Learning Fingerprints for Medical Time Series with Redundancy-Constrained Information Maximization

ICML
International Conference On Machine Learning

Huayu Li[1], ZhengXiao He[1], Xiwen Chen[2], Jingjing Wang[3], Siyuan Tian[4], Jinghao Wen[5], Ao Li[1]
[1]University of Arizona, [2]Morgan Stanley, [3]Clemson University, [4]Tsinghua University, [5]Villanova University

Keywords: *Medical Time Series, Representation Learning, Healthcare*

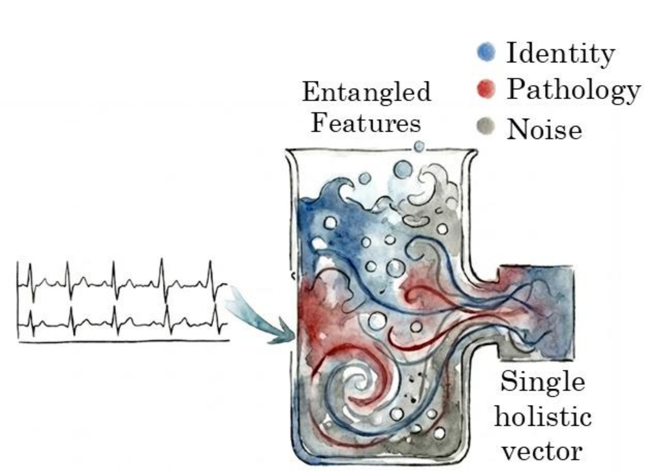
1. The Entangled Representation

Standard Medical Time Series (MedTS) encoders often learn variable-length, entangled representations and rely on heuristic aggregation strategies such as global average pooling or a designated [CLS] token.

Goal: learn compact, fixed-size, and semantically meaningful representations that generalize across diverse medical time series, including ECG, EEG, and activity signals.

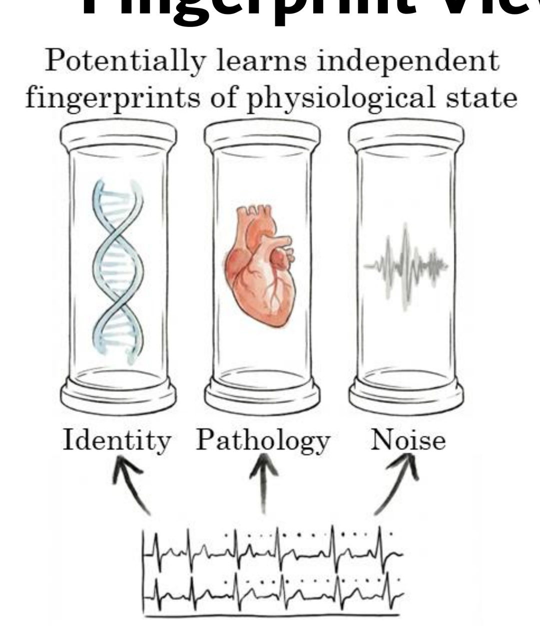
Why it matters:
- **Interpretability:** Disentangled factors enable clearer clinical insights.
- **Sample Efficiency:** Structured latent representations reduce the need for large labeled datasets.
- **Robustness:** Orthogonal fingerprint tokens improve generalization under noisy and heterogeneous physiological signals.

MAE Pitfall



Standard MAE: mixed factors are diffused across every embedding dimension.

