Time-Conditioned Foreseeing: An EHR-Specific Foundation Model for Irregular Dynamics and Calendrical Time

Bong Gyun Kang[1], Junyong Ahn[1], Hyeongrok Han[2], Sungroh Yoon[1,2]

[1] Interdisciplinary Program in Artificial Intelligence, Seoul National University [2] Department of Electrical and Computer Engineering, Seoul National University

Paper and Full Code Available

SEOUL NATIONAL UNIVERSITY

ICML International Conference On Machine Learning

TL;DR

- Conventional **Electronic Health Record (EHR)** foundation models often inherit LLM training paradigms, limiting their fit to EHR data structures.
- 1. **Value** representation - **Pathology-Focused Binning**, which allocates resolution to clinically important tail values instead of over-resolving the normal range as in quantile binning.
- 2. **Time** representation - **Dual-Calendar RoPE**, which captures the exact occurrence time of each event and the regularity between events, instead of assuming uniformly spaced events.
- 3. **Learning objective** - **Time-Conditioned Foresee,** a joint "when-and-what" prediction objective that trains to predict all events within a given future horizon.

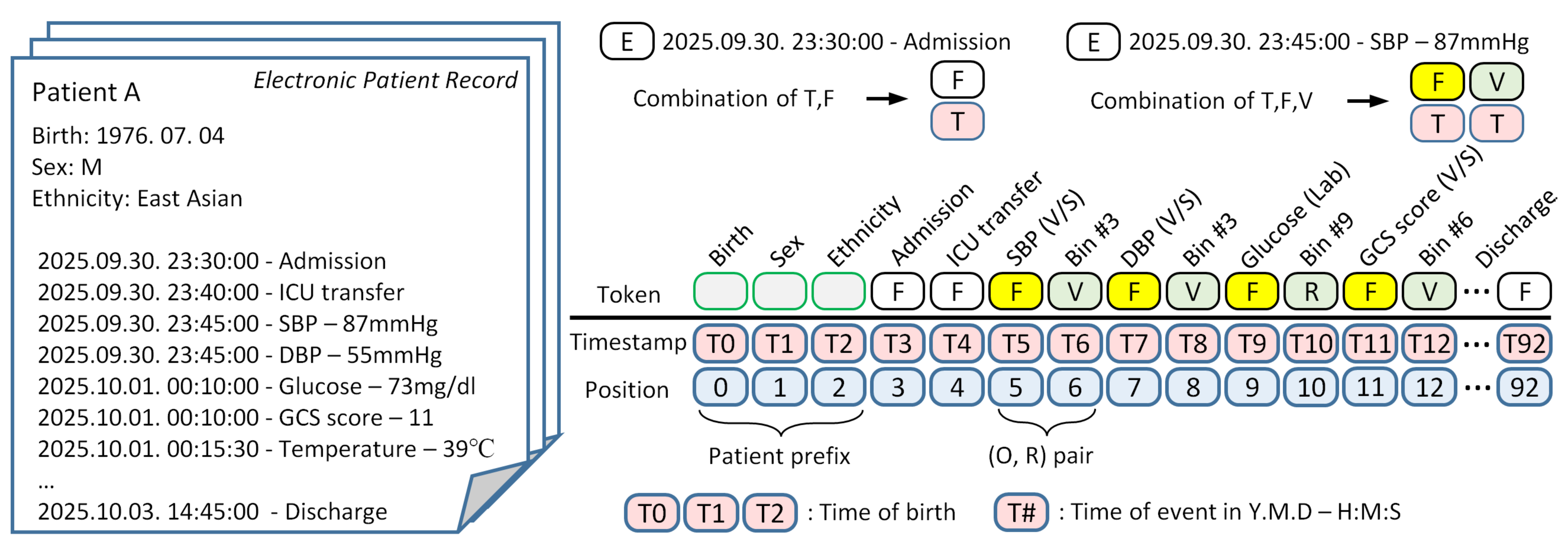


Figure 1. (Left) Raw patient data from the EHR database in chronological order. (Right) Tokenization of each event (E) with triplet representation. Patient information is placed at the beginning, `Features` and `Values` are tokenized, and timestamps remain continuous.

