OICYCLE: OBSERVE, INFER, AND INTERVENE CYCLE FOR BELIEF-STATE LEARNING IN LONGITUDINAL CLINICAL DECISION MAKING

Anonymous authors
Paper under double-blind review

ABSTRACT

Clinical care is a closed loop: clinicians observe partial evidence, infer the patient's state, and intervene, and each intervention determines what is observed next. This makes clinical decision making a partially observable Markov decision process (POMDP), in which a belief over a latent patient state must be updated from irregular, multimodal, and incomplete records. Yet most deep clinical models predict from a single snapshot or from coarsely aggregated history, while POMDP-based clinical methods need hand-specified discrete state spaces and transition models that real electronic health records cannot supply. We introduce **OICycle** (Observe, Infer, and Intervene Cycle), which learns a stochastic belief state directly from retrospective trajectories. At every episode, a learned recurrent operator fuses a sample of the previous belief with the new multimodal observations and the intervention that produced them. The resulting Gaussian belief is trained to predict clinicians' next interventions, the time to the next episode, and next-episode mortality. To enable reproducible evaluation, we release a benchmark derived from MIMIC-IV, MIMIC-IV-Note, and MIMIC-CXR that organizes 72,075 hospital admissions (5.4M laboratory, microbiology, radiology, prescription, and note records) into temporally ordered therapeutic–diagnostic episodes. Compared with an otherwise identical history-free model, OICycle improves next-intervention adaptive recall from 0.253 to 0.347, time-to-next-episode macro-F1 from 0.293 to 0.351, and next-episode mortality macro-F1 from 0.585 to 0.649. Analysis shows that the learned belief encodes intervention type and trajectory position and evolves smoothly across episodes, and that the benefit of history is concentrated in interventions whose ordering depends on prior findings.

1 INTRODUCTION

Longitudinal Clinical Decision Making (LDCM) is a sequential process in which clinicians act as a patient's state evolves, spanning single admissions to years of care. Figure 1 presents this process. The history of interventions is crucial but often overlooked by AI models that consider only isolated snapshots, leading to suboptimal decisions. Clinical data are also heterogeneous, irregular, incomplete, and only partially observe the true patient state, further complicating LDCM (Jayaraman et al.; Najafi et al.; Yu et al.).

Prior work on POMDP-inspired frameworks has largely focused on specific diagnostic or treatment tasks. It does not comprehensively address the broader complexity of real-world clinical trajectories, which are diverse and irregular, and often rely on regular, homogeneous data and limited intervention types. As a result, these methods do not fully capture the fine-grained temporal and multimodal complexity of real longitudinal patient care.

OICycle: Observe, Infer, and Intervene Cycle addresses this gap by introducing a unified, predictive belief-state learning framework that models the temporal evolution of patient trajectories through recurrent therapeutic-diagnostic cycles. Existing latent-dynamics models often learn latent representations by modeling or reconstructing observed clinical data. This objective can be limiting for clinical data for two reasons. First, observations span free text, images, and thousands of sparse codes, so reconstruction spends capacity on detail that does not matter for decisions. Second, an

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

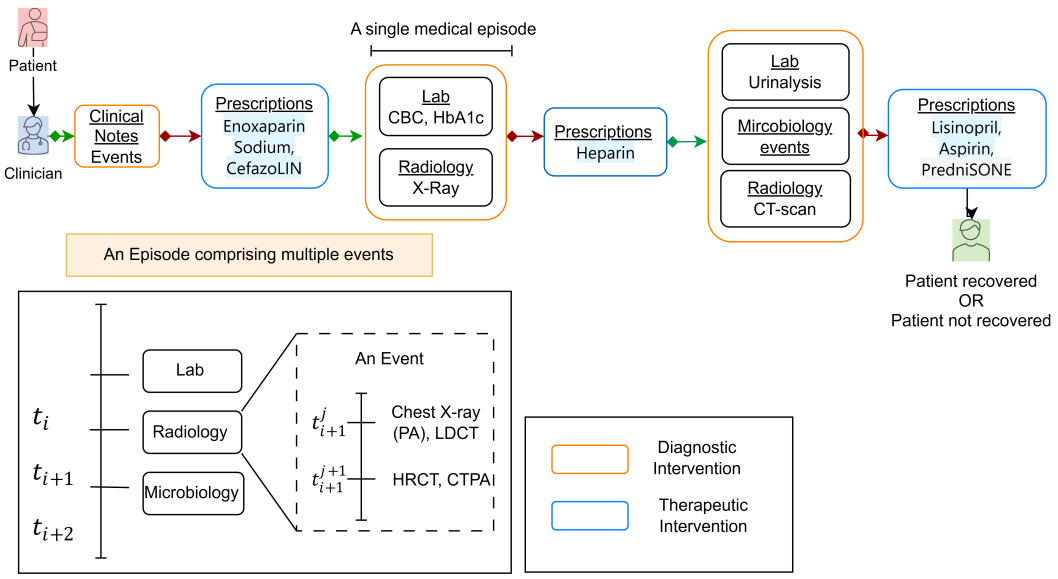


Figure 1: The clinical decision-making cycle often involves a follow-up diagnostic-therapeutic intervention cycle of episodes. Each episode can have multiple events, and each event can have multiple records.

observation exists only if a clinician chose to acquire it, so its absence is itself informative and an observation likelihood would have to model the ordering policy too. OICycle departs from these models in three ways that are, to our knowledge, not previously combined:

- It reformulates the clinical POMDP as predictive belief-state learning: the belief is shaped by its ability to forecast future clinical events rather than regenerate observations, removing the need for an explicit state space, transition model, or observation model.
- It makes the *action-conditioned therapeutic-diagnostic episode* the unit of belief update, so that every update is tied to the intervention that caused the new evidence. Sequence models that treat interventions as covariates, and clinical POMDPs defined over fixed time steps, both lack this structure
- It pairs the model with a benchmark that, unlike existing MIMIC-derived resources, exposes this action-observation structure across modalities, making such models evaluable.

