STATUS-AWARE SELF-SUPERVISED FORECASTING FOR IRREGULAR CLINICAL TIME SERIES

Kwanhyung Lee[1,2*] **Joohyung Lee**[1*] **Jongheon Kim**[2] **Hanhsang Chul**[1] **Eunho Yang**[1,2]
[1]AITRICS
[2]Korea Advanced Institute of Science and Technology (KAIST)

ABSTRACT

Electronic health record (EHR) time series are irregular event streams, and many clinical tasks require forecasting a patient's future state from their history. Because labeled outcomes in EHR are often limited and class-imbalanced, self-supervised learning (SSL) on large unlabeled cohorts is attractive. However, most existing EHR SSL methods first convert irregular event streams into discretized time grids, thereby losing the native event-set structure of observations. We propose a downstream task-aligned pretraining framework that models EHR trajectories as event sets and learns representations by forecasting future clinical states in latent space from the full past context. Our framework combines a momentum teacher for stable targets, a query-based Transformer decoder to predict variable future event sets, and an auxiliary masked event objective to improve local robustness. Across multiple ICU prediction tasks, our approach consistently improves downstream performance.

Track: Research

1 INTRODUCTION

Irregular multivariate EHR time series are naturally represented as event streams, where each observation is a triplet of value, timestamp, and variable type (Horn et al., 2020; Tipirneni & Reddy, 2022; Lee et al., 2023a; Xu et al., 2025). Many clinical prediction problems are inherently forward-looking: given a patient's history, the goal is to forecast future clinical status (e.g., CPR, mortality risk) (Harutyunyan et al., 2019; Lee et al., 2023b).

Because labeled outcomes are often limited and highly imbalanced, self-supervised learning (SSL) has emerged as a promising approach to leverage large collections of unlabeled trajectories. To be effective in this setting, an SSL objective should operate directly on irregular event streams and align its supervision with the forecasting nature of downstream clinical tasks.

However, many existing SSL approaches discretize irregular event streams onto fixed time grids (often with imputation), introducing preprocessing bias. On such representations, objectives typically learn overly global representations that blur evolving patient states (Lee et al., 2023b; Labach et al., 2023; Chowdhury et al., 2023; Yu et al., 2024). Moreover, existing forecasting pretext tasks in EHR are defined on such grids, predicting discretized future vectors rather than modeling native future event structure (Tipirneni & Reddy, 2022; Lee et al., 2023b).

Building on event-level representations, we propose a forecasting-aligned self-supervised objective that predicts future patient status directly from past observations (see Figure 1).

EHR Set Predictive Coding (EHR-SPC). We propose **EHR-SPC**, a momentum-target pretraining framework for irregular event sets that aggregates short event windows into *status* representations and predicts future statuses from past context in latent space. An online **Status Encoder** learns to match targets from an EMA **Momentum Status Encoder** (He et al., 2020).

Query Predictor. Future event streams vary in size and composition, making token-wise forecasting with fixed alignment inadequate. We employ a query-based Transformer **Query Predictor** with K

*Equal contribution.

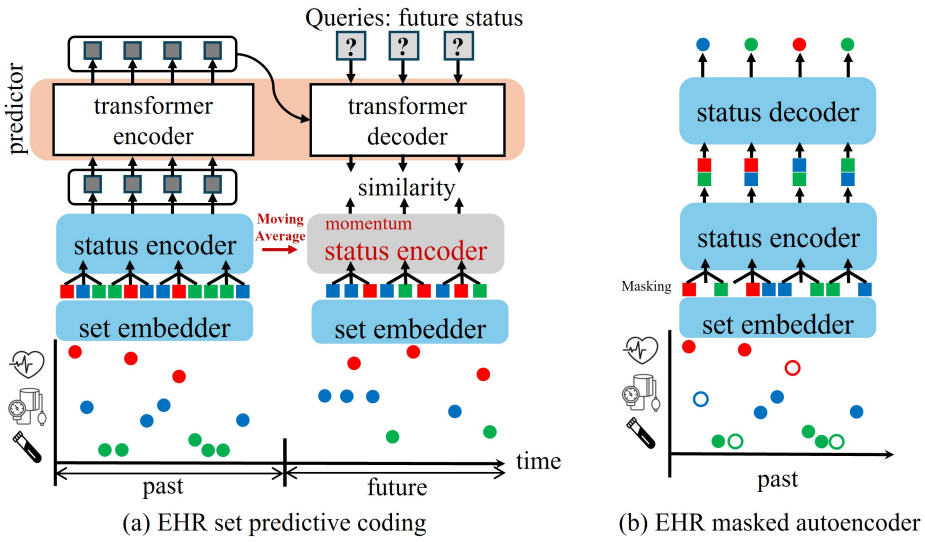

Figure 1: **Overview of our EHR pretraining framework.** (a) **EHR Set Predictive Coding:** Past triplets (v, t, f) are aggregated into status tokens and encoded into context. A query-based **Query Predictor** generates future status tokens, supervised by an EMA **Momentum Status Encoder**. (b) **EHR Masked Autoencoder:** Events in past segments are masked and reconstructed to improve local robustness.

