LREC-COLING 2024

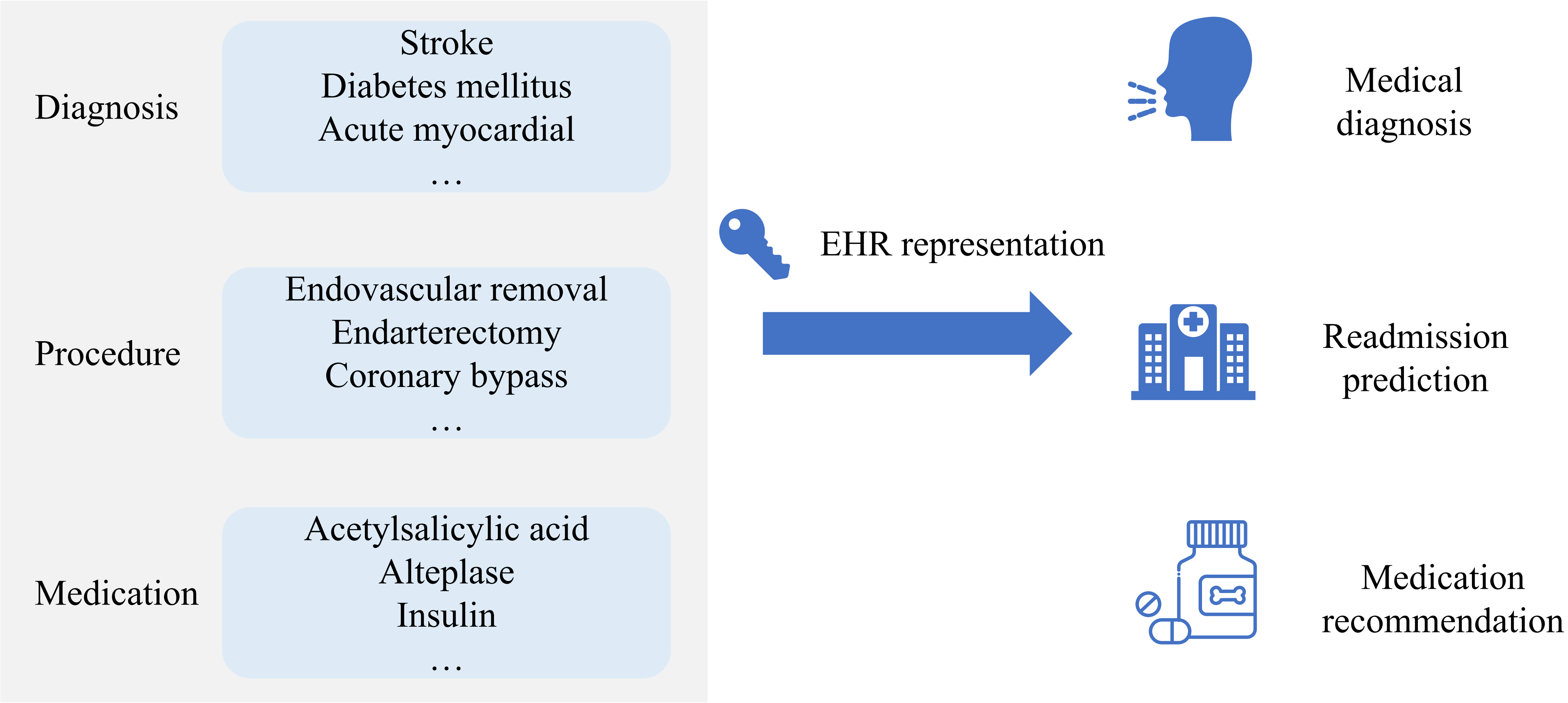
MHGRL: An Effective Representation Learning Model for Electronic Health Records

Feiyan Liu[1], Liangzhi Li[2], Xiaoli Wang[1], Chang Liu[1], Feng Luo[3], Jinsong Su[1], Yiming Qian[4]

[1]Xiamen University, [2]Meetyou AI Lab, [3]RMIT University,

[4]Institute of High Performance Computing

Background: EHR



An example of Electronic Health Record (EHR)

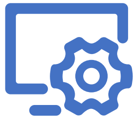
Downstream tasks

Challenge: Data Insufficiency

LREC-COLING 2024

- Data from a single healthcare system is often insufficient to train deep learning model

Privacy regulations

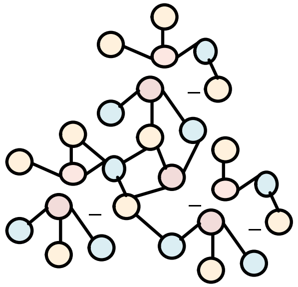


Incompatibility

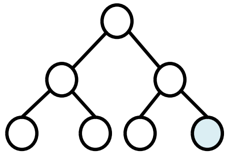
Rare diseases

Uncommon conditions (e.g ICU)

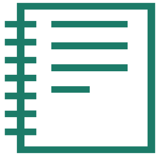
- Previous studies fail to exploit external information fully, typically focusing on a single type of data.



