From Indicators to Insights: Diversity-Optimized for Medical Series-Text Decoding via LLMs

Xiyuan Jin, Jing Wang*, Ziwei Lin, Qianru Jia, Yuqing Huang, Xiaojun Ning, Zhonghua Shi, Youfang Lin

Beijing Jiaotong University, China
Capital Medical University, China

2025.11.04

https://github.com/jinxyBJTU/InDiGO

Background and Motivations

Background

北京交通大學
BEIJING JIAOTONG UNIVERSITY

From Pattern Recognition to Physiological Understanding

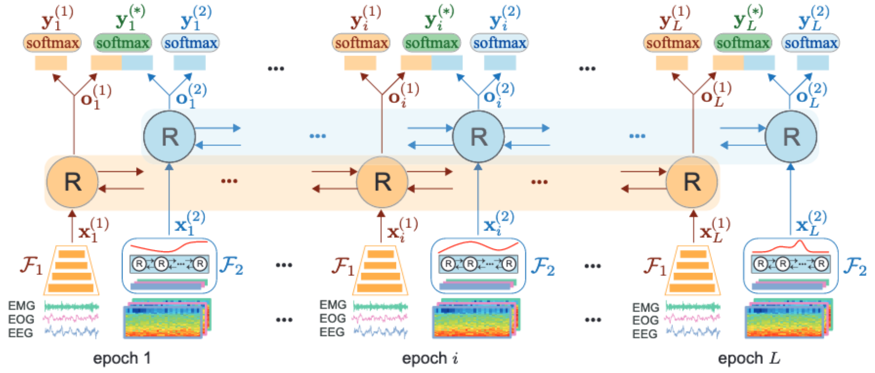
Regular Time Series	vs.	Medical Time Series
Physically or mechanically generated (e.g.,weather,traffic)	Nature of signals	Biophysiological and multi-source (e.g., EEG, ECG, EMG)
Stable statistical patterns, often stationary	Data consistency	High inter-subject variability, non-stationary dynamics
Approximate prediction acceptable	Error tolerance	Misinterpretation may lead to clinical risk

Medical time series transform pattern recognition into **knowledge-driven interpretation** — where data must be understood within its physiological context.

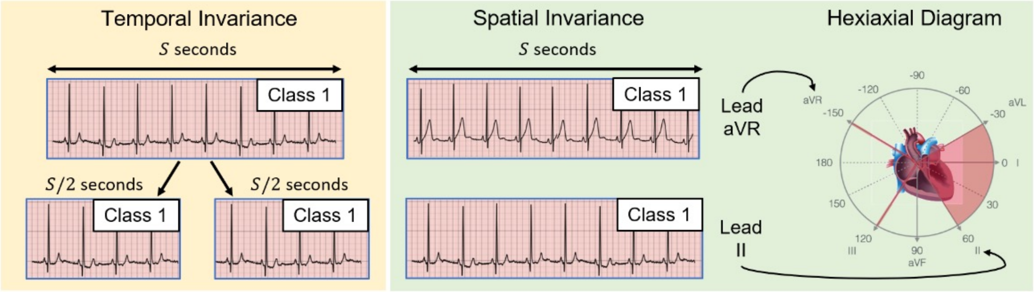
Motivations

BEIJING JIAOTONG UNIVERSITY

Early knowledge integration in medical modeling:

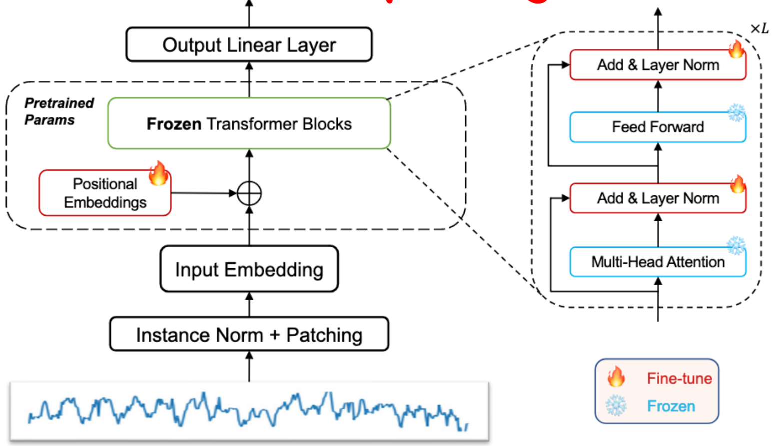


Task-specific module design

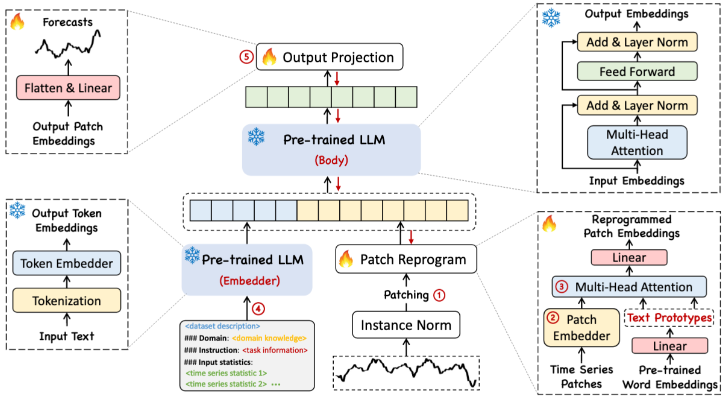


Task-agnostic loss design

LLMs have become a new paradigm due to their efficient ability to transform knowledge



