北京大学 健康医疗大数据国家研究院
NATIONAL INSTITUTE OF HEALTH DATA SCIENCE AT PEKING UNIVERSITY

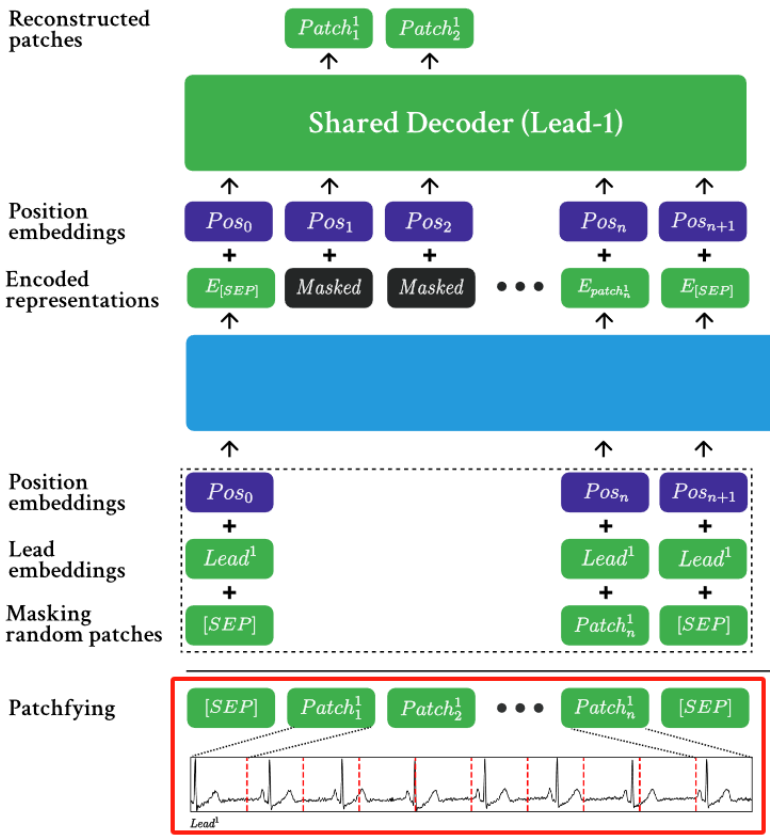
ICLR

Reading Your Heart: Learning ECG Words and Sentences via Pre-training ECG Language Model

Jiarui Jin

2025.3.25

Introduction



Two Significant Drawbacks:

- Ignoring Form and Rhythm Characteristics of ECG
- Ignoring Latent Semantic Relationships of ECG

[1] Na, Yeongyeon, et al. "Guiding Masked Representation Learning to Capture Spatio-Temporal Relationship of Electrocardiogram." *ICLR 2024*.

Introduction

- **New Perspective**: Segments ECG into "words" and "sentences”
- **HeartLang**: Self-supervised framework for ECG language processing
- **Largest Vocabulary**: Covering diverse heart conditions

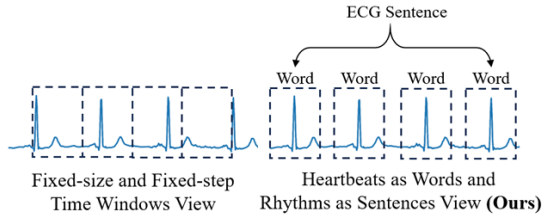


Figure 1: Two perspectives on ECG signals.

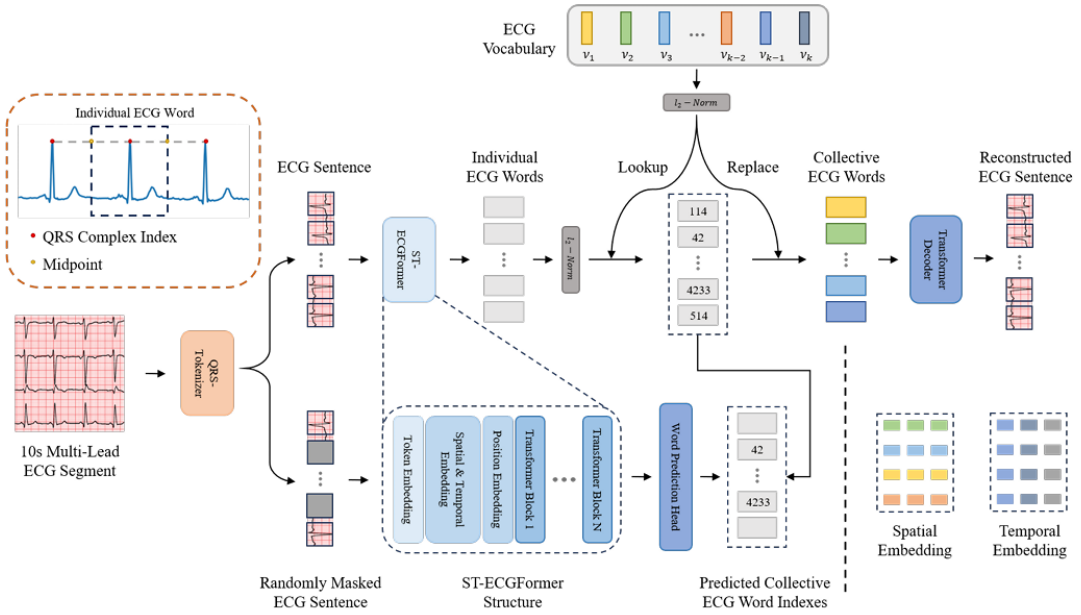


Figure 2: Framework of HeartLang.