Medical knowledge graph



Ontology

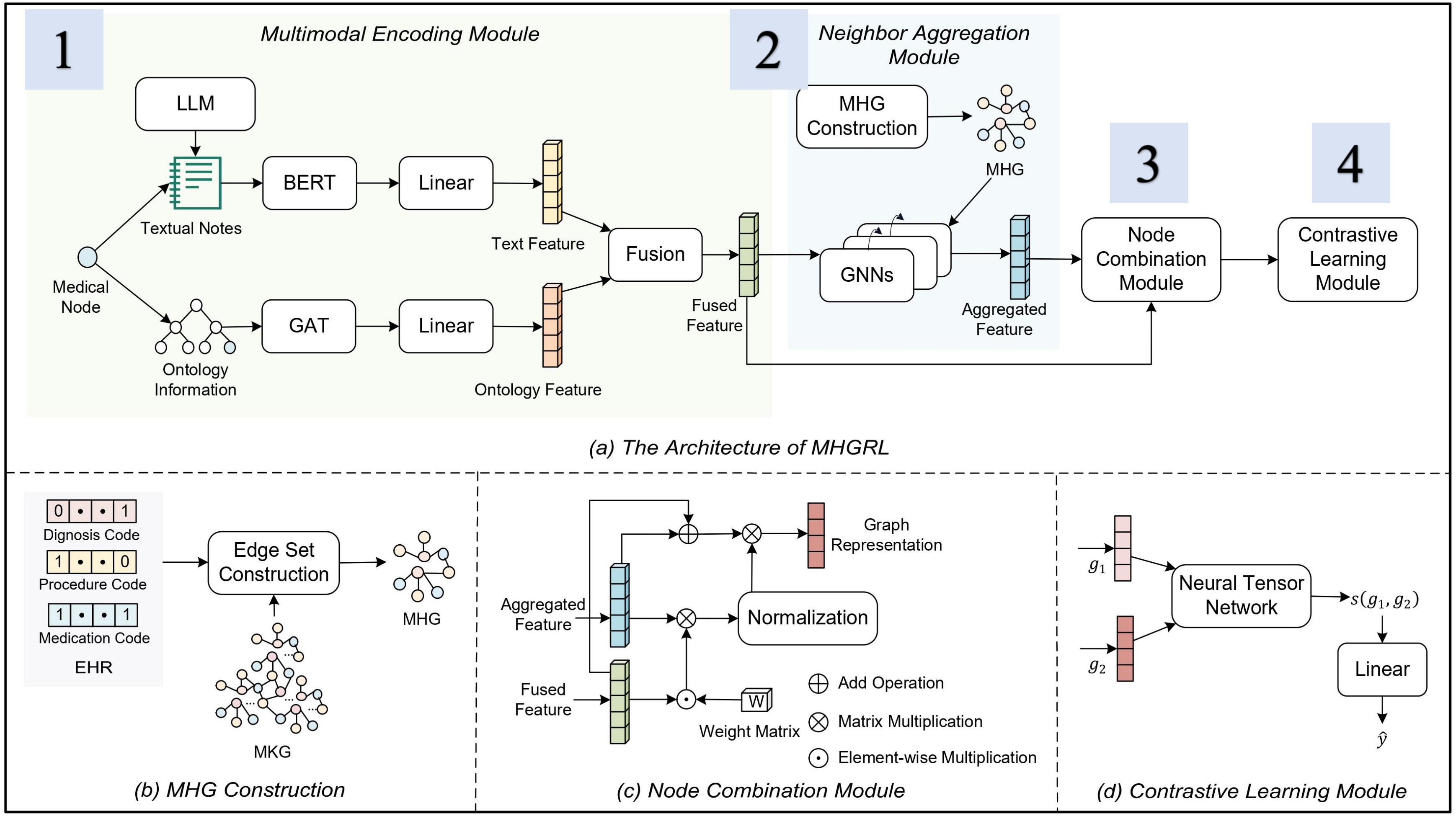


Textual notes

Methodology

LREC-COLING 2024

Multimodal Heterogeneous Graph-enhanced Representation Learning (MHGRL)