Dataset summary

- Used dataset – MIMIC-3,4 & eICU
- Train – ~160k patients / ~ 270m events
- Eval – ~18k patients / ~34m events
- Train to Eval shift: avg. 41 years

Table 2. Data summary. Parentheses indicate cases where bins are not shared. Refer to Appendix C for more details

Pre-training statistic	MIMIC-III Train	MIMIC-III Test	MIMIC-IV Train	MIMIC-IV Test
Total Patient #	28.7k	5.1k	159.0k	18.3k
Total Hospitalization #	35.7k	6.3k	455.6k	66.3k
Total Events #	38.6m	6.7m	272.6m	34.3m
Total Tokens #	77.1m	13.5m	534.2m	67.2m
Avg. length	2.7k	2.7k	3.4k	3.7k
Max length	393.3k	62.8k	254.1k	335.7k
Unique Tokens #	155 (1,208)		361 (2301)	
Token # bin	10		10	
Token # ethnicity	10		7 (grouped)	
Token # chart events	17		148	
Token # laboratory tests	100		120	
Token # other features	0		9	

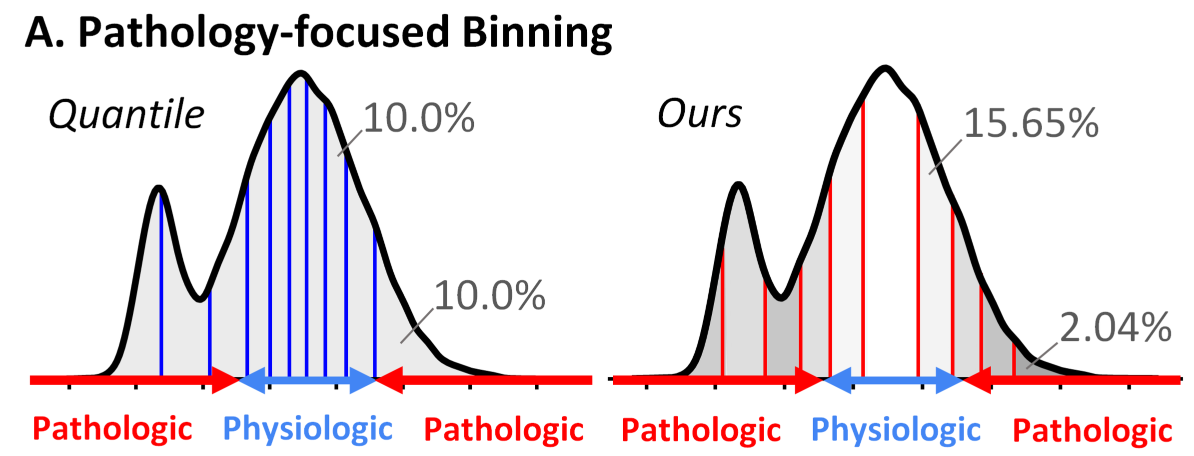
Table 1. Comparison of recent EHR models by architecture and training objective. **[Data usage]**: whether the model uses `Time` and `Value` information. **[Value binning]**: whether values are discretized using non-quantile binning (STraTS embeds values continuously) and whether bin tokens are shared across `Features`. **[Time addressing]**: whether the model considers relative time intervals, calendrical time, and uses an order-preserving tie-breaker for concurrent events. **[Learning objective]**: type of loss, **'Foresee'** - whether model forecasts beyond the next-token, and **'Temporal Generation'** - whether temporal generative modeling is possible. Abbreviations*: MEP; missing entity prediction, TTE; time to event prediction, TCF; time conditioned foreseeing. Refer to the related works section for details†.

Models	[Data usage] Event Timestamp	[Data usage] Numeric Value	[Value Binning] Non-quantile	[Value Binning] Bin sharing	[Time addressing] Relative interval	[Time addressing] Calendrical time	[Time addressing] Order-preserving	[Learning Objective] Type	[Learning Objective] Foresee	[Learning Objective] Temporal Generation
BEHRT (2020) Med-BERT (2021) Foresight (2024) ClinicalMamba (2024) EHR-BERT (2024)	X	X	-	-	X	X	-	NTP&MLM	X	X
HEART (2024) FM4EHR (2025)	X	**O**	X	X O	X	X	-	MEP* NTP	X	X
MOTOR (2024)	**O**	X	-	-	**O**	X	X	TTE*	**O**	X
STraTS (2022)	**O**	**O**	**O**	O	X	X	-	MSE	X	X
CLMBR-T-base (2023)	**O**	**O**	X	X	**O**	X	X	NTP	X	X
TRADE (2024)	**O**	**O**	**O**	X	X	X	-	MLM	X	X
EHRMamba (2025)	**O**	**O**	X	X	X	X	-	NTP&MLM	X	X
ETHOS (2024)	**O**	**O**	X	O	△†	X	-	NTP	X	△†
OURS	**O**	**O**	**O**	both	**O**	**O**	**O**	TCF*	**O**	**O**

