingchang Zheng, Panicos A. Kyriacou, and Mohamed Elgendi. Photoplethysmography (PPG): State-of-the-art methods and applications. *Physiological Measurement*, 42(10):100301, 2021. doi: 10.1088/1361-6579/ac2d82.

Vasilii Feofanov, Songkang Wen, Jianfeng Zhang, Lujia Pan, and Ievgen Redko. MantisV2: Closing the zero-shot gap in time series classification with synthetic data and test-time strategies. *arXiv preprint arXiv:2602.17868*, 2026. doi: 10.48550/arXiv.2602.17868. URL https://arxiv.org/abs/2602.17868.

Eloy Geenjaar, Vince D. Calhoun, Scott Daly, Gouthaman KV, Lie Lu, Trisha Mittal, and Daniel P. Darcy. A robust PPG foundation model using multimodal physiological supervision. In *Forty-third International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/64846.

Zongheng Guo, Tao Chen, Yang Jiao, Yi Pan, Xiao Hu, and Manuela Ferrario. SIGMA-PPG: Statistical-prior informed generative masking architecture for PPG foundation model. In *Forty-third International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/64268.

Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1930–1939, 2018. doi: 10.1145/3219819.3220007.

Jukka-Tapani Mäkinen, Jaakko Vartiainen, Olli Heikkinen, Aleksi Vornanen, Heidi Similä, Edoardo D'Anna, Kirsi Maansaari, and Mika Kangas. Smart sensing in Oura Ring 4: Advancing personalized health monitoring through adaptive sensor technology. In *Biophotonics in Exercise Science, Sports Medicine, Health Monitoring Technologies, and Wearables VII*, volume 13845 of *Proceedings of SPIE*, pp. 1384505, 2026. doi: 10.1117/12.3076768.

Wanting Mao, Maxwell A. Xu, Harish Haresamudram, Mithun Saha, Santosh Kumar, and James Matthew Rehg. Prototype-based self-supervised multimodal learning for PPG and accelerometry signals. *arXiv preprint arXiv:2510.09764*, 2025. doi: 10.48550/arXiv.2510.09764. URL https://arxiv.org/abs/2510.09764.

Manuel Meier, Berken Utku Demirel, and Christian Holz. Tri-spectral PPG: Robust reflective photoplethysmography by fusing multiple wavelengths for cardiac monitoring. In *2024 IEEE 20th International Conference on Body Sensor Networks*, pp. 1–4, 2024a. doi: 10.1109/BSN63547.2024.10780586.

Manuel Meier, Berken Utku Demirel, and Christian Holz. WildPPG: A real-world PPG dataset of long continuous recordings. In *Advances in Neural Information Processing Systems: Datasets and Benchmarks Track*, volume 37, 2024b. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/0433292a3eb101df8f4d72a63f5410d4-Abstract-Datasets_and_Benchmarks_Track.html.

Alessandro Montanari, Andrea Ferlini, Ananta Narayanan Balaji, Cecilia Mascolo, and Fahim Kawsar. EarSet: A multi-modal dataset for studying the impact of head and facial movements on in-ear PPG signals. *Scientific Data*, 10:850, 2023. doi: 10.1038/s41597-023-02762-3.

Julie Mordacq, Leo Milecki, Maria Vakalopoulou, Steve Oudot, and Vicky Kalogeiton. ADAPT: Multimodal learning for detecting physiological changes under missing modalities. In *Proceedings of the 7th International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research*, pp. 1040–1055. PMLR, 2024. URL https://proceedings.mlr.press/v250/mordacq24a.html.

Mohammad Moulaeifard, Peter H. Charlton, and Nils Strodthoff. MIMIC-III-Ext-PPG: A PPG benchmark dataset for cardiorespiratory analysis. PhysioNet, version 1.0.0, 2026a. URL https://physionet.org/content/mimic-iii-ext-ppg/1.0.0/.

Mohammad Moulaeifard, Peter H. Charlton, and Nils Strodthoff. MIMIC-III-Ext-PPG: A PPG benchmark dataset for cardiorespiratory analysis. PhysioNet, version 1.1.0, 2026b. URL https://physionet.org/content/mimic-iii-ext-ppg/1.1.0/.

Mohammad Moulaeifard, Marie Kutscher, Philip J. Aston, Peter H. Charlton, and Nils Strodthoff. MIMIC-III-Ext-PPG, a PPG-based benchmark dataset for cardiovascular and respiratory signal analysis. *Scientific Data*, 13:668, 2026c. doi: 10.1038/s41597-026-07335-8.

National Center for Health Statistics. *Physical Activity Monitor (PAM) Procedures Manual*. Centers for Disease Control and Prevention, January 2012. URL https://wwwn.cdc.gov/nchs/data/nhanes/public/2011/manuals/2012-Physicial-Activity-Monitor-Procedures-Manual-508.pdf.

Natalia Neverova, Christian Wolf, Graham W. Taylor, and Florian Nebout. ModDrop: Adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1692–1706, 2016. doi: 10.1109/TPAMI.2015.2461544.

Guangkun Nie, Xiaocheng Fang, Gongzheng Tang, Yujie Xiao, Jun Li, Bo Liu, Hongyan Li, and Shenda Hong. AnyPPG: An ECG-guided PPG foundation model trained on over 100,000 hours of recordings for holistic health profiling. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, 2026. doi: 10.1145/3770855.3818911.

Arvind Pillai, Dimitris Spathis, Fahim Kawsar, and Mohammad Malekzadeh. PaPaGei: Open foundation models for optical physiological signals. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=kYwTmlq6Vn.

Petra Poklukar, Miguel Vasco, Hang Yin, Francisco S. Melo, Ana Paiva, and Danica Kragic. Geometric multimodal contrastive representation learning. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17782–17800. PMLR, 2022. URL https://proceedings.mlr.press/v162/poklukar22a.html.

Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14):3079, 2019. doi: 10.3390/s19143079.

Mithun Saha, Maxwell A. Xu, Wanting Mao, Sameer Neupane, James M. Rehg, and Santosh Kumar. Pulse-PPG: An open-source field-trained PPG foundation model for wearable applications across lab and field settings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9(3):1–35, 2025. doi: 10.1145/3749494.

Megha Thukral, Cyrus Tanade, Simon A. Lee, Juhyeon Lee, Hao Zhou, Keum San Chun, Migyeong Gwak, Viswam Nathan, Md Mahbubur Rahman, Li Zhu, Mehrab Bin Morshed, Subramaniam Venkatraman, and Sharanya Arcot Desai. Wavelet-driven masked multiscale reconstruction for PPG foundation models. *arXiv preprint arXiv:2601.12215*, 2026. doi: 10.48550/arXiv.2601.12215. URL https://arxiv.org/abs/2601.12215.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. doi: 10.48550/arXiv.1807.03748.

Yuxuan Weng, Wenhan Luo, and Qijia Shao. VCR: Learning valid contextual representation for incomplete wearable signals. *arXiv preprint arXiv:2605.18837*, 2026. doi: 10.48550/arXiv.2605.18837. URL https://arxiv.org/abs/2605.18837.

Maxwell A. Xu, Girish Narayanswamy, Kumar Ayush, Dimitris Spathis, Shun Liao, Shyam A. Tailor, Ahmed Metwally, A. Ali Heydari, Yuwei Zhang, Jake Garrison, Samy Abdel-Ghaffar, Xuhai Xu, Ken Gu, Jacob Sunshine, Ming-Zher Poh, Yun Liu, Tim Althoff, Shrikanth Narayanan, Pushmeet Kohli, Mark Malhotra, Shwetak Patel, Yuzhe Yang, James M. Rehg, Xin Liu, and Daniel McDuff. LSM-2: Learning from incomplete wearable sensor data. *arXiv preprint arXiv:2506.05321*, 2025. doi: 10.48550/arXiv.2506.05321. URL https://arxiv.org/abs/2506.05321.

Sukwon Yun, Inyoung Choi, Jie Peng, Yangfan Wu, Jingxuan Bao, Qiyiwen Zhang, Jiayi Xin, Qi Long, and Tianlong Chen. Flex-MoE: Modeling arbitrary modality combination via the flexible mixture-of-experts. In *Advances in Neural Information Processing Systems*, volume 37, pp. 98782–98805, 2024. doi: 10.52202/079017-3135. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/b2f2af5403042b1344f4e93b35fb67d9-Abstract-Conference.html.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 12310–12320. PMLR, 2021. URL https://proceedings.mlr.press/v139/zbontar21a.html.

APPENDIX

A TEST-SET DIAGNOSTICS

These experiments use the fixed adapted models and the 2/4/2-participant test cohorts described in Section 6.3. A denotes aligned and R reconstruction-only training. Unless stated otherwise, changes are intervention minus the corresponding arm's reference; positive error changes indicate worse prediction. Displayed errors and error differences are rounded to three decimals. All frozen interventions retain the model parameters and prediction head, so their effects describe sensitivity of the fitted predictor rather than performance after retraining.

Evaluation uses bfloat16 with TF32 and cuDNN benchmarking enabled. The initial policy, availability, and wavelength evaluations used 32 GPUs; the complete policy grid used 32 GPUs for A and eight for R. Frozen routing and expert interventions used eight A100 GPUs. Reference replays agreed on all 54,546 windows per arm, and repeated second-best/worst endpoints reproduced the earlier metrics. This agreement covers the repeated conditions; performance and timing comparisons across execution settings are outside this analysis.

A.1 GROUPED EXPERT-ASSIGNMENT SENSITIVITY

Table 3: Joint expert-assignment permutation on matched test windows. Optical modules are cycled and the ACC/shared modules are swapped, while expert parameters, inputs, physical masks, and gate values remain fixed. Errors and changes are in bpm; changes are relative to each arm's reference. A: aligned; R: reconstruction-only.

Dataset	Arm	Reference MAE	Permuted MAE	Δ MAE	Δ macro MAE
WildPPG	A	9.748	29.254	+19.506	+19.542
WildPPG	R	8.951	11.522	+2.571	+2.581
EarSet	A	9.415	27.070	+17.655	+18.425
EarSet	R	7.930	8.499	+0.569	-0.035
PPG-DaLiA	A	8.291	24.201	+15.910	+15.889
PPG-DaLiA	R	7.918	9.449	+1.531	+1.525

We permute the assignment of expert modules to representation slots while holding waveforms, acquisition choices, gate values, physical masks, and parameter values fixed. The optical intervention cycles red/infrared/green modules; the ACC/shared intervention exchanges those two modules; the joint intervention combines both. Together with the reference, these yield 12 conditions per arm across the three cohorts. The joint permutation reproduces the results in Table 3.

Table 4: Grouped expert-assignment permutations on matched test cohorts. Optical cycles red/infrared/green modules; ACC/shared swaps those two modules; joint combines both. Changes are pooled MAE differences from each arm's reference (bpm). Interaction is joint minus optical minus ACC/shared plus reference in absolute MAE, quantifying nonadditivity of the two permutations. A: aligned; R: reconstruction-only.