1. Multimodal Encoding
2. Neighbor Aggregation
3. Node Combination
4. Contrastive Learning

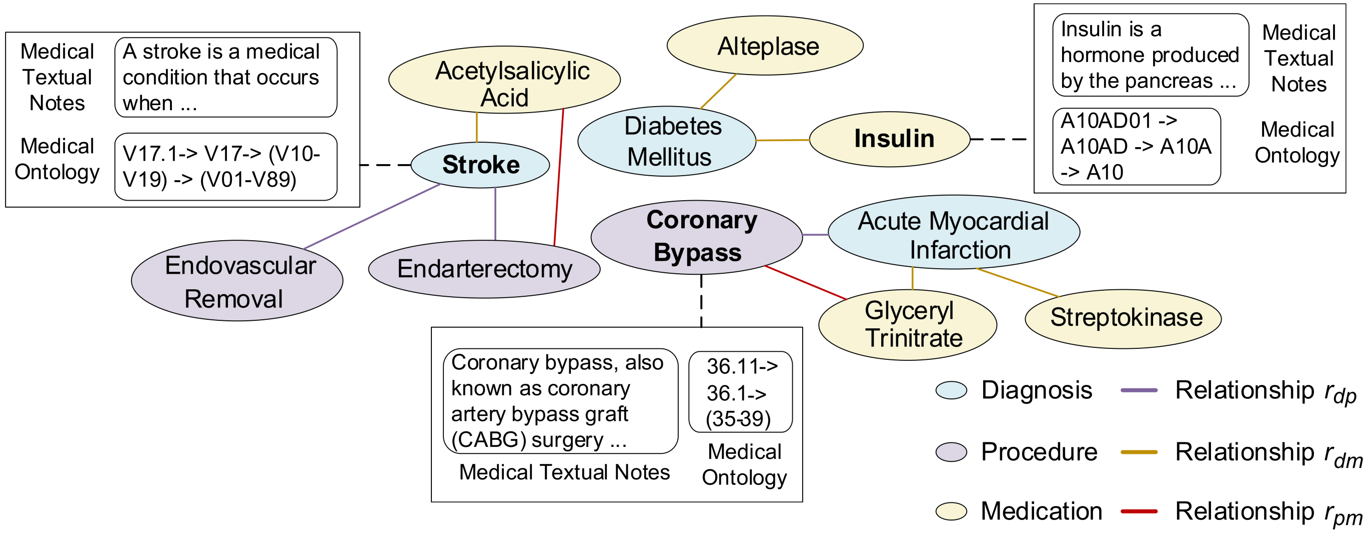
Methodology (con't)

Multimodal Heterogeneous Graph (MHG) Construction

Unordered sets of features → A graph considering internal structure and external knowledge

Diagnosis	Stroke Diabetes mellitus Acute myocardial …
Procedure	Endovascular removal Endarterectomy Coronary bypass …
Medication	Acetylsalicylic acid Alteplase Insulin …

