SCHOOL OF COMPUTING & DATA SCIENCE
The University of Hong Kong

ICML
International Conference On Machine Learning
2025

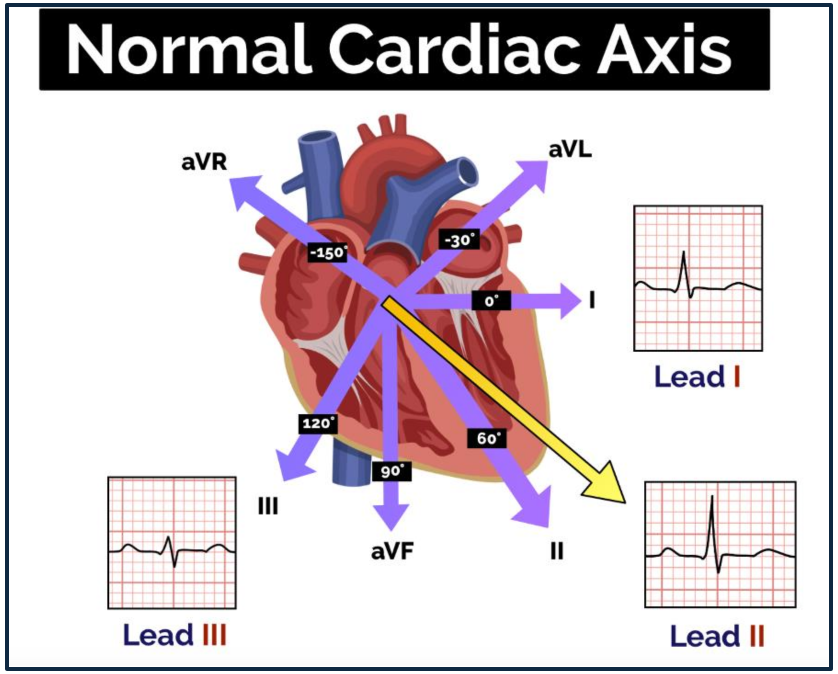
From Token to Rhythm: A Multi-Scale Approach for ECG-Language Pretraining

Fuying Wang[1*], Jiacheng Xu[1*], Lequan Yu[1]

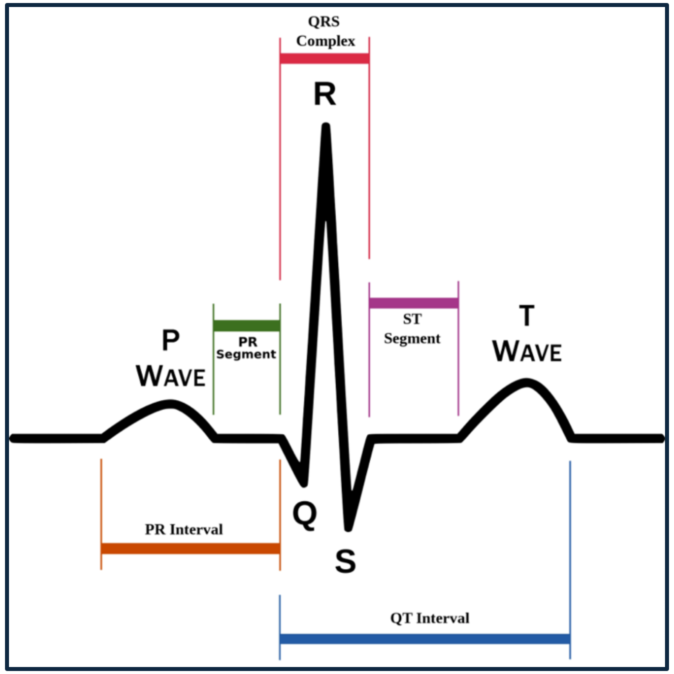
[1]School of Computing and Data Science, The University of Hong Kong

Presentation at ICML 2025

Introduction to Electrocardiograms (ECGs)



ECG illustration



ECG waveforms

- Electrocardiograms (ECGs) records the heart's electrical impulses during each heartbeat
- Each heartbeat generates distinct waveforms
- These characteristics are essential for detecting heart-related abnormalities.