learnable queries to generate future status representations. Inspired by DETR (Carion et al., 2020), which handles an unknown number of objects via query-based set prediction, our design accommodates variable numbers of future clinical statuses without requiring explicit token alignment.

Auxiliary local robustness. We additionally include an **EHR-MAE** objective that masks events within past segments and reconstructs their values, improving local robustness.

Our contributions are:

1. **Status-forecast SSL for irregular event sets.** We introduce **EHR-SPC**, a self-supervised objective that forecasts future *status* representations in latent space from irregular event sets without discretization, aligning the pretraining horizon with downstream tasks.
2. **Query-based set prediction.** We use a DETR-style (Carion et al., 2020) learnable-query Transformer for set forecasting, handling variable future event structure and cardinality.
3. **Consistent empirical gains.** Across multiple ICU prediction tasks, **EHR-SPC** demonstrates consistent performance improvements under fine-tuning.

2 Method

2.1 Problem Setup

A patient trajectory is a time-ordered event sequence $\mathcal{X} = \{(v_i, t_i, f_i)\}_{i=1}^{N}$, where i indexes observed events,$v_i \in \mathbb{R}$ is the value, $t_i \in \mathbb{R}^+$ the timestamp, and $f_i \in \{1, \ldots, F\}$ the variable type. We pretrain directly on observed triplets to learn representations for downstream forecasting.

2.2 Status-Level Hierarchical Encoding

We partition $\mathcal{X}$ into contiguous *status segments* of fixed size P: $\{\mathbf{S}_j\}_{j=1}^{M}$, where $\mathbf{S}_j = \{(v_{ij}, t_{ij}, f_{ij})\}_{i=1}^{P}$ is sorted by time. For each segment, we compute the mean timestamp $\bar{t}_j$ and define segment-wise relative time $\tilde{t}_{ij} = t_{ij} - \bar{t}_j$. Each event set is embedded as

$$\mathbf{e}_{ij} = \mathrm{MLP}_v(v_{ij}) + \mathrm{MLP}_t(\tilde{t}_{ij}) + \mathbf{E}_f[f_{ij}]. \tag{1}$$

A transformer **Status Encoder** aggregates events per segment via a learnable [STATUS] token:

$$\mathbf{h}_j = \mathrm{StatusEnc}\big([\mathbf{e}_{\text{status}}; \mathbf{e}_{j1}, \ldots, \mathbf{e}_{jP}]\big)_{\text{status}}. \tag{2}$$

A **Transformer Encoder** models long-range dynamics over status tokens, adding segment-time embeddings, static demographics, and a [CLS] token used for downstream prediction:

$$[\mathbf{c}_{\text{cls}}; \mathbf{c}_{\text{demo}}; \mathbf{c}_{1:M}] = \mathrm{TransformerEnc}\big([\mathbf{z}^0_{\text{cls}}; \mathbf{z}_{\text{demo}}; \{\mathbf{h}_j + \mathrm{MLP}_{\text{t}}(\bar{t}_j)\}_{j=1}^{M}]\big). \tag{3}$$

2.3 EHR-SPC: Future Status Prediction with Momentum Targets

We split segments into past and future: $1{:}M_{\text{past}}$ and $M_{\text{past}}{+}1{:}M$. Past segments are encoded by the online encoders to form $\mathbf{C}_{\text{past}} = [\mathbf{c}_{\text{cls}}; \mathbf{c}_{\text{demo}}; \mathbf{c}_1, \ldots, \mathbf{c}_{M_{\text{past}}}]$. Future status targets are produced by an EMA teacher (**Momentum Status Encoder**) updated as

$$\theta_{\text{mom}} \leftarrow m\,\theta_{\text{mom}} + (1-m)\,\theta_{\text{online}}, \quad m = 0.99, \tag{4}$$

following momentum-target learning (He et al., 2020).