Dataset	Arm	Ref. MAE	Δ optical	Δ ACC/shared	Δ joint	Interaction
WildPPG	A	9.748	+4.299	+20.844	+19.506	-5.638
WildPPG	R	8.951	+1.027	+1.456	+2.571	+0.088
EarSet	A	9.415	+7.828	+15.444	+17.655	-5.617
EarSet	R	7.930	-0.872	+1.512	+0.569	-0.072
PPG-DaLiA	A	8.291	+9.202	+21.509	+15.910	-14.801
PPG-DaLiA	R	7.918	+0.287	+1.405	+1.531	-0.161

For A, exchanging ACC/shared modules increases MAE more than cycling optical modules on every cohort (Table 4). The joint increment is smaller than the sum of the separate increments: the MAE interaction is -5.638, -5.617, and -14.801 bpm on WildPPG, EarSet, and PPG-DaLiA. For R, ACC/shared swaps increase MAE by 1.405–1.512 bpm; optical changes are +1.027, -0.872, and +0.287 bpm, with interactions from -0.161 to +0.088 bpm. Participant-macro MAE preserves the

directions of the separate interventions. A has higher pooled and participant-macro MAE in all 12 matched conditions.

The large aligned-model penalty therefore reflects sensitivity to motion/fused assignments as well as wavelength assignments. The nonlinear predictor and absolute-error metric prevent partitioning these effects into additive contribution percentages. The analysis uses one optical cycle and one ACC/shared exchange; the improvement of R on EarSet after optical cycling illustrates that the original assignment is not uniformly best even among these choices. PPG-DaLiA can change because its green representation is processed by a different module despite unchanged acquisition.

A.2 Availability and fixed-wavelength sensitivity

Table 5: Conditional ACC-removal effects in bpm (PPG-only minus each arm's reference). Reference wavelength choices are fixed before removing ACC waveforms and derived motion features, so the intervention measures downstream reliance on ACC after acquisition. A: aligned; R: reconstruction-only.

Dataset	Arm	Δ MAE	Δ RMSE	Δ macro MAE
WildPPG	A	+1.476	+1.464	+1.481
WildPPG	R	+2.171	+3.588	+2.128
EarSet	A	+0.601	-0.646	+0.189
EarSet	R	-0.644	+1.076	-0.225
PPG-DaLiA	A	+0.502	-0.194	+0.490
PPG-DaLiA	R	-0.824	-0.678	-0.829

All evaluated windows originally contain ACC. We remove its waveform and derived motion features after fixing the reference wavelength, measuring the predictor's use of motion conditional on acquisition. R's MAE improves after removal on EarSet and PPG-DaLiA, whereas A's MAE worsens on all cohorts (Table 5). On EarSet, R's MAE decreases by 0.644 bpm while RMSE increases by 1.076 bpm, indicating different responses of average absolute error and the metric more sensitive to large errors. ACC remains available to the preceding acquisition decision, so this experiment does not evaluate an ACC-free sensing policy.

Table 6: Fixed recorded-wavelength test evaluation with unchanged ACC (bpm). All supported conditions within a dataset use the same cohort; dashes denote unavailable wavelengths. Each recorded wavelength is evaluated separately, while the primary acquisition policy remains fixed. Macro denotes participant-macro MAE. A: aligned; R: reconstruction-only.

Dataset	PPG	MAE A	MAE R	RMSE A	RMSE R	Macro A	Macro R
WildPPG	Red	11.049	10.611	14.951	14.665	11.044	10.614
WildPPG	Infrared	11.027	9.895	15.034	14.196	11.022	9.901
WildPPG	Green	9.606	8.725	13.930	13.201	9.627	8.745
EarSet	Red	9.180	9.265	12.535	13.011	9.800	10.087
EarSet	Infrared	11.119	6.772	13.981	11.730	11.358	7.748
EarSet	Green	8.566	6.909	13.613	13.039	9.352	7.812
PPG-DaLiA	Red	–	–	–	–	–	–
PPG-DaLiA	Infrared	–	–	–	–	–	–
PPG-DaLiA	Green	8.291	7.918	12.402	12.505	8.305	7.933

A has higher MAE in all three WildPPG fixed-wavelength conditions. EarSet is heterogeneous: A has slightly lower red MAE (9.180 versus 9.265), but substantially higher infrared MAE (11.119 versus 6.772). The same cohort is used for every available wavelength within a dataset, allowing wavelength-dependent prediction errors to be separated from cohort composition. These controls support the policy decomposition below; the primary policy comparison retains its original selection rule. PPG-DaLiA supplies only the green condition.

Table 7: Paired policy endpoints ($\rho = 1$), relative to each arm's reference (bpm). Second-best and worst are defined by frozen surrogate scores. Every eligible wavelength choice changes in WildPPG and EarSet. PPG-DaLiA has no alternative wavelength, giving zero change in every metric. Macro denotes participant-macro MAE. A: aligned; R: reconstruction-only.

Dataset	Policy	Δ MAE A	Δ MAE R	Δ macro A	Δ macro R	Δ RMSE A	Δ RMSE R
WildPPG	second-best	+1.176	+0.990	+1.157	+0.978	+0.954	+0.839
WildPPG	worst	+1.259	+1.386	+1.238	+1.373	+0.955	+1.119
EarSet	second-best	+0.321	+0.200	-0.175	-0.045	-1.011	-1.001
EarSet	worst	+0.299	-1.042	-0.366	-1.472	-0.823	-1.927

A.3 Policy endpoints and wavelength-transition decomposition

Each policy supplies one recorded PPG wavelength and unchanged ACC, with identical choices in both arms. On PPG-DaLiA, the reference and both endpoint policies produce identical within-arm predictions because no alternative wavelength is available.

For six arm–dataset groups and 18 policy-condition comparisons, we reconstruct policy predictions from the corresponding fixed-wavelength predictions. All replays match exactly, yielding zero residual in the transition-wise error decomposition. On EarSet, worst-score selection changes green to infrared in 1,036 of 2,187 windows. This transition contributes +0.634 bpm to A's pooled MAE change and -0.644 bpm to R's; their 1.278-bpm difference accounts for most of the 1.341-bpm contrast between total policy effects. The decomposition attributes the observed error difference to specific wavelength transitions; it does not identify the training mechanism that produced those preferences.