Implicit knowledge modeling

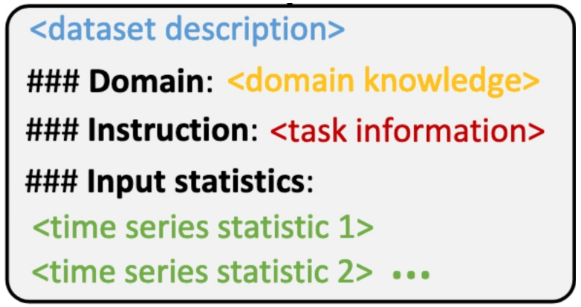


Explicit knowledge encoding

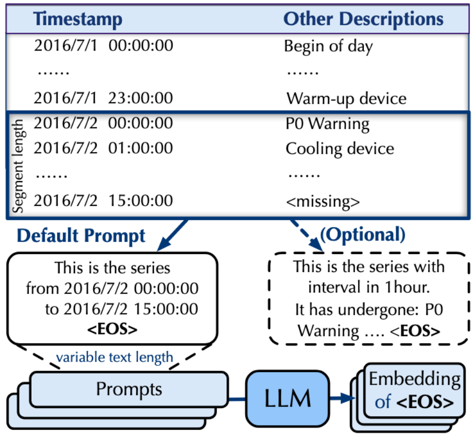
Motivations

北京交通大學
BEIJING JIAOTONG UNIVERSITY

Limited inspiration for decision-making

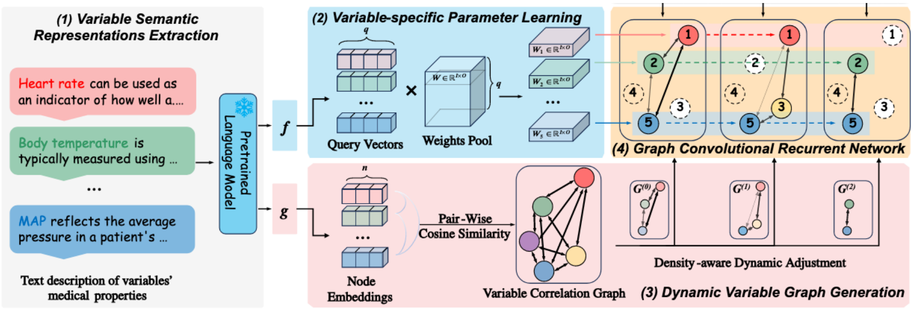


(a) dataset descriptions and sample statistics



(b) timestamp information

Gap



(c) time-varying inter-variable dependencies

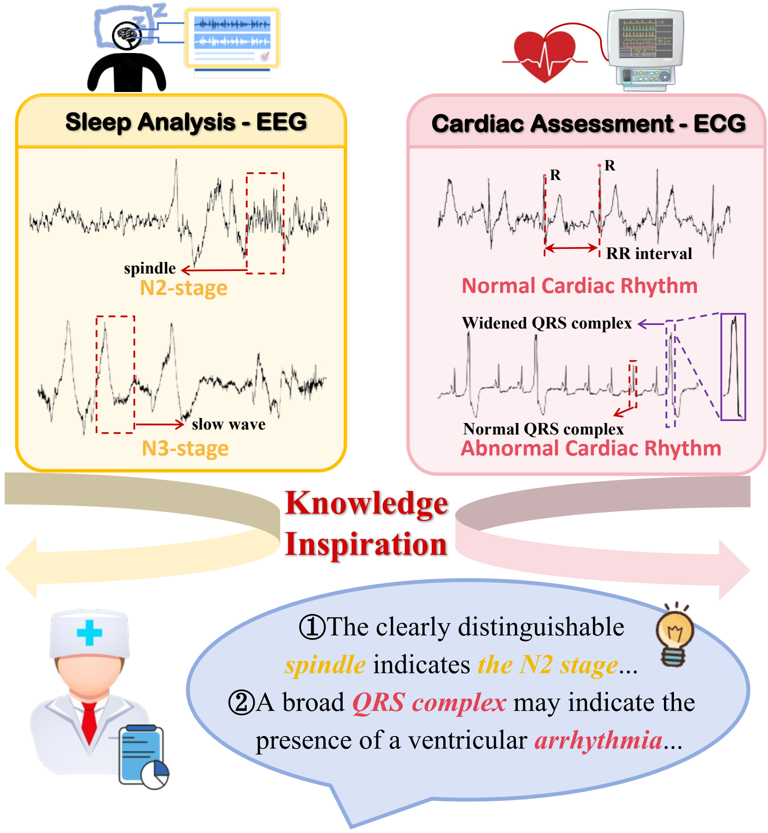


Figure 1: Task-relevant indicators provide critical cues for physiological state interpretation in sleep analysis and cardiac assessment

Motivations

BEIJING JIAOTONG UNIVERSITY

Chanllenges

- **What kinds of promptable knowledge are most effective for decoding medical time series?**
  - Existing prompts lack task-specific, discriminative cues that experts actually rely on. The challenge is to identify what type of knowledge truly drives physiological interpretation.
- **How to robustly integrate time series and suboptimal text prompts?**
  - Text prompts may be incomplete or inaccurate.
  - A robust model must still align and learn meaningful cross-modal representations even when the textual guidance is noisy or partially wrong.

InDiGO —— designed to integrate indicator-guided prompts and optimize their diversity and alignment through an evolutionary learning process.