1. Value - Pathology-Focused Binning

- *Assumption*: Common high-density range values are less informative, while low-density range values are more clinically important.
- *Method*: Estimate the density with a Gaussian KDE and assign inverse-density weights.

$$\hat{p}(g_k) = \frac{1}{N}\sum_{j=1}^{N} K_h(g_k - v_j),\ \text{where}\ K_h(u) = \exp\left(-\frac{u^2}{2h^2}\right) \qquad w(v) \propto \hat{p}(v)^{-\lambda}$$

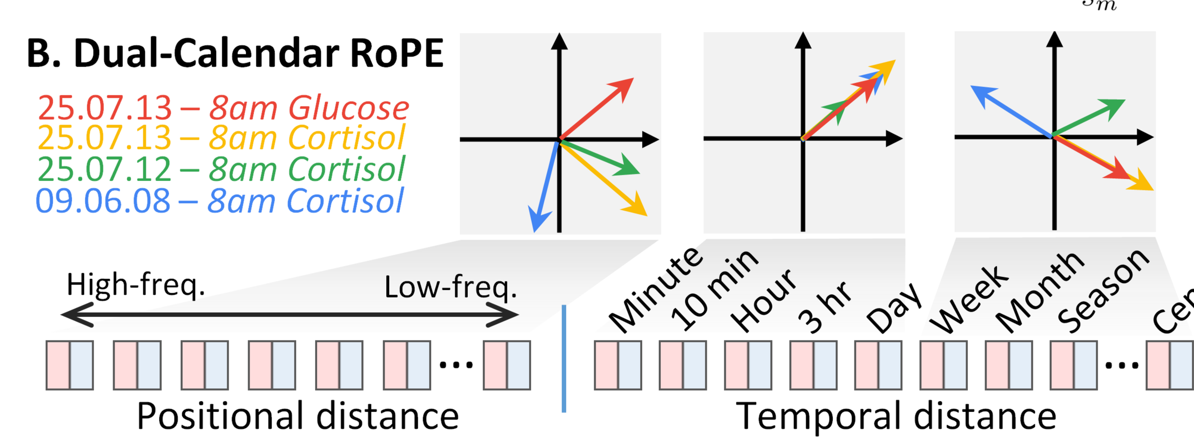


2. Time - Dual-Calendar RoPE

- *Assumption*: Explicitly representing the exact timing and periodicity of events improves the model's temporal representation learning.
- *Method*: Partition the RoPE dimensions into order-based and time-based components, with the time-based periods arranged in ascending calendrical order.