A.4 Complete paired policy grid

Table 8 complements Figure 3 with endpoint changes, effective change rates, and integrated excess error.

Table 8: Paired policy-grid experiments. Changed is the fraction of wavelength choices that differ from the reference at attempted $\rho = 1$; it is identical for both arms. Endpoint MAE changes and trapezoidal excess-error areas (bpm) are relative to each arm's reference. Areas integrate over attempted ρ; negative values denote lower error. A: aligned; R: reconstruction-only.

Dataset	Policy	Changed (%)	Δ A	Δ R	Area A	Area R
WildPPG	Second-best	100.00	+1.176	+0.990	+0.607	+0.500
WildPPG	Random	67.01	+0.803	+0.798	+0.407	+0.405
WildPPG	Worst	100.00	+1.259	+1.386	+0.632	+0.693
WildPPG	Noisy quality	50.24	+0.525	+0.528	+0.259	+0.257
WildPPG	Energy-biased	12.94	-0.185	-0.228	-0.090	-0.114
WildPPG	Motion-misclassified	14.75	+0.089	+0.155	+0.044	+0.079
WildPPG	Preference shift	81.30	+1.278	+0.944	+0.651	+0.484
EarSet	Second-best	100.00	+0.321	+0.200	+0.214	+0.091
EarSet	Random	67.35	+0.270	-0.307	+0.143	-0.138
EarSet	Worst	100.00	+0.299	-1.042	+0.135	-0.526
EarSet	Noisy quality	58.21	+0.018	-0.379	+0.061	-0.182
EarSet	Energy-biased	18.34	-0.524	-0.560	-0.300	-0.279
EarSet	Motion-misclassified	26.15	-0.265	-0.273	-0.112	-0.130
EarSet	Preference shift	94.97	+1.702	-1.156	+0.885	-0.605

Each arm has 43 conditions per dataset: one reference plus seven policies at six attempted rates, giving 129 evaluations per arm. Cohorts, wavelength choices, and actual change counts match in every pair. All 21 zero-rate checks reproduce the reference predictions, and all 43 PPG-DaLiA conditions coincide within each arm. These repeated conditions verify consistency but do not increase the number of distinct perturbations or independent participants.

A has higher pooled and participant-macro MAE in all 129 pairs. The largest pooled gap occurs at the EarSet preference-shift endpoint: 11.117 bpm for A versus 6.774 for R, with a reported difference of 4.344 bpm. Relative to their references, this policy changes MAE by +1.702 and -1.156 bpm, respectively; the excess-error areas are +0.885 and -0.605 bpm. Identical wavelength choices therefore worsen A while improving R, further demonstrating that a policy's downstream effect depends on the fitted predictor.

A has a smaller excess-MAE area than R for WildPPG worst-score and motion-misclassification policies and for EarSet energy bias; its area is larger in the other eleven dataset–policy combinations. The three smaller areas still accompany higher absolute A MAE at every measured rate. Effective intervention strength also varies: energy bias changes only 12.94% of WildPPG choices, whereas second-best and worst-score selection change 100%. Areas integrate excess error over attempted corruption by the trapezoidal rule and summarize these curves without uncertainty estimates. Energy bias refers to a simulated policy preference; the experiment measures prediction error, without measuring device energy consumption.

A.5 Frozen router-conditioning removal

Each arm has twelve evaluations: the reference and separate zeroing of task, availability, or sensing covariates on each cohort. Optical choices, ACC waveforms, and physical availability remain fixed. Covariates are zeroed before gate-input LayerNorm, so each effect includes the resulting normalization response. The physical hard mask is retained when availability covariates are removed, and invalid-expert gate mass remains zero throughout.

Table 9: Frozen router-conditioning removal. Each covariate is zeroed before gate-input LayerNorm while acquisition choices, physical masks, and model parameters remain fixed. Error changes are intervention minus each arm's reference (bpm); macro denotes participant-macro MAE. Top-1 is the percentage of windows whose highest-weight expert changes. A: aligned; R: reconstruction-only.

Dataset	Zeroed	Δ MAE A	Δ MAE R	Δ macro A	Δ macro R	Top-1 A (%)	Top-1 R (%)
WildPPG	Task	+0.065	+0.012	+0.068	+0.012	4.64	0.79
WildPPG	Availability	-0.071	-0.044	-0.078	-0.046	11.35	2.58
WildPPG	Sensing	+0.691	-0.123	+0.659	-0.133	43.60	19.50
EarSet	Task	+0.245	+0.023	+0.203	+0.014	7.77	0.82
EarSet	Availability	-0.495	-0.094	-0.419	-0.101	14.22	3.38
EarSet	Sensing	-1.691	-0.503	-1.529	-0.521	48.47	21.81
PPG-DaLiA	Task	+0.025	+0.016	+0.025	+0.016	4.92	0.07
PPG-DaLiA	Availability	-0.084	-0.053	-0.085	-0.053	9.59	1.27
PPG-DaLiA	Sensing	-0.541	-0.319	-0.548	-0.319	46.84	4.29

Task-token removal increases pooled MAE by 0.025–0.245 bpm in A and 0.012–0.023 bpm in R. Removing availability features lowers pooled and participant-macro MAE on all three cohorts in both arms. Sensing removal is dataset-dependent in A: pooled MAE changes by $+0.691$, -1.691, and -0.541 bpm on WildPPG, EarSet, and PPG-DaLiA; corresponding R changes are -0.123, -0.503, and -0.319 bpm. Participant-macro MAE agrees in direction. Thus sensing covariates help A on WildPPG but do not consistently improve MAE elsewhere. RMSE reveals additional variation: R's WildPPG RMSE rises by 0.006 bpm after sensing removal, and A's PPG-DaLiA RMSE rises by 0.065 bpm after availability removal despite lower MAE.