Fingerprint View

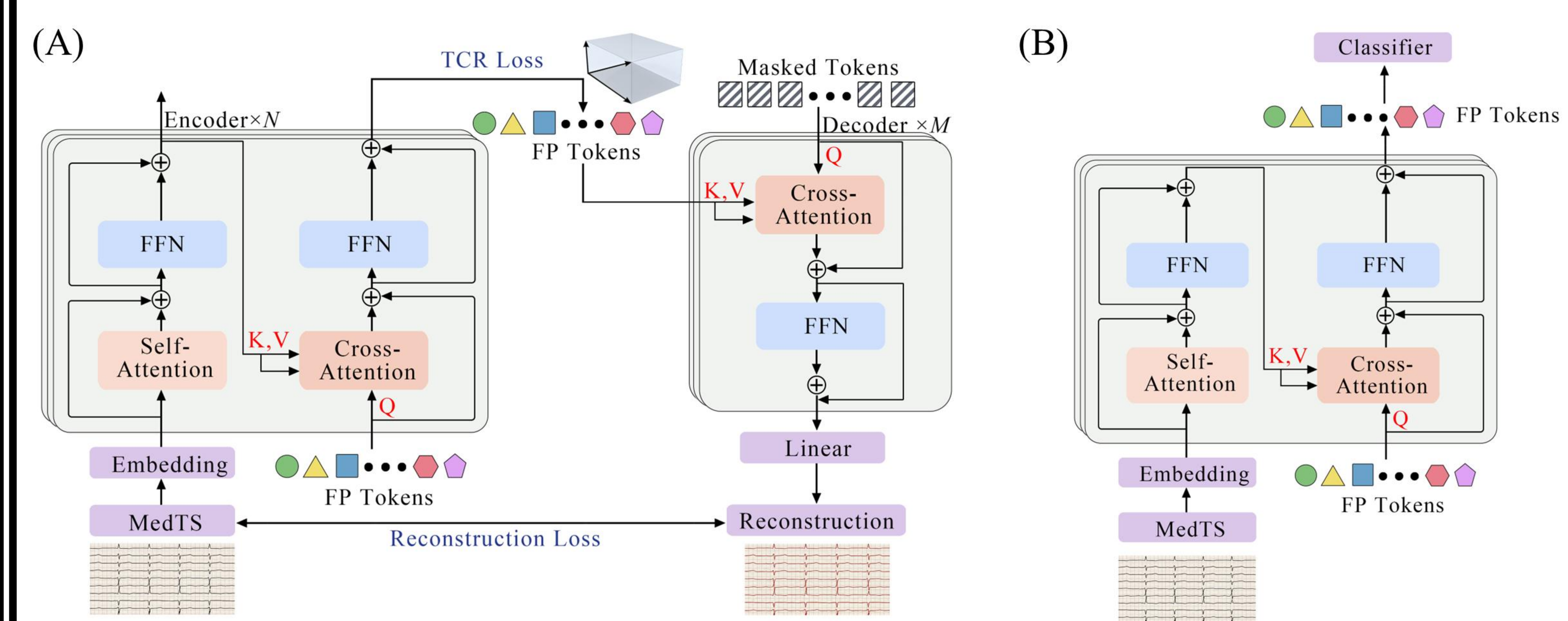


Fingerprint: each fingerprint token is encouraged to capture an independent physiological factor.

Bonus Result

Multi-scale embedding helps. A simple front-end upgrade further improves TS-Fingerprint on ADFTD: Acc 53.91 → 55.06 and F1 51.79 → 53.01. Easy extension.

2. Methodology: TS-Fingerprint



Information-Theoretic Objective

$$\max_{\theta} \underbrace{I(X;F')}_{Sufficiency} - \lambda \underbrace{TC(F')}_{Independence} \longrightarrow \mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{div}$$

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{x\sim\mathcal{D}}\left[\|x - D_\phi(T_{\text{mask}}, F')\|^2\right] \quad \mathcal{L}_{\text{div}} = -\frac{1}{2}\log\det\left(I + \frac{d}{k\epsilon^2}F'^{\top}F'\right)$$

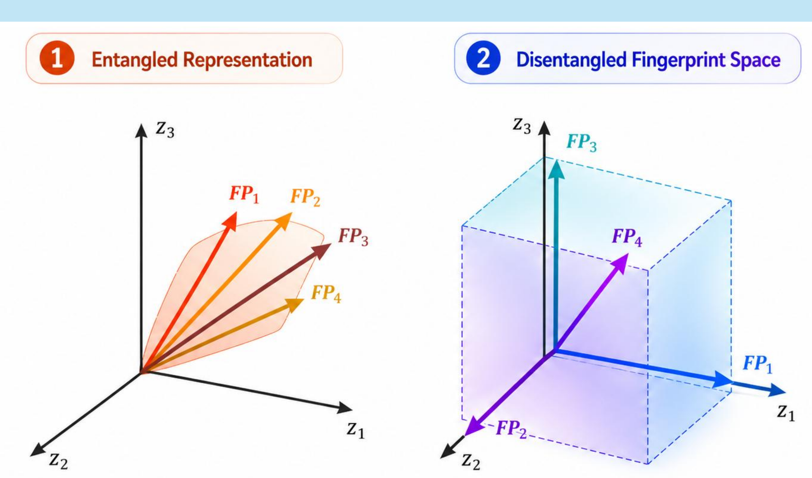
Reconstruction → Maximizing the lower bound of $I(X;F')$

Minimizing redundancy / Total correlation

Theoretical Insight

Minimizing the information-theoretic objective is equivalent to maximizing information sufficiency while reducing Total Correlation among latent tokens.

Geometric Intuition



Orthogonal fingerprint tokens maximize latent volume, reduce mode collapse, and encourage each token to capture a distinct factor of variation.