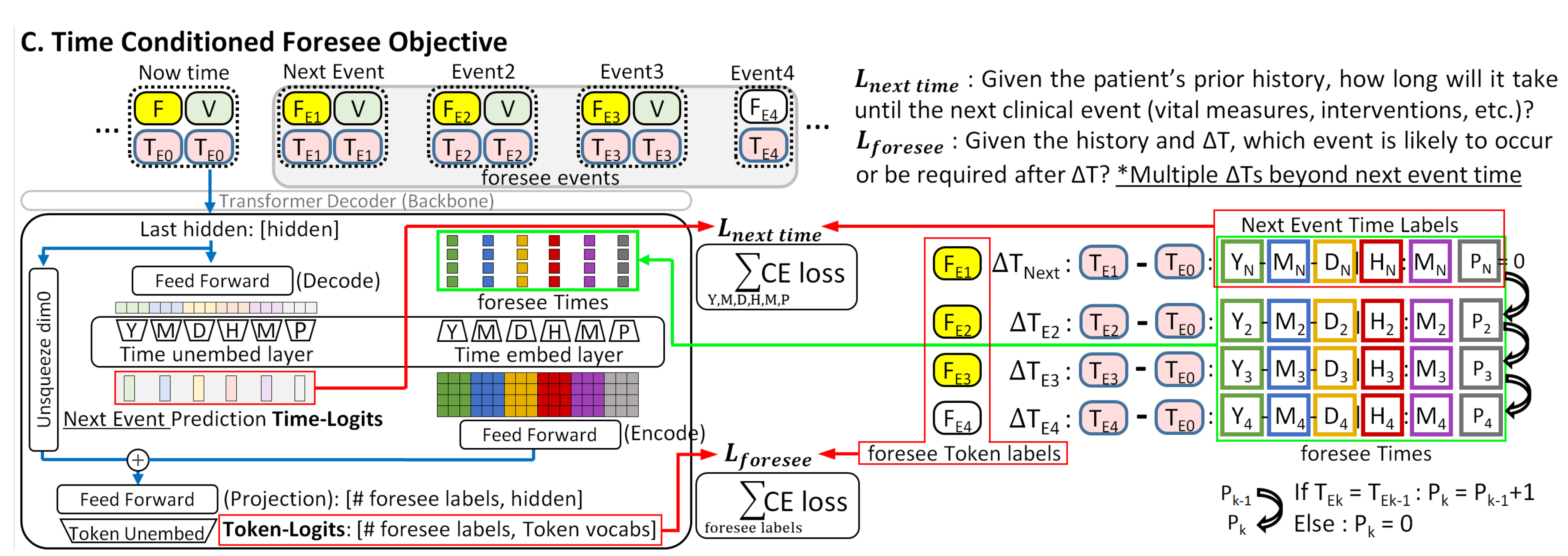
$$\mathbf{z}' = \mathrm{RoPE}(\mathbf{z}_{\text{pos}}, \Theta^{\text{pos}}(P)) \,\|\, \mathrm{RoPE}(\mathbf{z}_{\text{time}}, \Theta^{\text{time}}(T))$$

$$\theta_j^{\text{pos}}(P) = P \cdot (10000)^{-2j/d_{\text{pos}}}, \quad j \in \{0, \dots, \tfrac{d_{pos}}{2} - 1\}. \qquad \theta_m^{\text{time}}(T) = 2\pi \cdot \frac{T \pmod{s_m}}{s_m}, \quad m \in \{1, \dots, M\}.$$



3. Objective - Time-Conditioned Foreseeing

- *Assumption*: Models should jointly predict when and what events will occur, while also forecasting events over long horizons.
- *Method*: A. Predict the time-to-next-event interval. B. Predict which events will occur within a future horizon conditioned on a given time.



Experiments

- ***Downstream tasks performances***
- MIMIC-3: 7 downstream tasks

	Tasks	Total score (↓)			IHM		Phe		Dec-death		Dec-arrest		LOS-ICU		HUO		Vaso	
	Metric	117	17	6	ROC	PRC	macro	micro	ROC	PRC	ROC	PRC	Ma-f1	Kappa	macro	micro	ROC	PRC
No Value share	HEART	5.304	5.434	5.835	0.838	0.442	0.717	0.718	0.869	0.205	0.862	0.199	0.150	0.142	0.703	0.701	0.865	0.363
	MOTOR	4.889	5.159	5.645	0.880	0.552	0.770	0.773	0.908	0.274	0.896	0.268	0.177	0.166	0.753	0.748	0.891	0.438
	EHRSHOT	5.497	5.565	6.026	0.831	0.459	0.638	0.632	0.867	0.188	0.854	0.181	0.147	0.155	0.672	0.668	0.54	0.359
	TRADE	5.260	5.454	6.048	0.828	0.441	0.738	0.738	0.867	0.170	0.857	0.165	0.158	0.151	0.732	0.731	0.869	0.393
	EHRmamba	5.137	5.439	5.926	0.868	0.557	0.690	0.687	0.901	0.277	0.886	0.260	0.150	0.159	0.751	0.753	0.873	0.399
	Ours (No share)	**4.686**	**4.907**	**5.367**	**0.889**	**0.607**	**0.809**	**0.816**	**0.928**	**0.400**	**0.917**	**0.388**	**0.181**	**0.185**	**0.776**	**0.781**	**0.912**	**0.498**
Value share	FM4EHR	5.529	5.691	6.200	0.830	0.435	0.609	0.599	0.886	0.273	0.874	0.261	0.131	0.153	0.658	0.667	0.878	0.384
	ETHOS	4.971	5.248	5.572	0.859	0.530	0.739	0.746	0.900	0.311	0.890	0.304	0.170	0.165	0.721	0.731	0.890	0.437
	STraTS	5.786	5.812	6.071	0.759	0.311	0.656	0.661	0.840	0.141	0.804	0.103	0.123	0.121	0.590	0.598	0.864	0.331
	Ours (Share)	**4.879**	**5.043**	**5.561**	**0.876**	**0.559**	**0.781**	**0.784**	**0.910**	**0.319**	**0.902**	**0.310**	**0.173**	**0.170**	**0.749**	**0.755**	**0.906**	**0.470**