Sensing removal changes the highest-weight expert in 43.60%, 48.47%, and 46.84% of A windows, versus 19.50%, 21.81%, and 4.29% for R. A nevertheless has higher pooled and participant-macro MAE in all twelve paired conditions. On EarSet, sensing removal narrows the A-minus-R pooled MAE gap from 1.485 to 0.298 bpm without reversing it. The router's substantial response to sensing features therefore does not consistently improve prediction. These single-task HR interventions assess dependence on the existing conditioning; they do not test joint-task learning. PPG-DaLiA remains sensitive because its router can change the mixture of optical, ACC, and shared paths even with a fixed wavelength.

A.6 Frozen routing selection and weighting

Each arm has fifteen evaluations: the learned top-2 reference and four interventions per cohort. Optical selection, ACC input, hard masks, and router covariates are preserved. Equal top-2 retains the learned nonzero support with equal weights. Soft available applies softmax to the original logits over all available experts; equal available assigns them uniform weights. Top-1 gives unit weight to the original highest-logit expert. In every case, the available set consists of the recorded-wavelength, ACC, and shared experts; unrecorded optical channels remain excluded.

Table 10: Frozen routing selection and weighting relative to each arm's learned top-2 reference (bpm). Equal top-2 preserves the selected experts with uniform weights; soft/equal available use all physically available experts with softmax/uniform weights; top-1 retains only the highest-logit expert. Macro denotes participant-macro MAE. Model parameters remain fixed. A: aligned; R: reconstruction-only.

Dataset	Routing	Δ MAE A	Δ MAE R	Δ macro A	Δ macro R	Δ RMSE A	Δ RMSE R
WildPPG	Equal top-2	+0.043	+0.330	+0.039	+0.316	+0.152	+0.291
WildPPG	Soft available	+0.820	+0.155	+0.833	+0.160	+0.361	+0.010
WildPPG	Equal available	+1.526	+1.312	+1.546	+1.325	+0.908	+0.807
WildPPG	Top-1	+6.372	+1.064	+6.297	+1.071	+6.879	+1.222
EarSet	Equal top-2	+0.964	-0.439	+1.101	-0.399	+1.371	-0.753
EarSet	Soft available	+4.047	+0.395	+3.561	+0.364	+2.519	-0.057
EarSet	Equal available	+5.585	+0.283	+5.333	+0.216	+3.130	-1.669
EarSet	Top-1	+5.425	+0.341	+3.604	+0.309	+4.705	+0.495
PPG-DaLiA	Equal top-2	+0.184	-0.476	+0.185	-0.477	+0.432	-1.045
PPG-DaLiA	Soft available	+2.859	-0.037	+2.860	-0.037	+1.970	-0.379
PPG-DaLiA	Equal available	+2.388	-0.706	+2.391	-0.712	+0.557	-2.396
PPG-DaLiA	Top-1	+9.194	+0.095	+9.176	+0.089	+8.827	-0.671

At fixed top-2 support, equalizing weights raises A's pooled MAE by 0.043, 0.964, and 0.184 bpm on WildPPG, EarSet, and PPG-DaLiA. For R, it raises WildPPG MAE by 0.330 bpm but lowers EarSet and PPG-DaLiA MAE by 0.439 and 0.476 bpm. Learned relative weights therefore help A compared with equal weighting on the same support, while their effect in R depends on the cohort. Comparing the two equal-weight rules, expanding top-2 support to all available experts changes A's MAE by +1.483/+4.621/+2.204 bpm and R's by +0.983/+0.722/-0.230 bpm. This comparison changes both the support and the normalization of retained weights.

Top-1 causes substantially larger MAE increases in A (+6.372/+5.425/+9.194 bpm) than R (+1.064/+0.341/+0.095 bpm); participant-macro MAE agrees in direction. A nevertheless has higher pooled and participant-macro MAE in all fifteen paired conditions. Dependence on the learned mixture is therefore stronger in A, while absolute accuracy favors R. Other metric orderings remain mixed: R's PPG-DaLiA top-1 intervention increases MAE by 0.095 bpm while lowering RMSE by 0.671 bpm, and its EarSet equal-available intervention increases MAE by 0.283 bpm while lowering RMSE by 1.669 bpm.

The experts and head remain adapted to the original top-2 mixture throughout. The resulting penalties measure sensitivity to substitution within the fitted predictor. Uniform weighting also creates ties, so the index returned by argmax in these conditions has no interpretation as an expert preference. PPG-DaLiA retains three available paths despite its single optical wavelength.

A.7 Frozen expert-path deletion and forced paths

Each arm has 21 conditions: the reference, deletion of each optical/ACC/shared path, and forced use of each path on all three cohorts. Across both arms, these give 36 intervention-minus-reference differences and 21 A-minus-R comparisons.

Deletion zeros the specified path in the original top-2 and renormalizes surviving gates without activating another expert. A forced path receives unit weight even if the original router did not select it. Waveforms, acquired wavelength, physical availability, and router covariates remain fixed. Invalid-expert gate mass stays zero, and deleting an already inactive path leaves predictions unchanged.

Table 11: Frozen expert-path interventions relative to each arm's reference (bpm). Drop removes a path from the original top-2 and renormalizes surviving weights without selecting a replacement; Only assigns unit weight to the specified available path. Inputs and physical masks remain fixed. Macro denotes participant-macro MAE. A: aligned; R: reconstruction-only.