Method

Method

InDIGO — Indicator-informed Diversity-Guided Optimization

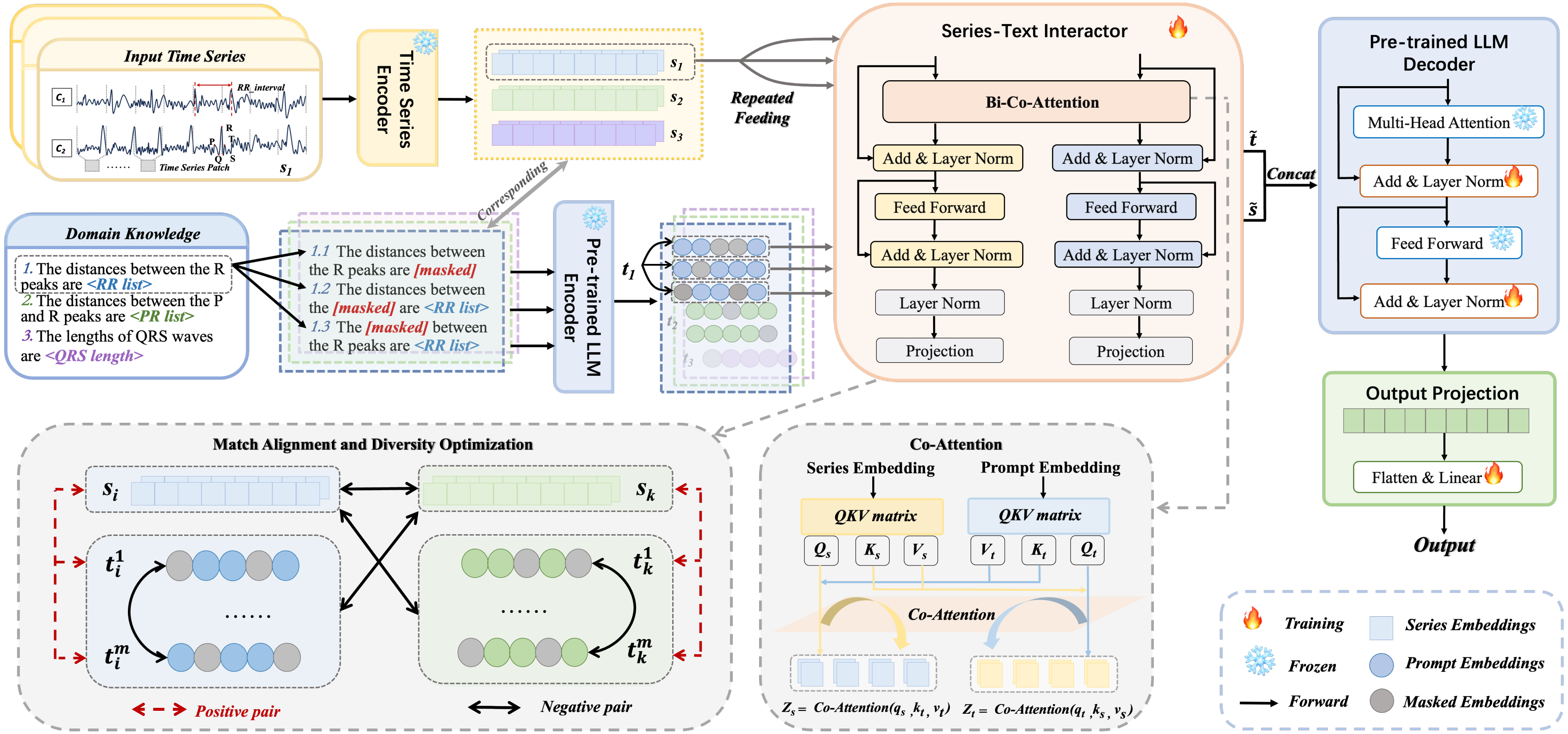


Figure 1: Overview of InDiGO. Given medical time series and initial texts, we apply mask-based importance sampling and pre-trained encoders to extract features. A series-text interactor captures relevance, followed by alignment and diversity optimization to identify the optimal combination.

Method

Step 1: Indicator-Guided Prompt Prototype Construction

LLM predicts the target conditioned on both the series and a text prompt :

$$P_{\text{LLM}}(Y) = \mathbb{E}_{(s,t)\sim P(s,t)}\left[P_{\text{LLM}}(Y|s,t)\right] \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{t\sim P(t|s_i)}\left[P_{\text{LLM}}(Y|s_i,t)\right]$$

In practice, we approximate the whole text distribution with a single suboptimal text → this introduces bias

$$P_{\text{LLM}}(Y) = \mathbb{E}_{(s,t)\sim P(s,t)}\left[P_{\text{LLM}}(Y|s,t)\right] \approx \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{t\sim P(t|s_i)}\left[P_{\text{LLM}}(Y|s_i,t)\right]$$

$$\text{Bias}(\hat{\mathcal{M}}) = \mathbb{E}[P_{LLM}(y_i|s_i,t)P(t|s_i)] - \int P_{LLM}(y_i|s_i,t)P(t|s_i)dt$$