Because future event streams have variable size and composition, we predict future *status* tokens using a query-based **Query Predictor** inspired by DETR (Carion et al., 2020). The Query Predictor is implemented as a transformer decoder that attends to the past context $\mathbf{C}_{\text{past}}$ via cross-attention.

Let $\mathbf{Q} = [\mathbf{q}_j]_{j=M_{\text{past}}+1}^{M}$ be learnable query tokens corresponding to future status segments. Each query is augmented with the mean timestamp $\bar{t}_j$ of its associated future segment:

$$\tilde{\mathbf{H}}_{\text{future}} = \mathrm{QueryPred}\big([\mathbf{q}_j{+}\mathrm{MLP}_{\text{t}}(\bar{t}_j)]_{j=M_{\text{past}}+1}^{M_{\text{past}}+M_{\text{future}}},\ \mathbf{C}_{\text{past}}\big),\ \hat{\mathbf{H}}_{\text{future}} = \mathrm{MLP}_{\text{pred}}(\tilde{\mathbf{H}}_{\text{future}}) = [\hat{\mathbf{h}}_j]_{j=1}^{M_{\text{future}}}. \tag{5}$$

We minimize cosine distance to stop-gradient teacher targets over valid future status segments. Let $\mathcal{V}_{\text{future}}$ denote the set of available future segments. The loss is defined as

$$\mathcal{L}_{\text{future}} = \frac{1}{|\mathcal{V}_{\text{future}}|} \sum_{j \in \mathcal{V}_{\text{future}}} \Big(1 - \cos\big(\hat{\mathbf{h}}_j,\ \mathrm{sg}(\mathbf{h}_j^{\text{mom}})\big)\Big), \tag{6}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operator.

2.4 EHR-MAE and Final Objective

As auxiliary supervision, we mask P_{mask} events within each past segment and reconstruct masked values using a lightweight transformer-based reconstruction head, yielding $\mathcal{L}_{\text{MAE}}$ (MSE on masked values). For masked events, timestamp and variable-type embeddings are preserved, and only the value embedding is replaced with a learned mask token; the model predicts the original values conditioned on time and type. The final objective is

$$\mathcal{L}_{\text{SSL}} = \mathcal{L}_{\text{future}} + \lambda_{\text{MAE}}\mathcal{L}_{\text{MAE}}. \tag{7}$$

Fine-tuning Interface and Future Directions. For downstream tasks, we fine-tune a task head on the global [CLS] representation $\mathbf{c}_{\text{cls}}$ from the **Context Encoder**.

3 Experiments

3.1 Dataset

We evaluate hourly online monitoring tasks on MIMIC-IV (CPR, Intubation, Mortality), predicting whether an event will occur within the next 12 hours (Johnson et al., 2023). CPR is highly class-imbalanced (Appendix Table 3), motivating the use of AUPRC alongside AUROC. Data preprocessing follows the protocol of Lee et al. (2023a). Detailed cohort statistics and implementation details are provided in Appendix A. For each method, we report AUROC and AUPRC on test sets. Models are selected based on the best validation AUROC across training epochs. For SSL pretraining, we select the checkpoint with the lowest SSL objective value, as no labels are used during pretraining. We compare against representative EHR SSL baselines, including STraTS, SMART, and DueTT (Tipirneni & Reddy, 2022; Yu et al., 2024; Labach et al., 2023).

Table 1: **Main performance comparison. No Pretrain** indicates training from random initialization. **SSL-pretrained** indicates initialization with weights obtained via the original SSL objective proposed by each method.

Model	CPR		Intubation		Mortality	
	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
No pretrain						
STraTS	0.5397	0.0018	0.8558	0.2734	0.8778	0.1457
SMART	0.5098	0.0005	0.8464	0.2357	0.8457	0.0941
DueTT	0.6478	0.0010	0.8642	0.2899	0.8967	0.1527
Ours	0.5330	**0.0030**	0.8493	0.2636	0.8836	0.1538
SSL-pretrained						
STraTS	0.6047	0.0019	0.8631	0.2832	0.9003	0.1783
SMART	0.6534	0.0021	0.8568	0.2753	0.8620	0.1369
DueTT	0.7228	0.0012	0.8706	0.2959	**0.9229**	**0.1784**
Ours	**0.7717**	**0.0030**	**0.8725**	**0.3079**	0.9118	0.1680

Table 2: **Ablation on SSL objectives. Random Init**: no pretraining. **Event-MAE Only**: masked event reconstruction objective. **Future Only**: future status prediction objective. **EHR-SPC (Full)**: combination of future prediction and Event-MAE.