Dataset	Path	Δ MAE A	Δ MAE R	Δ macro A	Δ macro R	Δ RMSE A	Δ RMSE R
WildPPG	Drop optical	+2.675	+3.020	+2.610	+2.955	+2.807	+3.139
WildPPG	Drop ACC	+6.210	+0.101	+6.169	+0.100	+6.305	+0.199
WildPPG	Drop shared	+3.489	+1.710	+3.459	+1.740	+4.673	+1.712
WildPPG	Only optical	+9.002	+1.773	+8.944	+1.801	+8.989	+1.828
WildPPG	Only ACC	+7.207	+9.652	+7.101	+9.701	+8.177	+9.039
WildPPG	Only shared	+1.716	+3.391	+1.722	+3.319	+1.265	+3.534
EarSet	Drop optical	-0.289	-0.522	-0.834	-0.775	-0.764	-0.845
EarSet	Drop ACC	+8.287	+0.229	+7.962	+0.414	+6.323	+0.470
EarSet	Drop shared	+1.441	+0.671	+0.423	+0.825	+2.080	+0.396
EarSet	Only optical	+12.419	+0.870	+11.710	+1.185	+9.712	+0.797
EarSet	Only ACC	+1.534	+6.544	+1.381	+6.639	+4.069	+4.389
EarSet	Only shared	+4.039	-1.247	+3.320	-1.812	+2.032	-1.655
PPG-DaLiA	Drop optical	+0.293	+0.052	+0.282	+0.036	-1.256	-1.619
PPG-DaLiA	Drop ACC	+10.339	-0.499	+10.350	-0.503	+10.148	-0.916
PPG-DaLiA	Drop shared	+4.575	+0.617	+4.557	+0.614	+5.621	+0.254
PPG-DaLiA	Only optical	+14.914	+0.117	+14.907	+0.112	+13.666	-0.642
PPG-DaLiA	Only ACC	+5.534	+4.112	+5.496	+4.107	+5.937	+2.547
PPG-DaLiA	Only shared	+1.671	+0.555	+1.673	+0.533	-0.082	-0.112

ACC-path deletion raises A's MAE by 6.210/8.287/10.339 bpm on WildPPG/EarSet/PPG-DaLiA, versus +0.101/+0.229/-0.499 for R. Participant-macro MAE agrees in direction. This stronger reliance on A's ACC path reflects both the effect when active and its frequency of selection, since inactive-path deletion leaves predictions unchanged. Shared-path deletion raises pooled and macro MAE in both arms on all cohorts. Optical-path deletion is heterogeneous: EarSet pooled MAE falls by 0.289 bpm in A and 0.522 in R, despite increasing on the other cohorts.

Forcing the optical path raises A's MAE by 9.002/12.419/14.914 bpm, much more than R's 1.773/0.870/0.117 bpm. A has lower pooled and macro MAE under only-ACC on WildPPG and EarSet and under only-shared on WildPPG; R leads in the other 18 of 21 pairs, including all references. EarSet only-ACC still has slightly higher A RMSE (+0.115 bpm), so its MAE advantage does not extend to all metrics. R's EarSet only-shared MAE is 6.683 bpm, 1.247 below its reference, whereas every forced-path condition increases A's MAE. Thus A benefits from combining paths within its fitted predictor even though R generally achieves lower absolute error.

Expert paths and sensor inputs differ: fused/shared representations can retain cross-sensor information when a path is deleted. Path deletion also renormalizes the mixture. These properties explain why path interventions and conditional ACC-input removal answer different questions and why their error changes are not additive sensor contributions. All comparisons remain conditional on the fitted models and test participants described above.

B Evaluation Details

This section documents the datasets, metric definitions, model adapters, and source information for the HR, BP, and AF results. The paired HR experiments compare A and R on matched windows and acquisition decisions. The seven-model PPG-only comparison evaluates representation transfer with separate task heads.

B.1 Datasets and stage allocation

Table 13 distinguishes published dataset sizes from the participants and windows used in the paired HR experiments. WildPPG's 216 hours count synchronized recording time across participants; summing its four simultaneous devices would count the same time repeatedly. EarSet's bilateral sensor

Table 12: Evaluation settings. All tasks use PPG alone in the seven-model comparison. Paired HR diagnostics additionally use ACC.

Task	Cohort	Observed model input	Primary endpoint
HR	WildPPG, EarSet, PPG-DaLiA	PPG only; selected PPG + ACC for diagnostics	Macro MAE (PPG-only); pooled MAE (diagnostics), in bpm
BP	MIMIC	PPG	SBP/DBP MAE (mmHg)
AF	MIMIC	PPG	Record-level AUPRC

channels likewise do not increase its participant count. Dataset sizes follow the original publications (Meier et al., 2024b; Montanari et al., 2023; Reiss et al., 2019).

MIMIC-III-Ext-PPG v1.1.0 contains 6,189 subjects and 6,399,754 30-second segments (Moulaeifard et al., 2026c;b). The earlier v1.0.0 release contains 6,131 patients and approximately 4.9 million segments (Moulaeifard et al., 2026a). These published source totals are distinct from the samples used for training or BP/AF evaluation. NHANES supplies motion-initialization data. Its Physical Activity Monitor component collects wearable motion measurements (National Center for Health Statistics, 2012); this citation describes the source's monitoring component, without specifying the training cohort.

Table 13: Published dataset sizes and stage assignments. The last column gives the paired HR test cohorts. MIMIC release totals describe v1.1.0; they are not the number of training or downstream evaluation examples. N/A indicates that the source is not an HR test cohort.

