PULSE: SELF-SUPERVISED CLINICAL PHASE DISCOVERY FOR MULTIMODAL ICU DISEASE PROGRESSION MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

ICU disease progression often evolve through clinically distinct phases, characterized by different physiological states and diagnostic and treatment needs. These phases, however, are not explicitly delineated in routinely collected data and must be inferred from asynchronous multimodal observations. Observation intensity also varies substantially over time, leaving some phases densely sampled and others sparsely observed. Models that directly aggregate observed events can therefore be biased to densely sampled phases in the learned trajectory representations, potentially obscuring the underlying clinical dynamics. We propose PULSE, a self-supervised framework that explicitly discovers latent clinical phases from asynchronous multimodal observations. These phases are shared across patients and modalities and provide a common structure for representing evolving patient states. We evaluate PULSE on MIMIC-IV and eICU across five dynamic prediction tasks: AKI progression, sepsis progression, vasopressor initiation, intubation need, and remaining length of stay. PULSE achieves the best overall performance among the evaluated baselines, supporting the value of phase-aware representations for modeling and predicting evolving patient states from irregular multimodal disease trajectories.

1 INTRODUCTION

Critically ill patients rarely deteriorate or recover along a smooth path. Instead, their ICU stays unfold through clinically distinct phases, such as deterioration, treatment escalation, stabilization, and recovery (Malbrain et al., 2018). Each phase calls for different care. During deterioration, clinicians intensify monitoring and may initiate fluid resuscitation or vasopressor support; during recovery, monitoring becomes less frequent as care shifts to ventilator weaning and discharge planning (Evans et al., 2021). Identifying the current phase is therefore central to predicting disease progression and upcoming interventions (Lyons et al., 2022). Inferring phases from routine ICU records, however, poses two challenges.

(1) Latent clinical phases. Phases are not annotated in ICU records and must be discovered without supervision from asynchronous multimodal observations, such as hourly vital signs and laboratory tests, and intermittent chest radiographs (CXR), electrocardiograms (ECG), and clinical notes. Existing latent-state and prototype models (Li et al., 2020; Wang et al., 2025) learn such states through task-driven or generic alignment objectives, without clinical knowledge of which phases exist or how they relate. The learned states are therefore not guaranteed to correspond to clinical phases or to be comparable across patients.

(2) Observation-intensity bias. Observation intensity reflects not only the patient's state but also clinical workflow and institutional practice (Agniel et al., 2018), so similar intensities can arise in different phases. For example, two patients may both undergo frequent portable chest radiography, one to verify endotracheal tube placement during treatment escalation and the other to evaluate suspected pneumonia during respiratory deterioration. Event-based multimodal models (Lee et al., 2023; Zhang et al., 2023; Susetzky et al., 2025) aggregate observations at the event level, so each period contributes in proportion to its number of observations, and the learned representation conflates observation intensity with patient state. Indeed, in three of our ten dataset–task settings, an EHR-only

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.

Transformer outperforms all seven multimodal baselines (Table 1), indicating that naively adding asynchronous modalities can degrade prediction.

These challenges compound: without phase labels, a model can organize patient states only through the observations themselves, whose intensity is biased. Our key idea is to represent each stay by the latent phase of every hour, where phases are anchored by clinical knowledge and shared across patients, and every hour carries equal weight in phase assignment. Observation intensity then determines how much evidence a period provides, but not how much the period counts, and sparsely observed periods can be interpreted through phases learned across the cohort.

We instantiate this idea in PULSE, a self-supervised framework for discovering shared latent clinical phases from asynchronous multimodal ICU trajectories. PULSE proceeds in two stages. **Stage 1: Self-supervised multimodal phasing pretraining.** *Knowledge-informed phase prototyping* (Sec. 2.2) initializes a shared phase bank from medical-LLM descriptions of ICU disease phases, addressing latent clinical phases. *Multimodal phase alignment* (Sec. 2.3) fuses asynchronous observations onto an hourly grid and assigns hours to phases by unbalanced fused Gromov-Wasserstein optimal transport, which gives every hour equal mass while preserving temporal and clinical structure, addressing observation-intensity bias. The resulting assignments serve as pseudo-labels for both hourly and event representations. **Stage 2: Phase-aware dynamic prediction** (Sec. 2.4). The pretrained representations are fine-tuned for downstream tasks, with the learned phases guiding temporal attention. Our contributions are summarized as follows:

- We identify *observation-intensity bias* in multimodal ICU disease progression modeling, where event-level aggregation weights each period by its number of observations, and propose to represent stays in a shared latent phase space in which every hour carries equal weight.
- We propose PULSE, which discovers clinical phases without phase labels by aligning multimodal trajectories to knowledge-informed phase geometry with unbalanced fused Gromov-Wasserstein optimal transport, and uses the learned phases for dynamic prediction.
- Experiments on MIMIC-IV and eICU across five dynamic prediction tasks show that PULSE achieves the best performance in most settings. PULSE also remains more stable as input histories become longer and more irregular, and in case studies its discovered phases align with clinically meaningful transitions.

2 METHOD

Method overview. PULSE discovers shared latent clinical phases without phase labels and uses them for dynamic prediction (Figure 1). Stage 1 is self-supervised pretraining with two components, each addressing one challenge raised in the introduction. **Knowledge-informed phase geometry** (Section 2.2) addresses *latent clinical phases*: a medical LLM defines the clinical relations among phases, which both initialize the shared phase prototypes and constrain how each stay is aligned to them. **Intensity-balanced phase distillation** (Section 2.3) addresses *observation-intensity bias*: an unbalanced fused Gromov-Wasserstein transport assigns every hour of a stay to phases with equal mass, so an hour counts the same regardless of how many events it contains; these hour-level assignments are then distilled into the fused hourly representations and into the representations of individual, often sparse, clinical events. Stage 2 fine-tunes the pretrained model for downstream tasks and uses the learned phases to guide temporal attention (Section 2.4).

2.1 PROBLEM FORMULATION

Multimodal ICU stays. For each ICU stay, routine electronic health record (EHR) measurements are aggregated into L hourly bins, $\mathbf{X}^{\mathrm{EHR}} = [\mathbf{x}_0^{\mathrm{EHR}}, \ldots, \mathbf{x}_{L-1}^{\mathrm{EHR}}] \in \mathbb{R}^{L \times d_{\mathrm{EHR}}}$, where d_{EHR} is the number of EHR variables and $\ell \in \{0, \ldots, L-1\}$ indexes an hour. Clinical events such as CXRs, ECGs, and clinical notes are kept at their acquisition times, $\mathbf{X}^{\mathrm{Evt}} = \{e_i = (m_i, \mathbf{x}_i^{\mathrm{Evt}}, t_i)\}_{i=1}^{N^{\mathrm{Evt}}}$, where m_i is the modality, $\mathbf{x}_i^{\mathrm{Evt}}$ is the content, N^{Evt} is the number of the events, and $t_i \in [0, L)$ is the acquisition time in hours from admission.

Latent phases and dynamic prediction. We assume that the hours of all stays are governed by K latent clinical phases shared across patients, for which no labels are available. PULSE learns a bank

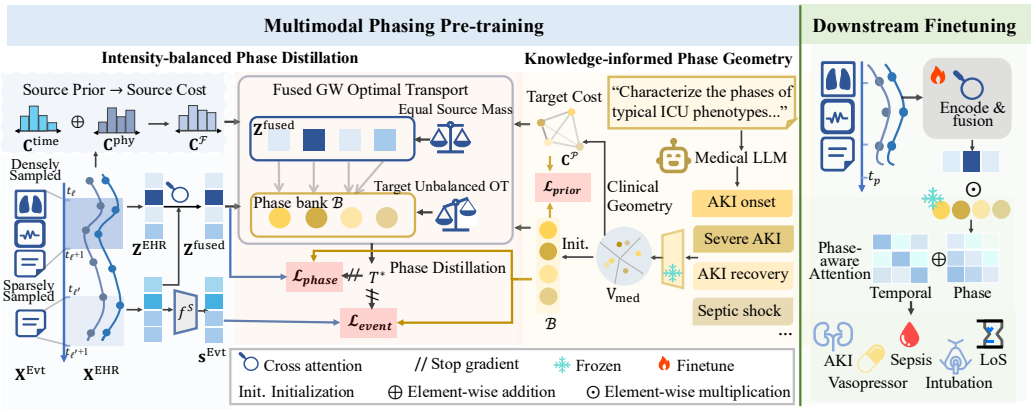

Figure 1: **Overview of PULSE.** PULSE comprises multimodal phasing pre-training and downstream fine-tuning. Stage 1 contains two components. **Knowledge-informed Phase Geometry** uses a medical LLM to describe disease phases; the resulting clinical text embeddings initialize shared phase prototypes and regularize their geometry. **Intensity-balanced Phase Distillation** fuses hourly EHR data with irregular clinical events and aligns equal-mass hours to the phase bank using target-unbalanced fused Gromov–Wasserstein optimal transport, whose coupling provides pseudo-labels for hourly fused and event representations. During **Downstream Finetuning**, the frozen phase bank guides phase-aware temporal attention for five dynamic clinical prediction tasks.

of K phase prototypes and a soft assignment of each hour to them (Sections 2.2 and 2.3). Given a prediction time t_p, the downstream task is to predict a future clinical outcome y within a task-specific horizon from the observations before t_p, $\mathbf{X}^{\mathrm{EHR}}_{t<t_p}$ and $\mathbf{X}^{\mathrm{Evt}}_{t<t_p}$.

2.2 Knowledge-Informed Phase Geometry

Without phase labels, prototypes learned from data alone are not guaranteed to correspond to clinical phases or to be comparable across patients. We therefore anchor the phases in clinical knowledge elicited from a medical LLM, which specifies both which phases exist and how they relate. The phase descriptions initialize a phase bank shared across patients, and their relations form a clinical geometry that regularizes the bank and serves as the target structure when each stay is aligned to the phases by optimal transport (Section 2.3).

Clinical phase geometry. For each ICU phenotype in a standard clinical benchmark (Harutyunyan et al., 2019), we prompt Med-Gemini (Yang et al., 2024) to describe the phases of its disease course, each with its progression status and distinguishing clinical evidence. We encode each description with MedEmbed (Balachandran, 2024) and, after ℓ_2-normalization, cluster the embeddings on the sphere into K centers $\mathbf{V}_{\mathrm{med}} = [\mathbf{v}_1^\top; \ldots; \mathbf{v}_K^\top] \in \mathbb{R}^{K \times d_{\mathrm{med}}}$. Their pairwise distances define the clinical geometry among phases:

$$\mathbf{C}^{\mathcal{P}}_{kk'} = \frac{1 - \mathbf{v}_k^\top \mathbf{v}_{k'}}{2}. \tag{1}$$

This geometry plays two roles: (1) the target structure onto which the hours of each stay are aligned (Section 2.3), so that hours close in time and physiology are assigned to clinically related phases; (2) the anchor for the learnable phase bank described next.

Knowledge-anchored phase bank. We project $\mathbf{V}_{\mathrm{med}}$ to the model dimension d to initialize a shared learnable phase bank $\mathcal{B} = [\mathbf{p}_1^\top; \ldots; \mathbf{p}_K^\top] \in \mathbb{R}^{K \times d}$. Because an initialization alone fades as the prototypes are updated on patient data, we also keep the geometry of the bank close to the clinical geometry:

$$\mathbf{C}^{\mathcal{B}}_{kk'} = \frac{1 - \mathbf{p}_k^\top \mathbf{p}_{k'}}{2}, \qquad \mathcal{L}_{\mathrm{prior}} = \frac{1}{K^2} \left\| \mathbf{C}^{\mathcal{B}} - \mathbf{C}^{\mathcal{P}} \right\|_F^2. \tag{2}$$

The prototypes can thus adapt to observed trajectories while preserving the clinical relations among phases.

2.3 INTENSITY-BALANCED PHASE DISTILLATION

Event-level aggregation lets each period contribute in proportion to its number of observations, so the learned representation conflates observation intensity with patient state. We instead assign phases per hour: asynchronous observations are first fused onto the hourly grid, and each hour is then aligned to the phase bank by unbalanced fused Gromov-Wasserstein (FGW) optimal transport (OT) with equal mass per hour. The resulting hour-level assignments serve as pseudo-labels that are distilled into both the fused hourly representations and the representations of individual clinical events.

Hourly multimodal fusion. EHR measurements track the patient's physiological state every hour, whereas CXRs, ECGs, and clinical notes provide diagnostic evidence at irregular times. Fusing them onto the hourly grid makes the hour, rather than the event, the unit of phase assignment. We encode the hourly EHR sequence with a Transformer encoder (Vaswani et al., 2017) to obtain $\mathbf{Z}^{\mathrm{EHR}} \in \mathbb{R}^{L\times d}$, and embed each clinical event e_i of modality m_i with a unified multimodal set embedding (Lee et al., 2023):

$$\mathbf{u}_i^{\mathrm{Evt}} = F_{m_i}(\mathbf{x}_i^{\mathrm{Evt}}), \qquad \mathbf{z}_i^{\mathrm{Evt}} = W_{m_i}^{\mathrm{Evt}}\mathbf{u}_i^{\mathrm{Evt}} + \mathbf{e}_{\mathrm{time}}(t_i) + \mathbf{e}_{\mathrm{mod}}(m_i), \tag{3}$$

where F_{m_i} is a modality-specific encoder, $W_{m_i}^{\mathrm{Evt}}$ maps the event to the common width d, and $\mathbf{e}_{\mathrm{time}}$ and $\mathbf{e}_{\mathrm{mod}}$ are time and modality embeddings. Stacking the event tokens gives $\mathbf{Z}^{\mathrm{Evt}} \in \mathbb{R}^{N^{\mathrm{Evt}}\times d}$, over which the hourly EHR features attend:

$$\mathbf{Z}^{\mathrm{fused}} = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}, \quad \mathbf{Q} = \mathbf{Z}^{\mathrm{EHR}}W_Q, \quad \mathbf{K} = \mathbf{Z}^{\mathrm{Evt}}W_K, \quad \mathbf{V} = \mathbf{Z}^{\mathrm{Evt}}W_V, \tag{4}$$

where W_Q, W_K, and W_V are learnable projections, and $\mathbf{Z}^{\mathrm{fused}} \in \mathbb{R}^{L\times d}$ becomes an hourly representation that combines routine physiology with clinical-event evidence.