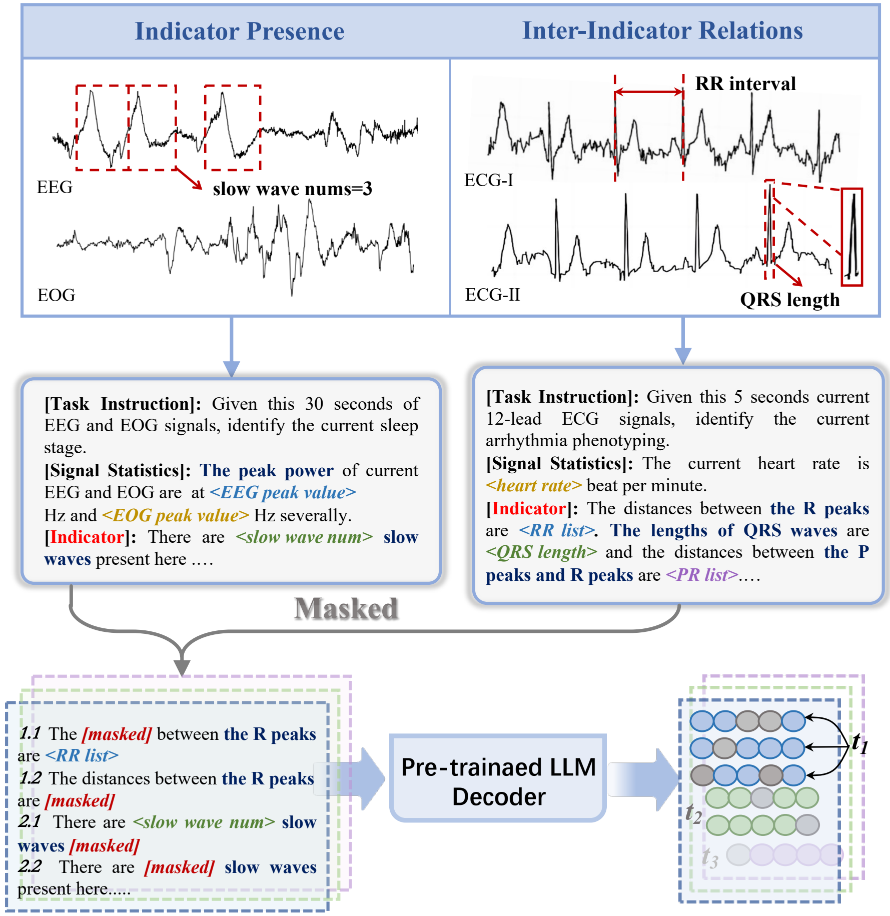


Figure 3: Scene-specific indicator prompting via semi-automated extraction.

Method

Step 2: Masked Monte Carlo Importance Sampling

Marginal likelihood: $$\hat{\mathcal{M}} = \int P_{\text{LLM}}(y_i|s_i,t)q(t|t_i^0)\frac{P(t|s_i)}{q(t|t_i^0)}dt \approx \frac{1}{m}\sum_{j=1}^{m} P_{\text{LLM}}(y_i|s_i,t_i^j)\frac{P(t_i^j|s_i)}{q(t_i^j|t_i^0)}$$

Optimal Text Prompt: $$t_i^* = \underset{t_i}{\text{argmax}}[\log P_{\text{LLM}}(y_i|s_i,t_i) + \log P(t_i|s_i) - \log q(t_i|t_i^0)]$$

Random Masking Operation: $$\mathcal{D}^{t_i} = \{t_i^j|t_i^{j,1}, t_i^{j,2}, \ldots, t_i^{j,L_{t_i}}; j = 1, \ldots, m\}$$