Source	Published source scale	Stage / role	HR test: people / windows
MIMIC-III-Ext-PPG	6,189 subjects; 6.4M 30-s segments (v1.1)	I: optical; III/IV: BP, AF	N/A
NHANES	Wearable activity-monitoring survey	I: motion	N/A
WildPPG	16 people; 216 h	II; III/IV: HR	2 / 44,223
EarSet	30 people; ∼17 h	II; III/IV: HR	4 / 2,187
PPG-DaLiA	15 people	II; III/IV: HR	2 / 8,136

WildPPG and EarSet provide the paired optical channels used for cross-wavelength calibration. PPG-DaLiA provides green PPG–ACC pairs. The paired HR test set contains 54,546 windows from eight participants: two from WildPPG, four from EarSet, and two from PPG-DaLiA. The window count measures prediction instances, whereas the participant count measures independent people. The counts in this table apply to the paired HR experiments and are not assigned to the separate seven-model comparison.

B.2 Metrics and aggregation

For a regression target, let P be the number of participants, $e_{ij} = \hat{y}_{ij} - y_{ij}$ the prediction error for window j of participant i, n_i the number of windows for participant i, and $N = \sum_{i=1}^{P} n_i$. The reported aggregations are

$$\mathrm{MAE}_{\mathrm{pooled}} = \frac{1}{N}\sum_{i=1}^{P}\sum_{j=1}^{n_i}|e_{ij}|, \qquad \mathrm{MAE}_{\mathrm{macro}} = \frac{1}{P}\sum_{i=1}^{P}\frac{1}{n_i}\sum_{j=1}^{n_i}|e_{ij}|, \tag{11}$$

$$\mathrm{RMSE}_{\mathrm{pooled}} = \sqrt{\frac{1}{N}\sum_{i=1}^{P}\sum_{j=1}^{n_i}e_{ij}^2}, \qquad \mathrm{Bias}_{\mathrm{pooled}} = \frac{1}{N}\sum_{i=1}^{P}\sum_{j=1}^{n_i}e_{ij}. \tag{12}$$

HR errors are measured in bpm and BP errors in mmHg. Positive bias denotes overprediction and negative bias denotes underprediction; a smaller absolute bias indicates a smaller mean offset. SBP and DBP are evaluated separately. In the paired experiments, an intervention effect is its metric minus the same arm's reference metric; a between-arm difference is A minus R. For MAE and RMSE, negative differences indicate lower error for the intervention or for A, respectively.

AF uses record-level AUPRC and AUROC consistently across all seven models. AUPRC summarizes precision and recall for AF-positive records; AUROC summarizes the true-positive/false-positive-rate relationship. These metrics measure discrimination across thresholds. Threshold-dependent classification scores and probability calibration are not evaluated in Table 1.

B.3 Acquisition policies and frozen interventions

The HR reference policy computes a quality–energy–motion score from all candidate PPG waveforms recorded over the same window, selects one wavelength, and exposes only that waveform and ACC to the predictor. The simulated energy costs are policy parameters. The seven interventions select the second-best, a random, or the worst-scoring wavelength, or modify quality scores, energy preferences, motion classification, or wavelength preferences. At attempted corruption rate ρ, the challenger replaces the reference decision with probability ρ. A replacement can leave the wavelength unchanged, so attempted and effective change rates are reported separately. The two multi-wavelength datasets measure responses to changed optical choices; PPG-DaLiA provides a no-change control for this experiment.

Router-conditioning, expert-assignment, mixture-weight, and path interventions preserve the selected waveform and model parameters. Their differences measure responses of a fitted predictor to the specified intervention. Deleting an ACC expert is distinct from deleting ACC input: path deletion changes the mixture while fused representations still contain motion information. The routing experiments compare fixed configurations within A and R; neither model is retrained after an intervention. Details and all per-dataset results appear in Appendix A.

B.4 Model adapters and comparison protocol

The seven-model roster comprises supervised ResNet1D, MantisV2, PaPaGei-S, Pulse-PPG, AnyPPG, SIGMA-PPG Pro, and aligned MSMA-PPG. In the reported PPG-only protocol, pretrained encoders are frozen, routing is disabled, and separate affine heads are fitted for HR, paired SBP/DBP outputs, and AF. The models use separate task datasets and labels. ResNet1D is optimized end to end and matches MSMA-PPG's stored encoder parameter count within 2%; this match concerns parameters rather than executed operations.

Training–test separation. All foundation-model encoders in this comparison are pretrained anew under our data splits. The test examples are excluded from every training stage, including MSMA-PPG's channel initialization, paired alignment, and task adaptation, as well as baseline pretraining and downstream head fitting. Validation data are used for checkpoint selection; the held-out test data are used for the reported seven-model results and paired HR diagnostics. The evaluated checkpoints therefore follow the study's training–test partition throughout pretraining and task fitting. The source-scale figures in Table 13 are not the amounts used for pretraining. Appendix B.5 describes the MIMIC source checks.

Published pretraining sources. The original model publications describe different pretraining corpora. MantisV2 uses two million synthetic CauKer sequences (Feofanov et al., 2026). PaPaGei includes cleaned MIMIC-III Waveform Matched Subset data (Pillai et al., 2025). Pulse-PPG's main model uses MOODS field recordings, with a separate MIMIC-based clinical-pretraining comparison (Saha et al., 2025). AnyPPG includes PulseDB, which combines MIMIC-III and VitalDB sources (Nie et al., 2026). SIGMA-PPG reports VitalDB and the MIMIC-III Waveform Matched Subset (Guo et al., 2026). MIMIC-III-Ext-PPG is itself derived from the Matched Subset (Moulaeifard et al., 2026c); sharing an upstream source does not establish use of the full Ext-PPG release. These published corpora describe the original models, while our evaluated encoders follow the study-specific retraining and test-exclusion protocol above.

For AF, each source interval contains 1,024 samples at 128 Hz, corresponding to eight seconds. Both the SIGMA-PPG Pro and Pulse-PPG adapters resample this interval to 400 samples at 50 Hz. Pulse-PPG therefore uses the same input sampling rate as its published evaluation. The SIGMA-PPG Pro adapter produces four patches; its pretraining checkpoint accepts 12,000 samples, so the adapted eight-second input is a separate downstream setting. Scores in Table 1 refer to these adapters and task settings.