In summary, this paper makes the following key contributions:

- We introduce OICycle as a sequential predictive latent belief-state learning framework for longitudinal clinical decision-making.
- We release the data preprocessing code to provide open access to a derived MIMIC dataset that organizes patients' clinical trajectories into meaningful clinical episodes
- We evaluate our results on three predictive tasks: next-intervention selection, time-to-next-episode forecasting, and next-episode mortality prediction, compare them against a single-snapshot clinical model, and analyze the predictive belief-state representation.

2 Related Works

Early works (Li et al. (c); Lee et al.; Huang et al.; Choi et al.; Rasmy et al.; Peng et al.) primarily learn local clinical context using BERT-based models (Devlin et al.) or use a graph model to structure relationships between medical concepts. However, these approaches do not account for the patient's medical trajectory.

Closest to OICycle are decision-making frameworks like DECIDE-Lab (Allenbrand), which uses Value of Information for diagnostic selection and maintains a Bayesian posterior over explicitly de-

fined clinical states. Similarly, DPWM (Li et al. (a)) learns intervention-conditioned dynamics from recorded clinical trajectories to simulate counterfactual clinical scenarios, but focuses on preoperative risk assessment and intervention evaluation. By contrast, OICycle is designed for retrospective longitudinal data and learns a unified probabilistic predictive belief-state over a latent clinical representation that integrates observations from both historical interventions and diverse clinical events. While FORBES (Chen et al.) also models a probabilistic belief state, it does so in a general problem framework, not in a clinical setting. It does not address the challenges present in clinical data. This distinction underlines OICycle's focus on realistically structured longitudinal clinical trajectories.

(Li et al. (b); Krishnan et al.) learn probabilistic belief states within POMDP/RL frameworks but operate at coarse temporal resolution (inter-episode). In contrast, OICycle models fine-grained, multimodal intra-episode data, which is prevalent in real clinical scenarios (Tsoukalas et al.; Futoma et al.), and updates Bayesian beliefs using discrete, manually defined states with estimated transition/observation probabilities. OICycle does not require a predefined discrete state space or explicit transition and observation probabilities; instead, it learns the belief representation directly from observed retrospective multimodal trajectories. DeepCare (Pham et al.) predicts outcomes using coarse temporal windows (multiple admissions), missing single-admission intra-episode complexities. The POMDP approach in (Hauskrecht & Fraser) also relies on discrete, explicitly defined state spaces, unlike OICycle's continuous latent modeling.

In summary, prior research has established key foundations, but challenges remain in jointly modeling fine-grained intra-episode temporal structure and heterogeneous modalities. Most methods do not fully leverage temporal dependencies between sequential interventions and subsequent observations. OICycle addresses these by unifying belief-state learning and predictive modeling across multimodal, asynchronous, longitudinal data.

Public longitudinal multimodal datasets are scarce and face challenges like high dimensionality, privacy, and trade-offs between resolution and storage (Acosta et al.). MIMIC-IV (Johnson et al. (c)) is widely used but requires processing. HiRID-ICU-benchmark (Yèche et al.) and EHRSHOT (Wornow et al.) offer benchmarking pipelines but contain only ICU or EHR data. Various derived datasets exist for specific tasks, while preprocessing tools (Yang et al.; Wang et al.) enable standardized pipelines but are not tailored for fine-grained multimodal episode modeling.

Thus, while existing resources provide longitudinal and multimodal clinical data, OICycle's derived dataset organizes heterogeneous observations and interventions into clinically meaningful episodes, providing a standardized basis for benchmarking sequential clinical decision-making models.

3 PRELIMINARIES

3.1 PROBLEM FORMULATION: CLINICAL DECISION MAKING PROCESS

The clinical decision-making process can be formulated as a partially observable Markov Decision Process (POMDP) (Kaelbling et al.). A POMDP is formally defined as a 6-tuple:

$$< \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathbb{T}, \mathcal{R}, \Omega > \tag{1}$$

where, $\mathcal{S}, \mathcal{A}, \mathcal{O}$: is a set of states, actions, and observations

$\mathbb{T}$: is the transition probability function $\mathbb{T}(s_{t+1}|s_t, a_t) : \mathcal{S} \times \mathcal{S} \times \mathcal{A}$

$\mathcal{R}$: is the reward function $\mathcal{R}(s_t, a_t) : \mathcal{S} \times \mathcal{A}$

Ω : is the observation probability function $\Omega(o_{t+1}|s_{t+1}, a_t) : \mathcal{O} \times \mathcal{S} \times \mathcal{A}$

However, real-world clinical data rarely provide the necessary information to derive these explicit transition and observation functions. Therefore, inspired by the POMDP framework, we abstracted the OICycle clinical decision-making process as sequential predictive belief-state learning, defined as a 4-tuple:

$$< \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{F}_\theta > \tag{2}$$

where, $\mathcal{S}, \mathcal{A}, \mathcal{O}$ is a set of latent clinical patient states, interventions, and observations and $\mathcal{F}_\theta$ is the learned recurrent operator that updates the latent belief state from observations and interventions. At each prediction point, the model has access to the clinical history accumulated up to that point (H_t), defined as

$$H_t = \{(a_0^*, o_1), (a_1^*, o_2) \dots (a_{t-1}^*, o_t)\} \tag{3}$$

where, a^* is the realized/ground truth clinical intervention(action), a_0^* is the initial clinical assessment by the clinician, a_{t-1}^* is the realized clinical intervention at episode $t-1$, and o_t is the clinical observations at episode t.