Related Work

Contrastive-based methods:

- CLOCS [ICML 2021]
- ISL [AAAI 2022]
- BTFS [ICML 2022]
- ASTCL [TNNLS 2024]

Reconstruction-based methods:

- MaeFE [TIM 2023]
- CRT [TNNLS 2023]
- ST-MEM [ICLR 2024]

- Treating them as ordinary time-series data
- Focus on **spatio-temporal** or **time-frequency** domain representation learning
- Ignore **morphological** and **semantic** representation learning

Related Work

Computer Methods and Programs in Biomedicine 202 (2021) 105959

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

ECG Language processing (ELP): A new technique to analyze ECG signals☆

Sajad Mousavi[a,*], Fatemeh Afghah[a], Fatemeh Khadem[a], U. Rajendra Acharya[b,c,d]

[a] School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA
[b] School of Engineering, Ngee Ann Polytechnic, Singapore
[c] School of Science and Technology, Singapore University of Social Sciences, 463 Clementi Road, 599494, Singapore
[d] Department Bioinformatics and Medical Engineering, Asia University, Taiwan

ECGBERT: Understanding Hidden Language of ECGs with Self-Supervised Representation Learning

Seokmin Choi[1,2,†] Sajad Mousavi[1,†*] Phillip Si[1,3,†]
Haben G. Yhdego[1] Fatemeh Khadem[1] Fatemeh Afghah[4]
[1]*CardioPhi LLC*, CA, USA
[2]University at Buffalo, SUNY, NY, USA
[3]Carnegie Mellon University, PA, USA
[4]Clemson University, SC, USA
{seokmin.choi,sajad.mousavi,phillip.si}@cardiophi.com
{haben.yhdego,fatemeh.khadem}@cardiophi.com
fatemeh.afghah@clemson.edu

) Jun 2023

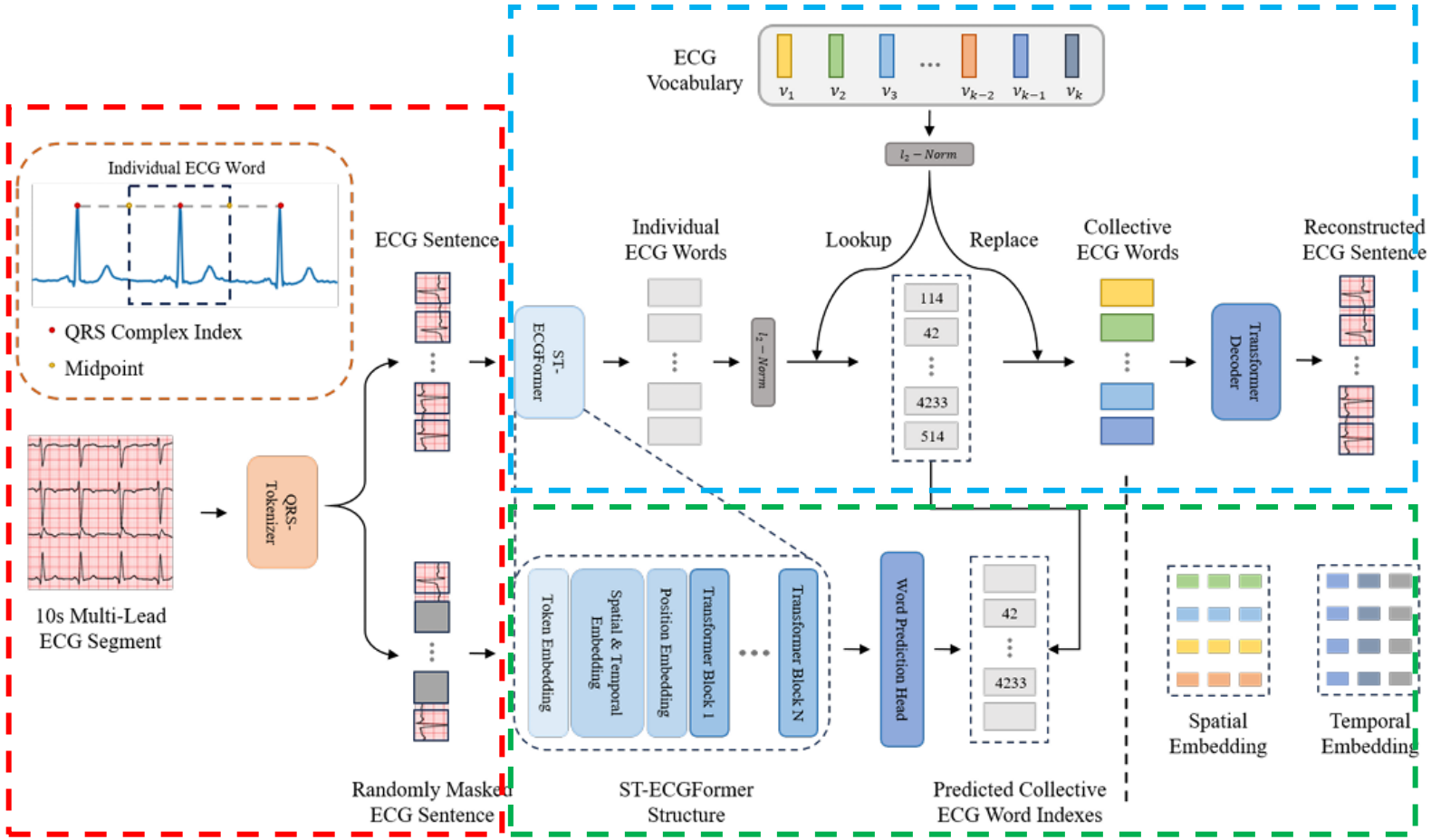
- Heartbeats have clear semantics in the time
- Only two papers in ECG language processing
- The vocabulary is constructed **based on the waves** in the heartbeat and **without contextual information**
- Vocabulary does not exceed 70 words (clusters)