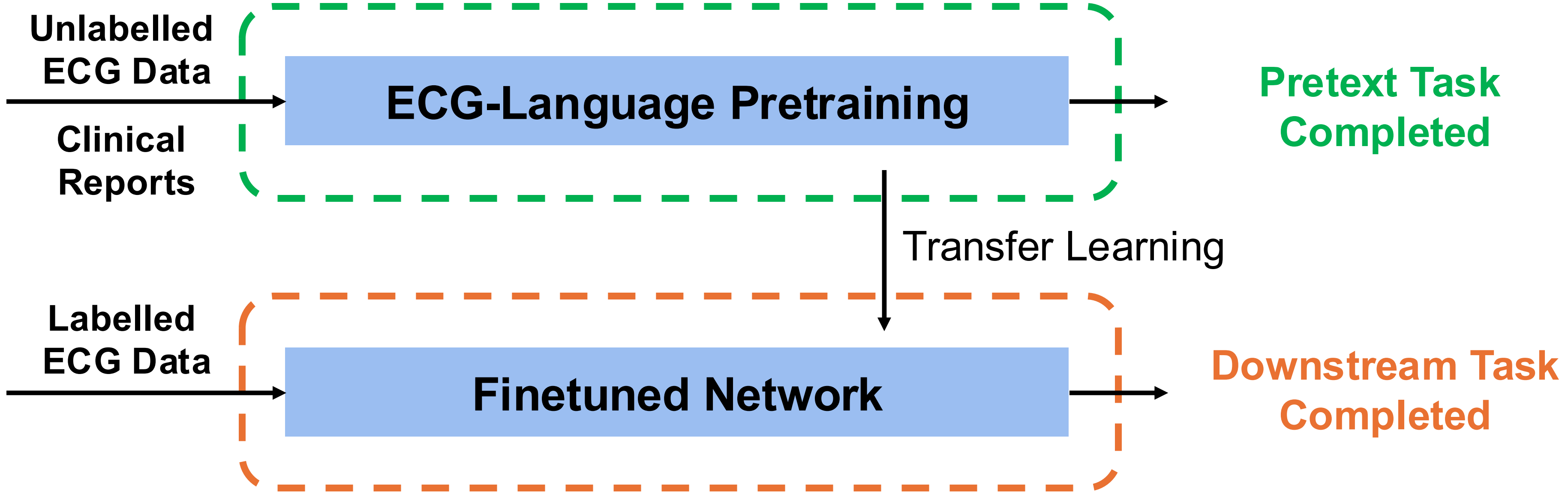
Image Source: https://geekymedics.com/how-to-read-an-ecg/

ECG-Language Pretraining

- **Motivation**: Analyzing ECG signals requires large-scale annotations

ECG-Language Pretraining

- **Motivation**: Analyzing ECG signals requires large-scale annotations

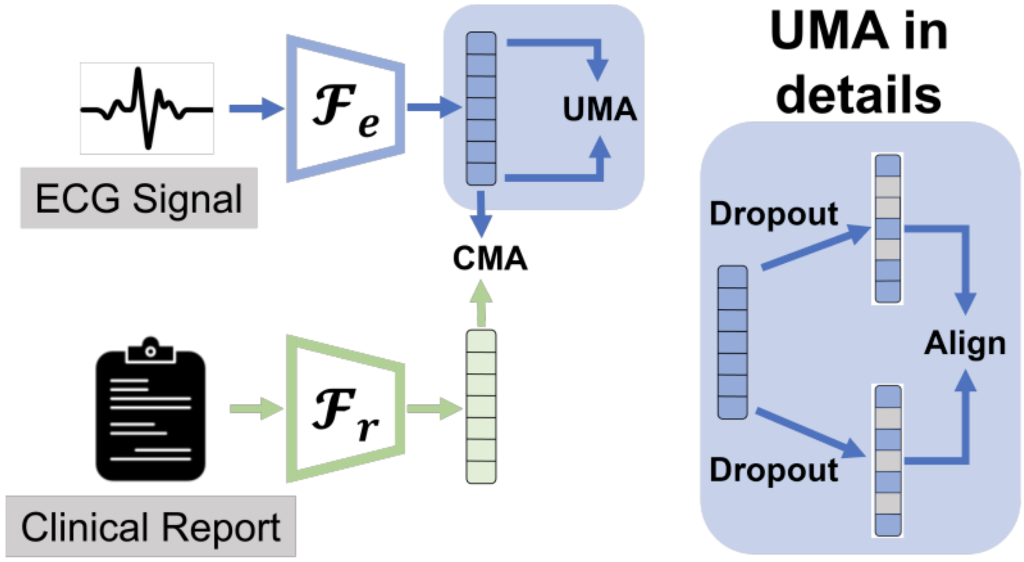


- Learning from paired clinical reports is a promising research direction for learning effective ECG representations.