3.2 Clinical Episode Embedding

At each episode, the model receives multimodal observations. These observations are acquired after the realized clinical action taken in the preceding episode and are therefore temporally coupled with that action. We construct the episode embedding X_t as

$$X_t = E(o_t, a_{t-1}^*); \text{where, } E \text{ denotes the multimodal episode encoder.} \tag{4}$$

3.3 Learned Recursive Update Operator

We then combined the sampled predictive belief from the previous episode with the embedding of the current clinical episode using a recurrent operator F_θ. In our implementation, F_θ is instantiated as a Gated Recurrent Unit (GRU) (Cho et al.), defined as

$$r_t = F_\theta(z'_{t-1}, X_t) \tag{5}$$

where, r_t is the recurrent update representation produced by the operator F_θ at episode t and z'_{t-1} denotes a sample from the learned predictive belief distribution at episode $t-1$.

3.4 Probabilistic Belief-state parametrization

We introduced a small belief network g_ϕ as a two-layer fully connected neural network that transforms the recurrent update representation into the parameters of learned predictive distribution over the latent belief-state representation. It is defined as

$$m_t = g_\phi(r_t); \mu_t = w_\mu m_t + b_\mu; \log \sigma_t^2 = w_\sigma m_t + b_\sigma \tag{6}$$

where $\mu, \log \sigma_t^2 \in \mathbb{R}^{512}$ are mean and log variance of the learned predictive belief-state distribution. We model the learned predictive belief-state distribution q_ϕ defined as

$$q_\phi(z_t|H_t) \approx \mathcal{N}(\mu_t, diag(\sigma_t^2)) \tag{7}$$

where, z_t denotes the random variable for the learned latent predictive belief-state representation, $z_t \in \mathbb{R}^{512}$.

3.5 Differentiable latent belief sampling

To obtain a sampled latent belief representation z'_t from the learned predictive belief-state distribution q_ϕ while preserving differentiability, we use the reparameterization trick (Kingma & Welling) as defined as:

$$z'_t \sim q_\phi(z_t|H_t)$$
$$\text{Let } \epsilon_t \sim \mathcal{N}(0, I_{512}), \text{ then } z'_t = \mu_t + \sigma_t \odot \epsilon_t \tag{8}$$

where $\sigma_t = \exp(\frac{1}{2}\log \sigma_t^2)$, z_t is the latent predictive belief-state random variable at episode t, and z'_t is the sampled latent belief representation at episode t.

3.6 Latent belief-guided prediction

Then the sampled latent belief representation is used to predict the next episode's clinical interventions (actions) as

$$p_\psi(a_{t+1}|z'_t); a_{t+1} \in \mathcal{A} \tag{9}$$

The prediction losses (defined in Eq. 10) are computed against the corresponding realized clinical actions (a_{t+1}^*) and backpropagated through the sampled latent belief representation. The process is recursively applied across the patient's clinical episodes until the episode sequence terminates, resulting in a POMDP-inspired sequential predictive belief-state learning process OICycle.

$$\mathcal{L}_{int}^t = \lambda_{ty}\mathcal{L}_{ty}^t + \lambda_{sty}\mathcal{L}_{sty}^t \tag{10}$$

$$\text{where, } \mathcal{L}_{sty} = -\sum_k w_k^{sty}[y_{t+1,k}^{sty}\, log(\hat{y}_{t+1,k}^{sty}) + (1 - y_{t+1,k}^{sty}) log(1 - \hat{y}_{t+1,k}^{sty})]$$

$$\mathcal{L}_{ty} = -\sum_k w_k^{ty}[y_{t+1,k}^{ty}\, log(\hat{y}_{t+1,k}^{ty}) + (1 - y_{t+1,k}^{ty}) log(1 - \hat{y}_{t+1,k}^{ty})]$$

Here, $\mathcal{L}_{int}^t$ is the total loss, $\mathcal{L}_{ty}^t$ and $\mathcal{L}_{sty}^t$ are losses for intervention types and subtypes, w_k^{ty} and w_k^{sty} are class weights for type and subtype imbalance, and λ_{ty}, λ_{sty} control the respective loss terms.

4 METHODOLOGY

4.1 CLINICAL DATASET AND EPISODE CONSTRUCTION

We constructed a multimodal longitudinal episodic dataset by integrating MIMIC-IV Johnson et al. (c), MIMIC-IV-Note Johnson et al. (a), and MIMIC-CXR-JPG Johnson et al. (b). From 364,627 patients in MIMIC-IV, we selected each patient's first hospital admission with both a MIMIC-IV-Note and at least one DRG code, restricting to one admission per patient for consistency and preventing data leakage. The final cohort includes 72,075 patients with 349,489 unstructured text records, 5,057,180 structured clinical records, and 15,044 radiological images. Appendix A.2 summarizes dataset statistics and episode distributions.

4.1.1 CLINICAL EPISODE CONSTRUCTION

A *clinical intervention* is the smallest data unit, categorized as either *Therapeutic Intervention* (TI; prescriptions) or *Diagnostic Intervention* (DI; notes, labs, microbiology, radiology). Each may include multiple subtypes, such as several lab tests within a laboratory intervention.

A medical episode is a temporal unit containing either a TI or DI. TI and DI typically alternate but may occur consecutively; note-only DI episodes can appear at the start of a trajectory. Modality-specific temporal windows group interventions into episodes primarily by intervention type, and few conflicts were resolved when long gaps separate multiple single interventions by using a fixed temporal window per intervention type.

We define a *TI-DI episode* as a temporally bounded clinical unit consisting of a TI followed by the subsequent DI, where the TI represents a realized therapeutic intervention and the DI represents the resulting clinical observations. In addition, organizing data into TI-DI episodes enables modeling of the sequential cause-and-effect relationships that are central to clinical reasoning. Consequently, this construct serves as the basic temporal unit for our sequential predictive belief-state learning framework.

4.1.2 PREDICTION TASKS

Our episodic representation supports three prediction tasks. First, the primary task predicts the intervention type and subtype of the next clinical episode, providing the main learning signal for belief-state representation. Second, we predict the elapsed time to the next episode by discretizing intervals into temporal bins, evaluating whether the representation captures expected timing. Only episodes with an observed subsequent episode are included. Third, we predict episode-level mortality in the subsequent episode. We exclude final episodes from episode-level mortality.