[1] Mousavi, Sajad, et al. "ECG Language processing (ELP): A new technique to analyze ECG signals." *Computer methods and programs in biomedicine* 202 (2021): 105959.
[2] Choi, Seokmin, et al. "ECGBERT: Understanding hidden language of ECGs with self-supervised representation learning." *arXiv preprint arXiv:2306.06340* (2023).

Method

Step 1: Generating ECG Sentences Using the QRS-Tokenizer

Step 2: Vector-Quantized Heartbeat Reconstruction Training

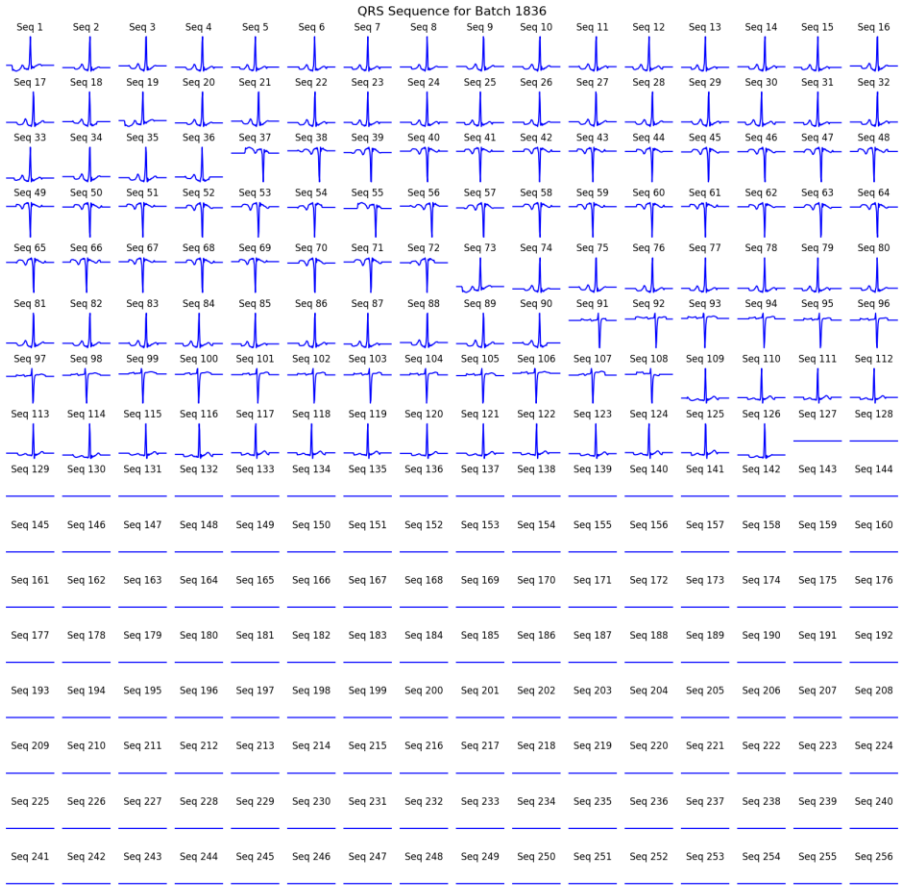


Step 3: Masked ECG Sentence Pre-Training

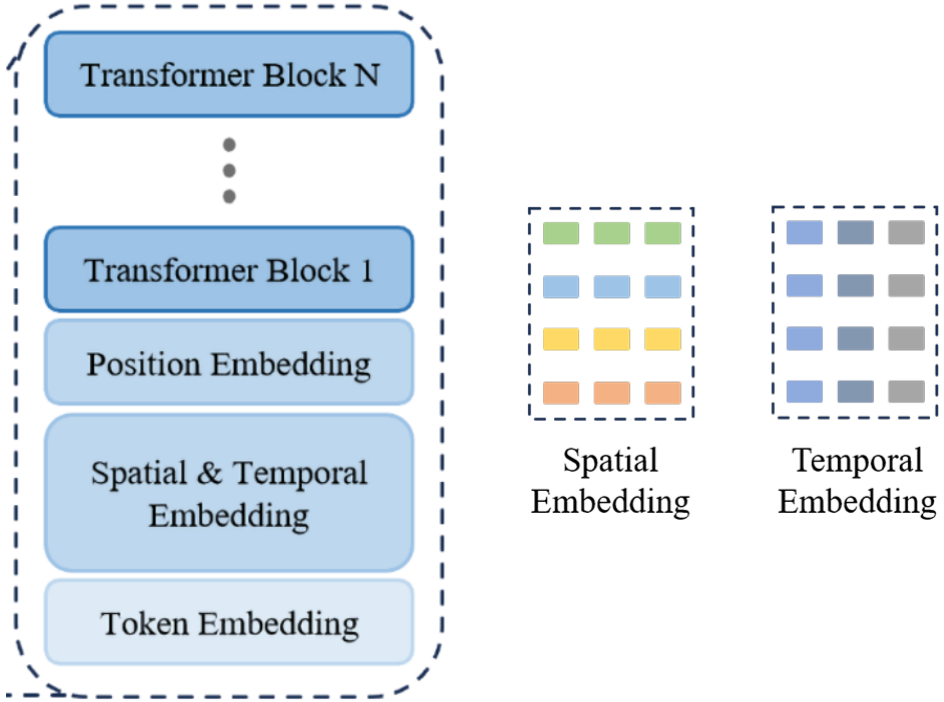