Related Work

- MERL [1] propose multimodal ECG representation learning framework, and first introduce zero-shot ECG classification.

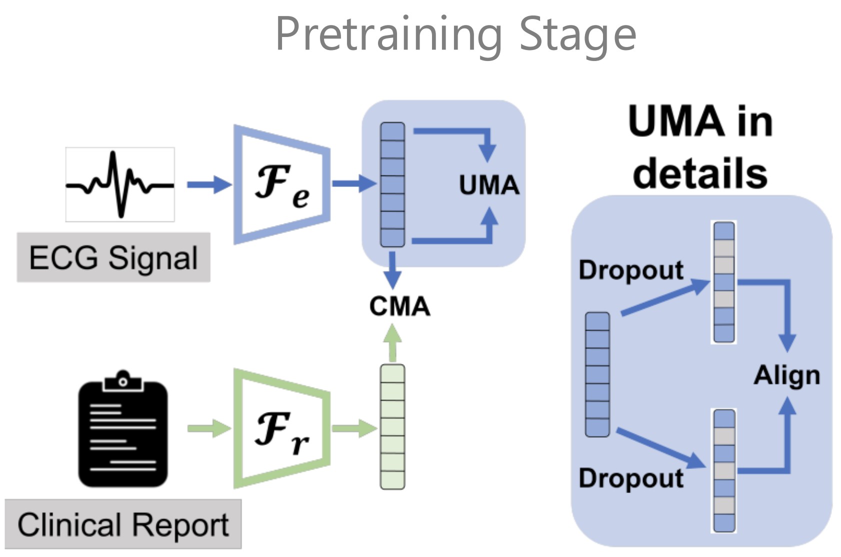
Pretraining Stage



[1]. Liu, Che, Zhongwei Wan, Cheng Ouyang, Anand Shah, Wenjia Bai, and Rossella Arcucci. "Zero-shot ecg classification with multimodal learning and test-time clinical knowledge enhancement." *arXiv preprint arXiv:2403.06659* (2024).

Related Work

- MERL [1] propose multimodal ECG representation learning framework, and first introduce zero-shot ECG classification.



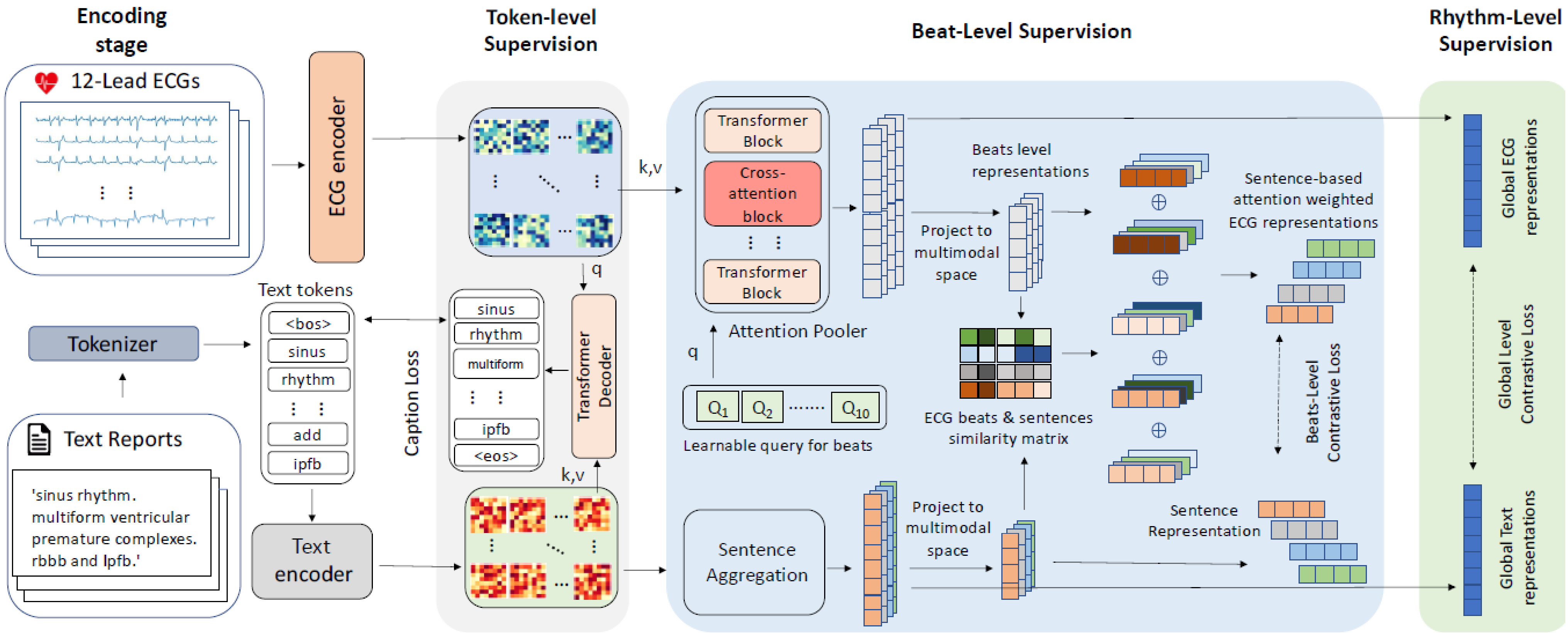
- **Challenge**: Focus solely on global signal, overlooking fine-grained waveform characteristics