Equal-mass phase assignment. Given the fused features $\mathbf{Z}^{\mathrm{fused}}$ and the shared phase bank $\mathcal{B}$, we use FGW optimal transport to assign each hourly feature to the latent phases (See Appendix A.2 for a detailed overview of FGW). The source set consists of the L hourly nodes, and the target set consists of the K phase prototypes. To encode both physiological similarity and temporal continuity, we define as $\mathbf{C}^{\mathrm{phy}}_{\ell\ell'} = \frac{1}{2}\left(1 - \frac{(\mathbf{x}_\ell^{\mathrm{EHR}})^\top \mathbf{x}_{\ell'}^{\mathrm{EHR}}}{\|\mathbf{x}_\ell^{\mathrm{EHR}}\|_2\|\mathbf{x}_{\ell'}^{\mathrm{EHR}}\|_2}\right), \mathbf{C}^{\mathrm{time}}_{\ell\ell'} = 1 - \exp\left(-\frac{(\ell-\ell')^2}{2\sigma^2_{\mathrm{time}}}\right)$, where σ_{time} controls the temporal bandwidth. Their combination gives the source relational cost

$$\mathbf{C}^{\mathcal{F}}_{\ell\ell'} = \beta\mathbf{C}^{\mathrm{time}}_{\ell\ell'} + (1-\beta)\mathbf{C}^{\mathrm{phy}}_{\ell\ell'}, \qquad \beta \in [0,1]. \tag{5}$$

The temporal term encourages locally coherent phase transitions, while the physiological term connects hours with similar patient states even when they are not adjacent.

Let $\mathbf{T} \in \mathbb{R}_+^{L\times K}$ be the feature-to-phase coupling. The cross-set matching cost is the cosine distance $\mathbf{C}^{\mathrm{match}}_{\ell j} = (1 - \bar{\mathbf{z}}_\ell^{\mathrm{fused}\top}\mathbf{p}_j)/2$, where $\bar{\mathbf{z}}_\ell^{\mathrm{fused}} = \mathbf{z}_\ell^{\mathrm{fused}}/\|\mathbf{z}_\ell^{\mathrm{fused}}\|_2$. The target relational cost $\mathbf{C}^{\mathcal{P}}$ is the medical phase prior in Eq. 1. Given parameter $\alpha \in [0,1]$, and letting $\mathbf{C} := \{\mathbf{C}^{\mathrm{match}}, \mathbf{C}^{\mathcal{F}}, \mathbf{C}^{\mathcal{P}}\}$, the FGW objective is:

$$\mathcal{F}_{\mathrm{FGW}}(\mathbf{C},\mathbf{T}) = (1-\alpha)\sum_{\ell,j}\mathbf{C}^{\mathrm{match}}_{\ell j}\mathbf{T}_{\ell j} + \alpha\sum_{\ell,\ell',j,j'}\left|\mathbf{C}^{\mathcal{F}}_{\ell\ell'} - \mathbf{C}^{\mathcal{P}}_{jj'}\right|^2\mathbf{T}_{\ell j}\mathbf{T}_{\ell' j'}. \tag{6}$$

The structure term assigns hours that are related in time and physiology to phases that are clinically related. Every hour receives the same source mass, $\mu_\ell = 1/L$, regardless of how many events it contains. Observation intensity then determines how much evidence an hour provides, through its fused feature, but not how much the hour counts in the phase assignment; the normalization by L also keeps long stays from dominating training. We fix only this source marginal (Séjourné et al., 2020),

$$\mathcal{T}_{\boldsymbol{\mu}} = \left\{\mathbf{T} \in \mathbb{R}_+^{L\times K} \;\middle|\; \mathbf{T}\mathbf{1}_K = \boldsymbol{\mu}\right\}, \qquad \boldsymbol{\mu} = [\mu_0,\ldots,\mu_{L-1}]^\top, \tag{7}$$

and relax the target marginal with a Kullback–Leibler (KL) penalty:

$$\mathbf{T}^* = \arg\min_{\mathbf{T}\in\mathcal{T}_{\boldsymbol{\mu}}} \mathcal{F}_{\mathrm{FGW}}(\mathbf{C},\mathbf{T}) + \lambda_\pi D_{\mathrm{KL}}\left(\mathbf{T}^\top\mathbf{1}_L \,\middle\|\, \boldsymbol{\pi}\right), \tag{8}$$

where $\boldsymbol{\pi} = \mathbf{1}_K/K$ is a uniform phase prior. The KL term keeps the coupling from collapsing onto a few prototypes, while not enforcing the target marginal exactly allows each stay to occupy only the phases it actually visits.

Phase distillation. The coupling $\mathbf{T}^*$ serves as a self-supervisory signal. Within a stay, training the hourly and event representations to predict their phases makes observations of the same phase consistent across modalities and acquisition times; across stays, because the phase bank $\mathcal{B}$ is shared, phases learned where observations are rich also serve to interpret sparsely observed periods. Normalizing each row of $\mathbf{T}^*$ by its mass gives a per-hour phase posterior:

$$R^*_{\ell j} = \frac{\mathbf{T}^*_{\ell j}}{\mu_\ell}, \qquad \sum_{j=1}^{K} R^*_{\ell j} = 1. \tag{9}$$

For any representation $\mathbf{z}$, the soft phase assignment is

$$P_j(\mathbf{z}; \mathcal{B}) = \frac{\exp\left(\cos(\mathbf{z}, \mathbf{p}_j)/\tau_p\right)}{\sum_{j'=1}^{K} \exp\left(\cos(\mathbf{z}, \mathbf{p}_{j'})/\tau_p\right)}, \tag{10}$$

where $\cos(\cdot,\cdot)$ is the cosine similarity and $\tau_p > 0$ controls the sharpness. We distill the hour-level posteriors into the fused hourly representations with the OT-derived targets by minimizing:

$$\mathcal{L}_{\text{phase}} = -\sum_{\ell=0}^{L-1} \mu_\ell \sum_{j=1}^{K} \text{sg}(R^*_{\ell j}) \log P_j(\mathbf{z}^{\text{fused}}_\ell; \mathcal{B}), \tag{11}$$

where $\text{sg}(\cdot)$ denotes stop-gradient. During pretraining, the coupling is recomputed from the current representations and phase bank, so that pseudo-labels, representations, and prototypes are refined together. We further distill the phases into individual clinical events. Each event e_i inherits the posterior of its nearest hourly bin, $r_{ij} = R^*_{\ell(i)j}$ with $\ell(i) = \arg\min_\ell |t_i - \ell|$. We encode the event content with a content projector f^s, $\mathbf{s}^{\text{Evt}}_i = f^s(\mathbf{u}^{\text{Evt}}_i)$, so that the event representation must infer its phase from its clinical content, and minimize

$$\mathcal{L}_{\text{event}} = -\sum_{i=1}^{N^{\text{Evt}}} \sum_{j=1}^{K} \text{sg}(r_{ij}) \log P_j(\mathbf{s}^{\text{Evt}}_i; \mathcal{B}). \tag{12}$$

As the phase bank $\mathcal{B}$ is shared across the cohort, phases learned from richly observed periods can then be recognized from the content of a single CXR or note in a sparsely observed period. The complete pretraining objective is

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{phase}} + \lambda_{\text{event}} \mathcal{L}_{\text{event}} + \lambda_{\text{prior}} \mathcal{L}_{\text{prior}}, \tag{13}$$

where λ_{event} and λ_{prior} weight the event-distillation and prior geometry terms respectively.

2.4 Phase-Aware Dynamic Prediction

In Stage 2, the pretrained model is fine-tuned for each downstream task on the observations before the prediction time t_p. We freeze the phase bank $\mathcal{B}$, so that fine-tuning does not alter the phase space learned in pretraining, and fine-tune the encoders and the fusion module to obtain the fused hourly representations $\mathbf{Z}^{\text{fused}}_{t<t_p}$ (Section 2.3).

Phase-aware temporal attention. The learned phases guide how evidence is aggregated over time: hours assigned to similar phases attend more strongly to each other, so that each hour, including a sparsely observed one, can draw on evidence from other hours of the same phase. For each fused hour ℓ, we compute its soft phase assignment $\mathbf{q}_\ell = \text{softmax}([\,\mathbf{p}_j^\top \mathbf{z}^{\text{fused}}_\ell\,]_{j=1}^K)$ and form the attention bias

$$A_{\ell\ell'} = \mathbf{q}_\ell^\top \mathbf{q}_{\ell'} + \frac{\ell - \ell'}{\max(t_p - 1, 1)}. \tag{14}$$

The phase term increases the attention between time points with similar phase assignments, while the relative-position term gives greater weight to earlier observations. We add A to the attention logits of the final multi-head self-attention (MHA) layer:

$$\widetilde{\mathbf{Z}}^{\text{fused}}_{t<t_p} = \text{MHA}\left(\mathbf{Z}^{\text{fused}}_{t<t_p}; A\right). \tag{15}$$

Prediction objective. A prediction head maps $\widetilde{\mathbf{Z}}^{\text{fused}}_{t<t_p}$ to $\widehat{y}^{\text{fused}}$. For each modality m in the set $\mathcal{S}$ of modalities observed before t_p, an auxiliary head produces $\widehat{y}^m$, which encourages every available modality to retain task-relevant information. The downstream objective is

$$\mathcal{L}_{\text{down}} = \text{CE}(\widehat{y}^{\text{fused}}, y) + \frac{1}{|\mathcal{S}|} \sum_{m \in \mathcal{S}} \lambda_m \, \text{CE}(\widehat{y}^m, y), \tag{16}$$

where λ_m weights modality m. Training and implementation details are provided in Section A.5.

3 EXPERIMENTS

We evaluate PULSE on two public ICU databases, MIMIC-IV and eICU, across five dynamic prediction tasks that span disease progression, clinical intervention, and ICU outcome. Our experiments are designed to answer four questions:

Q1 Does PULSE improve dynamic prediction over multimodal longitudinal baselines? (Section 3.2)
Q2 Does PULSE benefit from asynchronous modalities without suffering from observation-intensity bias? (Section 3.2)
Q3 Do the discovered phases correspond to clinically meaningful states? (Section 3.3)
Q4 How does each component contribute? (Section 3.4)

3.1 EXPERIMENTAL SETUP

Datasets and modalities. We use MIMIC-IV (Johnson et al., 2023c) and eICU (Pollard et al., 2018), and both datasets provide multimodal longitudinal records, combining regularly sampled EHR measurements with asynchronously acquired clinical events for dynamic prediction. For **MIMIC-IV**, we extract hourly EHR variables, including vital signs and laboratory tests, and three event modalities: chest radiographs from MIMIC-CXR (Johnson et al., 2019), ECGs from MIMIC-IV-ECG (Gow et al., 2023), and clinical notes (chief complaints, past medical history, and radiology reports) from MIMIC-IV-Note (Johnson et al., 2023a). **eICU** is a multicenter database from ICUs across the United States; following Van De Water et al. (2024) and Pollard et al., we extract hourly EHR variables and use past-history and health-status notes as events.

Tasks and metrics. We evaluate five tasks in three families. (1) *Disease progression*: following Van De Water et al. (2024), **acute kidney injury (AKI) progression** and **sepsis progression** classify each 12-hour transition as improved, unchanged, or worsened; we report macro-F1 because the three classes are imbalanced. (2) *Clinical intervention*: following Lee et al. (2023), **vasopressor need** and **intubation need** predict whether the patient will receive vasopressor support or undergo invasive airway intervention within the next 12 hours; we report AUPRC for these imbalanced binary tasks. (3) *ICU outcome*: following Liu et al. (2025), **remaining length of stay (LoS)** is a seven-class ordinal classification, evaluated with quadratic-weighted Cohen's κ. We split the data by patient into training, validation, and test sets (7:1:2), so that all stays of a patient fall into the same partition. Pretraining uses training patients only. We report the mean $\pm$ standard deviation over three random seeds. Cohort statistics and label definitions are given in Appendix A.4.