An example of EHR



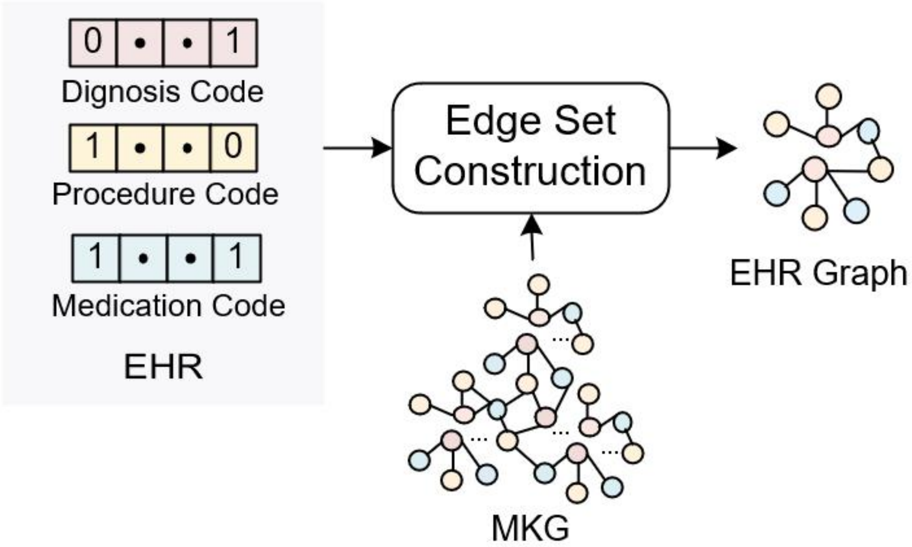
An example of MHG

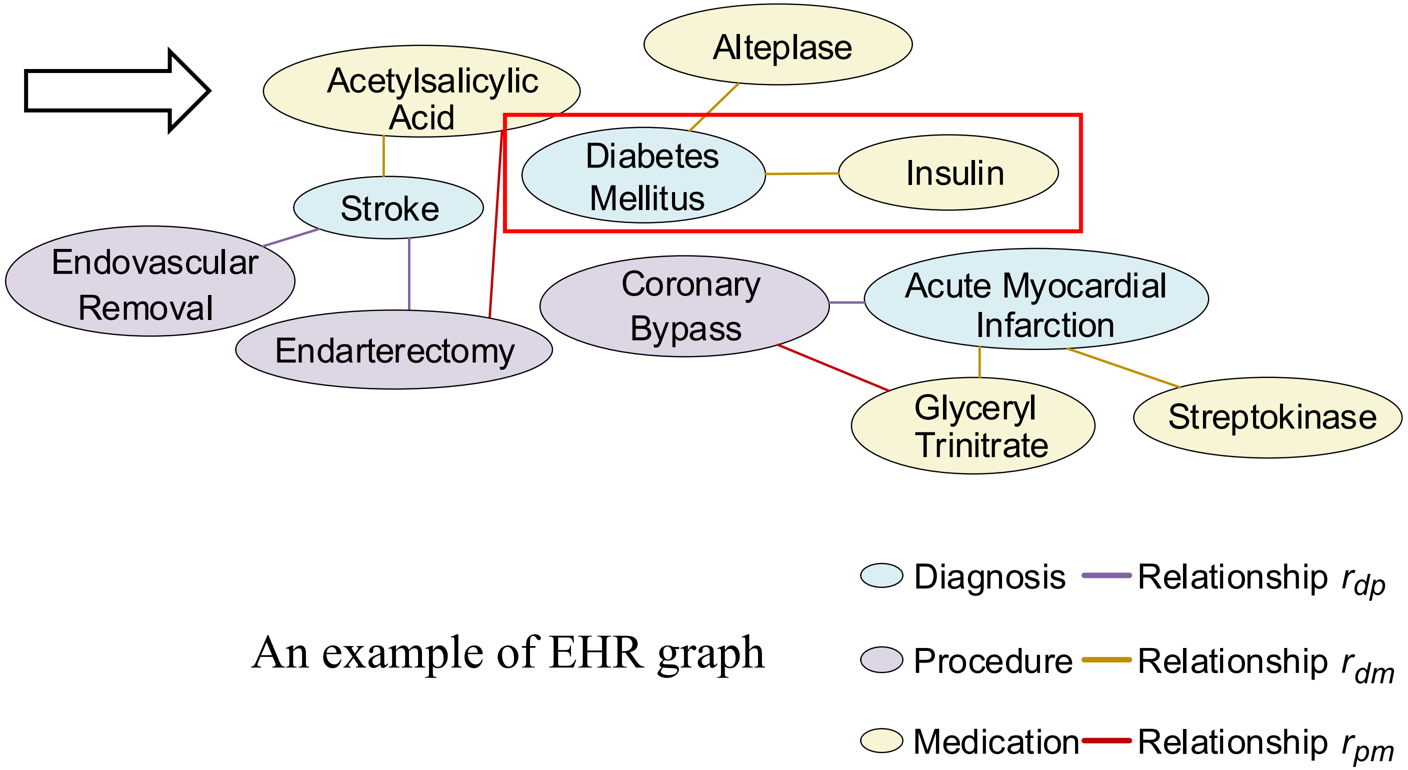
Methodology (con't)

LREC-COLING 2024

Multimodal Heterogeneous Graph (MHG) Construction

Step 1. Given an EHR, we employ the medical knowledge graph (MKG) to construct EHR graph