4.2 MODEL ARCHITECTURE

Figure 2 presents an overview of the OICycle architecture. Initially, the model receives the initial state z_0 and clinical notes X_1, which are input into Q_ϕ (the learned predictive belief-state distribution). Within Q_ϕ, F_θ—the recurrent update operator implemented using a GRU—processes the inputs sequentially. The output of this sequence passes to a small belief network that produces the parameters of the learned predictive belief-state distribution. Next, an intermediate state $z'_{i,1}$ is sampled, where i refers to the TI intervention within a TI-DI episode, facilitating the propagation of state information. At each prediction point, the model generates predictions and backpropagates

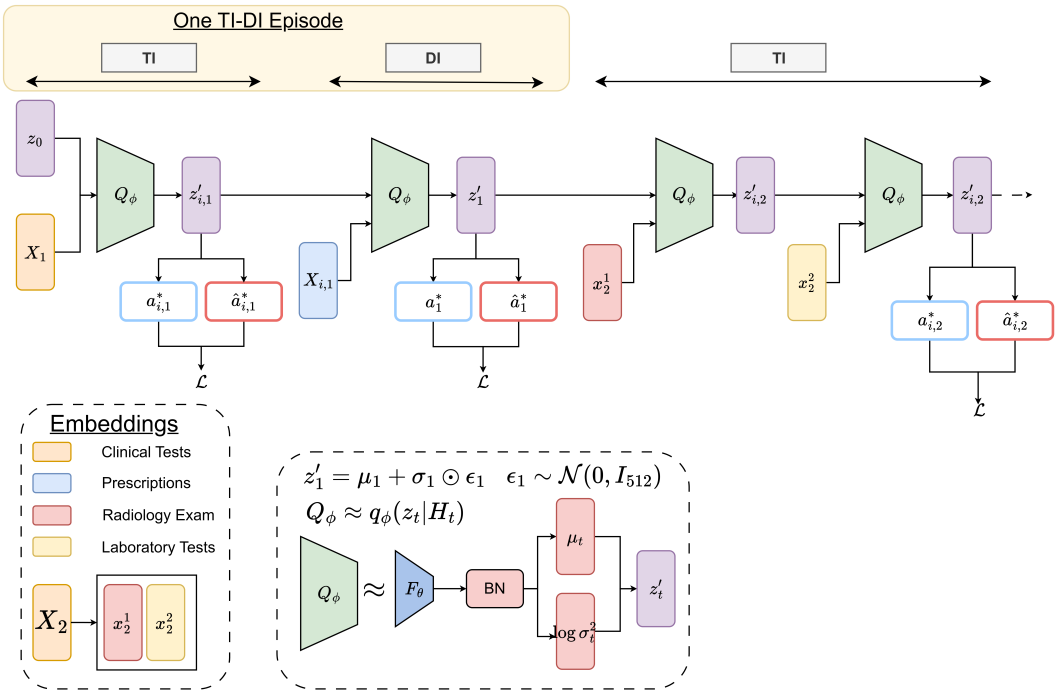


Figure 2: Overview of OICycle architecture. z_0 is the initial random state from where the sequential belief-state learning starts followed by a series of TI and DI episode.

the associated loss. If a single episode comprises multiple events (for example, X_2 includes a radiology exam embedding x_2^1 and a laboratory test embedding x_2^2), the system processes each event embedding in sequence and generates the prediction only after processing the final embedding. For medical text embeddings, used for both clinical notes and radiology notes, we use BioClinicalBERT Ling to obtain the embeddings, followed by a projection layer to map them back to the common embedding dimension. For some patients, radiology images are also available; in these cases, we use image embeddings instead of radiology note embeddings.

5 Experiments & Results

5.1 Experimental Setup

We ran two models: our proposed model, OICycle that learns from patient trajectory history, and OICycle-Single (OICycle-S), which predicts only from current-episode observations. Both share the same architecture, except that OICycle incorporates patient history. The initial state is a parametrized object initialized with Xavier uniform initialization, and the hyperparameters used are shown in the Appendix (A.1). We ran these models for 10 epochs on an NVIDIA RTX 5000 Ada Generation GPU. We used our derived OICycle dataset for training, validation, and testing, split at a 70:15:15 ratio. The dataset used multilabel stratification on *patient_id*, *drg_code*, and task-related predictive targets. A data point in the dataset represents a single patient with all its episodic data. As mentioned in (Section 4.1.2), we have three predictive tasks: Task 1, Task 2, and Task 3. For Task 1, intervention and sub-intervention type prediction, we used a class-imbalance-aware BCE loss defined in Eq. (10). For Task 2, the temporal prediction task, where the next-episode time interval belongs to one of C mutually exclusive temporal bins, we use multiclass cross-entropy:

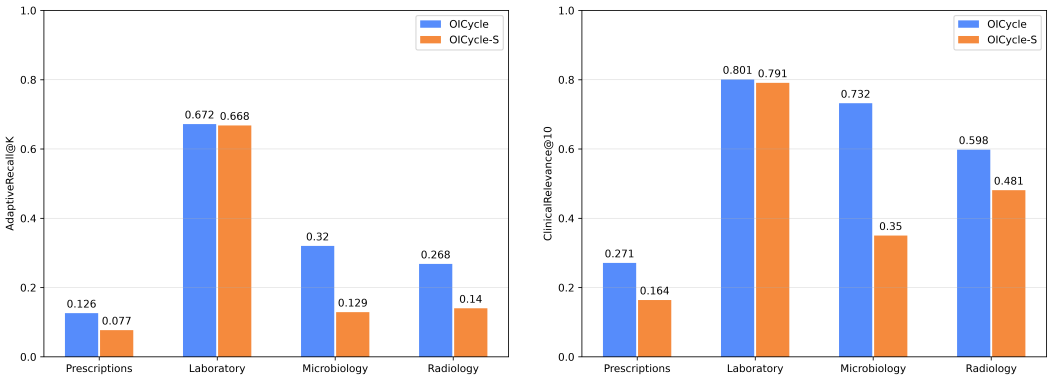
$$\mathcal{L}_{time}^{t} = -\sum_{c=1}^{C} y_{t+1,c}^{time} log(\hat{y}_{t+1,c}^{time}) \tag{11}$$

And for Task 3: mortality prediction, we use binary cross-entropy:

$$\mathcal{L}_{m}^{t} = -[y_{t+1}^{m} log(\hat{y}_{t+1}^{m}) + (1 - y_{t+1}^{m}) log(1 - \hat{y}_{t+1}^{m})] \tag{12}$$

Table 1: Results of OICycle and OICycle-S across all three predictive tasks. Values are reported with 95% confidence intervals estimated using bootstrap resampling for statistical uncertainty assessment.