Baselines. We compare PULSE with a diverse set of multimodal longitudinal models designed to handle asynchronous observations, irregular sampling, or incomplete modalities. **UMSE** (Lee et al., 2023) combines triplet-structured set embeddings with modality-aware attention to handle missing data and integrate multiple modalities. **UTDE** (Zhang et al., 2023) models asynchronous longitudinal data using a gated, attention-based imputation framework. **DiPro** (Liu et al., 2025) disentangles static anatomy from dynamic pathology and performs multimodal alignment across multiple temporal scales. **FuseMoE** (Han et al., 2024) models irregularly sampled trajectories with a mixture-of-experts framework equipped with a Laplace-gated Gaussian router. **Flex-MoE** (Yun et al., 2024) addresses missing modalities using a missing-modality bank and a sparse mixture-of-experts architecture. **TAMME** (Susetzky et al., 2025) handles irregular, long-term multimodal sequences using time-aware embeddings and piecewise sequence summarization. **CTPD** (Wang et al., 2025) discovers cross-modal temporal patterns by refining shared representations with slot attention and contrastive alignment. We additionally include **Uni-EHR**, an EHR-only Transformer (Vaswani et al., 2017), as the reference for whether adding asynchronous modalities improves or degrades prediction.

Table 1: **Performance comparison on the MIMIC and eICU datasets.** We evaluate the proposed method and various baselines on Acute Kidney Injury (AKI) progression, Sepsis progression, Vasopressor need, Intubation need, and Remaining Length of Stay (LoS). Results are reported as mean ± standard deviation. The best results are highlighted in **bold**, and the second-best results are underlined. More results are provided in Tables 5 to 8 in Appendix.

	AKI	Sepsis	Vasopressor	Intubation	LoS
Models	Macro-F1	Macro-F1	AUPRC	AUPRC	Kappa
MIMIC Dataset					
Uni-EHR	0.458±0.003	0.341±0.003	0.560±0.003	0.911±0.004	0.402±0.006
UMSE (Lee et al., 2023)	0.436±0.003	0.348±0.006	0.557±0.002	0.899±0.002	0.392±0.006
UTDE (Zhang et al., 2023)	0.446±0.003	0.361±0.003	**0.586**±0.003	0.911±0.003	0.391±0.003
DiPro (Liu et al., 2025)	0.439±0.003	0.350±0.010	0.556±0.006	0.902±0.002	0.399±0.009
FuseMoE (Han et al., 2024)	0.432±0.001	0.328±0.010	0.561±0.008	0.898±0.002	0.376±0.003
Flex-MoE (Yun et al., 2024)	0.434±0.004	0.338±0.007	0.543±0.011	0.888±0.004	0.386±0.001
TAMME (Susetzky et al., 2025)	0.426±0.003	0.334±0.008	0.573±0.002	0.905±0.007	0.387±0.007
CTPD (Wang et al., 2025)	0.441±0.001	0.359±0.008	0.558±0.001	0.903±0.003	0.380±0.001
PULSE	**0.478**±0.002	**0.380**±0.010	0.576±0.005	**0.920**±0.004	**0.411**±0.008
eICU Dataset					
Uni-EHR	0.433±0.006	0.365±0.013	0.496±0.005	0.903±0.002	0.475±0.005
UMSE (Lee et al., 2023)	0.407±0.003	0.343±0.012	0.513±0.003	0.925±0.000	0.448±0.007
UTDE (Zhang et al., 2023)	0.442±0.004	0.362±0.004	0.524±0.002	0.924±0.002	0.476±0.003
DiPro (Liu et al., 2025)	0.413±0.002	0.331±0.013	0.497±0.005	0.920±0.002	0.456±0.004
FuseMoE (Han et al., 2024)	0.414±0.002	0.355±0.013	0.500±0.004	0.921±0.001	0.461±0.002
Flex-MoE (Yun et al., 2024)	0.416±0.003	0.320±0.037	0.504±0.004	0.920±0.002	0.462±0.005
TAMME (Susetzky et al., 2025)	0.399±0.003	0.364±0.002	0.517±0.004	0.925±0.002	0.457±0.001
CTPD (Wang et al., 2025)	0.415±0.003	0.330±0.006	0.493±0.010	0.915±0.001	0.462±0.009
PULSE	**0.494**±0.009	**0.381**±0.019	**0.534**±0.010	**0.932**±0.002	**0.492**±0.004

3.2 Main Results

PULSE outperforms multimodal baselines, most clearly on disease progression (Q1). Table 1 reports results on both datasets. PULSE achieves the best result in nine of the ten dataset–task settings and ranks second only on MIMIC-IV vasopressor need. The margin depends on the task. On disease progression, PULSE improves macro-F1 over the strongest multimodal baseline by 7.2% (AKI) and 5.3% (sepsis) on MIMIC-IV, and by 11.8% and 4.7% on eICU. On intervention need and LoS, the gains are smaller (0.8%–3.4%). This pattern matches what phases are designed to capture. A progression label records whether the patient moves between clinical stages within the next 12 hours, so predicting it requires knowing which state the patient is currently in and how that state relates to the others. PULSE represents each hour by its position in a phase geometry derived from clinical knowledge, which provides exactly this information.

PULSE benefits from asynchronous modalities where event-level models may not (Q2). Adding asynchronous modalities does not reliably help existing models. As in Table 1, of the 70 multimodal baseline results, 47 fall below the EHR-only Uni-EHR, and in three settings Uni-EHR outperforms all seven baselines. The drop concentrates on disease progression and LoS (36 of 42 results), whose labels depend on the patient's state over long windows. On the eICU intervention tasks, in contrast, most baselines exceed Uni-EHR, so the events do carry predictive information. The problem therefore lies in how the events are used rather than in the events themselves. This is consistent with observation-intensity bias: event-level models let densely documented periods dominate the representation, even though documentation density partly reflects care routines rather than patient state. PULSE outperforms Uni-EHR in all ten settings, with gains of up to 14.1% on eICU AKI, where six of the seven baselines fall below Uni-EHR, and 11.4% on MIMIC-IV sepsis. Because every hour receives equal mass in phase assignment, events refine the estimate of each hour's state without deciding how much that hour counts.

PULSE remains stable as histories lengthen and observation intensity varies more (Q2). If observation intensity biases a model, the effect should grow when the input history mixes densely

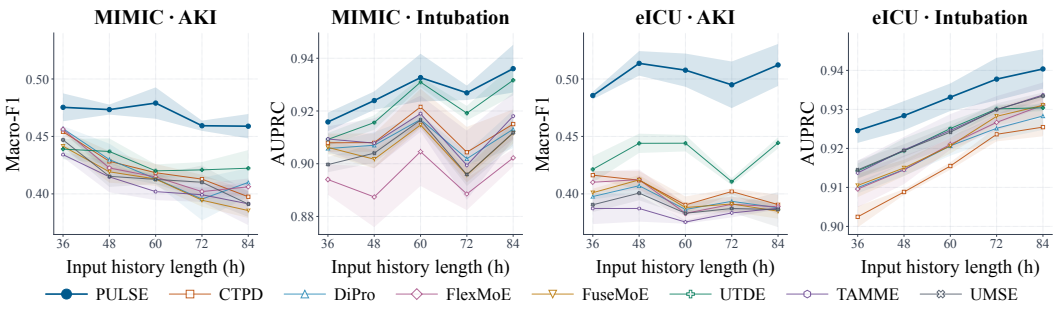

Figure 2: **Performance as histories lengthen and observation intensity becomes more uneven.** Each point evaluates models on test stays with at least h hours before the prediction time, using exactly h hours of history as input. From left to right, the panels show AKI progression and intubation need on MIMIC-IV and eICU. Longer histories have more uneven observation intensity. Curves show the mean over three seeds, and shaded regions show the standard deviation. On MIMIC-IV AKI, most baselines degrade as histories lengthen, whereas PULSE remains stable and performs best at every history length.

and sparsely observed periods. To test this, we vary the input history length h from 36 to 96 hours in 12-hour steps. For each h, we evaluate on test stays with at least h hours before t_p and feed each model exactly the most recent h hours. We report AKI progression and intubation need at each length (Figure 2). To quantify how unevenly the asynchronous events are observed, we compute Δ_σ, the standard deviation of the time gaps between consecutive events (e.g., $|t_i - t_{i-1}|$) within the history window. A small Δ_σ indicates regularly spaced events, whereas a large Δ_σ indicates bursts of documentation separated by long unobserved periods. Δ_σ increases steadily with history length and roughly doubles from the shortest to the longest history in all four dataset–task settings, e.g., from 5.3 to 10.2 hours on MIMIC-IV AKI (Table 9). Longer histories therefore contain more information about the patient, but also more uneven observation intensity. On MIMIC-IV AKI, most baselines degrade as the history grows, which suggests that the uneven intensity outweighs the additional information for these models. PULSE instead remains stable and consistently performs best. On intubation, PULSE further improves as more history becomes available. This demonstrates that PULSE makes use of long histories exactly where observation intensity is most uneven, while event-level baselines lose accuracy there.

3.3 Clinical Validity of Discovered Phases

Discovered phases align with documented clinical transitions (Q3). PULSE discovers phases without phase labels, so we examine whether its phase assignments correspond to clinical states in two MIMIC-IV stays with distinct clinical courses. For each discovered phase, an LLM (Codex 5.6) generates a clinical interpretation from the raw EHR measurements, clinical notes, and discharge summary, and a clinician reviewed the resulting interpretations. In a patient with respiratory failure (Figure 3), the transition into pulmonary edema with difficult weaning coincides with sustained increases in heart rate and respiratory rate and less stable oxygenation, and the subsequent pre-extubation phase coincides with improved oxygenation and reduced respiratory burden. In a patient after coronary artery bypass grafting (CABG; Appendix, Figure 4), one phase coincides with postoperative AKI, reflected by rising creatinine, blood urea nitrogen (BUN), and potassium, and a later phase with postoperative atrial fibrillation, marked by a sharp increase in heart rate. In both cases, phase boundaries fall where the documented clinical state changes rather than at regular time intervals, indicating that PULSE discovers clinically coherent states of disease progression, complications, and recovery.

3.4 Ablation Study

Every component contributes, and the clinical geometry contributes most (Q4). Table 2 removes one component at a time on MIMIC-IV. Every removal lowers performance on all five tasks, and the same holds on eICU (Appendix, Tables 12 and 13). Removing the clinical geometry (V1), which serves as both the FGW target structure and the anchor of the phase bank, causes the largest drop on four of the five tasks, e.g., 10.5% on AKI and 12.8% on vasopressor need. Within intensity-balanced phase distillation, removing the physiological cost (V2) relates hours only by time and lowers all tasks, while removing $\mathcal{L}_{\text{event}}$ (V3) reduces macro-F1 by 8.4% on AKI and 4.5% on sepsis, showing that sparse events benefit from phase information distilled from the hourly

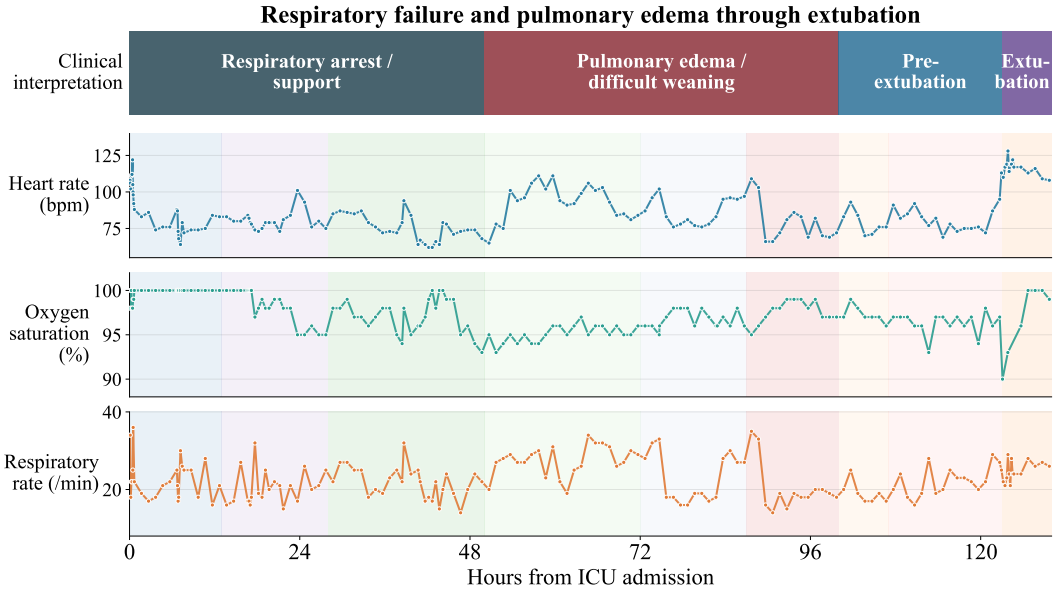

Figure 3: **Discovered phases in a respiratory failure case.** Background colors indicate the phases discovered by PULSE, and the top band gives their post-hoc interpretation based on the EHR measurements, clinical notes, and discharge summary. The lower panels show heart rate, oxygen saturation, and respiratory rate. Phase transitions align with documented changes in the clinical course, from respiratory support and stabilization, through pulmonary edema with difficult weaning, to pre-extubation recovery and extubation.

Table 2: **Ablation Study on MIMIC dataset.** This table presents the results of ablating major designs to assess their contribution to overall performance. Variants "V1"–"V4" correspond to variants of PULSE with each individual module removed. More results are provided in Table 10 and Table 11.