[1]. Liu, Che, Zhongwei Wan, Cheng Ouyang, Anand Shah, Wenjia Bai, and Rossella Arcucci. "Zero-shot ecg classification with multimodal learning and test-time clinical knowledge enhancement." *arXiv preprint arXiv:2403.06659* (2024).

Motivation

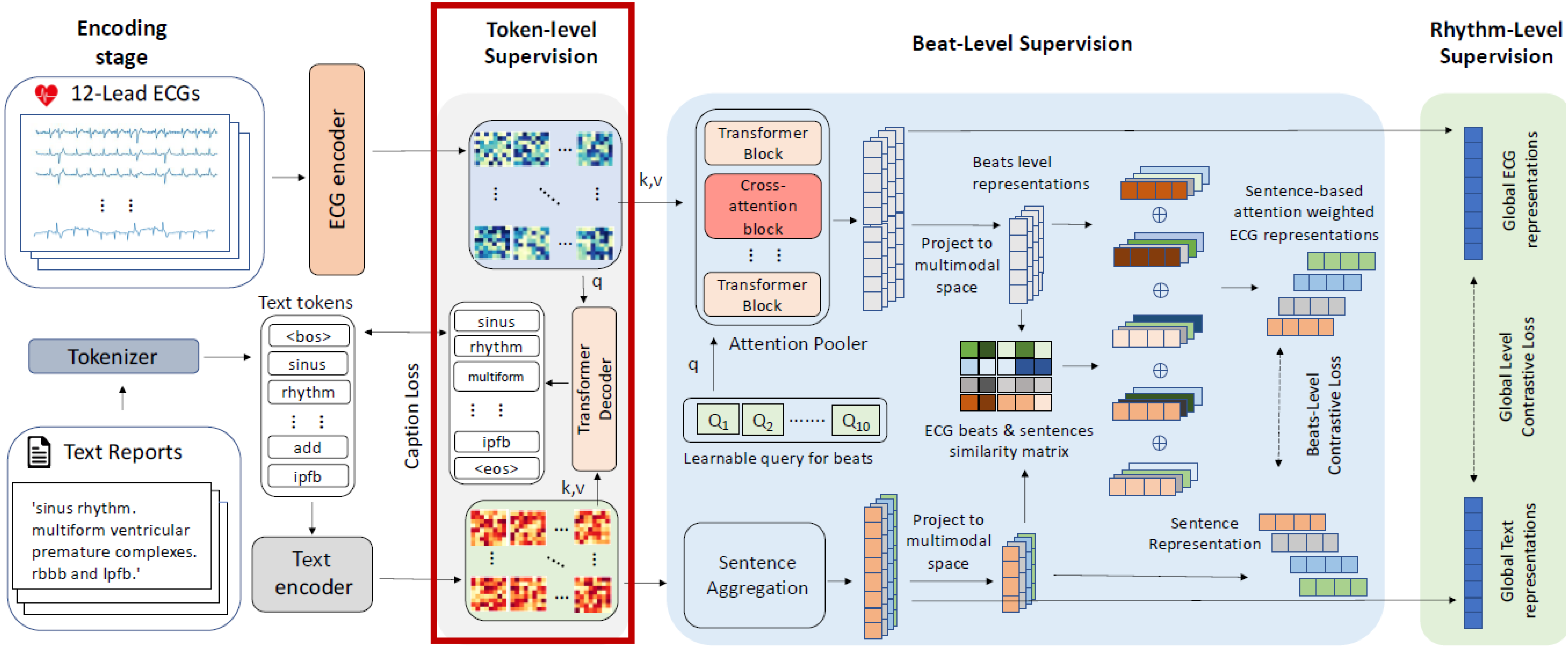
- ECG signals can be interpreted in a hierarchical manner:
  - **Rhythm-level**: Capture the overall rhythm pattern across the entire ECG signal.
  - **Beat-level**: Focuses on individual heartbeats as complete units.
  - **Token-level**: Examines fine-grained waveform components, such as P waves and QRS complexes.