Figure 2: Framework of HeartLang.

Generating ECG Sentences Using the QRS-Tokenizer

- Step 1: **QRS Detection**
- Step 2: **Generating ECG Sentences**



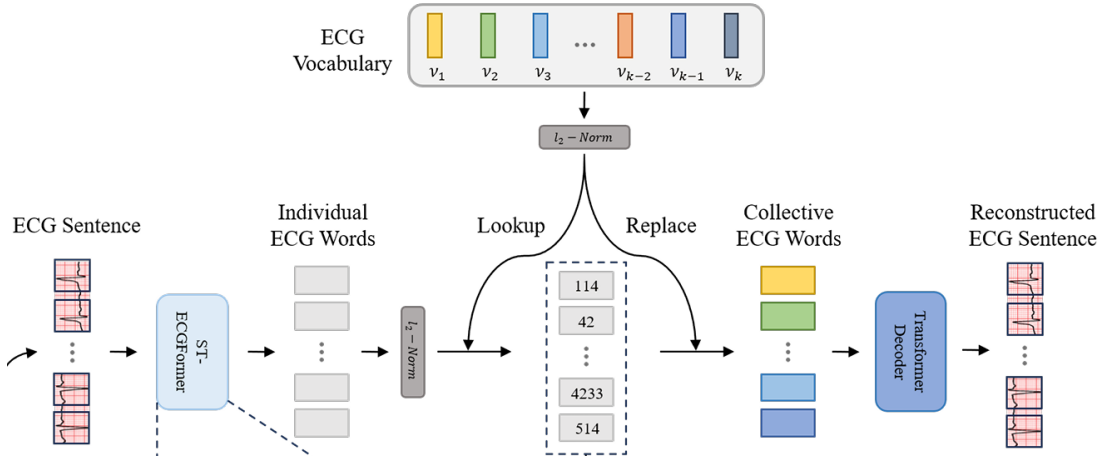
ST-ECGFormer Backbone Network



- **Spatial Embedding**: Assign an index to each individual ECG word based on its lead position
- **Temporal Embedding**: Divide the original 10s signal into 10 intervals. Assign each individual ECG word an index based on the interval position of its QRS index.

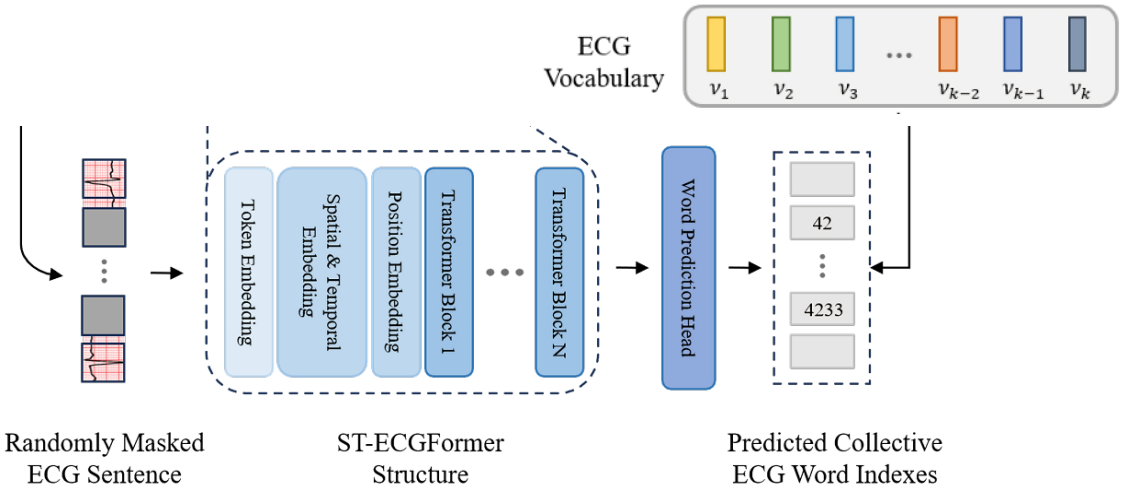
Vector-Quantized Heartbeat Reconstruction Training