Algorithms	Task 1		Task 2	Task 3
	Mean AR	Mean CR	Mean F1	Mean F1
OICycle	**0.347**(0.344-0.350)	**0.601**(0.597-0.604)	**0.351**(0.346-0.356)	**0.649**(0.622-0.674)
OICycle-S	0.253(0.252-0.255)	0.446((0.443-0.449))	0.293(0.292-0.295)	0.585(0.571-0.599)



(a) Adaptive Recall@K

(b) Clinical Relevance@10

Figure 3: Comparison of intervention-level prediction performance between OICycle & OICycle-S

Hence, the overall loss is then defined as:

$$\mathcal{L} = \sum_t [\lambda_{int}\mathcal{L}^t_{int} + \lambda_{time}\mathcal{L}^t_{time} + \lambda_m\mathcal{L}^t_m] \tag{13}$$

where λ_{int}, λ_{time}, and λ_m control the relative contributions of the three predictive tasks.

For Task 1 evaluation, because a clinical episode may contain multiple intervention subtypes and the number of relevant interventions varies across episodes, we use both adaptive and fixed-K retrieval metrics. Therefore, we introduce two derived metrics: *AdaptiveRecall@K* (AR) and *ClinicalRelevance@10* (CR). AdaptiveRecall@K measures the proportion of relevant interventions recovered within a prediction set whose size adapts to the number of relevant interventions in the ground truth of that particular episode. Whereas ClinicalRelevance@10 evaluates the clinical relevance of the top-10 predicted interventions. Their formulations are given in Eqs: 14, 15

$$\text{AdaptiveRecall@K} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|\mathcal{Y}_t \cap \hat{\mathcal{Y}}_t^{K_t}\right|}{|\mathcal{Y}_t|}, \qquad K_t = |\mathcal{Y}_t|. \tag{14}$$

$$\text{ClinicalRelevance@10} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|\hat{\mathcal{Y}}_t^{10} \cap \mathcal{Y}_t\right|}{\min(10, |\mathcal{Y}_t|)}, \tag{15}$$

Here, T is the number of evaluated episodes, $\mathcal{Y}_t$ is the set of ground-truth interventions for the episode t, and $\hat{\mathcal{Y}}_t^{K_t}$ contains the model's top-K_t predictions, $\hat{\mathcal{Y}}_t^{10}$ denotes the top-10 predicted interventions. For Task 2 and Task 3, we use the macro-F1 score as the primary evaluation metric.

5.2 Main Results

Table 1 summarizes OICycle and OICycle-S results on all three tasks. OICycle outperforms on Task 1, with a mean AR of 0.34 and a mean CR of 0.60, compared with OICycle-S, which achieves a mean AR of 0.22 and a mean CR of 0.48. For Task 2, OICycle-L leads by 0.19 points with an F1 score of 0.35; for Task 3, it leads by 0.15 points with an F1 score of 0.69.

Figure 3 shows intervention-level performance across categories. Laboratory interventions achieve the highest performance, while prescriptions are lowest, with microbiology and radiology in between. This likely reflects the structured, numerical nature of laboratory test data, which supports more accurate predictions. In contrast, prescriptions involve indirect and confounded relationships between orders and outcomes, making prediction more challenging. Microbiology and radiology data include more unstructured or qualitative elements, such as free-text reports, further complicating predictive modeling. These results highlight that input data structure and interpretability are key determinants of predictive accuracy across clinical intervention types.

6 Analysis & Discussion

We analyze the learned predictive belief state from three perspectives: representational structure, state dynamics, and the role of history accumulation (Figure 4).

First, we examine what clinically relevant information is encoded in the belief state, including intervention structure and temporal progression (Panel A). Panel A (a) examines the clinical intervention structure in the latent belief state using Principal Component Analysis (PCA). The representations show clear separation by clinical intervention type, particularly between TI and DI interventions. This separation suggests that the belief state distinguishes treatment-related interventions from diagnostic investigations. Panel A(b) examines the temporal structure by grouping representations according to normalized episode progress. The projected representations show differentiation from top to bottom at 0-100% normalized episode progress, suggesting a structure tied to temporal position in the latent space.

Second, we examine state dynamics and differences between OICycle and OICycle-S by looking at history dependence and consecutive-state similarity. Panel B(a) examines the relationship between history accumulation and the difference in belief state between OICycle and OICycle-S. As OICycle accumulates more history, the difference between the two models increases, suggesting that OICycle is not just learning a different representation but that this consistently drives differentiation for patients. Panel B(b) examines the cosine similarity between belief states at consecutive episodes for both models. OICycle belief states show a high degree of similarity with previous belief states, reflecting temporal dependency rather than random belief states at each episode. Unlike OICycle, OICycle-S has little similarity, which is trivial, as it does not accumulate history.

Lastly, we examine the relationship between episode progression and performance in Panel C. Panel C(a) & (b) plot the normalized episode progression against task metrics and intervention metrics. Task 1 metrics and the intervention subtypes are minimally affected by episode progression. On the other hand, task 2 & task 3 are clearly affected, which can be seen after 60-70% episode progression, where OICycle and OICycle-S have different slopes for degraded performance compared to early episodes, where their slopes are similar. This suggests that as episodes progress, task2 & task3 will have a major performance impact, as also seen in the quantitative evaluation. Note that for Task 2 & 3, macro-F1 is calculated at the episode level, not the dataset level.