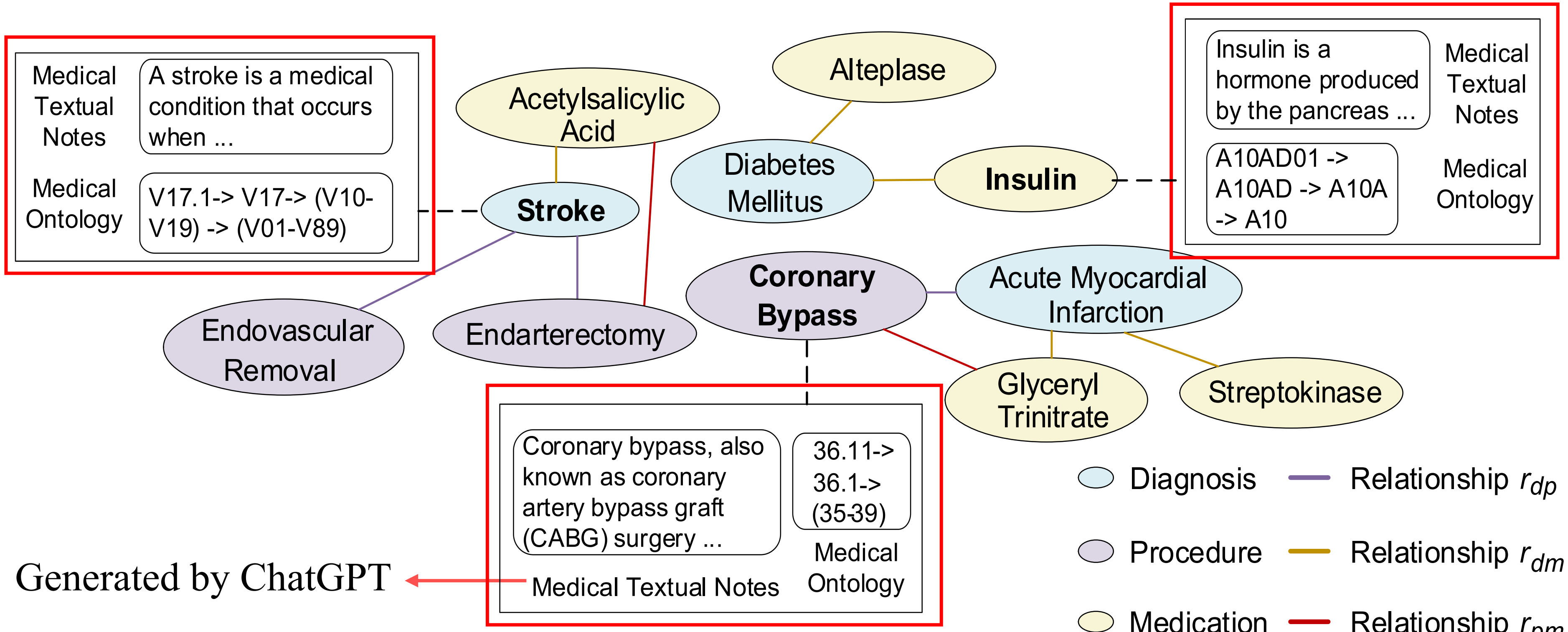


An example of EHR graph

Methodology (con't)

Multimodal Heterogeneous Graph (MHG) Construction

Step 2. Add multimodal information (ontology & textual notes) for each node



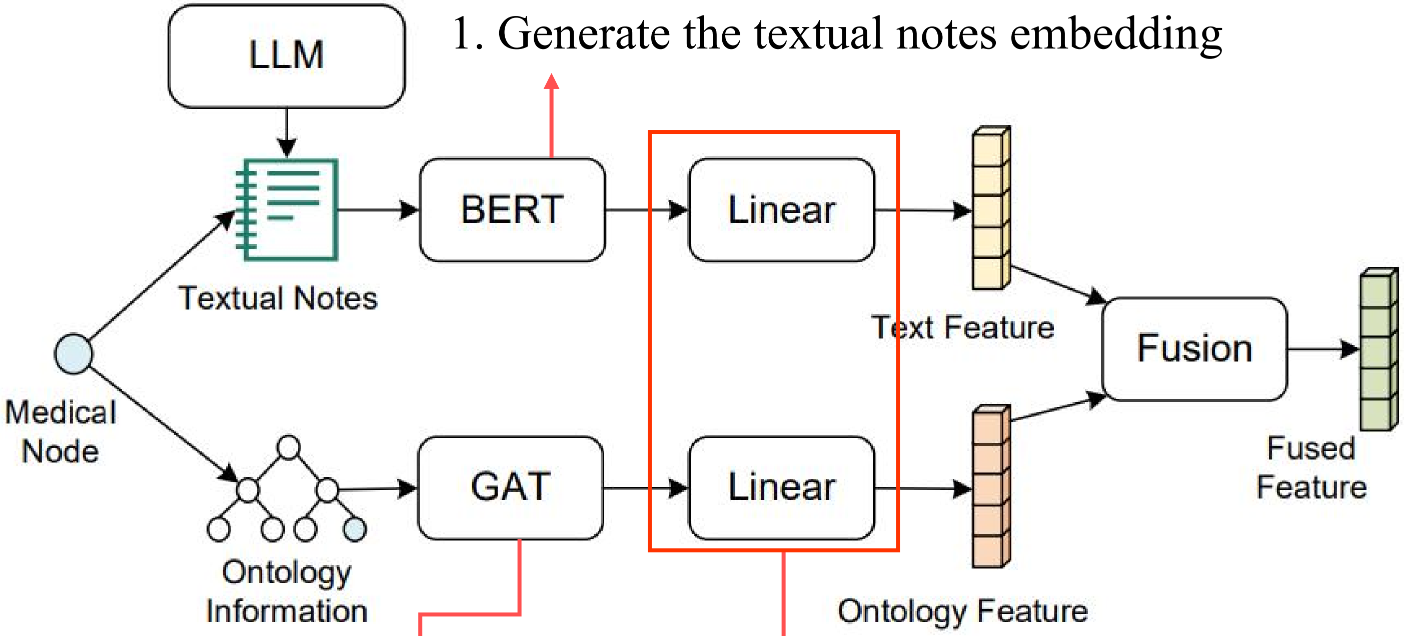
An example of MHG

Methodology (con't)

LREC-COLING 2024

1. Multimodal Encoding Module

Generate a multimodal embedding for each node in MHG



2. Capture the structural information

$$e'_i = \sum_{j \in ch(i)} \alpha_{ij} W e_j \qquad \alpha_{ij} = \frac{\exp(LReL(a^T[We_i || We_j]))}{\sum_{k \in ch(i)} \exp(LReL(a^T[We_i || We_k]))}$$