Pretrain	Variant	CPR		Intubation		Mortality	
		AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
No	Random Init	0.5330	0.0030	0.8493	0.2636	0.8836	0.1538
Yes	Event-MAE Only	0.7275	0.0023	0.8723	0.3070	0.8952	0.1396
Yes	Future Only	0.7477	**0.0036**	**0.8731**	**0.3195**	0.9107	0.1594
Yes	EHR-SPC (Full)	**0.7717**	0.0030	0.8725	0.3079	**0.9118**	**0.1680**

3.2 Main Results and Ablation

Table 1 reports downstream performance. EHR-SPC achieves the strongest gains on CPR and Intubation and remains competitive on Mortality. The largest improvements appear on CPR, the most class-imbalanced task. Across models, SSL pretraining consistently outperforms random initialization, confirming the benefit of representation learning.

We further ablate the pretraining objectives: (i) random initialization, (ii) Event-MAE only, (iii) future status prediction only, and (iv) the full model. Table 2 shows that combining both objectives yields the most consistent gains.

4 Discussion

The improvements suggest that forecasting-aligned pretraining yields representations better suited for clinical prediction. By modeling observed triplets and aggregating them into status-level abstractions, EHR-SPC preserves irregular event structure while capturing short-term dynamics. Momentum-based future prediction and Event-MAE jointly provide stable and consistent gains across forecasting tasks.

5 Conclusion

We presented **EHR-SPC**, an event-aligned self-supervised framework for irregular EHR time series that avoids discretization and imputation. By hierarchically modeling value-time-type triplets and combining momentum-based future status prediction with an auxiliary Event-MAE objective, EHR-SPC learns representations tailored to clinical forecasting. Across multiple ICU tasks, it consistently improves downstream performance, supporting the value of future-status-oriented supervision for event-based clinical data. The status-level abstraction may also provide a structured interface for aligning EHR representations with large language models in future multimodal clinical systems.

REFERENCES

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pp. 213–229. Springer, 2020.

Ranak Roy Chowdhury, Jiacheng Li, Xiyuan Zhang, Dezhi Hong, Rajesh K Gupta, and Jingbo Shang. Primenet: Pre-training for irregular multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7184–7192, 2023.

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. URL https://www.nature.com/articles/s41597-019-0103-9.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. Set functions for time series. In *Proceedings of ICML*, pp. 4353–4363. PMLR, 2020. URL https://proceedings.mlr.press/v119/horn20a.html.

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, 2023.

Alex Labach, Aslesha Pokhrel, Xiao Shi Huang, Saba Zuberi, Seung Eun Yi, Maksims Volkovs, Tomi Poutanen, and Rahul G Krishnan. Duett: dual event time transformer for electronic health records. In *Machine Learning for Healthcare Conference*, pp. 403–422. PMLR, 2023.

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 *Proceedings of MLHC*, pp. 423–442. PMLR, 2023a. URL https://proceedings.mlr.press/v219/lee23a.html.

Kwanhyung Lee, John Won, Heejung Hyun, Sangchul Hahn, Edward Choi, and Joohyung Lee. Self-supervised predictive coding with multimodal fusion for patient deterioration prediction in fine-grained time resolution. In *Proceedings of ICLR 2023 Workshop on Trustworthy Machine Learning for Healthcare*, 2023b. URL https://openreview.net/forum?id=3aqPxh5YjP.

Sindhu Tipirneni and Chandan K Reddy. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 16(6):1–17, 2022. URL https://dl.acm.org/doi/10.1145/3516367.

Justin Xu et al. Aces: Automatic cohort extraction system for event-stream datasets. In *Proceedings of ICLR*, 2025. URL https://openreview.net/forum?id=P4XmKjXTrM.

Zhihao Yu, Chu Xu, Yujie Jin, Yasha Wang, and Junfeng Zhao. Smart: Towards pre-trained missing-aware model for patient health status prediction. *Advances in Neural Information Processing Systems*, 37:63986–64009, 2024.

A Dataset and Implementation Details

A.1 Dataset and Cohort Construction

We use MIMIC-IV ICU stays and follow the preprocessing protocol of Lee et al. (2023a). We evaluate three prediction tasks: CPR, Intubation, and Mortality. Unless otherwise stated, we split patients into train/validation/test sets with no patient overlap. Models are selected by the best validation AUROC and evaluated once on the held-out test set.