Multi-scale ECG-Language Pretraining



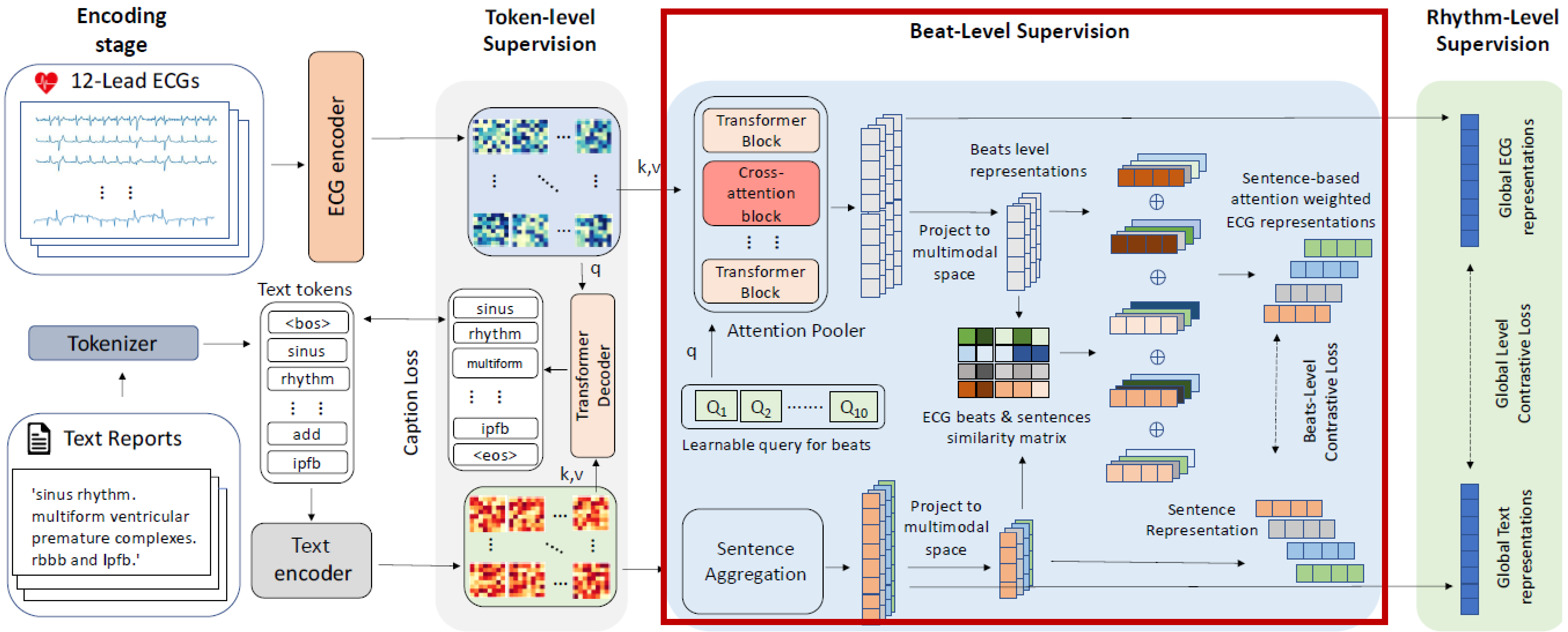
- We introduce a Multi-scale ECG-Language Pretraining (MELP), a framework that leverages multi-scale supervision and generalizes well across diverse downstream tasks.