- Step 1: **Vector Quantization**
- Step 2: **Heartbeat Reconstruction**



Masked ECG Sentence Pre-Training

- Step 1: **Individual ECG Words Masking**
- Step 2: **Collective ECG Words Prediction**



Experiments

4.1 VQ-HBR Training and Pre-training Configuration

MIMIC-IV-ECG. This publicly accessible dataset (Gow et al., 2023) contains 800,035 12-lead ECG recordings from 161,352 subjects. Each ECG recording was sampled at 500 Hz and lasted for 10 seconds. To prepare the pretraining dataset, we replaced the "NaN" and "Inf" values in the ECG recordings with the average of six neighboring points.

Implementation. Before VQ-HBR training and pre-training stage, we first downsampled all records in the dataset to 100 Hz and used the QRS-Tokenizer to transform the raw ECG recordings into a unified ECG sentence. We split the training and validation sets into 9:1, with the validation set data used for VQ-HBR training. In the VQ-HBR training stage, we set the learning rate to 5×10^{-5} and trained for 100 epochs, with an ECG vocabulary size of 8,192 and a collective ECG word dimension of 128. In the pre-training stage, we set the learning rate to 5×10^{-4}, trained for 200 epochs, and applied a random masking rate of 50%. For both stages, a randomly initialized ST-ECGFormer was used as the backbone network, the AdamW optimizer was selected, and cosine annealing was applied for learning rate scheduling. All experiments were conducted on 8 NVIDIA GeForce RTX 4090 GPUs, with a batch size of 64 per GPU. We set the random seed to 0 to ensure the reproducibility of all results. More experimental details are provided in the *appendix*.

4.2 Downstream Tasks Configuration

We evaluated our method on the three widely used public datasets listed below, which cover over 100 types of cardiac conditions. Detailed information on the data split can be found in the appendix.

PTB-XL. This publicly accessible dataset (Wagner et al., 2020) contains 21,837 12-lead ECG recordings collected from 18,885 patients. Each ECG recording was sampled at 500 Hz and lasted for 10 seconds. Based on the SCP-ECG protocol, the multi-class classification task has four subsets: **Superclass** (5 classes), **Subclass** (23 classes), **Form** (19 classes), and **Rhythm** (12 classes). We followed the official data split (Strodthoff et al., 2021) for training, validation, and testing.

CPSC2018. This publicly accessible dataset (Liu et al., 2018) contains 6,877 12-lead ECG recordings. Each recording was sampled at 500 Hz, with durations ranging from 6 to 60 seconds. The dataset is annotated with 9 different labels. We split the dataset into 70%:10%:20% for training, validation, and testing.

Chapman-Shaoxing-Ningbo (CSN). This publicly accessible dataset (Zheng et al., 2020; 2022) contains 45,152 12-lead ECG recordings. Each ECG recording was sampled at 500 Hz and lasted for 10 seconds. Following the configuration provided by MERL, we removed ECG records with "unknown" annotations. The refined version of the dataset contains 23,026 ECG recordings with 38 distinct labels. We split the dataset into 70%, 10%, 20% for training, validation, and testing.

Implementation. Before fine-tuning in downstream tasks, we first downsampled all records in the dataset to 100 Hz and used the QRS-Tokenizer to transform the raw ECG recordings into a unified ECG sentence. For linear probing, we kept the ST-ECGFormer backbone network frozen and only trained the randomly initialized parameters of the linear classifier. To explore the performance of our method under low-resource conditions, we conducted linear probing using 1%, 10%, and 100% of the training data for each task. We set the learning rate to 5×10^{-3} and trained for 100 epochs. For the CPSC2018 and CSN datasets, we scaled the ECG recordings to the range of $[-3, 3]$ to enhance QRS detection. All test results were obtained from the best validation model, rather than testing the model on the test set after each epoch and reporting the highest result. For all downstream tasks, we used the macro AUC as the evaluation metric. We set the random seed to 0 to ensure the reproducibility of all results. More experimental details are provided in the *appendix*.

Results and Discussions

Table 1: Linear probing results of HeartLang and other eSSL methods. The best results are **bolded**, with gray indicating the second highest.