In summary, OICycle belief states are not optimized random states but have meaningful structure, relationships with episode progression, and impact due to episode progression. However, several limitations and caveats remain. Because OICycle is trained on observational EHR data, the learned latent dynamics should not be interpreted as uniquely identifying the underlying physiological disease dynamics. Rather, the learned state transition may capture a mixture of patient-related dynamics and clinician behavioral or intervention-application patterns. In addition, the dataset is still smaller than the multimodal scenarios clinicians use for decision-making.

7 Conclusion & Future Work

In summary, OICycle offers a POMDP-inspired framework for sequential clinical decision-making, updating belief states as interventions and observations unfold. It outperforms OICycle-S, especially for longitudinal outcomes like near-term mortality, and its open dataset establishes a benchmark for future research. Future work includes expanding multimodal data sources (e.g., electronic medication, ECG), clinically validating top predictions, enhancing explainability, and developing tools for clinician training and scenario analysis based on OICycle.

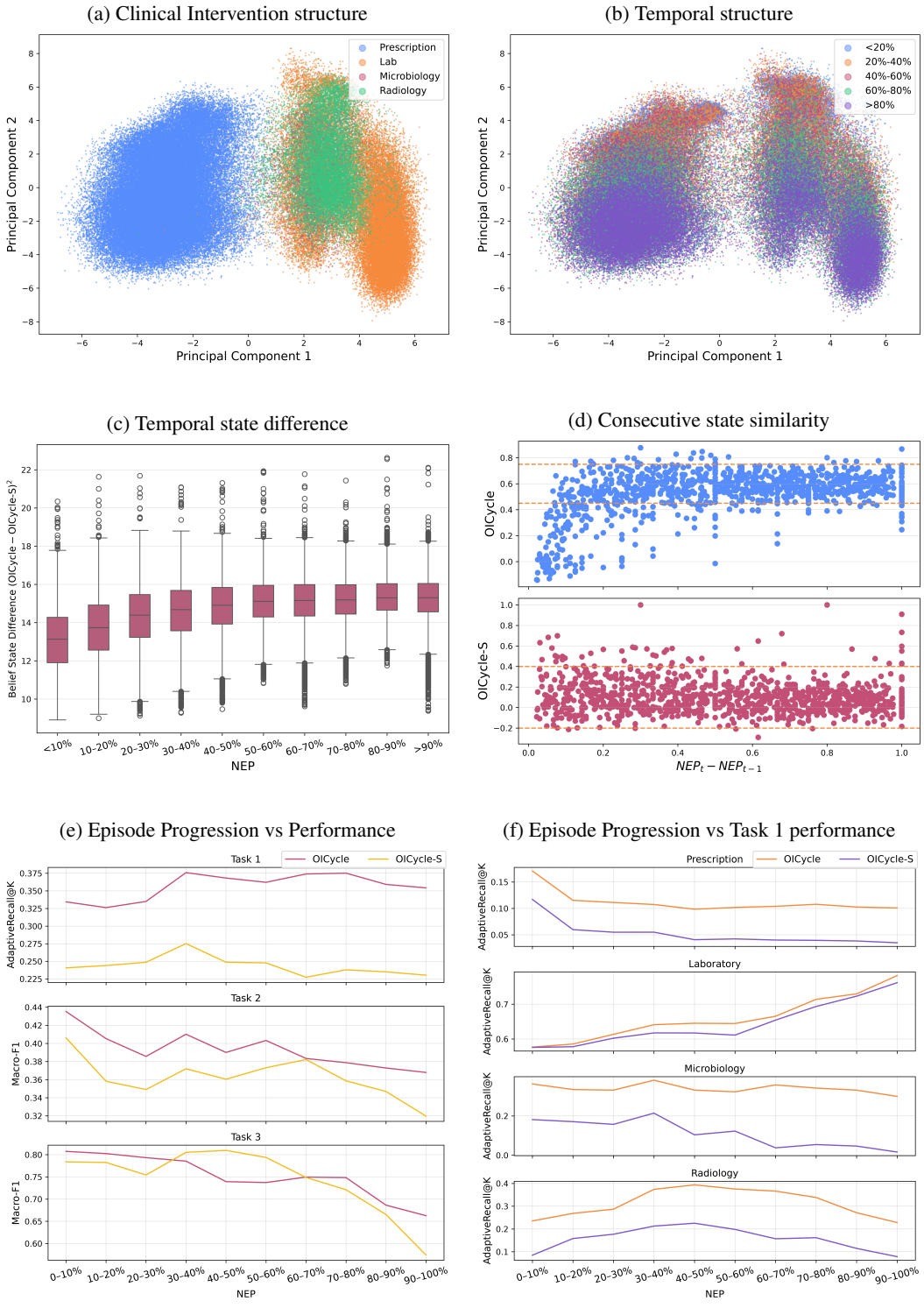


Figure 4: Figure (a)(b) examines the clinical intervention and temporal structure of the belief state, Fig. (c)(d) examines the evolving nature of the belief state, and Fig. (e)(f) examines the predictive utility of history accumulation in the belief state. Here, NEP means Normalized Episode Progression

AI USE STATEMENT

We use ChatGPT (LLM) and Grammarly LLM for correcting grammar, using right tone and check any word that does not fit technically. In addition, LLM was also used for reduce the literature search and checks for potential plagiarism in conjunction with fine-read of those papers. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

Due to the restrictions on dataset release by PhysioNet, we cannot publicly release our train/val/tests set. However, we have provided the complete data preprocessing pipeline with instructions. Researchers can register in Physionet *https://physionet.org/* to get the raw data and use the data preprocessing pipeline to get the processed final data.