Token-level Supervision



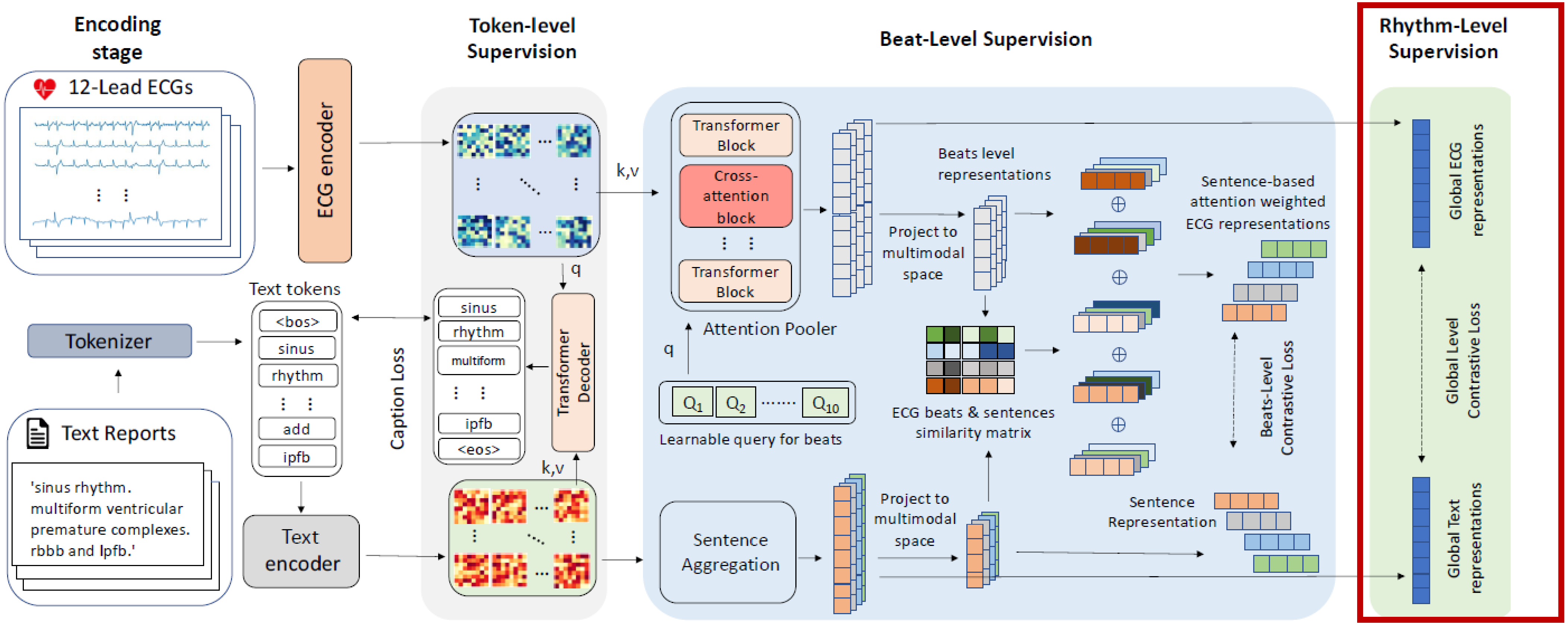
- Learning to generate ECG reports from token-level embedding

Multi-scale ECG-Language Pretraining



- Attention-weighted local contrastive loss for beat-sentence alignment.

Multi-scale ECG-Language Pretraining



- Global contrastive loss for global ECG signal and global report alignment.