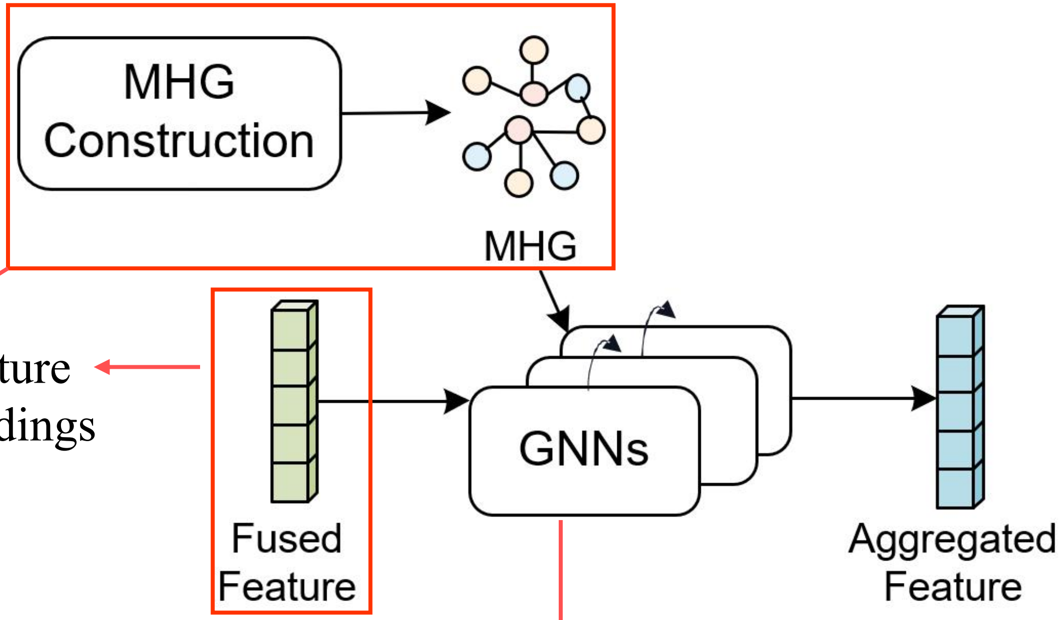
3. Normalize two embeddings into a shared space

$$o'_n = f(o_n, W_2, b_2) \qquad t'_n = f(t_n, W_1, b_1)$$

Methodology (con’t)

2. Neighbor Aggregation Module

Get high-order neighborhood information



1. Input the graph structure and multimodal embeddings

2. Get neighborhood information with L layers of GNNs

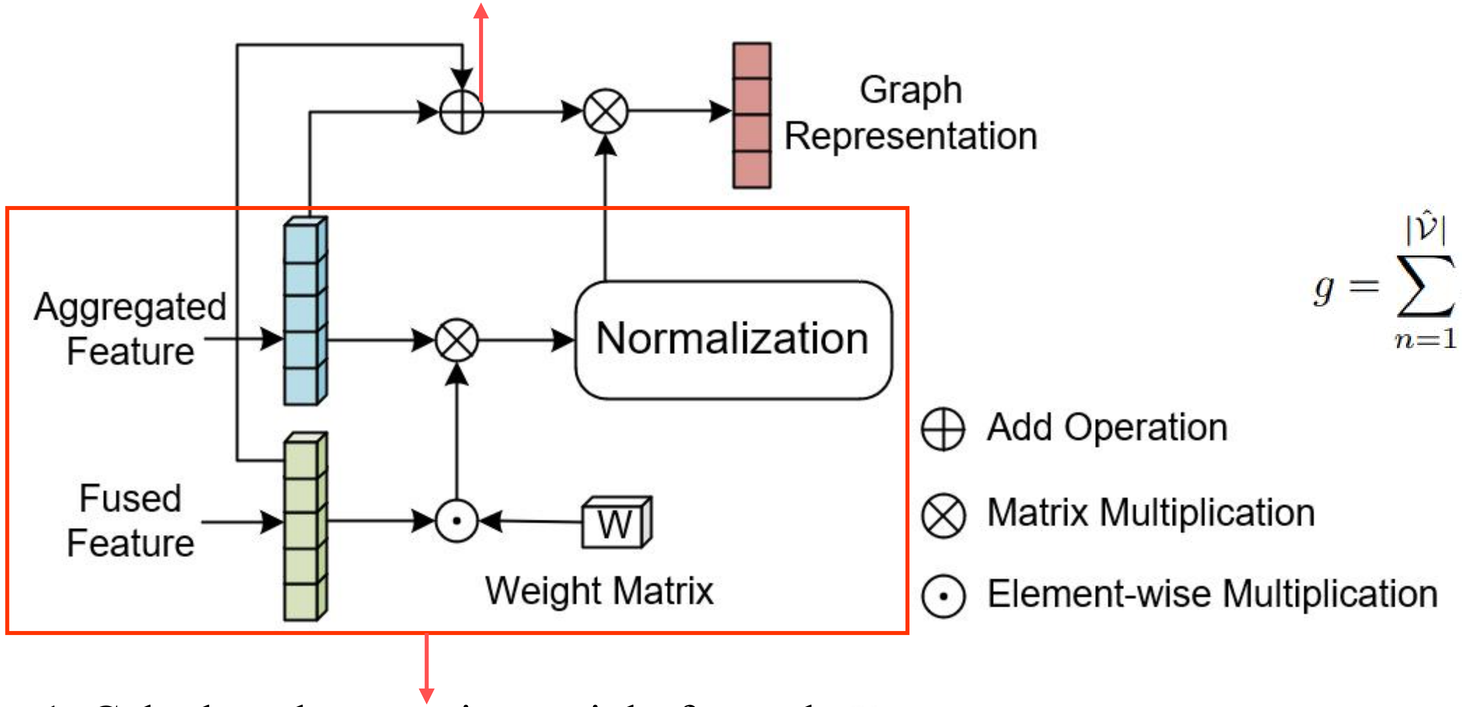
$$H^{(l)} = \sigma(UpdateFunction(H^{(l-1)}, A))$$