- MIMIC-4: 11 downstream tasks

Task	Metric	No Share MOTOR	No Share Ours	Share ETHOS	Share Ours
IHM	ROC	0.908	**0.916**	0.876	0.894
	PRC	0.385	**0.419**	0.284	0.336
Dec-death	ROC	0.944	**0.953**	0.922	0.952
	PRC	0.188	**0.302**	0.162	0.278
Dec-arrest	ROC	0.937	**0.947**	0.918	**0.947**
	PRC	0.183	**0.281**	0.154	0.259
Dec-ICU	ROC	0.845	**0.863**	0.786	0.862
	PRC	0.160	0.213	0.116	**0.233**
30d-Readm.	ROC	0.659	**0.668**	0.611	0.667
	PRC	0.365	0.381	0.311	**0.389**
30d-death	ROC	0.879	**0.892**	0.818	0.881
	PRC	0.309	**0.330**	0.145	0.296
LOS-Adm.	Ma-F1	0.181	**0.188**	0.158	0.184
	Kappa	0.184	**0.193**	0.154	0.190
LOS-ICU	Ma-F1	0.190	**0.200**	0.161	0.192
	Kappa	0.173	**0.183**	0.126	0.166
Phe.	macro	0.816	**0.840**	0.719	0.816
	micro	0.829	**0.852**	0.724	0.825
Vaso	ROC	0.963	**0.971**	0.952	0.959
	PRC	0.691	**0.732**	0.602	0.657
HUO	macro	0.760	**0.822**	0.700	0.754
	micro	0.761	**0.823**	0.702	0.754