B.5 MIMIC SOURCE OVERLAP CHECKS

MIMIC provides Stage-I optical initialization data and downstream BP/AF data. Its processed PLETH aliases refer to one physical waveform, even when they occupy multiple compatibility slots. These aliases supply neither independent optical measurements nor cross-wavelength supervision.

The split protocol excludes downstream test data from Stage-I training. Cross-split checks reported zero overlap in the reviewed identity, time, and source-window comparisons. Table 14 reports an additional 64-bit sketch census and a sampled waveform review. Among 330 sketch-hit pairs, none are exact processed matches, one has correlation $r \geq 0.995$, and 318 have $0.9 \leq r < 0.995$. Among 96 nonhit controls, none are exact or high-similarity matches and one lies in the intermediate range. These are similarity checks: the census identifies candidate pairs, and the waveform review describes only the sampled pairs. Similarity counts are not counts of reused test examples and do not replace the training–test partition rule.

Table 14: MIMIC cross-split checks. The candidate census counts pairs satisfying the sketch-distance threshold, not confirmed duplicate waveforms or unique patients. The waveform review describes the separately sampled pairs shown in panel (b).

(a) Candidate census: 64-bit sketch, Hamming distance ≤ 4		
Split comparison	Candidate pairs	Per million possible pairs
Train / validation	85,912,322	1.805174
Train / test	74,205,758	1.984799
Validation / test	14,713,538	1.947410
(b) Processed-waveform review: sampled pairs only		
Outcome	Sketch hits ($n = 330$)	Nonhit controls ($n = 96$)
Exact processed match	0	0
High similarity ($r \geq 0.995$)	1	0
Intermediate similarity ($0.9 \leq r < 0.995$)	318	1
Weak similarity ($r < 0.9$)	11	95

C TRAINING OBJECTIVE DETAILS

Task adaptation. For task t, let $\mathcal{I}_t$ contain target elements with available labels and finite values. Each nonempty task loss is averaged over its eligible elements, and the general objective combines them with weights α_t:

$$\mathcal{L}_{\text{task}} = \sum_{t:\,|\mathcal{I}_t|>0} \alpha_t \frac{1}{|\mathcal{I}_t|} \sum_{j \in \mathcal{I}_t} \ell_t^{\text{train}}(\hat{y}_j^t, y_j^t). \tag{13}$$

The training loss ℓ_t^{train} is distinct from the evaluation loss in Equation 2. Regression uses smooth-L_1 after dividing predictions and targets by 5 bpm for HR, 8 mmHg for SBP, and 5 mmHg for DBP. Classification uses masked cross-entropy. The tasks are fitted separately, so this notation does not imply joint optimization across HR, BP, and AF. These losses describe adaptation; the separate frozen affine-head comparison follows Appendix B.4.

Paired representation alignment. For standardized feature matrices Z^a, Z^b over valid synchronized windows, let C^{ab} be their cross-correlation matrix. Following Barlow Twins (Zbontar et al., 2021), define

$$\mathcal{B}(Z^a, Z^b) = \sum_j (1 - C_{jj}^{ab})^2 + \eta \sum_{j \neq k} (C_{jk}^{ab})^2. \tag{14}$$

The diagonal term encourages agreement between corresponding features, while the off-diagonal term, weighted by η, discourages redundant coordinates. Spectral consistency $\mathcal{L}_{\mathrm{spec}}$ averages $\mathcal{B}$ over eligible wavelength pairs. PPG–motion alignment $\mathcal{L}_{\mathrm{pm}}$ uses the same criterion by default, with a configurable symmetrized InfoNCE alternative (van den Oord et al., 2018). In the latter, synchronized optical and motion embeddings form positive pairs, with the matched pair included in each softmax denominator.

Full-to-subset consistency matches normalized fused embeddings:

$$\mathcal{L}_{\mathrm{cons}} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \left(1 - \frac{\langle z_i^S, \mathrm{sg}(z_i^F) \rangle}{\|z_i^S\|_2 \, \|\mathrm{sg}(z_i^F)\|_2} \right), \tag{15}$$

where $\mathcal{I}$ contains valid paired examples and normalization uses a numerical stability floor. Gradients update the subset branch through embedding consistency, with the full-view embedding providing a stopped-gradient target.

Reconstruction and combined objective. Temporal reconstruction $\mathcal{L}_{\mathrm{temp}}$ and channel reconstruction $\mathcal{L}_{\mathrm{chan}}$ use smooth-L_1 on artificially hidden patches and channels that were present in the source recording. Naturally missing channels supply no reconstruction targets. In the equation below, $\mathcal{L}_{\mathrm{gate}}$ denotes entropy and load-balance regularization of the two-way optical/motion reconstruction gate; it does not regularize the downstream five-expert CA-MoE router. The Stage-II objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{II}} = \lambda_{\mathrm{temp}} \mathcal{L}_{\mathrm{temp}} + \lambda_{\mathrm{chan}} \mathcal{L}_{\mathrm{chan}} + \lambda_{\mathrm{spec}} \mathcal{L}_{\mathrm{spec}} \\ + \lambda_{\mathrm{pm}} \mathcal{L}_{\mathrm{pm}} + \lambda_{\mathrm{cons}} \mathcal{L}_{\mathrm{cons}} + \mathcal{L}_{\mathrm{gate}}. \end{aligned} \tag{16}$$

The coefficients $\lambda_{\cdot}$ control the relative contributions of reconstruction and representation alignment. Stage I uses unimodal reconstruction, while Stage III combines the valid-label task loss with the configured auxiliary terms. The A-versus-R comparison consequently evaluates the two training recipes as a whole.