		AKI	**Sepsis**	**Vasopressor**	**Intubation**	**LoS**
ID	Models	Macro-F1	Macro-F1	AUPRC	AUPRC	Kappa
	PULSE	**0.478**$\pm$0.002	**0.380**$\pm$0.010	**0.576**$\pm$0.005	**0.920**$\pm$0.004	**0.411**$\pm$0.008
V1	w/o Clinical geometry	0.428$\pm$0.004	0.370$\pm$0.005	0.502$\pm$0.018	0.887$\pm$0.006	0.375$\pm$0.007
V2	w/o Physiological cost	0.461$\pm$0.001	0.370$\pm$0.015	0.551$\pm$0.006	0.910$\pm$0.009	0.384$\pm$0.007
V3	w/o $\mathcal{L}_{event}$	0.438$\pm$0.004	0.363$\pm$0.020	0.569$\pm$0.008	0.914$\pm$0.002	0.380$\pm$0.005
V4	w/o Phase-aware attention	0.468$\pm$0.002	0.369$\pm$0.015	0.571$\pm$0.002	0.913$\pm$0.005	0.404$\pm$0.009

representations. Replacing phase-aware attention with standard self-attention (V4) causes the smallest drops, and V4 still outperforms the strongest multimodal baseline on four of the five tasks, so most of the gain comes from the pretrained phases.

4 Conclusion

In this paper, we presented PULSE, a self-supervised framework for discovering latent clinical phases from asynchronous multimodal ICU trajectories. By combining knowledge-informed phase geometry with intensity-balanced phase distillation, PULSE learns phases that are anchored in clinical knowledge and mitigates observation-intensity bias. Experiments on MIMIC-IV and eICU across five dynamic prediction tasks show that PULSE achieves the best performance in nine of the ten settings, and case studies show that the discovered phases align with documented clinical transitions. By providing a phase-level view of patient trajectories, PULSE has the potential to support early warning of disease progression and timely clinical intervention in the ICU. In future work, we plan to extend PULSE to longitudinal settings beyond the ICU, such as chronic disease management and cancer treatment planning, where disease also unfolds through clinically distinct phases.

AI USE STATEMENT

We used the medical large language model Med-Gemini (Yang et al., 2024) to generate clinically informed descriptions and relations among broad ICU clinical phases. These descriptions were converted into medical prior knowledge for initializing and regularizing the shared phase-prototype bank. Med-Gemini was not used to generate experimental data, labels, or results.

We also used generative AI tools to improve the clarity and grammar of the manuscript. All AI-assisted outputs, including the generated medical descriptions and edited text, were manually reviewed and revised by the us for clinical validity, technical correctness, and consistency with the experimental evidence. We take full responsibility for the final content of this work, including all text, claims, methods, and results.

ETHICS STATEMENT

This study uses the publicly available MIMIC (Johnson et al., 2023c) and eICU Collaborative Research databases (Pollard et al., 2018). Both datasets contain de-identified patient records and were released for research use under their respective data-access and usage policies. We conducted no new data collection or interaction with human participants, and made no attempt to re-identify individuals. We followed the required access, documentation, and citation procedures for both datasets.

REPRODUCIBILITY STATEMENT

The appendix provides the experimental and implementation details needed to reproduce our results, including data preprocessing, label extraction, model implementation, and evaluation procedures. We plan to release the code repository upon acceptance of the paper.

REFERENCES

Denis Agniel, Isaac S. Kohane, and Griffin M. Weber. Biases in electronic health record data due to processes within the healthcare system: Retrospective observational study. *BMJ*, 361:k1479, 2018. doi: 10.1136/bmj.k1479.

Abhinand Balachandran. Medembed: Medical-focused embedding models, 2024. URL `https://github.com/abhinand5/MedEmbed`.

Karla L Caballero Barajas and Ram Akella. Dynamically modeling patient's health state from electronic medical records: a time series approach. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 69–78, 2015.

Laura Evans, Andrew Rhodes, Waleed Alhazzani, Massimo Antonelli, Craig M Coopersmith, Craig French, Flávia R Machado, Lauralyn Mcintyre, Marlies Ostermann, Hallie C Prescott, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive care medicine*, 47(11):1181–1247, 2021.

Brian Gow, Tom Pollard, Larry A Nathanson, Alistair Johnson, Benjamin Moody, Chrystinne Fernandes, Nathaniel Greenbaum, Jonathan W Waks, Parastou Eslami, Tanner Carbonati, Ashish Chaudhari, Elizabeth Herbst, Dana Moukheiber, Seth Berkowitz, Roger Mark, and Steven Horng. MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset. *PhysioNet*, September 2023. doi: 10.13026/4nqg-sb35. URL `https://doi.org/10.13026/4nqg-sb35`. Version 1.0.

Xing Han, Huy Nguyen, Carl Harris, Nhat Ho, and Suchi Saria. Fusemoe: Mixture-of-experts transformers for fleximodal fusion. *Advances in Neural Information Processing Systems*, 37: 67850–67900, 2024.

Hrayr Harutyunyan, Hrant Khachatrian, David C Kale, Greg Ver Steeg, and Aram Galstyan. Multitask learning and benchmarking with clinical time series data. *Scientific Data*, 6(1):96, 2019.

Nasir Hayat, Krzysztof J Geras, and Farah E Shamout. MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images. In *Machine Learning for Healthcare Conference*, pp. 479–503. PMLR, 2022.

Alistair Johnson, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV-Note: Deidentified free-text clinical notes. *PhysioNet*, January 2023a. doi: 10.13026/1n74-ne17. URL https://doi.org/10.13026/1n74-ne17. Version 2.2.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10:1, 2023b. doi: 10.1038/s41597-022-01899-x.

Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*, 6(1):317, 2019.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, 2023c.

Kwanhyung Lee, Soojeong Lee, Sangchul Hahn, Heejung Hyun, Edward Choi, Byungeun Ahn, and Joohyung Lee. Learning missing modal electronic health records with unified multi-modal data embedding and modality-aware attention. In *Machine Learning for Healthcare Conference*, pp. 423–442. PMLR, 2023.

Yue Li, Pratheeksha Nair, Xing Han Lu, Zhi Wen, Yuening Wang, Amir Ardalan Kalantari Dehaghi, Yan Miao, Weiqi Liu, Tamas Ordog, Joanna M Biernacka, et al. Inferring multimodal latent topics from electronic health records. *Nature communications*, 11(1):2536, 2020.

Chen Liu, Wenfang Yao, Kejing Yin, William K Cheung, and Jing Qin. Multimodal disease progression modeling via spatiotemporal disentanglement and multiscale alignment. *arXiv preprint arXiv:2510.11112*, 2025.

Patrick G Lyons, Aaloke Mody, Alice F Bewley, Morgan Schoer, Bharat Neelam Raju, Elvin Geng, Philip RO Payne, Pratik Sinha, and Anitha Vijayan. Multistate modeling of clinical trajectories and outcomes in the icu: A proof-of-concept evaluation of acute kidney injury among critically ill patients with covid-19. *Critical Care Explorations*, 4(12):e0784, 2022.

Manu LNG Malbrain, Niels Van Regenmortel, Bernd Saugel, Brecht De Tavernier, Pieter-Jan Van Gaal, Olivier Joannes-Boyau, Jean-Louis Teboul, Todd W Rice, Monty Mythen, and Xavier Monnet. Principles of fluid management and stewardship in septic shock: it is time to consider the four d's and the four phases of fluid therapy. *Annals of intensive care*, 8(1):66, 2018.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In Maria-Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pp. 2664–2672. JMLR.org, 2016. URL http://proceedings.mlr.press/v48/peyre16.html.

Tom Pollard, Alistair Johnson, Matthieu Komorowski, Jesse Raffa, et al. Mit-lcp/eicu-code: eicu-crd code repository v1. 0. *Zenodo*.

Tom J Pollard, Alistair EW Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. The eicu collaborative research database, a freely available multi-center database for critical care research. *Scientific data*, 5(1):180178, 2018.

Ilya O. Ryzhov, Raghu Pasupathy, and Harsha Honnappa. Introduction to optimal transport. In *Winter Simulation Conference, WSC 2024, Orlando, FL, USA, December 15-18, 2024*, pp. 101–115. IEEE, 2024. doi: 10.1109/WSC63780.2024.10838746. URL https://doi.org/10.1109/WSC63780.2024.10838746.

Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. The unbalanced gromov wasserstein distance: Conic formulation and relaxation. In *Neural Information Processing Systems*, 2020. URL https://api.semanticscholar.org/CorpusID:221557962.

Tobias Susetzky, Huaqi Qiu, Rickmer Braren, and Daniel Rueckert. A holistic time-aware classification model for multimodal longitudinal patient data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 24–33. Springer, 2025.

Robin Van De Water, Hendrik Schmidt, Paul Elbers, Patrick Thoral, Bert Arnrich, and Patrick Rockenschaub. Yet another icu benchmark: A flexible multi-center framework for clinical ml. In *International Conference on Learning Representations*, volume 2024, pp. 44638–44693, 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Fuying Wang, Feng Wu, Yihan Tang, and Lequan Yu. Ctpd: Cross-modal temporal pattern discovery for enhanced multimodal electronic health records analysis. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 6783–6799, 2025.

Lin Yang, Shawn Xu, Andrew B. Sellergren, Timo Kohlberger, Yuchen Zhou, Ira Ktena, Atilla P. Kiraly, Faruk Ahmed, Farhad Hormozdiari, Tiam Jaroensri, Eric Wang, Ellery Wulczyn, Fayaz Jamil, Theo Guidroz, Charles Lau, Siyuan Qiao, Yun Liu, Akshay Goel, Kendall Park, Arnav Agharwal, Nick George, Yang Wang, Ryutaro Tanno, David G. T. Barrett, Wei-Hung Weng, S. Sara Mahdavi, Khaled Saab, Tao Tu, Sreenivasa Raju Kalidindi, Mozziyar Etemadi, Jorge A. Cuadros, Gregory Sorensen, Yossi Matias, Katherine Chou, Greg S. Corrado, Joelle K. Barral, Shravya Shetty, David Fleet, S. M. Ali Eslami, Daniel Tse, Shruthi Prabhakara, Cory Y. McLean, David Steiner, Rory Pilgrim, Christopher Kelly, Shekoofeh Azizi, and Daniel Golden. Advancing multimodal medical capabilities of gemini. *ArXiv*, abs/2405.03162, 2024. URL `https://api.semanticscholar.org/CorpusID:269605403`.

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. *Advances in Neural Information Processing Systems*, 37:98782–98805, 2024.

Aaron Zalewski, William Long, Alistair EW Johnson, Roger G Mark, and Li-wei H Lehman. Estimating patient's health state using latent structure inferred from clinical time series and text. In *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 449–452. IEEE, 2017.

Xinlu Zhang, Shiyang Li, Zhiyu Chen, Xifeng Yan, and Linda Ruth Petzold. Improving medical predictions by irregular multimodal electronic health records modeling. In *International Conference on Machine Learning*, pp. 41300–41313. PMLR, 2023.

A APPENDIX

A.1 RELATED WORKS

Irregular and Multimodal ICU Trajectory Modeling. ICU trajectories combine routinely recorded EHR variables with notes, ECGs, and images acquired at irregular times. Integrating these sources is important for dynamic prediction because changes in a patient's condition may be reflected in different modalities at different times. UTDE uses gated attention-based imputation to model irregular longitudinal measurements (Zhang et al., 2023), whereas UMSE represents multimodal records as triplet-structured sets and uses modality-aware attention for fusion (Lee et al., 2023). TAMME introduces time-aware embeddings and sequence summarization to handle long, irregular trajectories (Susetzky et al., 2025). For EHR and sequential chest radiographs, DiPro separates static anatomical information from dynamic pathology and aligns the modalities across multiple timescales (Liu et al., 2025). FuseMoE and Flex-MoE instead use mixture-of-experts architectures to accommodate irregular trajectories and missing modalities (Han et al., 2024; Yun et al., 2024). While these methods improve multimodal modeling, the frequency of recorded events can reflect clinical workflow and institutional practice as well as patient state (Agniel et al., 2018). Aggregating observations at the event level can therefore give densely observed periods disproportionate influence. `PULSE` assigns equal mass to each hour during phase alignment and distills the resulting assignments into fused and event representations, reducing the direct influence of event counts.

Latent State Discovery for Clinical Trajectories. Latent-state discovery offers a way to summarize evolving clinical conditions from many heterogeneous measurements. Early probabilistic dynamic models estimate time-varying patient health states from longitudinal EHR records (Caballero Barajas & Akella, 2015). Zalewski et al. combine physiological time series with unstructured ICU nursing notes to infer latent health states (Zalewski et al., 2017). Extending latent-variable modeling to heterogeneous EHR modalities, MixEHR uses a Bayesian multi-view topic model to discover shared patient phenotypes (Li et al., 2020). More recently, CTPD uses slot attention and contrastive alignment to learn cross-modal temporal patterns for clinical prediction (Wang et al., 2025). While these methods capture latent clinical structure, they do not explicitly account for observation-intensity bias when discovering states or patterns, nor do they use medical knowledge to define relationships among shared phases. `PULSE` addresses both issues through clinically informed phase geometry and equal-mass hourly phase alignment.

A.2 FUSED GROMOV-WASSERSTEIN OPTIMAL TRANSPORT.