Method	PTBXL-Super			PTBXL-Sub			PTBXL-Form			PTBXL-Rhythm			CPSC2018			CSN		
	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%
SimCLR (Chen et al., 2020)	63.41	69.77	73.53	60.84	68.27	73.39	54.98	56.97	62.52	51.41	69.44	77.73	59.78	68.52	76.54	59.02	67.26	73.20
BYOL (Grill et al., 2020)	71.70	73.83	76.45	57.16	67.44	71.64	48.73	61.63	70.82	41.99	74.40	77.17	60.88	74.42	78.75	54.20	71.92	74.69
BarlowTwins (Zbontar et al., 2021)	72.87	75.96	78.41	62.57	70.84	74.34	52.12	60.39	66.14	50.12	73.54	77.62	55.12	72.75	78.39	**60.72**	71.64	77.43
MoCo-v3 (Chen et al., 2021)	73.19	76.65	78.26	55.88	69.21	76.69	50.32	63.71	71.31	51.38	71.66	74.33	**62.13**	76.74	75.29	54.61	**74.26**	77.68
SimSiam (Chen & He, 2021)	73.15	72.70	75.63	62.52	69.31	76.38	55.16	62.91	71.31	49.30	69.47	75.92	58.35	72.89	75.31	58.25	68.61	77.41
TS-TCC (Eldele et al., 2021)	70.73	75.88	78.91	53.54	66.98	77.87	48.04	61.79	71.18	43.34	69.48	78.23	57.07	73.62	78.72	55.26	68.48	76.79
CLOCS (Kiyasseh et al., 2021)	68.94	73.36	76.31	57.94	72.55	76.24	51.97	57.96	72.65	47.19	71.88	76.31	59.59	**77.78**	77.49	54.38	71.93	76.13
ASTCL (Wang et al., 2024)	72.51	77.31	81.02	61.86	68.77	76.51	44.14	60.93	66.99	52.38	71.98	76.05	57.90	77.01	79.51	56.40	70.87	75.79
CRT (Zhang et al., 2023c)	69.68	78.24	77.24	61.98	70.82	78.67	46.41	59.49	68.73	47.44	73.52	74.41	58.01	76.43	**82.03**	56.21	73.70	78.80
ST-MEM (Na et al., 2024)	61.12	66.87	71.36	54.12	57.86	63.59	55.71	59.99	66.07	51.12	65.44	74.85	56.69	63.32	70.39	59.77	66.87	71.36
HeartLang (Ours)	**78.94**	**85.59**	**87.52**	**64.68**	**79.34**	**88.91**	**58.70**	**63.99**	**80.23**	**62.08**	**76.22**	**90.34**	60.44	66.26	77.87	57.94	68.93	**82.49**

Evaluation on Signal Slicing Perspective

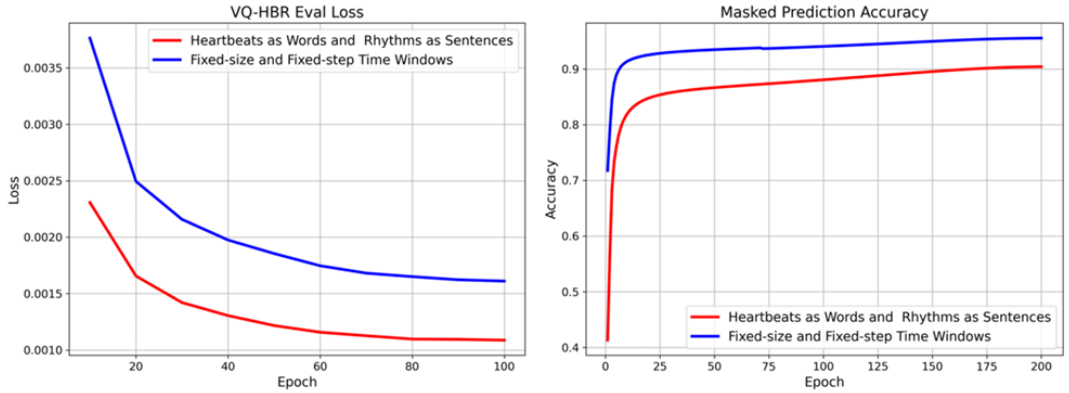


Figure 3: The validation loss curve during VQ-HBR training (left) and the prediction accuracy curve during masked ECG sentence pre-training (right), shown from two perspectives.

Table 2: Linear probing results of two signal slicing perspectives. The best results are **bolded**, while improved values are marked in **green**, and decreased values are marked in **red**.

Perspective	PTBXL-Super			PTBXL-Sub			PTBXL-Form			PTBXL-Rhythm		
	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%
Fixed-size and Fixed-step Time Windows	66.41	77.89	80.51	60.87	70.72	78.32	**59.94**	**65.52**	76.03	55.83	74.09	86.05
Heartbeats as Words and Rhythms as Sentences	**78.94**	**85.59**	**87.52**	**64.68**	**79.34**	**88.91**	58.70	63.99	**80.23**	**62.08**	**76.22**	**90.34**
Improvement	**12.52**	**7.71**	**7.01**	**3.81**	**8.62**	**10.58**	**-1.24**	**-1.53**	**4.20**	**6.26**	**2.14**	**4.29**