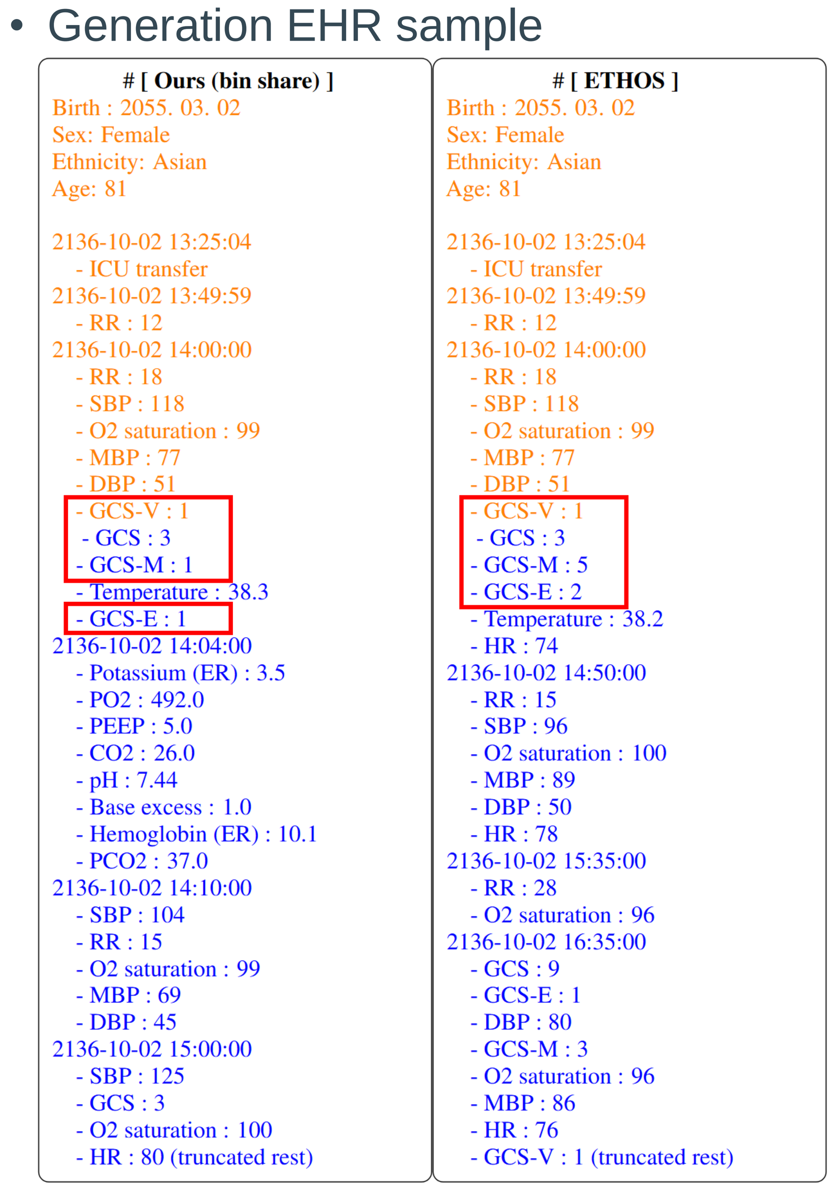
- ML models

Model	AUROC (95% CI)	AUPRC (95% CI)
Dataset: MIMIC-III (117 features)		
Logistic Regression	0.780 (0.752–0.803)	0.392 (0.342–0.447)
Decision Tree	0.634 (0.609–0.658)	0.190 (0.165–0.215)
Random Forest	0.854 (0.834–0.872)	0.501 (0.448–0.549)
XGBoost (2016)	0.860 (0.841–0.879)	0.509 (0.456–0.566)
LightGBM (2017)	0.881 (0.866–0.896)	0.556 (0.510–0.603)
CatBoost (2018)	0.883 (0.869–0.900)	0.565 (0.522–0.614)
MOTOR	0.880 (0.855–0.890)	0.552 (0.500–0.591)
OURS	**0.889 (0.874–0.905)**	**0.607 (0.566–0.649)**
Dataset: MIMIC-IV (195 features)		
Logistic Regression	0.848 (0.836–0.859)	0.256 (0.234–0.282)
Decision Tree	0.607 (0.596–0.619)	0.076 (0.068–0.086)
Random Forest	0.878 (0.868–0.889)	0.321 (0.295–0.352)
XGBoost	0.892 (0.884–0.901)	0.344 (0.319–0.371)
LightGBM	0.900 (0.893–0.910)	0.343 (0.314–0.375)
CatBoost	0.906 (0.898–0.914)	0.359 (0.334–0.387)
MOTOR	0.908 (0.899–0.914)	0.385 (0.365–0.407)
OURS	**0.916 (0.908–0.923)**	**0.419 (0.391–0.446)**

- External Validation (eICU)

Task	Metric	No Share MOTOR	No Share Ours	Share ETHOS	Share Ours
IHM	ROC	0.757	**0.786**	0.735	0.771
	PRC	0.324	**0.355**	0.298	0.346
ICUM	ROC	0.779	**0.811**	0.765	0.795
	PRC	0.256	**0.286**	0.234	0.284

- ***Generative performances***
- Generation EHR sample



- Generation vs ground truth

(a) Timestamp Generation						
Interval (Minutes)	ETHOS RMSE	ETHOS MAE	OURS RMSE	OURS MAE	Improv. (%) RMSE	Improv. (%) MAE
0–1	13426	163	**140**	**1.6**	99	99
1–10	13913	244	**1077**	**11**	92	95
10–100	50409	2404	**645**	**31**	99	99
100–1k	61858	7939	**1796**	**331**	97	96

(b) Token Generation						
Token (Normalized)	ETHOS RMSE	ETHOS MAE	OURS RMSE	OURS MAE	Improv. (%) RMSE	Improv. (%) MAE
$\lvert\sigma\rvert < 1$ sd	0.520	0.307	**0.438**	**0.265**	16	14
$\lvert\sigma\rvert > 2$ sd	2.127	1.279	**1.972**	**1.170**	7	9
Overall	0.680	0.368	**0.623**	**0.337**	8	9

- Temporal generation