Fused Gromov-Wasserstein (FGW) optimal transport provides a soft correspondence between two structured discrete sets by jointly considering cross-set feature similarity and within-set relational structure (Peyré et al., 2016; Ryzhov et al., 2024). This makes FGW suitable for matching sets with different numbers of elements or different feature spaces. Let $\mathcal{X}$ and $\mathcal{Y}$ be two weighted sets with relational cost matrices $\mathbf{C}^x \in \mathbb{R}^{n\times n}$ and $\mathbf{C}^y \in \mathbb{R}^{m\times m}$, cross-set feature cost matrix $\mathbf{C}^{\text{match}} \in \mathbb{R}^{n\times m}$, and normalized mass vectors $\mathbf{p} \in \Delta^{n-1}$ and $\mathbf{q} \in \Delta^{m-1}$, where $\Delta^{d-1} = \{\mathbf{a} \in \mathbb{R}^d_+ : \mathbf{1}_d^\top \mathbf{a} = 1\}$ is the probability simplex. Their transport polytope is

$$\mathcal{T}(\mathbf{p}, \mathbf{q}) = \left\{ \mathbf{T} \in \mathbb{R}^{n\times m}_+ \;\middle|\; \mathbf{T}\mathbf{1}_m = \mathbf{p},\ \mathbf{T}^\top \mathbf{1}_n = \mathbf{q} \right\}. \tag{17}$$

FGW seeks the coupling

$$\mathbf{T}^* = \arg\min_{\mathbf{T}\in\mathcal{T}(\mathbf{p},\mathbf{q})} \left[(1-\alpha) \sum_{i,j} \mathbf{C}^{\text{match}}_{ij} \mathbf{T}_{ij} + \alpha \sum_{i,k,j,l} \left| \mathbf{C}^x_{ik} - \mathbf{C}^y_{jl} \right|^2 \mathbf{T}_{ij} \mathbf{T}_{kl} \right], \tag{18}$$

where $\alpha \in [0, 1]$ controls the trade-off between feature similarity and relational structure. The first term favors matching elements with similar features, whereas the second term favors couplings that preserve pairwise relationships. The resulting coupling is a soft correspondence, and its objective value measures the fused structural discrepancy.

PROMPT PROVIDED FOR CLINICAL PHENOTYPE PHASING

The following system prompt was provided to the Med-Germini (Yang et al., 2024) to generate the clinical phenotype descriptions and phase segmentations:

You are a critical-care physician and clinical knowledge curator. For each ICU phenotype below, produce a concise, evidence-based description of the patient's clinically meaningful disease course. The output will be used as a fixed medical prior for multimodal longitudinal modelling. Describe an archetypal ICU trajectory, while acknowledging when a phenotype is too broad or heterogeneous for reliable phasing.

Clinical reasoning requirements

1. Decide whether the phenotype has a clinically meaningful ordered course during an ICU stay.
2. If it is phaseable, define the smallest useful set of ordered phases (normally 2–4), such as onset, progression, peak severity, stabilization, recovery, or relapse. Do not force phases that are not supported by medical knowledge.
3. If it is not naturally phaseable, set *phaseability* to *not_naturally_phaseable* and provide one or more clinically justified *state* segments only when they represent defensible distinctions. If no useful distinction exists, provide one typical ICU state.
4. Each phase/state must describe a patient during that time interval, not a textbook definition. Explain what the patient is experiencing, why the state occurs, and how it is manifested. Each segment must represent exactly one clinically coherent state.
5. Keep every phase/state *description* to one paragraph and at most 500 words. Do not split it into bullet points or subfields.
6. Do not combine temporally or clinically incompatible states in one segment. A segment name and its description must refer to one state only. Never use combined labels such as *severe pneumonia and recovery*, *onset/stabilization*, or *progression and resolution*.
7. A recovery segment is optional. Create one only when the phenotype has a clinically meaningful and observable recovery trajectory during an ICU stay. If recovery cannot be distinguished reliably, do not force a recovery segment; use the acute/severe state or mark the phenotype as *not_naturally_phaseable*.
8. When a disease can recover or remain severely impaired, keep severe disease, stabilization, and recovery as separate states when each is clinically distinguishable. Record alternative outcomes in *transition_cues* or *confidence*; never merge severe disease and recovery into one segment.

Content of each phase/state description Each paragraph must integrate the following elements in natural prose:

- the clinical status and severity (early, worsening, peak, improving, or persistent);
- the pathophysiology, precipitant, or mechanism explaining why this state occurs;
- the most informative symptoms, examination findings, and organ dysfunction;
- expected direction or trajectory of the key available EHR variables;
- imaging or physiologic findings only when medically relevant.

Notes:

- Recovery is disease-specific and must not follow a generic template. For each recovery segment, select the variables and findings that actually demonstrate improvement for that phenotype. For example, renal recovery is reflected by creatinine, blood urea nitrogen, urine output, bicarbonate, pH, and fluid balance; pneumonia recovery by oxygenation, respiratory rate, ventilatory support, temperature, white blood cell count, and relevant imaging; stroke recovery primarily by neurologic examinations, airway protection, and respiratory status. Chronic conditions such as lipid disorders, essential hypertension, or stable chronic kidney disease generally should not receive an ICU recovery phase unless an acute complication has a clearly observable resolution.

- Do not assume that every phaseable phenotype must have onset, deterioration, and recovery. Use only the states supported by the natural history of that condition and the available observations. A stabilization phase may be more appropriate than a recovery phase when organ dysfunction persists but is no longer worsening.
- Mention CXR only when it contributes to the phenotype (for example, pulmonary edema, consolidation, pleural effusion, pneumothorax, or atelectasis). If CXR is usually normal or not clinically useful, omit it entirely. CT, MRI, ultrasound, echocardiography, coronary imaging, or other imaging may be described when that modality is clinically relevant to the phenotype. Mention ECG or waveform findings when they help identify the state, its severity, or a complication; omit them when there is no meaningful expected finding. Never invent a finding that is not characteristic or documented.
- Select only the most relevant measurements from the permitted EHR variables below. Describe qualitative trends rather than invented patient-specific values. Mark a variable as supportive or nonspecific when it cannot establish the diagnosis by itself. Consider baseline values, missingness, fluid dilution, dialysis, medication effects, and other treatment confounding when relevant.

The full prompt has been attached to the supplementary file.

The generated phase description samples are as follows:

AKI onset: The patient develops a new fall in kidney filtration after hypoperfusion, sepsis, nephrotoxin exposure, obstruction, or intrinsic renal injury. Creatinine and blood urea nitrogen begin to rise from baseline and urine output may decline; bicarbonate and pH can fall when metabolic acidosis develops, while blood pressure and mean arterial pressure may be low in hypoperfusion. Renal ultrasound or CT is used only when obstruction or structural disease is suspected. These measurements are supportive and require baseline and clinical context.

Severe AKI: Kidney dysfunction is established and causes clinically important azotemia, oliguria, acidosis, or volume overload because injured nephrons cannot maintain water, acid-base, and solute homeostasis. Creatinine and blood urea nitrogen remain high or rise, urine output stays low, bicarbonate and pH may decrease, and daily weight may increase. Pulmonary edema can lower oxygen saturation and partial pressure of oxygen and increase fraction of inspired oxygen or positive end-expiratory pressure; severe uremia may reduce Glasgow Coma Scale components. Chest imaging is relevant only when overload is suspected.

Renal recovery: The acute insult is resolving and filtration and fluid handling move toward the patient's baseline. Creatinine and blood urea nitrogen plateau and then fall, urine output increases, bicarbonate and pH improve, and daily weight and oxygen support decline if volume overload resolves. Creatinine may lag behind physiologic recovery, and dialysis or fluid administration can obscure trends. Persistent abnormalities indicate incomplete recovery or acute-on-chronic disease.

Developing shock: The patient develops inadequate effective tissue perfusion from hypovolemia, pump failure, vasodilation, or obstructed circulation, initially with compensatory responses. Heart rate rises and systolic or mean arterial pressure may fall; lactic acid increases, urine output falls, and Glasgow Coma Scale may decline. Oxygen saturation, partial pressure of oxygen, pH, creatinine, hemoglobin, and hematocrit identify associated respiratory, metabolic, renal, or hemorrhagic consequences. Imaging and ECG are used only to identify a suspected subtype or cause.

Established shock: Persistent hypoperfusion causes organ dysfunction and requires vasoactive, mechanical, or other intensive support. Low mean arterial pressure, rising lactic acid, tachycardia or bradycardia, oliguria, increasing creatinine, low Glasgow Coma Scale, metabolic acidosis, and worsening oxygen or ventilatory requirements may coexist. Hemoglobin, hematocrit, platelet count, prothrombin time, and partial thromboplastin time help identify hemorrhagic or coagulopathic contexts. CXR, CT, ultrasound, echocardiography, and ECG findings depend on the subtype.

Shock reversal: The cause of hypoperfusion is controlled and organ perfusion is recovering. Lactic acid falls, mean arterial pressure stabilizes without escalating support, urine output increases,

and creatinine, Glasgow Coma Scale, pH, and oxygenation improve. Cause-related imaging may resolve or remain abnormal, and ECG may retain underlying disease. Persistent lactate or organ dysfunction indicates ongoing injury rather than reversal.

A.3 Details of Data Preprocessing

We used the MIMIC-IV (Johnson et al., 2023b) and eICU Collaborative Research Database (eICU-CRD) (Pollard et al., 2018) datasets. All timestamps were converted to hours relative to ICU admission.

MIMIC Cohort Construction Following Hayat et al. (2022); Harutyunyan et al. (2019), we retained at most one ICU stay per hospital admission to prevent multiple correlated examples from the same admission. Patients younger than 18 years were excluded. For the multimodal analysis, a stay was retained when at least one of ECG, AP-view CXR, or clinical-note event was available in the eligibility window from 24 hours before ICU admission through the first 24 hours of the stay. Then we split the stay in to train/validation/test (7:1:2) using 'subject_id' to avoid data leakage.

MIMIC EHR preprocesssing Following Liu et al. (2025), We incorporate vital signs, laboratory measurements, and clinical scores, resulting in a total of 38 clinical time-series variables, including alanine aminotransferase, albumin, alkaline phosphate, anion gap, asparate aminotransferase, bicarbonate, bilirubin, blood urea nitrogen, chloride, creatinine, diastolic blood pressure, fraction inspired oxygen, Glasgow coma scale (eye opening, motor response, and verbal response), glucose, heart rate, height, hematocrit, hemoglobin, magnesium, mean blood pressure, oxygen saturation, partial pressure of carbon dioxide, partial thromboplastin time, platelets, positive end-expiratory pressure, potassium, prothrombin time, respiratory rate, sodium, systolic blood pressure, temperature, troponin-T, urine output, daily weight, white blood cell count, and pH. We applied identical discretization and standardization procedures.

MIMIC-IV ECG, CXR, and Clinical Notes For chest radiographs, we kept AP-view images, selected one deterministic image per study after timestamp ordering, and aligned each image to an ICU stay by its acquisition time. Images were converted to RGB and resized while preserving aspect ratio so that the shorter edge was 256 pixels. The resulting image identifier and relative acquisition time were stored with the stay. ECG records were restricted to 12-lead waveforms sampled at 500 Hz. Each record was checked for non-finite or all-zero signals, and represented as a fixed 5,000-sample waveform (10 seconds at 500 Hz) together with its relative measurement time.

Clinical notes were assembled from MIMIC-IV discharge and radiology sources. For discharge notes, we extracted the Chief Complaint and Past Medical History sections. For radiology notes, we retained the Comparison, Findings, and Impression sections when present. Section markers, repeated underscores, and list-prefix artifacts were normalized, whitespace was collapsed, and notes shorter than ten characters after cleaning were removed. Duplicate note identifiers were resolved deterministically, and each remaining note was associated with its source chart time. Admission-summary sections were aligned to ICU admission, whereas other notes used their original chart time.

eICU-CRD Cohort Construction For eICU-CRD, the patient table supplied the ICU-stay identifier, unique patient identifier, ICU discharge offset, and discharge status. Following Van De Water et al. (2024), We retained stays with at least 1,440 minutes of ICU time and required at least two distinct documentation episodes in the designated 24-hour pre-ICU note window. The latter criterion was applied before generating dynamic landmarks. Hospital discharge status was used only to construct the optional outcome metadata and was not included as an input feature.

eICU EHR preprocesssing Followng Van De Water et al. (2024)T, the eICU hourly EHR we extracted comprises 48 variables: albumin (alb), alkaline phosphatase (alp), alanine aminotransferase (alt), aspartate aminotransferase (ast), bicarbonate (bicar), total bilirubin (bili), direct bilirubin (bili_dir), band neutrophils (bnd), blood urea nitrogen (bun), total calcium (ca), ionized calcium (cai), creatine kinase (ck), creatine kinase-MB (ckmb), chloride (cl), creatinine (crea), C-reactive protein (crp), diastolic blood pressure (dbp), fibrinogen (fgn), fraction of inspired oxygen (fio2), glucose (glu), hemoglobin (hgb), heart rate (hr), international normalized ratio (inr_pt), potassium

(k), lymphocytes (lymph), mean arterial pressure (map), mean corpuscular hemoglobin (mch), mean corpuscular hemoglobin concentration (mchc), mean corpuscular volume (mcv), methemoglobin (methb), magnesium (mg), sodium (na), neutrophils (neut), oxygen saturation (o2sat), phosphate (phos), platelets (plt), partial thromboplastin time (ptt), respiratory rate (resp), systolic blood pressure (sbp), temperature (temp), urine output (urine), white blood cell count (wbc), lactic acid, partial pressure of carbon dioxide, troponin-T, pH, partial pressure of oxygen, and base excess.

eICU Clinical Notes Following Pollard et al., we extract eICU note stream from `note.csv`, `pastHistory.csv`, and `nurseAssessment.csv`. Interface-only records such as "Copies", "View", "Save", and "Print" were discarded. Repeated path/value/text items were deduplicated within an episode and across CSV chunks. Nurse-assessment cells with the same clinical timestamp were concatenated in deterministic path order. Each resulting episode stores its clinical time, availability time, source type, and a stable event identifier; the availability time is used when enforcing the observed-prefix constraint. Cleaned episodes were embedded with the frozen clinical-language encoder.