REPRODUCIBILITY STATEMENT

To enable reproducibility, we have released the code at *https://github.com/JitindraFartiyal/OICycle/tree/main*. This code contains both the data preprocessing code and the main framework code as well. The instructions are further provided in the README section and Issues can be raise for any doubts.

ACKNOWLEDGMENTS

This research was funded by the European Union's Horizon Europe project BETTER (grant agreement No 101136262). SKT acknowledges support of the UK Multidisciplinary Centre for Neuromorphic Computing (UKRI982).

REFERENCES

Julián N. Acosta, Guido J. Falcone, Pranav Rajpurkar, and Eric J. Topol. Multimodal biomedical AI. 28(9):1773–1784. ISSN 1546-170X. doi: 10.1038/s41591-022-01981-2. URL https://www.nature.com/articles/s41591-022-01981-2. Publisher: Nature Publishing Group.

Corban Allenbrand. DECIDE-lab: A value-of-information and POMDP framework for diagnostic laboratory test selection. 16(15):2356. ISSN 2075-4418. doi: 10.3390/diagnostics16152356. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC13464772/.

Xiaoyu Chen, Yao Mark Mu, Ping Luo, Shengbo Li, and Jianyu Chen. Flow-based recurrent belief state learning for POMDPs. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 3444–3468. PMLR. URL https://proceedings.mlr.press/v162/chen22q.html. ISSN: 2640-3498.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. GRAM: Graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pp. 787–795. Association for Computing Machinery. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098126. URL https://dl.acm.org/doi/10.1145/3097983.3098126.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran,

and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

Joseph Futoma, Michael Hughes, and Finale Doshi-Velez. POPCORN: Partially observed prediction constrained reinforcement learning. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pp. 3578–3588. PMLR. URL https://proceedings.mlr.press/v108/futoma20a.html. ISSN: 2640-3498.

Milos Hauskrecht and Hamish Fraser. Planning treatment of ischemic heart disease with partially observable markov decision processes. 18(3):221–244. ISSN 0933-3657. doi: 10.1016/S0933-3657(99)00042-1. URL https://www.sciencedirect.com/science/article/pii/S0933365799000421.

Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. ClinicalBERT: Modeling clinical notes and predicting hospital readmission. URL http://arxiv.org/abs/1904.05342.

Pushkala Jayaraman, Jacob Desman, Moein Sabounchi, Girish N. Nadkarni, and Ankit Sakhuja. A primer on reinforcement learning in medicine for clinicians. 7:337. ISSN 2398-6352. doi: 10.1038/s41746-024-01316-0. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11599275/.

Alistair Johnson, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV-note: Deidentified free-text clinical notes, a. URL https://physionet.org/content/mimic-iv-note/2.2/.

Alistair Johnson, Tom Pollard, Roger Mark, Seth Berkowitz, and Steven Horng. MIMIC-CXR database, b. URL https://physionet.org/content/mimic-cxr/2.1.0/.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo A. Celi, and Roger G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. 10(1):1, c. ISSN 2052-4463. doi: 10.1038/s41597-022-01899-x. URL https://www.nature.com/articles/s41597-022-01899-x. Publisher: Nature Publishing Group.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. 101(1):99–134. ISSN 0004-3702. doi: 10.1016/S0004-3702(98)00023-X. URL https://www.sciencedirect.com/science/article/pii/S000437029800023X.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. URL http://arxiv.org/abs/1312.6114.

Rahul G. Krishnan, Uri Shalit, and David Sontag. Deep kalman filters. URL http://arxiv.org/abs/1511.05121.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. 36(4):1234–1240. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz682. URL https://doi.org/10.1093/bioinformatics/btz682.

Jiawen Li, Xuchao Fang, Ping Yu, and Lanfang Jiang. Beyond static risk scores: dynamic world models simulating patient-specific trajectories to inform preoperative risk mitigation strategies. 13, a. ISSN 2296-858X. doi: 10.3389/fmed.2026.1913206. URL https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2026.1913206/full. Publisher: Frontiers.

Luchen Li, Matthieu Komorowski, and Aldo A. Faisal. Optimizing sequential medical treatments with auto-encoding heuristic search in POMDPs, b. URL http://arxiv.org/abs/1905.07465.

Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. BEHRT: Transformer for electronic health records. 10:7155, c. ISSN 2045-2322. doi: 10.1038/s41598-020-62922-y. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7189231/.

Yue Ling. Bio+clinical BERT, BERT base, and CNN performance comparison for predicting drug-review satisfaction. URL http://arxiv.org/abs/2308.03782.

Ali Najafi, Amirfarhad Farhadi, and Azadeh Zamanifar. Advancements in reinforcement learning for clinical decision-making in healthcare: a systematic review. 4(1):13. ISSN 2731-6955. doi: 10.1007/s44248-026-00112-3. URL https://doi.org/10.1007/s44248-026-00112-3.

Yifan Peng, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. URL http://arxiv.org/abs/1906.05474.

Trang Pham, Truyen Tran, Dinh Phung, and Svetha Venkatesh. Predicting healthcare trajectories from medical records: A deep learning approach. 69:218–229. ISSN 1532-0464. doi: 10.1016/j.jbi.2017.04.001. URL https://www.sciencedirect.com/science/article/pii/S1532046417300710.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. 4 (1):86. ISSN 2398-6352. doi: 10.1038/s41746-021-00455-y. URL https://www.nature.com/articles/s41746-021-00455-y. Publisher: Nature Publishing Group.

Athanasios Tsoukalas, Timothy Albertson, and Ilias Tagkopoulos. From data to optimal decision making: A data-driven, probabilistic machine learning approach to decision support for patients with sepsis. 3(1):e11. ISSN 2291-9694. doi: 10.2196/medinform.3445. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4376114/.