Step 3: Match Alignment + Diversity Optimization

Series-Text Interaction: $$(\mathbf{z}_s, \mathbf{z}_t) = \text{BiCoAttn}(\mathbf{s}, \mathbf{t}) = \left(\text{softmax}\left(\frac{\mathbf{q}_s\mathbf{k}_t^\top}{\sqrt{d}}\right)\mathbf{v}_t,\ \text{softmax}\left(\frac{\mathbf{q}_t\mathbf{k}_s^\top}{\sqrt{d}}\right)\mathbf{v}_s\right)$$

Alignment and Diversity Optimization: $$\mathcal{L}_{\text{objective}} = \mathbb{E}_{t_i^j \sim q(t|t_i^0)}[-\log P(s_i|t_i^j)] - H(q(t|t_i^0))$$

Results

Results

北京交通大學 BEIJING JIAOTONG UNIVERSITY

Methods	Sleep-EDF-20			Sleep-EDF-78		
	Acc.	Macro F1	Kappa	Acc.	Macro F1	Kappa
TF-C [48]	55.42 ±1.39	26.04 ±0.21	30.74 ±1.52	53.90 ±4.03	26.00 ±2.09	29.32 ±6.43
SimMTM [7]	66.91 ±1.89	53.21 ±1.95	53.25 ±2.02	63.06 ±2.67	57.07 ±2.13	53.07 ±3.42
OneFitsAll [49]	72.60 ±1.51	61.61 ±5.80	61.81 ±3.50	68.50 ±2.19	54.24 ±1.96	55.21 ±3.07
Time-LLM [12]	80.31 ±2.63	71.64 ±3.02	70.22 ±2.84	78.08 ±2.96	66.09 ±3.25	68.04 ±3.14
KEDGN [24]	74.89 ±3.86	64.29 ±3.36	64.90 ±5.46	70.34 ±1.85	58.59 ±2.74	57.47 ±2.56
MiniRocket [5]	81.60 ±1.55	72.82 ±2.01	72.79 ±1.96	78.36 ±1.93	70.18 ±2.35	69.46 ±2.46
BIOT [45]	81.86 ±4.41	75.29 ±4.47	75.14 ±6.00	77.15 ±3.04	69.36 ±4.13	68.26 ±4.36
TinySleepNet [38]	83.64 ±2.31	77.54 ±2.55	77.63 ±2.29	83.49 ±2.24	76.64 ±2.61	76.41 ±2.59
XSleepNet [32]	80.93 ±2.34	76.71 ±2.59	74.31 ±2.32	81.83 ±2.30	75.28 ±2.66	75.44 ±2.37
L-SeqSleepNet [33]	82.90 ±2.12	74.90 ±2.22	76.47 ±2.24	80.84 ±2.18	72.67 ±2.38	74.94 ±2.51
SleepHGNN [10]	81.15 ±1.96	72.88 ±2.17	73.35 ±2.16	77.35 ±2.13	69.56 ±2.39	68.65 ±2.41
SleepKD [21]	82.44 ±2.40	74.11 ±2.72	76.87 ±2.63	80.19 ±2.85	72.65 ±2.84	74.86 ±2.93
SleepDG [42]	81.92 ±2.27	74.74 ±2.53	76.43 ±2.47	79.95 ±2.42	72.21 ±2.59	74.16 ±2.68
Brant-X [47]	84.58 ±1.98	77.63 ±2.13	79.29 ±2.18	82.84 ±2.21	77.04 ±2.30	76.67 ±2.49
InDiGO	**89.04** ±1.80	**80.53** ±1.77	**84.91** ±2.51	**86.79** ±1.90	**81.12** ±1.88	**81.60** ±2.89