A.4 Details of label extraction

Landmarks and Dynamic Labels AKI, Sepsis, and LOS follow the YAIB definitions (Van De Water et al., 2024), whereas the vasopressor and intubation evidence follows the UMSE definition (Lee et al., 2023). In both datasets, labels are computed from raw, non-imputed events. Standardization, forward filling, mean imputation, and the model acquisition mask are never used to create a label.

For a stay with ICU admission time a, landmarks are

$$t_j = a + 24\,\mathrm{h} + 12j\,\mathrm{h}, \qquad 0 \le t_j - a \le 168\,\mathrm{h}.$$

The model history is the prefix before t_j (implemented as $[a, t_j)$). The past phase for trajectory tasks is $(t_j - 12\,\mathrm{h}, t_j]$ and the future phase is $(t_j, t_j + 12\,\mathrm{h}]$. AKI, Sepsis, vasopressor, and airway labels require a complete future ICU phase; LOS only requires that the landmark occur before ICU discharge. Events at exactly t_j belong to the past phase and are not future evidence. A phase without sufficient raw evidence is marked unknown/unassessable and is never converted to a negative class. Each output row is keyed by $(\text{stay_id}, t_j)$.

MIMIC-IV AKI. We extracted serum creatinine from `hosp/labevents` (`itemid=50912`) and urine-output events from `icu/outputevents`. The implementation uses the verified urine cells `226557`, `226558`, `226559`, `226560`, `226561`, `226563`, `226564`, `226565`, `226567`, `226584`, `227488`, and `227489`; irrigant input is subtracted from the paired urine-output volume. Admission or causal prior weight is used for urine normalization, with the YAIB 75-kg fallback when no valid weight is available. Creatinine and urine output are staged with the YAIB-compatible KDIGO rules (RRT is not an independent Stage-3 trigger), and the maximum stage in each adjacent 12-hour phase is compared. The primary MIMIC target column is `trajectory` with values `improved`, `unchanged`, and `worsened`.

MIMIC-IV Sepsis. Suspected infection is extracted from actual antibiotic administrations in `icu/inputevents` and culture requests in `hosp/microbiologyevents`, using the YAIB-compatible antibiotic–culture time pairing. Organ dysfunction is computed from raw measurements and interventions, including PaO_2/FiO_2 and respiratory support, platelets, bilirubin, MAP and vasopressors, GCS, creatinine, and urine output. We form the six-component hourly SOFA score and compare peak SOFA in the past and future phases. A change of at least two points is required for `improved` or `worsened`; smaller changes are `unchanged`.

MIMIC-IV vasopressor exposure. The UMSE reference identifies norepinephrine (`221906`), dopamine (`221662`), dobutamine (`221653`; correcting the UMSE item-ID typo), and epinephrine (`221289`) in `icu/inputevents`. A label is positive when a valid administration overlaps $(t_j, t_j + 12\,\mathrm{h}]$, including an infusion that started before t_j and remains active.

MIMIC-IV intubation needed. We use invasive-ventilation procedure evidence `icu/procedureevents.itemid=225792` and corroborating airway evidence: direct

Table 3: Dynamic-label cohort size. Split cells report the number of landmarks followed by the number of unique stays.

Dataset	Task	Stays	Landmarks	Train	Validation	Test	Assessable	Unknown
MIMIC-IV	AKI	23,711	106,856	75,054/16,587	10,366/2,323	21,436/4,801	101,152	5,704
MIMIC-IV	Sepsis	5,909	41,390	29,515/4,237	3,968/551	7,907/1,121	41,377	13
MIMIC-IV	Vasopressor	30,095	123,187	86,703/21,138	11,986/2,943	24,498/6,014	123,187	0
MIMIC-IV	Intubation	30,095	123,187	86,703/21,138	11,986/2,943	24,498/6,014	123,187	0
MIMIC-IV	LOS	41,560	163,378	114,875/29,146	15,980/4,138	32,523/8,276	163,378	0
eICU-CRD	AKI	63,775	357,432	250,209/44,641	34,922/6,225	72,301/12,909	158,455	198,977
eICU-CRD	Sepsis	38,024	189,633	132,725/26,614	18,562/3,762	38,346/7,648	5,061	184,572
eICU-CRD	Vasopressor	89,959	445,584	312,393/63,055	43,683/8,889	89,508/18,015	445,584	0
eICU-CRD	Intubation	89,959	445,584	312,393/63,055	43,683/8,889	89,508/18,015	445,584	0
eICU-CRD	LOS	114,603	548,111	384,161/80,210	53,954/11,451	109,996/22,942	548,111	0

intubation procedure `224385`, and chart items `223059` (intubation details), `225307` (oral ETT), `225308` (nasal ETT), and `226188` (intubation timeout). Procedure items `225468` and `225477` document unplanned extubation and are not interpreted as new intubations. Ventilation intervals and high-specificity airway points are combined; any overlap with the future phase is positive.

MIMIC-IV remaining LOS. Using `icustays.intime` and `icustays.outtime`, we compute remaining ICU hours at each landmark and assign the seven ordered classes $(0, 12]$, $(12, 24]$, $(24, 48]$, $(48, 72]$, $(72, 120]$, $(120, 168]$, and > 168 hours.

eICU-CRD AKI . Creatinine is extracted from `lab.csv`, urine output from `intakeOutput.csv`, and admission weight and ICU offsets from `patient.csv`. The YAIB-compatible creatinine and urine-output KDIGO stages are computed from raw offsets, the maximum stage in each 12-hour phase is compared, and the difference is mapped to `improved`, `unchanged`, or `worsened`. The target is `aki_trajectory` and the explicit evidence flag is `aki_assessable`.

eICU-CRD Sepsis. Antibiotic starts are extracted from `infusionDrug.csv` and `medication.csv`. Because local eICU microbiology coverage is insufficient for the primary product, suspected infection uses continuous antibiotic starts alone; microbiology-paired definitions are sensitivity analyses. The SOFA evidence comes from raw laboratory values, vital signs, respiratory care/charting, nurse-charted GCS, vasopressor administrations, and urine output.

eICU-CRD vasopressor and intubation need. For vasopressors, `infusionDrug.csv` is matched to norepinephrine, epinephrine, dopamine, phenylephrine, and vasopressin families; dobutamine is excluded from this target. Positive infusion snapshots are reconstructed into exposure intervals, and overlap with the future phase gives `vasopressor_exposure_next_12h`. For the airway task, `respiratoryCare.csv` and `respiratoryCharting.csv` provide mechanical/invasive ventilation intervals, while airway-type and treatment concepts provide high-specificity invasive-airway points. Their union defines `invasive_airway_exposure_next_12h`.

eICU-CRD remaining LOS. Remaining ICU time is computed from `patient.unitdischargeoffset` and uses the same seven class boundaries and target column `remaining_los_class` as MIMIC-IV.

Cohort Size and Label Distribution Each *landmark* is one $(\text{stay_id}, t_j)$ pair, whereas the stay column counts unique ICU stays. In Table 3, each split entry is reported as "landmarks/stays". "Assessable" denotes a landmark with a valid target under the task-specific assessability rule; "unknown" denotes a landmark excluded because the required outcome phase was not observable. A stay can contribute multiple landmarks, so landmark counts are the sample counts used by the dynamic tasks.

Table 4 gives the class distribution for each task. Percentages use the number of assessable landmarks for the same dataset and task as the denominator; unknown landmarks are therefore not silently treated as negative examples. For AKI and sepsis, the unknown rows are particularly important

Table 4: Dynamic-label class distribution. Percentages are among assessable landmarks.

Dataset	Task	Class	Count	Percent
MIMIC-IV	AKI	Improved	21,692	21.44%
MIMIC-IV	AKI	Unchanged	55,591	54.96%
MIMIC-IV	AKI	Worsened	23,869	23.60%
MIMIC-IV	Sepsis	Improved	6,448	15.58%
MIMIC-IV	Sepsis	Unchanged	31,306	75.66%
MIMIC-IV	Sepsis	Worsened	3,623	8.76%
MIMIC-IV	Vasopressor	No exposure	102,868	83.51%
MIMIC-IV	Vasopressor	Exposure	20,319	16.49%
MIMIC-IV	Intubation	No exposure	92,253	74.89%
MIMIC-IV	Intubation	Exposure	30,934	25.11%
MIMIC-IV	LOS	0–12 h	40,191	24.60%
MIMIC-IV	LOS	12–24 h	29,124	17.83%
MIMIC-IV	LOS	24–48 h	38,163	23.36%
MIMIC-IV	LOS	48–72 h	22,744	13.92%
MIMIC-IV	LOS	72–120 h	22,547	13.80%
MIMIC-IV	LOS	120–168 h	8,693	5.32%
MIMIC-IV	LOS	>168 h	1,916	1.17%
eICU-CRD	AKI	Improved	29,456	18.59%
eICU-CRD	AKI	Unchanged	93,808	59.20%
eICU-CRD	AKI	Worsened	35,191	22.21%
eICU-CRD	Sepsis	Improved	697	13.77%
eICU-CRD	Sepsis	Unchanged	3,777	74.63%
eICU-CRD	Sepsis	Worsened	587	11.60%
eICU-CRD	Vasopressor	No exposure	396,080	88.89%
eICU-CRD	Vasopressor	Exposure	49,504	11.11%
eICU-CRD	Intubation	No exposure	225,953	50.71%
eICU-CRD	Intubation	Exposure	219,631	49.29%
eICU-CRD	LOS	0–12 h	102,693	18.74%
eICU-CRD	LOS	12–24 h	79,330	14.47%
eICU-CRD	LOS	24–48 h	103,889	18.95%
eICU-CRD	LOS	48–72 h	64,025	11.68%
eICU-CRD	LOS	72–120 h	73,196	13.35%
eICU-CRD	LOS	120–168 h	39,898	7.28%
eICU-CRD	LOS	>168 h	85,080	15.52%

in eICU-CRD because the future phase may not have sufficient creatinine/urine-output or SOFA evidence for a valid transition.

A.5 Implementation Details

Architecture. Each example contains the observed EHR history before a prediction landmark. Hourly EHR values and observation masks are concatenated; timestamped ECG, chest X-ray (CXR), and note events are retained at their observed times. MIMIC uses 39 EHR variables and all three sparse modalities, whereas eICU uses 48 variables and notes. Missing modalities are masked. A two-layer, eight-head EHR Transformer with hidden width 256 encodes the hourly stream. ECGFounder and CheXNet prefix features pass through their trainable final blocks, while notes use pretrained BioClinicalBERT Zhang et al. (2023) embeddings. Each modality is projected to width 128. Two eight-head cross-attention layers use EHR tokens to query the sparse events. We initialize 12 phase tokens from a medical prior constructed from 53 clinical-state descriptions.

Pretraining and prediction. We pretrain on unlabeled training stays of at most 168 hours. The reported model uses one observed view: unbalanced fused Gromov–Wasserstein transport aligns the temporal and physiological EHR graph with the medical phase graph. Detached transport assignments supervise hourly phase prediction and event-to-phase alignment. The objective is

$$\mathcal{L}_{\mathrm{pre}} = \mathcal{L}_{\mathrm{phase}} + \lambda_{\mathrm{cross}}\mathcal{L}_{\mathrm{cross}} + \lambda_{\mathrm{prior}}\mathcal{L}_{\mathrm{prior}}.$$

We set $\lambda_{\mathrm{prior}} = 10$ and $\lambda_{\mathrm{cross}} = 0.2$ on MIMIC or 1.5 on eICU. The transport solver uses 30 iterations, structural weight 0.8, and marginal penalty 0.05. For downstream tasks, we freeze the phase bank and fine-tune the encoders and fusion module. A temporal self-attention layer combines phase-assignment similarity with a relative-time bias; the last valid token predicts the label. No transport solve is needed at inference. Auxiliary EHR and sparse-modality heads have loss weights λ_{E} and λ_{S}.

A.5.1 BASELINE ADAPTATION

The seven baselines use the same landmark examples, available modality streams, cached ECG/CXR prefixes, note embeddings, timestamps, and task-specific labels as MOT. Their shared input interface supplies EHR, ECG, CXR, and notes on MIMIC, and EHR and notes on eICU. We implement the characteristic fusion mechanism of each method with these inputs, with the minimum modification to fit our data.

CTPD (Wang et al., 2025) Separate slot-attention modules group EHR and sparse-event tokens, followed by slot fusion and an alignment loss. The adapter omits the original reconstruction decoder and loss.

UTDE (Zhang et al., 2023) EHR tokens separately attend to each available sparse modality before temporal self-attention and classification. Irregular event times are preserved.

UMSE (Lee et al., 2023) We retain bottleneck-token fusion, using separate EHR, ECG, CXR, and note streams on MIMIC, or EHR and note streams on eICU. Learned null tokens support missing sparse modalities.

TAMME (Susetzky et al., 2025) A time- and modality-aware Transformer jointly encodes EHR and sparse events. Its auxiliary decoder reconstructs 15% of masked valid token representations during training.

FuseMoE(Han et al., 2024) Masked modality summaries pass through a cross-modal Transformer; a four-expert, top-1 router processes the fused patient representation. A routing-balance loss is retained.