ECG Vocabulary Visualization

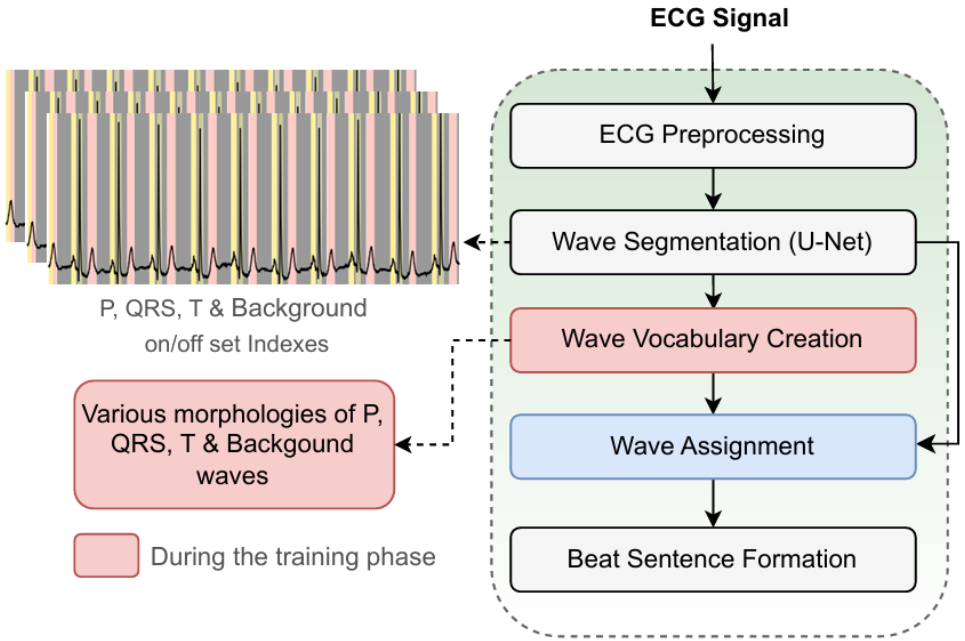


Figure 1: Holistic view of ECG language processing (ELP)

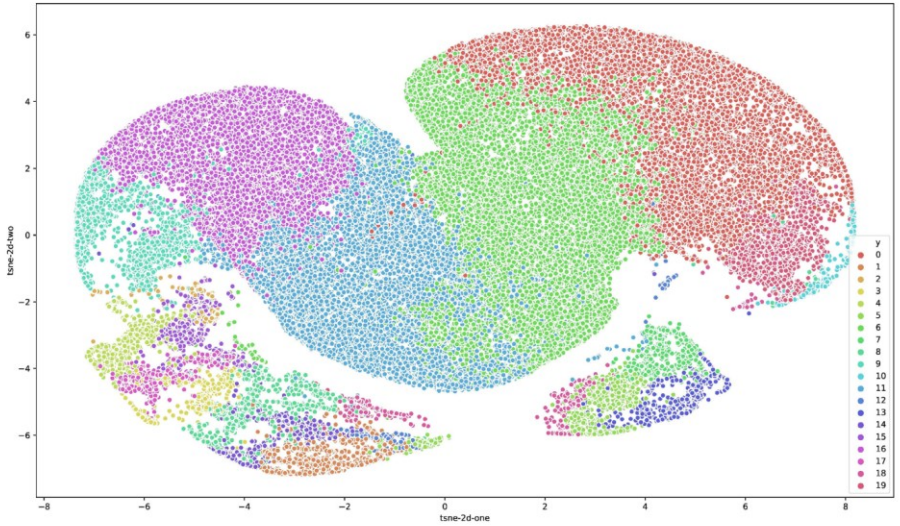
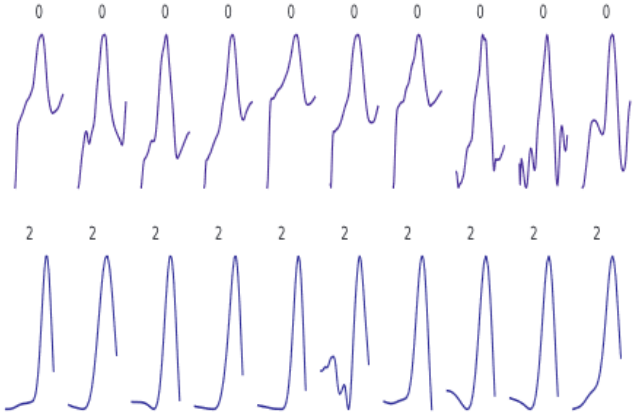