Experiment

Pretraining datasets:
- MIMIC-IV-ECG v1.0 database

Downstream datasets:
- PTB-XL
- CSN
- CPSC2018

Tasks
- Linear Probing ECG Classification
- Zero-shot ECG Classification

Database	#.Samples			
MIMIC-IV-ECG	760,618			
	Class	Train	Val	Test
PTB-XL	NORM	7254	916	913
	CD	2048	234	256
	HYP	1353	172	184
	MI	416	64	56
	STTC	1907	256	243
	Total	12 978	1642	1652
CPSC2018	NORM	1213	197	365
	AF	1289	168	342
	I-AVB	889	101	251
	LBBB	243	33	56
	RBBB	1964	292	589
	PAC	864	130	274
	PVG	1084	146	308
	STD	1148	178	345
	STE	264	58	68
	Total	8958	1303	2598
CSN	AF	1583	186	449
	GSVT	1639	189	472
	SB	2804	315	769
	SR	1625	161	436
	Total	7651	851	2126

ECG Classification Results

Table 2. Linear probing performance (AUC [%]) of MELP and baseline models across multiple datasets. Results are reported for different training data proportions (1%, 10%, and 100%). The best and second-best results are highlighted in bold and underlined, respectively.

Methods	PTBXL-Rhythm			PTBXL-Sub			PTBXL-Form			PTBXL-Super			CPSC2018			CSN		
Training ratio	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%
SimCLR (Chen et al., 2020)	51.41	69.44	77.73	60.84	68.27	73.39	54.98	56.97	62.52	63.41	69.77	73.53	59.78	68.52	76.54	59.02	67.26	73.20
BYOL (Grill et al., 2020)	41.99	74.40	77.17	57.16	67.44	71.64	48.73	61.63	70.82	71.70	73.83	76.45	60.88	74.42	78.75	54.20	71.92	74.69
BarlowTwins (Zbontar et al., 2021)	50.12	73.54	77.62	62.57	70.84	74.34	52.12	60.39	66.14	72.87	75.96	78.41	55.12	72.75	78.39	60.72	71.64	77.43
MoCo-v3 (Chen et al., 2021)	51.38	71.66	74.33	55.88	69.21	76.69	50.32	63.71	71.31	73.19	76.65	78.26	62.13	76.74	75.29	54.61	74.26	77.68
SimSiam (Chen & He, 2021)	49.30	69.47	75.92	62.52	69.31	76.38	55.16	62.91	71.31	73.15	72.70	75.63	58.35	72.89	75.31	58.25	68.61	77.41
TS-TCC (Eldele et al., 2021)	43.34	69.48	78.23	53.54	66.98	77.87	48.04	61.79	71.18	70.73	75.88	78.91	57.07	73.62	78.72	55.26	68.48	76.79
CLOCS (Kiyasseh et al., 2021)	47.19	71.88	76.31	57.94	72.55	76.24	51.97	57.79	72.65	68.94	73.36	76.31	59.59	77.78	77.49	54.38	71.93	76.13
ASTCL (Wang et al., 2024)	52.38	71.98	76.05	61.86	68.77	76.51	44.14	60.93	66.99	72.51	77.31	81.02	57.90	77.01	79.51	56.40	70.87	75.79
CRT (Zhang et al., 2023)	47.44	73.52	74.41	61.98	70.82	78.67	46.41	59.49	68.73	69.68	78.24	77.24	58.01	76.43	82.03	56.21	73.70	78.80
ST-MEM (Na et al., 2024)	51.12	65.44	74.85	54.12	57.86	63.59	55.71	59.99	66.07	61.12	66.87	71.36	56.69	63.32	70.39	59.77	66.87	71.36
HeartLang (Jin et al.)	62.08	76.22	90.34	64.68	79.34	**88.91**	58.70	63.99	80.23	78.94	85.59	87.52	60.44	66.26	77.87	57.94	68.93	82.49
MERL (Liu et al., 2024)	53.33	82.88	88.34	64.90	80.56	84.72	58.26	72.43	79.65	82.39	86.27	**88.67**	70.33	85.32	90.57	66.60	82.74	87.95
MELP (Ours)	**88.83**	**94.65**	**96.91**	**79.22**	**84.40**	87.46	**63.41**	**76.71**	**83.30**	**85.82**	**87.61**	87.87	**88.54**	**91.75**	**94.32**	**78.25**	**84.83**	**90.17**

Best AUC in 16 out of 18 settings.