Flex-MoE (Yun et al., 2024) A learned representation bank fills missing modality summaries according to the observed-modality combination, which also conditions a four-expert, top-1 router. There are eight combinations on MIMIC and two on eICU; no raw measurements are imputed.

DiPro (Liu et al., 2025) First and last sparse-event representations yield static and dynamic features, which attend locally and globally to EHR tokens. We retain static–dynamic orthogonality regularization, but omit anatomical-region processing and the original progression-aware and temporal-consistency losses due to no disease progression labels.

A.5.2 TRAINING AND HYPERPARAMETER TUNING

We use AdamW, mixed precision, and gradient clipping at 1.0. MOT is pretrained for 30 epochs with batch size 128, encoder learning rate 4×10^{-4}, phase-bank learning rate 10^{-4}, and weight decay 10^{-4}. Fine-tuning runs for up to 50 epochs with batch size 128, 5% warm-up, cosine decay, and patience 10. Most MIMIC baseline runs use 50 epochs and batch size 128 or 256; eICU baseline runs use 30 epochs and batch size 256. Baselines generally use learning rate 10^{-4} and weight decay 10^{-4}; MIMIC MultimodalMIMIC uses learning rate 10^{-5}, 10 epochs, and batch size 64 for vasopressor, intubation, and length of stay (LOS). Supervised models use training-class-weighted cross-entropy. Within each run, checkpoints are selected by validation macro-F1 for AKI/sepsis, PR-AUC for vasopressor/intubation, or quadratic-weighted κ for LOS. We report results over downstream seeds 2027, 2028, and 2029.

The configured task-specific grid searches varied downstream learning rate η, weight decay ω, and auxiliary-loss weights $\lambda_{\mathrm{E}}, \lambda_{\mathrm{S}}$. For MIMIC intervention tasks, the candidate sets were $\eta \in \{6 \times 10^{-5}, 8 \times 10^{-5}\}$, $\omega \in \{10^{-3}, 10^{-2}\}$, $\lambda_{\mathrm{E}} \in \{0, 0.1, 0.5\}$, and $\lambda_{\mathrm{S}} \in \{0.5, 1.0\}$. For MIMIC LOS, they were $\eta \in \{8 \times 10^{-5}, 10^{-4}\}$, $\omega \in \{10^{-3}, 10^{-2}\}$, $\lambda_{\mathrm{E}} \in \{0.6, 0.8, 1.0\}$, and $\lambda_{\mathrm{S}} \in \{0, 0.2\}$. For eICU sepsis, they were $\eta \in \{10^{-4}, 3 \times 10^{-4}, 10^{-3}\}$, $\omega \in \{10^{-3}, 10^{-2}\}$, $\lambda_{\mathrm{E}} \in \{0, 0.1, 0.5\}$, and $\lambda_{\mathrm{S}} \in \{0, 0.2\}$. These are task-specific candidate sets, not a common grid applied to every model.

Table 5: **Detailed results on MIMIC dataset.** This table presents the average performance (± standard deviation) on three MIMIC dataset tasks: AKI, Sepsis and Length of Stay.

Models	AKI			Sepsis			LoS			
	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Kappa
Uin-EHR	0.473 ± 0.003	0.484 ± 0.004	0.458 ± 0.003	0.416 ± 0.009	0.421 ± 0.004	0.341 ± 0.003	0.274 ± 0.004	**0.249 ± 0.008**	0.220 ± 0.003	0.402 ± 0.006
DiPro (Liu et al., 2025)	0.463 ± 0.005	0.453 ± 0.001	0.439 ± 0.003	0.462 ± 0.028	0.403 ± 0.011	0.350 ± 0.010	0.272 ± 0.002	0.237 ± 0.017	0.216 ± 0.003	0.399 ± 0.009
Flex-MoE (Yun et al., 2024)	0.453 ± 0.005	0.451 ± 0.004	0.434 ± 0.004	0.464 ± 0.024	0.379 ± 0.009	0.338 ± 0.007	0.269 ± 0.004	0.233 ± 0.003	0.210 ± 0.005	0.386 ± 0.001
CPTD (Wang et al., 2025)	0.467 ± 0.008	0.456 ± 0.004	0.441 ± 0.001	**0.497 ± 0.025**	0.393 ± 0.002	0.359 ± 0.008	0.264 ± 0.021	0.246 ± 0.004	0.204 ± 0.004	0.380 ± 0.001
UMSE (Lee et al., 2023)	0.462 ± 0.009	0.448 ± 0.002	0.436 ± 0.003	0.449 ± 0.013	0.407 ± 0.003	0.348 ± 0.006	0.260 ± 0.005	0.243 ± 0.008	0.212 ± 0.005	0.392 ± 0.006
UMSE (Zhang et al., 2023)	0.463 ± 0.012	0.470 ± 0.006	0.446 ± 0.003	0.455 ± 0.008	0.428 ± 0.002	0.361 ± 0.003	**0.280 ± 0.002**	0.251 ± 0.006	0.215 ± 0.001	0.391 ± 0.003
TAMME (Susetzky et al., 2025)	0.446 ± 0.007	0.445 ± 0.004	0.426 ± 0.003	0.435 ± 0.015	0.387 ± 0.007	0.334 ± 0.008	0.278 ± 0.007	0.236 ± 0.013	0.208 ± 0.003	0.387 ± 0.007
FuseMoE (Han et al., 2024)	0.460 ± 0.009	0.444 ± 0.005	0.432 ± 0.001	0.476 ± 0.005	0.388 ± 0.030	0.328 ± 0.010	0.277 ± 0.005	0.228 ± 0.004	0.198 ± 0.002	0.376 ± 0.003
PULSE	**0.501 ± 0.002**	**0.491 ± 0.001**	**0.478 ± 0.002**	**0.497 ± 0.039**	**0.436 ± 0.011**	**0.380 ± 0.010**	0.274 ± 0.005	0.241 ± 0.001	**0.224 ± 0.003**	**0.411 ± 0.008**

Table 6: **Detailed results on MIMIC dataset.** This table presents the average performance (± standard deviation) on two MIMIC dataset tasks: Vasopressor and Intubation.

Models	Vasopressor					Intubation				
	Acc	BAcc	F1	AUPRC	ROC-AUC	Acc	BAcc	F1	AUPRC	ROC-AUC
Uni-EHR	**0.788 ± 0.005**	0.775 ± 0.003	0.697 ± 0.004	0.560 ± 0.003	0.855 ± 0.005	0.930 ± 0.003	0.915 ± 0.003	0.906 ± 0.003	0.911 ± 0.004	0.960 ± 0.001
DiPro (Liu et al., 2025)	0.785 ± 0.027	0.770 ± 0.007	0.693 ± 0.018	0.556 ± 0.006	0.856 ± 0.003	0.927 ± 0.002	0.905 ± 0.004	0.902 ± 0.003	0.902 ± 0.002	0.956 ± 0.000
Flex-MoE (Yun et al., 2024)	0.787 ± 0.017	0.763 ± 0.008	0.692 ± 0.009	0.543 ± 0.011	0.847 ± 0.005	0.926 ± 0.001	0.899 ± 0.005	0.899 ± 0.002	0.888 ± 0.004	0.946 ± 0.002
CPTD (Wang et al., 2025)	0.805 ± 0.020	0.763 ± 0.012	**0.705 ± 0.011**	0.558 ± 0.001	0.855 ± 0.002	0.926 ± 0.002	0.901 ± 0.007	0.900 ± 0.004	0.903 ± 0.003	0.955 ± 0.001
UMSE (Lee et al., 2023)	0.792 ± 0.023	0.766 ± 0.011	0.697 ± 0.014	0.557 ± 0.002	0.854 ± 0.002	0.925 ± 0.004	0.906 ± 0.005	0.899 ± 0.004	0.899 ± 0.002	0.955 ± 0.002
UMSE (Zhang et al., 2023)	0.788 ± 0.009	0.777 ± 0.001	0.698 ± 0.006	**0.586 ± 0.003**	**0.864 ± 0.000**	0.931 ± 0.003	**0.918 ± 0.001**	0.909 ± 0.003	0.911 ± 0.002	**0.962 ± 0.001**
TAMME (Susetzky et al., 2025)	0.786 ± 0.012	0.778 ± 0.002	0.697 ± 0.010	0.573 ± 0.002	0.859 ± 0.000	0.927 ± 0.005	0.904 ± 0.007	0.901 ± 0.005	0.905 ± 0.007	0.957 ± 0.001
FuseMoE (Han et al., 2024)	0.782 ± 0.011	0.770 ± 0.008	0.691 ± 0.005	0.561 ± 0.008	0.853 ± 0.001	0.918 ± 0.001	0.904 ± 0.002	0.892 ± 0.001	0.898 ± 0.002	0.954 ± 0.003
PULSE	0.769 ± 0.020	**0.782 ± 0.000**	0.686 ± 0.014	0.576 ± 0.005	**0.864 ± 0.001**	**0.935 ± 0.002**	0.911 ± 0.006	**0.911 ± 0.003**	**0.920 ± 0.004**	**0.962 ± 0.002**

Table 7: **Detailed results on eICU dataset.** This table presents the average performance (± standard deviation) on three eICU dataset tasks: AKI, Sepsis and Length of Stay.

Models	AKI			Sepsis			LoS			
	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Kappa
Uni-EHR	0.467 ± 0.001	0.462 ± 0.009	0.433 ± 0.006	0.485 ± 0.011	0.397 ± 0.015	0.365 ± 0.013	0.258 ± 0.006	0.254 ± 0.003	0.226 ± 0.005	0.475 ± 0.005
DiPro (Liu et al., 2025)	0.456 ± 0.007	0.434 ± 0.004	0.413 ± 0.002	0.463 ± 0.084	0.363 ± 0.017	0.331 ± 0.013	0.254 ± 0.009	0.245 ± 0.003	0.214 ± 0.003	0.456 ± 0.004
Flex-MoE (Yun et al., 2024)	0.457 ± 0.015	0.439 ± 0.008	0.416 ± 0.003	0.443 ± 0.083	0.355 ± 0.002	0.320 ± 0.037	0.262 ± 0.005	0.246 ± 0.000	0.217 ± 0.008	0.462 ± 0.005
CPTD (Wang et al., 2025)	0.454 ± 0.011	0.441 ± 0.009	0.415 ± 0.003	0.489 ± 0.100	0.370 ± 0.002	0.330 ± 0.006	0.264 ± 0.006	0.241 ± 0.004	0.214 ± 0.016	0.462 ± 0.009
UMSE (Lee et al., 2023)	0.454 ± 0.010	0.427 ± 0.005	0.407 ± 0.003	0.452 ± 0.033	0.385 ± 0.012	0.343 ± 0.012	0.256 ± 0.004	0.246 ± 0.001	0.218 ± 0.006	0.448 ± 0.007
UMSE (Zhang et al., 2023)	0.468 ± 0.005	0.477 ± 0.004	0.442 ± 0.004	0.480 ± 0.029	0.397 ± 0.003	0.362± 0.004	0.269 ± 0.003	0.253 ± 0.001	0.225 ± 0.000	0.476 ± 0.003
TAMME (Susetzky et al., 2025)	0.428 ± 0.003	0.428 ± 0.008	0.399 ± 0.003	0.543 ± 0.032	0.390 ± 0.005	0.364 ± 0.002	0.260 ± 0.005	0.248 ± 0.001	0.219 ± 0.005	0.457 ± 0.001
FuseMoE (Han et al., 2024)	0.460 ± 0.010	0.434 ± 0.005	0.414 ± 0.002	0.496 ± 0.069	0.388 ± 0.009	0.355 ± 0.013	0.252 ± 0.001	0.240 ± 0.002	0.215 ± 0.002	0.461 ± 0.002
PULSE	**0.516 ± 0.006**	**0.534 ± 0.013**	**0.494 ± 0.009**	**0.554 ± 0.043**	**0.393 ± 0.017**	**0.381 ± 0.019**	**0.294 ± 0.003**	**0.275 ± 0.003**	**0.235 ± 0.003**	**0.492 ± 0.004**

Table 8: **Detailed results on eICU dataset.** This table presents the average performance (± standard deviation) on two eICU dataset tasks: Vasopressor and Intubation.