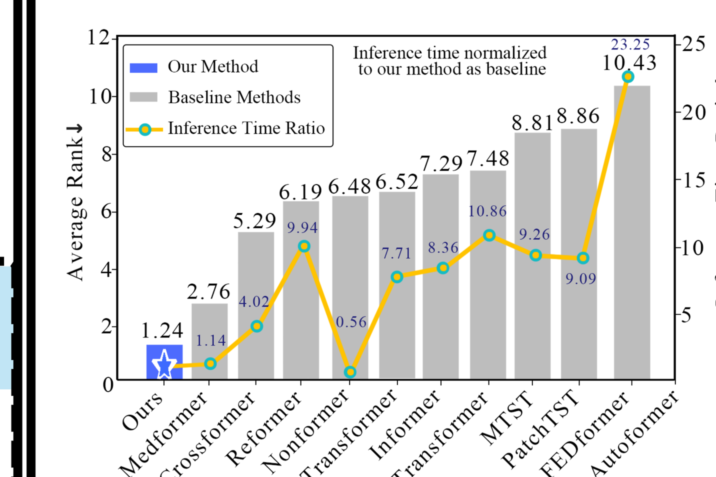
Conclusion

- Compact, disentangled, and interpretable representations for medical time series.
- Strong performance across diverse physiological benchmarks while providing token-level evidence for model decisions.
- Medical time-series AI toward more transparent, robust, and clinically meaningful digital biomarkers.

3. Experimental Results

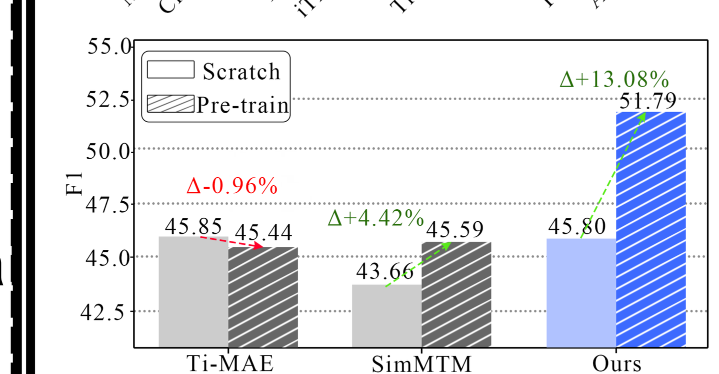
Benchmark: *ADFTD, PTB, PTB-XL, APAVA, SleepEDF, FLAAP, and UCI-HAR.*

Baseline: *Autoformer, Cross-former, FEDformer, Informer, MTST, iTransformer, Nonformer, PatchTST, Reformer, Medformer, Ti-MAE, SimMTM*



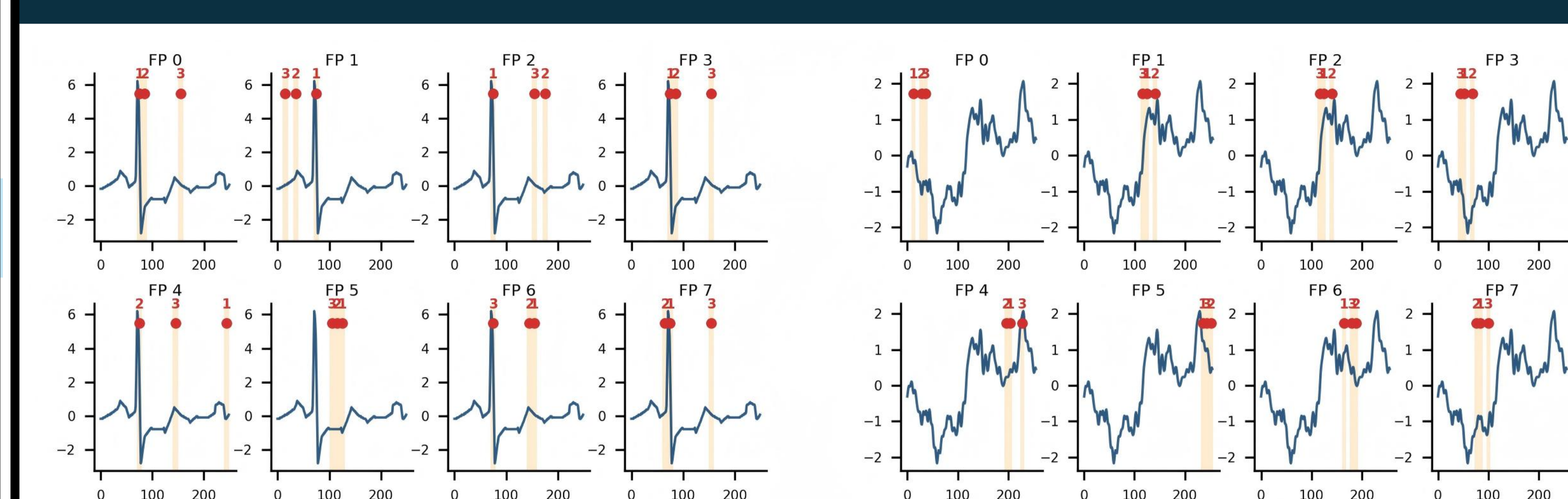
Average Ranking: 1.24

14% Faster than Medformer



13.08% gain from pretraining

4. Interpretability



Fingerprint tokens provide traceable physiological evidence. Each token attends to specific temporal regions, revealing which signal patterns contribute to the prediction.
ECG: sparse and focused. Diagnostic cues concentrate in a few dominant tokens.
EEG: distributed and diverse. Multiple tokens are activated to capture more complex neural dynamics.

Transparent physiological evidence