Task definitions. We formulate each task as an hourly online binary classification problem, where risk is re-estimated at every hour using the preceding observation window. For each task, we define an observation window and a prediction horizon following Lee et al. (2023a). Each prediction time requires at least 40 past observations within a 48-hour window and 20 future observations within a 12-hour window. The model outputs a binary risk prediction indicating whether the target clinical event will occur within the specified prediction horizon.

- **CPR:** observation window = 48 hours (min 40 events), prediction horizon = 12 hours.
- **Intubation:** observation window = 48 hours (min 40 events), prediction horizon = 12 hours.
- **Mortality:** observation window = 48 hours (min 40 events), prediction horizon = 12 hours.

Cohort statistics. Table 3 reports cohort sizes and class balance for each task.

Architecture. Our model consists of an event embedder, a local patch encoder (Status Encoder), a global sequence encoder (Context Encoder), a cross-attention transformer decoder (Query Predictor), and a lightweight reconstruction head for EHR-MAE. Table 4 summarizes the key architecture hyperparameters.

Pretraining. We pretrain with the combined objective $\mathcal{L}_{\text{SSL}} = \mathcal{L}_{\text{future}} + \lambda_{\text{MAE}}\mathcal{L}_{\text{MAE}}$. Table 5 lists training hyperparameters for pretraining and fine-tuning.

For all methods (both SSL pretraining and supervised fine-tuning), we perform a learning rate sweep over 10^{-3} to 10^{-6} and report the best-performing configuration.

B Limitations and Future Directions

Limitations.

- **Fixed-size status segmentation.** We use fixed event-count segments (P events), which may under-utilize extremely dense periods or oversimplify very sparse periods.
- **Query budget.** We use a fixed query budget (K), and unused queries do not contribute to the loss when fewer future segments are available.
- **Evaluation scope.** Our experiments focus on ICU forecasting tasks and a single preprocessing pipeline; broader settings (e.g., other cohorts or hospitals) remain to be explored.

Future directions.

- **Adaptive segmentation.** Replace fixed event-count segments with adaptive or learned segmentation while preserving event-aligned supervision.
- **Label efficiency analysis.** Systematically evaluate performance under progressively reduced labeled fine-tuning data to quantify transfer robustness in low-label settings.
- **Multimodal and LLM alignment.** Extend status representations to broader event modalities (e.g., medications, notes) and explore alignment with clinical language models, leveraging the structured status abstraction as an interface.

Table 3: **Cohort statistics (MIMIC-IV).**

Task	Split	# Patients / Stays	# Pos / # Neg
CPR	Train	29,544 / 37,156	561 / 1,578,395
	Val	6,341 / 8,160	54 / 31,345
	Test	6,292 / 7,934	151 / 326,508
Intubation	Train	29,544 / 37,156	70,804 / 1,508,152
	Val	6,341 / 8,160	6,864 / 29,017
	Test	6,292 / 7,934	14,960 / 311,699
Mortality	Train	29,544 / 37,156	7,705 / 1,571,251
	Val	6,341 / 8,160	751 / 31,304
	Test	6,292 / 7,934	1,706 / 324,953

Table 4: **Architecture hyperparameters.**

Component	Setting
Event embedding dim d	128
Value MLP MLP_v (layers / hidden)	2 / 128 (Tanh)
Time MLP MLP_t (layers / hidden)	2 / 128 (Tanh)
Type embedding $\mathbf{E}_f$	F=18, dim = 128
Status segment size P	16
MAE enlarged segment size P'	20
MAE mask count P_{mask}	4
Status Encoder (layers / heads / dropout)	3 / 4 / 0.1
Context Encoder (layers / heads / dropout)	3 / 4 / 0.1
Query Predictor (decoder layers / heads)	3 / 4
Query token	learnable (dim=128)
Future prediction head (layers)	2-layer MLP (128→128, GELU)
MAE reconstruction head (layers / heads / dim)	3 / 4 / 64
MAE prediction head	Linear (64→1)

Table 5: **Training hyperparameters.**

Hyperparameter	Pretraining	Fine-tuning
Optimizer	AdamW	AdamW
Learning rate	3e-4	task-dependent
Weight decay	1e-6	1e-6
Batch size	512	512
Epochs	20	30
LR schedule / warmup	OneCycleLR (pct_start=0.2)	OneCycleLR (pct_start=0.2)
EMA momentum m	0.99 (cosine schedule)	N/A
MAE mask count P_{mask}	4	N/A
Loss weight λ_{MAE}	3.0	N/A
Mixed precision	FP16 (AMP)	FP16 (AMP)