Methodology (con't)

3. Node Combination Module

Design an attention mechanism to aggregate node embeddings into a graph representation

2. Sum the initial multimodal embedding and the aggregated embedding



$$g = \sum_{n=1}^{|\hat{\mathcal{V}}|} \underbrace{sigmoid((m_n \odot w)^T h_n)}_{\text{Step 1}} \underbrace{(m_n + h_n)}_{\text{Step 2}}$$

1. Calculate the attention weight for node v_n

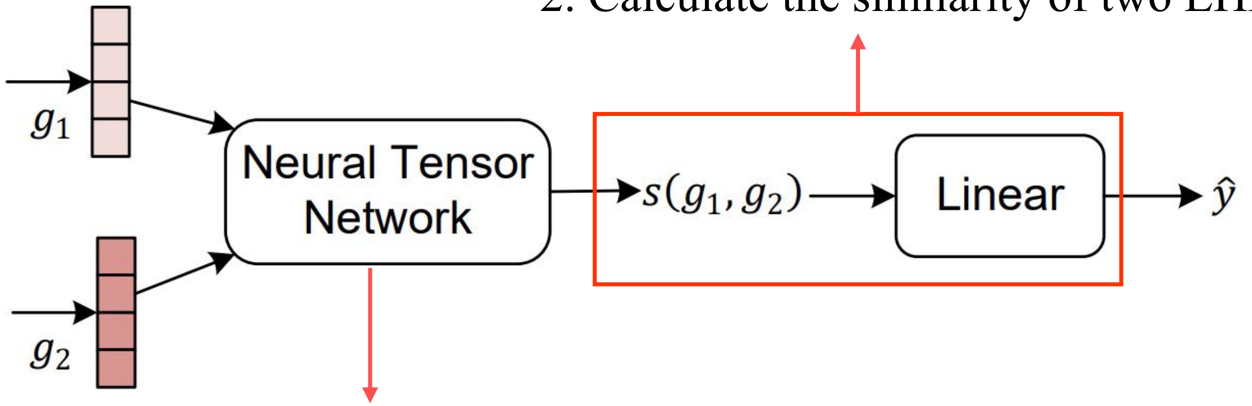
Methodology (con't)

LREC-COLING 2024

4. Contrastive Learning Module

Implement contrastive learning to ensure consistency in representations among similar EHRs

2. Calculate the similarity of two EHRs



1. Capture the interaction signals

$$s(g_1, g_2) = \sigma(g_1{}^T W_3 g_2 + W_4 \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + b_3)$$

Loss function

$$\mathcal{L}(y, \hat{y}) = -[ylog(\hat{y}) + (1 - y)log(1 - \hat{y})]$$

If two EHRs are in the same cohort, label y=1, otherwise y=0

Experiments

LREC-COLING 2024

EHR clustering: perform the k-means clustering based on EHR representations

Each EHR has a label that indicates the cohort to which it belongs.