Shirly Wang, Matthew B. A. McDermott, Geeticka Chauhan, Marzyeh Ghassemi, Michael C. Hughes, and Tristan Naumann. MIMIC-extract: a data extraction, preprocessing, and representation pipeline for MIMIC-III. In *Proceedings of the ACM Conference on Health, Inference, and Learning*, pp. 222–235. ACM. ISBN 978-1-4503-7046-2. doi: 10.1145/3368555.3384469. URL https://dl.acm.org/doi/10.1145/3368555.3384469.

Michael Wornow, Rahul Thapa, Ethan Steinberg, Jason A Fries, and Nigam H Shah. EHRSHOT: An EHR benchmark for few-shot evaluation of foundation models.

Chaoqi Yang, Zhenbang Wu, Patrick Jiang, Zhen Lin, Junyi Gao, Benjamin P. Danek, and Jimeng Sun. PyHealth: A deep learning toolkit for healthcare applications. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5788–5789. ACM. ISBN 979-8-4007-0103-0. doi: 10.1145/3580305.3599178. URL https://dl.acm.org/doi/10.1145/3580305.3599178.

Chao Yu, Jiming Liu, Shamim Nemati, and Guosheng Yin. Reinforcement learning in healthcare: A survey. 55(1):5:1–5:36. ISSN 0360-0300. doi: 10.1145/3477600. URL https://dl.acm.org/doi/10.1145/3477600.

Hugo Yèche, Rita Kuznetsova, Marc Zimmermann, Matthias Hüser, Xinrui Lyu, Martin Faltys, and Gunnar Rätsch. HiRID-ICU-benchmark — a comprehensive machine learning benchmark on high-resolution ICU data.

A Appendix

You may include other additional sections here.

Table 2: Hyper parameters table

Learning Rate	1e-3
Batch Size	1
Hidden dimension	512
Intervention Embedding dimension	128
Embedding dimension	512
Threshold	0.5
Accumulation steps	128
loss weights	1. for all

A.1 HYPER PARAMETERS

Hidden dimension is the dimension of the neural network used to project embeddings into a common embedding. The intervention embedding dimension embeds lab results, microbiology test IDs, radiology exam codes, and prescription drugs. The embedding dimension is the final common embedding used to train the state model. The threshold is used for predictions. Accumulation steps: how many patient data points we process before updating the weights.

A.2 DATASET SUMMARY & PLOTS

Table 3: Summary statistics of the training, validation, and test sets.

Summary		Train	Val	Test
Unique DRG		815	733	756
Unique Subjects		52088	9193	10794
Gender	Female	26115	4667	5347
	Male	25973	4526	5447
Age Distribution (mean/std)		59.42/19.46	60.21/19.19	59.81/19.27
Clinical Intervention	Laboratory Tests	2016432	346875	423007
	Microbiology Tests	99909	15860	19067
	Radiology Exams	153258	26452	31573
	Clinical Notes	52088	9193	10794
	Prescriptions	1636653	287133	347080

Intervention	Unique class
Prescriptions	4197
Lab Events	204
Microbiology Events	106
Radiology Exam	1989

Table 4: Unique classes of different intervention types.

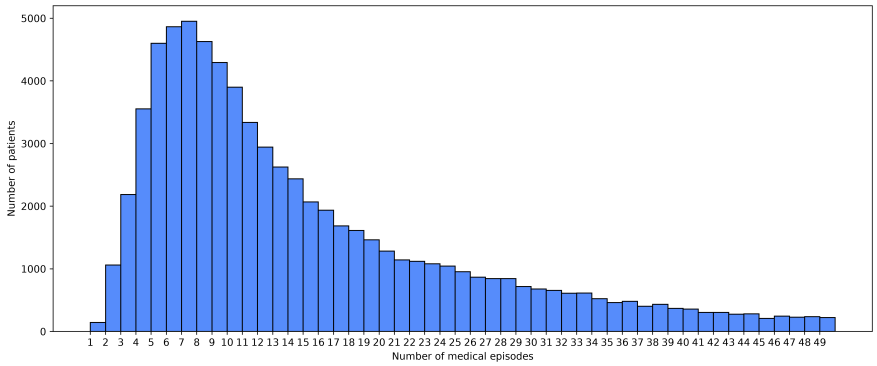


Figure 5: Distribution of Medical Episodes per Patient

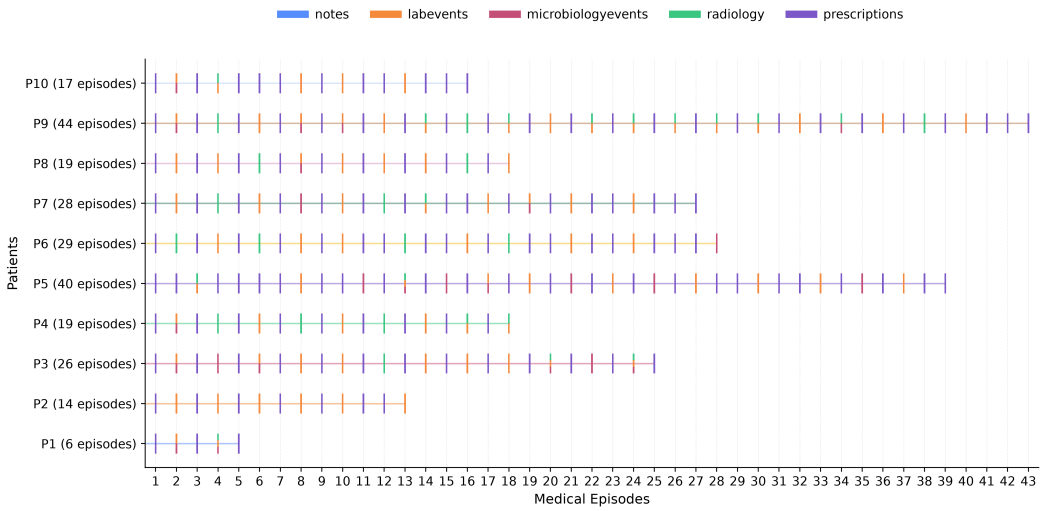


Figure 6: Sample patients raster plot for presenting heterogeneous episodic nature