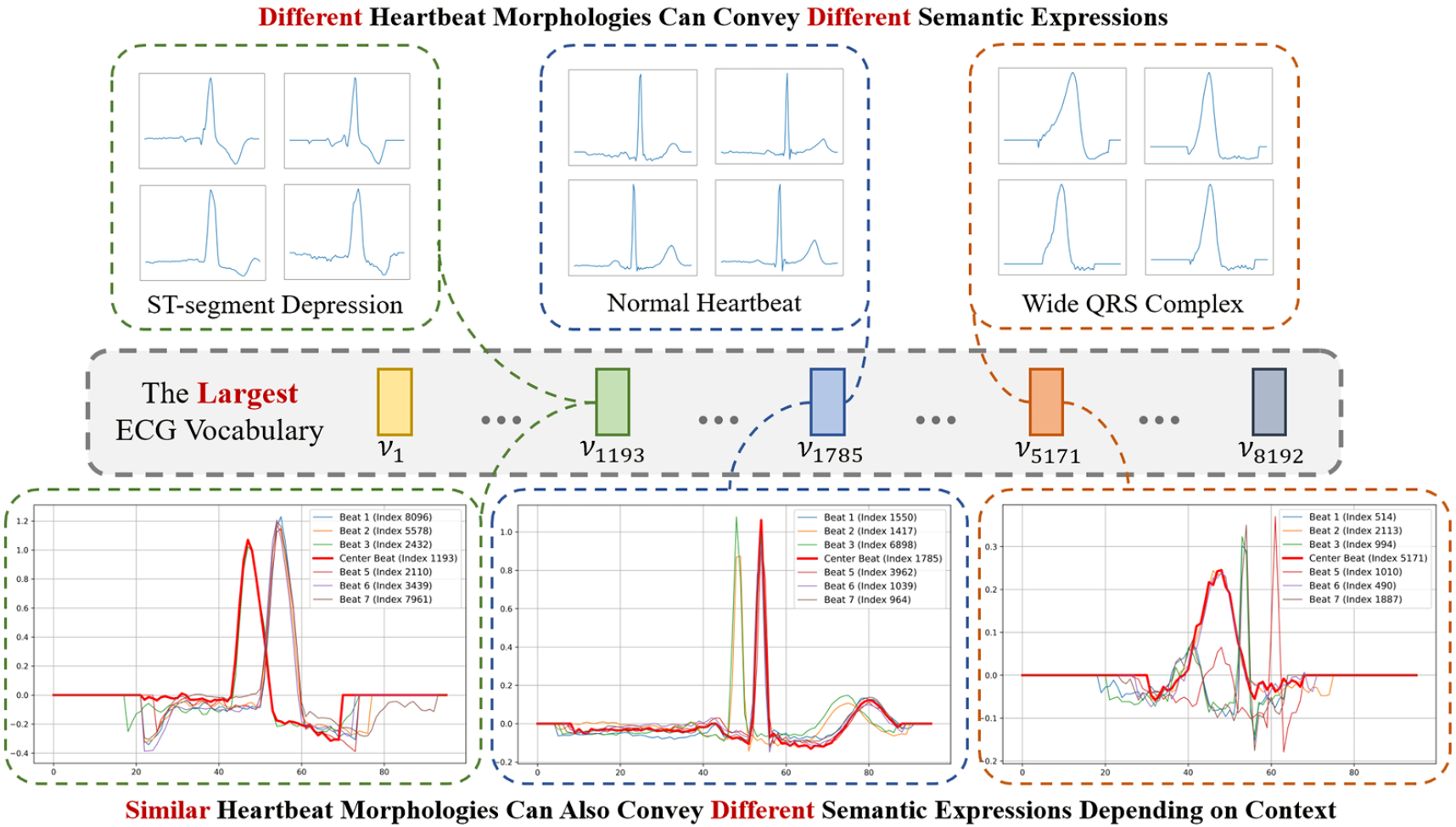
Fig. 4. Visualizations of thousands of extracted waves along with their 20 clusters on the 2017 PhysioNet/CinC Challenge database. The K-means clustering algorithm, $K = 20$ as the number of clusters, is used to cluster the extracted waves from the ECG signals and t-SNE technique is used to show how waves are presented in a high-dimensional space.

hah, F. Khadem et al.



[1] Mousavi, Sajad, et al. "ECG Language processing (ELP): A new technique to analyze ECG signals." *Computer methods and programs in biomedicine* 202 (2021): 105959.
[2] Choi, Seokmin, et al. "ECGBERT: Understanding hidden language of ECGs with self-supervised representation learning." *arXiv preprint arXiv:2306.06340* (2023).

ECG Vocabulary Visualization



Ablation Study

Table 3: Linear probing results of the ablation study. The best results are **bolded**, with gray indicating the second highest.

Method	PTBXL-Super			PTBXL-Sub			PTBXL-Form			PTBXL-Rhythm		
	1%	10%	100%	1%	10%	100%	1%	10%	100%	1%	10%	100%
w/o ECG Vocabulary	59.23	61.51	79.68	49.34	50.55	73.16	49.33	46.88	63.25	42.87	43.87	77.65
w/o Pre-training	70.34	77.82	80.31	55.70	67.64	80.57	53.73	57.67	66.70	**65.98**	76.49	81.40
w/o Spatio-temporal Embedding	69.78	81.79	85.12	58.06	73.76	87.33	55.09	63.37	73.90	61.79	74.83	84.93
w/o Spatial Embedding	78.74	85.32	86.87	**67.82**	79.31	88.66	60.76	**67.66**	79.10	63.54	79.00	89.25
w/o Temporal Embedding	77.87	84.86	85.85	64.56	77.48	87.50	**61.41**	67.54	77.50	65.44	**84.74**	87.81
HeartLang (Ours)	**78.94**	**85.59**	**87.52**	64.68	**79.34**	**88.91**	58.70	63.99	**80.23**	62.08	76.22	**90.34**

Conclusion

- **New ECG Signal Processing Perspective**: heartbeats serve as words and rhythms as sentences

- **Introduction of QRS-Tokenizer**: segment ECG signals into semantically meaningful ECG sentences

- **Development of HeartLang**: learn form-level representations through vector-quantized heartbeat reconstruction and rhythm-level representations via masked ECG sentence pre-training.

- **Creation of the Largest ECG Vocabulary**: capturing a wide range of heartbeat morphological patterns across diverse cardiac conditions to enhance ECG language processing.

- **Performance Evaluation**: HeartLang is evaluated on six public ECG datasets, showing strong performance against other self-supervised learning (SSL) methods.

- **Goal and Impact**: The study aims to inspire ECG research, particularly in ECG language processing, by introducing this innovative perspective and framework.

Thanks for listening