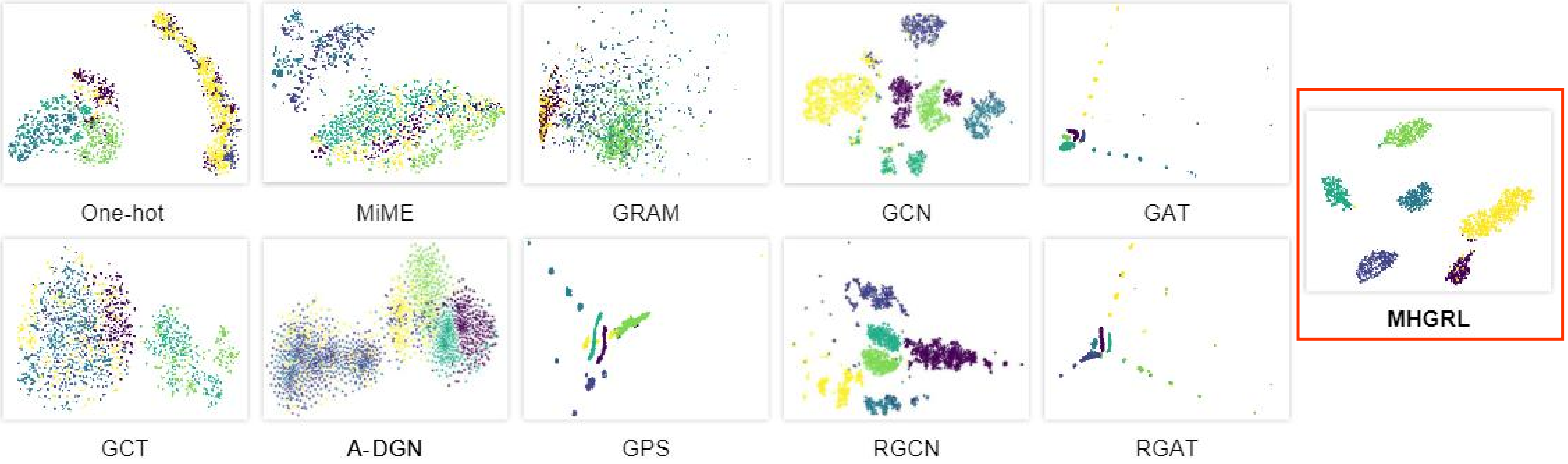
Category	Model	MIMIC-III			MIMIC-IV		
		Purity	NMI	RI	Purity	NMI	RI
Static	One-hot	0.6052	0.5023	0.7895	0.6093	0.4840	0.7587
RL	MiME	0.4368	0.2219	0.7041	0.5666	0.3773	0.7346
	GRAM	0.4840	0.3315	0.7589	0.6050	0.4517	0.7615
	GCT	0.6224	0.5137	0.8102	0.7473	0.5789	0.8142
GNN-based	GCN	0.9265	0.8243	0.9489	0.7464	0.6208	0.8738
	GAT	0.7957	0.7345	0.8753	0.7935	0.6905	0.8901
	RGCN	0.9472	0.8694	0.9590	0.7799	0.6744	0.8866
	RGAT	0.9170	0.8373	0.9331	0.7823	0.6711	0.8817
	A-DGN	0.6285	0.5101	0.8015	0.7410	0.6079	0.8681
	GPS	0.9334	0.8455	0.9530	0.8416	0.7301	**0.9146**
	MHGRL	**0.9761**	**0.9390**	**0.9811**	**0.8455**	**0.7515**	0.9141

Experiments (con’t)

LREC-COLING 2024

EHR clustering: perform the k-means clustering based on EHR representations

We use t-SNE to visualize the high-dimensional EHR representations in the test set on MIMIC-III.



Experiments (con't)

Disease prediction: predict the cohort to which each EHR belongs

1. Retrieve the *K* most similar EHRs from the training set.
2. The final predicted disease label is determined through a voting mechanism.

Category	Model	MIMIC-III			MIMIC-IV		
		K=1	K=3	K=5	K=1	K=3	K=5
Static	One-hot	0.6524	0.6732	0.6795	0.4315	0.5010	0.5272
RL	GRAM	0.5368	0.5493	0.5644	0.4898	0.5078	0.5292
GNN-based	GCN	0.6952	0.7165	0.7197	0.4820	0.5194	0.5408
	GAT	<u>0.7033</u>	0.7285	0.7234	0.4913	0.5287	0.5603
	RGCN	0.6920	<u>0.7291</u>	<u>0.7454</u>	0.4810	0.5078	0.5384
	RGAT	0.7008	0.7285	0.7341	0.5005	0.5282	<u>0.5646</u>
	A-DGN	0.5632	0.5820	0.5833	0.4456	0.4830	0.5131
	GPS	0.6983	0.7040	0.7209	<u>0.5102</u>	<u>0.5292</u>	0.5539
LLM	ChatGPT	0.3206	0.3226	0.3116	0.0782	0.0779	0.0771
	MHGRL	**0.7376**	**0.7514**	**0.7657**	**0.5398**	**0.5661**	**0.5855**

Experiments (con't)

Ablation study on MIMIC-III

Model	EHR clustering			Disease prediction		
	Purity	NMI	RI	K=1	K=3	K=5
MHGRL	**0.9761**	**0.9390**	**0.9811**	**0.7376**	**0.7514**	**0.7657**
w/o multimodal information	0.6562	0.5590	0.8135	0.5931	0.6126	0.6010
w/o medical textual notes	0.9755	0.9343	0.9805	0.7319	0.7469	0.7531
w/o medical ontology	0.9617	0.9089	0.9719	0.7131	0.7236	0.7382
w/o attention mechanism	0.9679	0.9177	0.9753	0.6714	0.6942	0.7003
w/o contrastive learning	0.7142	0.6670	0.8564	0.7344	0.7502	0.7542

1. All modalities of data contribute to EHR representation learning.
2. Using the attention mechanism can further boost the performance.
3. Contrastive learning also contributes to performance.

Conclusion

- We construct a novel MHG to accurately model EHR.

- We propose an EHR representation learning model, MHGRL, to incorporate both the internal structure and external knowledge information.

- We conduct comprehensive experiments on two real-world EHR datasets, demonstrating that MHGRL can learn effective EHR representations.

- Future work

  Incorporate additional information into our model, such as medical images, demographic information and temporal clinical data.

Thanks for listening!

feiyanliu@stu.xmu.edu.cn