Table 3. Zero-shot classification performance (AUC [%]) of MELP and baseline models across multiple datasets.

Methods	CSN	PTBXL-Rhythm	PTBXL-Form	PTBXL-Sub	PTBXL-Super	CPSC2018	Average
MERL (Liu et al., 2024a)	74.4	78.5	65.9	75.7	74.2	82.8	75.3
MELP (Ours)	**77.6**	**85.4**	**69.1**	**81.2**	**76.2**	**84.2**	**79.0**
Gains	**+3.2**	**+6.9**	**+3.2**	**+5.5**	**+2.0**	**+1.4**	**+3.7**

Improve average AUC by +3.7%

Ablation Studies

Table 5. Ablation results of loss functions on 6 linear probing tasks. The first row indicates training with only the instance-level contrastive loss $\mathcal{L}_g$. The **Best** and Second-best results are shown in **Bold** and underlined.

$\mathcal{L}_g$	$\mathcal{L}_{LM}$	$\mathcal{L}_{Local}$	PTBXL-Rhythm			PTBXL-Form			PTBXL-Sub			PTBXL-Super			CPSC2018			CSN			Average
			1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	
✓			83.78	88.44	94.98	57.93	72.14	82.07	77.32	81.97	84.36	84.55	87.24	87.52	78.52	87.07	92.57	75.94	82.04	86.66	82.51
	✓		77.64	79.44	85.21	52.95	63.80	76.91	71.41	76.67	82.97	78.73	82.80	85.18	64.19	73.05	85.26	69.81	79.37	84.41	76.10
		✓	81.04	89.88	96.67	49.81	67.82	81.41	66.14	81.38	84.76	79.94	87.49	87.73	64.18	84.08	93.17	55.89	79.77	88.79	78.89
✓	✓		83.25	89.87	94.86	56.58	72.71	81.99	78.61	82.14	85.84	84.62	87.18	87.56	83.74	88.40	92.77	74.86	80.48	87.11	82.92
✓		✓	84.36	88.44	95.29	57.22	72.07	82.96	**81.20**	82.89	85.42	84.80	87.25	87.57	76.97	86.31	92.26	73.77	81.43	81.50	82.32
✓	✓	✓	**88.83**	**94.65**	**96.91**	**63.41**	**76.71**	**83.30**	79.22	**84.40**	**87.46**	**85.82**	**87.61**	**87.87**	**88.54**	**91.75**	**94.32**	**78.25**	**84.83**	**90.17**	**85.78**

Best AUC when combing all three-scale supervision

Table 7. Ablation results of ECG encoders. We have used RLM as the augmentation technique for ECG by default. CMSC can't easily integrate into our model since it needs to split the ECG into two parts and performs contrastive learning.

ECG encoder	PTBXL-Rhythm			PTBXL-Form			PTBXL-Sub			PTBXL-Super			CPSC2018			CSN			Average
	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	
ResNet-18	85.10	90.11	94.31	62.82	73.59	79.23	75.59	81.72	85.70	**85.84**	86.99	87.24	83.31	89.75	93.35	68.79	82.12	89.71	83.07
Wav2Vec 2.0	**88.83**	**94.65**	**96.91**	**63.41**	**76.71**	**83.30**	**79.22**	**84.40**	**87.46**	85.82	**87.61**	**87.87**	**88.54**	**91.75**	**94.32**	**78.25**	**84.83**	**90.17**	**85.78**
Wav2Vec 2.0 + CMSC	83.15	88.25	94.82	62.07	75.55	82.57	77.21	82.29	84.85	85.14	87.52	87.64	80.69	88.40	92.91	71.89	81.00	87.42	82.97

Best results using Wave2Vec 2.0

Conclusion

- We introduce a Multi-scale ECG-Language Pretraining (MELP) framework that leverages multi-scale supervision for ECG representation Learning.
- (Future Work) Encoding prior medical knowledge into ECG foundation model for better interpretability
- (Future Work) Conduct instruction tuning for more unified ECG tasks via LLMs

SCHOOL OF COMPUTING & DATA SCIENCE
The University of Hong Kong

ICML International Conference On Machine Learning 2025

Thank you!

Code: https://github.com/HKU-MedAI/MELP

HuggingFace: https://huggingface.co/fuyingw/MELP_Encoder