Table 1: 5-fold cross-validated average results for sleep stage classification

Results

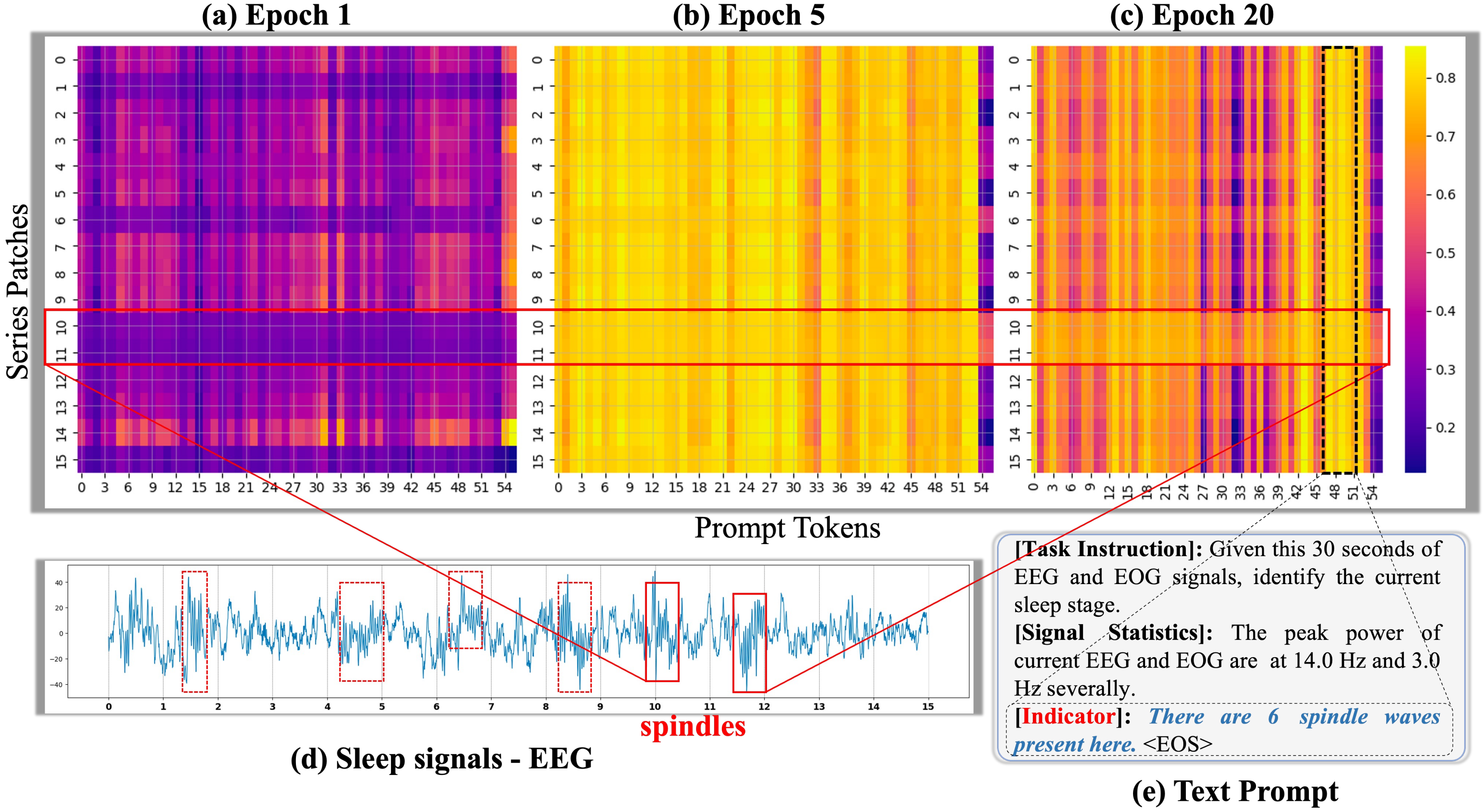


Figure 6: A showcase of series-text correlation with network training. (a)-(c) represent the co-attention scores of the series-text pair, (d) and (e) represent the series and text samples.

Thank you for your listening !

Xiyuan Jin
xiyuanjin@bjtu.edu.cn

2025.11.04