Models	Vasopressor					Intubation				
	Acc	BAcc	F1	AUPRC	ROC-AUC	Acc	BAcc	F1	AUPRC	ROC-AUC
Uni-EHR	0.804 ± 0.014	0.792 ± 0.003	0.673 ± 0.010	0.496 ± 0.005	0.873 ± 0.001	0.838 ± 0.003	0.839 ± 0.002	0.838 ± 0.003	0.903 ± 0.002	0.908 ± 0.002
DiPro (Liu et al., 2025)	0.796 ± 0.015	0.802 ± 0.001	0.671 ± 0.011	0.497 ± 0.005	0.879 ± 0.001	0.852 ± 0.002	0.852 ± 0.002	0.852 ± 0.002	0.920 ± 0.002	0.923 ± 0.002
Flex-MoE (Yun et al., 2024)	0.794 ± 0.010	0.804 ± 0.001	0.669 ± 0.008	0.504 ± 0.004	0.879 ± 0.001	0.850 ± 0.005	0.849 ± 0.005	0.850 ± 0.005	0.920 ± 0.002	0.923 ± 0.003
CPTD (Wang et al., 2025)	0.765 ± 0.032	0.799 ± 0.001	0.647 ± 0.023	0.493 ± 0.010	0.876 ± 0.001	0.846 ± 0.004	0.845 ± 0.004	0.845 ± 0.004	0.915 ± 0.001	0.919 ± 0.001
UMSE (Lee et al., 2023)	0.804 ± 0.017	0.802 ± 0.002	0.676 ± 0.013	0.513 ± 0.003	0.880 ± 0.001	0.860 ± 0.003	0.860 ± 0.003	0.859 ± 0.003	0.925 ± 0.000	0.927 ± 0.001
UMSE (Zhang et al., 2023)	0.793 ± 0.016	**0.808 ± 0.001**	0.670 ± 0.012	0.524 ± 0.002	**0.885 ± 0.001**	0.859 ± 0.001	0.858 ± 0.001	0.858 ± 0.001	0.924 ± 0.002	0.927 ± 0.001
TAMME (Susetzky et al., 2025)	0.786 ± 0.013	0.804 ± 0.004	0.663 ± 0.009	0.517 ± 0.004	0.881 ± 0.003	0.858 ± 0.004	0.858 ± 0.005	0.858 ± 0.005	0.925 ± 0.002	0.928 ± 0.002
FuseMoE (Han et al., 2024)	0.789 ± 0.028	0.799 ± 0.009	0.665 ± 0.020	0.500 ± 0.004	0.876 ± 0.005	0.856 ± 0.004	0.856 ± 0.005	0.856 ± 0.005	0.921 ± 0.001	0.926 ± 0.002
PULSE	**0.832 ± 0.025**	0.795 ± 0.019	**0.697 ± 0.015**	**0.534 ± 0.010**	0.882 ± 0.007	**0.871 ± 0.001**	**0.870 ± 0.001**	**0.870 ± 0.001**	**0.932 ± 0.002**	**0.934 ± 0.001**

Table 9: Test-sample count and inter-event-gap standard deviation Δ_σ by input history length.

	MIMIC AKI		MIMIC Intubation		eICU AKI		eICU Intubation	
Input history length (h)	Count	Δ_σ (h)	Count	Δ_σ (h)	Count	Δ_σ (h)	Count	Δ_σ (h)
36	3,627	5.27	4,483	5.25	5,108	3.58	13,632	3.94
48	2,777	6.71	3,312	6.70	4,049	4.68	10,991	5.15
60	2,197	7.97	2,578	7.95	3,067	5.60	8,675	6.11
72	1,703	9.02	1,966	9.03	2,572	6.23	7,265	7.01
84	1,362	9.75	1,550	9.80	2,048	7.06	5,984	7.89
96	1,070	10.21	1,206	10.30	1,758	8.03	5,181	9.02

Table 10: **Detailed Ablation Study results on MIMIC dataset.** This table presents the average performance (± standard deviation) on three MIMIC dataset tasks: AKI, Sepsis and Length of Stay.

Models	AKI			Sepsis			LoS			
	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Kappa
PULSE	**0.501 ± 0.002**	**0.491 ± 0.001**	**0.478 ± 0.002**	0.497 ± 0.039	**0.436 ± 0.011**	**0.380 ± 0.010**	0.274 ± 0.005	0.241 ± 0.001	**0.224 ± 0.003**	**0.411 ± 0.008**
w/o Medical prior	0.459 ±0.006	0.437 ±0.004	0.428 ±0.004	**0.535 ±0.020**	0.393 ±0.004	0.370 ±0.005	0.283 ±0.010	0.215 ±0.002	0.200 ±0.000	0.375 ±0.007
w/o Physiological prior	0.494 ±0.001	0.469 ±0.001	0.461 ±0.001	0.474 ±0.029	0.430 ±0.013	0.370 ±0.015	0.270 ±0.004	0.230 ±0.002	0.213 ±0.003	0.384 ±0.007
w/o $\mathcal{L}_{event}$	0.451 ±0.010	0.467 ±0.005	0.438 ±0.004	0.458 ± 0.028	0.430 ± 0.023	0.363 ± 0.020	**0.254 ±0.009**	0.189 ±0.006	0.380 ±0.005	
w/o Phase-aware attention	0.492 ±0.005	0.484 ±0.007	0.468 ±0.002	0.462 ±0.021	**0.438 ±0.015**	0.369 ±0.015	**0.289 ±0.006**	0.242 ±0.004	0.223 ±0.000	0.404 ±0.009

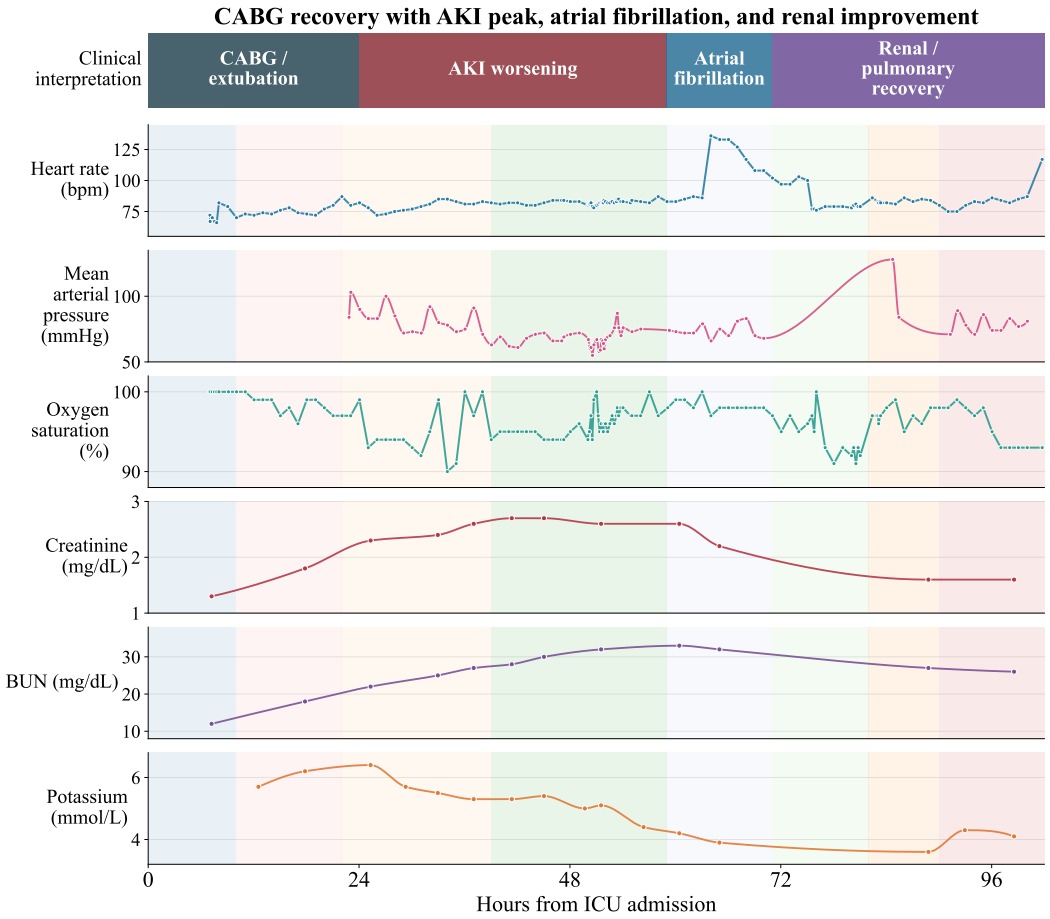

Figure 4: **Qualitative phase discovery in a postoperative CABG case.** The top band shows the clinical interpretation of the trajectory, while the lower panels display heart rate, mean arterial pressure, oxygen saturation, creatinine, BUN, and potassium. The background colors indicate the phases discovered by PULSE. The inferred phase transitions are consistent with the patient's clinical evolution: the transition from CABG and extubation to AKI worsening coincides with increasing creatinine and BUN and elevated potassium; the atrial fibrillation phase is marked by a sharp increase in heart rate; and the final renal and pulmonary recovery phase is accompanied by decreasing creatinine, BUN, and potassium with relatively stable oxygenation. The fine-grained phase changes within these broader clinical states suggest that PULSE captures gradual physiological progression rather than merely separating the trajectory according to time.

Table 11: **Detailed Ablation Study results on MIMIC dataset.** This table presents the average performance (± standard deviation) on two MIMIC dataset tasks: Vasopressor and Intubation.

Models	Vasopressor					Intubation				
	Acc	BAcc	F1	AUPRC	ROC-AUC	Acc	BAcc	F1	AUPRC	ROC-AUC
PULSE	0.769 ± 0.020	**0.782 ± 0.000**	0.686 ± 0.014	**0.576 ± 0.005**	**0.864 ± 0.001**	**0.935 ± 0.002**	0.911 ± 0.006	**0.911 ± 0.003**	**0.920 ± 0.004**	**0.962 ± 0.002**
w/o Medical prior	0.759 ±0.012	0.756 ±0.002	0.670 ±0.009	0.502 ±0.018	0.834 ±0.003	0.923 ±0.001	0.907 ±0.002	0.898 ±0.001	0.887 ±0.006	0.946 ±0.004
w/o Physiological prior	0.783 ±0.009	0.771 ±0.001	0.692 ±0.006	0.551 ±0.006	0.850 ±0.001	0.932 ±0.004	0.913 ±0.008	0.908 ±0.005	0.910 ±0.009	0.957 ±0.005
w/o $\mathcal{L}_{event}$	0.730 ±0.055	0.774 ±0.005	0.657 ±0.040	0.569 ±0.008	0.862 ±0.002	0.929 ±0.003	**0.923 ±0.001**	0.907 ±0.004	0.914 ±0.002	**0.962 ±0.001**
w/o Phase-aware attention	**0.774 ±0.008**	**0.782 ±0.000**	**0.689 ±0.006**	0.571 ±0.002	0.862 ±0.002	0.933 ±0.001	0.914 ±0.003	0.909 ±0.002	0.913 ±0.005	0.959 ±0.003

Table 12: **Detailed ablation study results on eICU dataset.** This table presents the average performance (± standard deviation) on three eICU dataset tasks: AKI, Sepsis and Length of Stay.

Models	AKI			Sepsis			LoS			
	Acc	BAcc	F1	Acc	BAcc	F1	Acc	BAcc	F1	Kappa
PULSE	**0.516 ± 0.006**	**0.534 ± 0.013**	**0.494 ± 0.009**	**0.554 ± 0.043**	**0.393 ± 0.017**	**0.381 ± 0.019**	**0.294 ± 0.003**	**0.275 ± 0.003**	**0.235 ± 0.003**	**0.492 ± 0.004**
w/o Medical prior	0.457 ± 0.002	0.464 ± 0.006	0.430 ± 0.003	0.534 ± 0.050	0.361 ± 0.032	0.341 ± 0.030	0.274 ± 0.004	0.262 ± 0.003	0.234 ± 0.008	0.489 ± 0.000
w/o Physiological prior	0.512 ± 0.007	0.518 ± 0.013	0.483 ± 0.004	0.460 ± 0.014	0.363 ± 0.022	0.332 ± 0.015	0.272 ± 0.008	0.251 ± 0.007	0.212 ± 0.017	0.481 ± 0.017
w/o $\mathcal{L}_{event}$	0.494 ± 0.006	0.516 ± 0.006	0.472 ± 0.005	0.560 ± 0.037	0.391 ± 0.015	0.370 ± 0.005	0.280 ± 0.007	0.262 ± 0.007	0.234 ± 0.009	0.487 ± 0.003
w/o Phase-aware attention	0.506 ± 0.010	0.528 ± 0.012	0.485 ± 0.011	0.555 ± 0.063	0.381 ± 0.024	0.366 ± 0.002	0.256 ± 0.009	0.241 ± 0.003	0.206 ± 0.018	0.479 ± 0.008

Table 13: **Detailed ablation study results on eICU dataset.** This table presents the average performance ($\pm$ standard deviation) on two eICU dataset tasks: Vasopressor and Intubation.

Models	Vasopressor					Intubation				
	Acc	BAcc	F1	AUPRC	ROC-AUC	Acc	BAcc	F1	AUPRC	ROC-AUC
PULSE	**0.832 ± 0.025**	0.795 ± 0.019	**0.697 ± 0.015**	**0.534 ± 0.010**	0.882 ± 0.007	**0.871 ± 0.001**	**0.870 ± 0.001**	**0.870 ± 0.001**	**0.932 ± 0.002**	**0.934 ± 0.001**
w/o Medical prior	0.804 ± 0.036	0.797 ± 0.009	0.676 ± 0.027	0.525 ± 0.002	0.881 ± 0.001	0.862 ± 0.002	0.862 ± 0.002	0.862 ± 0.002	0.926 ± 0.002	0.930 ± 0.001
w/o Physiological prior	0.819 ± 0.032	0.797 ± 0.015	0.688 ± 0.022	0.530 ± 0.006	0.882 ± 0.006	0.868 ± 0.002	0.869 ± 0.002	0.868 ± 0.002	0.929 ± 0.001	0.933 ± 0.000
w/o $\mathcal{L}_{event}$	0.792 ± 0.019	0.801 ± 0.001	0.667 ± 0.015	0.512 ± 0.004	0.879 ± 0.001	0.868 ± 0.002	0.868 ± 0.002	0.868 ± 0.002	0.929 ± 0.000	0.933 ± 0.001
w/o Phase-aware attention	0.814 ± 0.019	0.802 ± 0.004	0.684 ± 0.014	0.533 ± 0.003	0.885 ± 0.001	0.867 ± 0.003	0.868 ± 0.003	0.867 ± 0.003	0.929 ± 0.003	0.933 ± 0.002