Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network

Xi Zhang, Jingyuan Chou, Fei Wang

Weill Cornell Medicine

Cornell University

Outline

- Introduction
- Method
- Experiments
- Summary

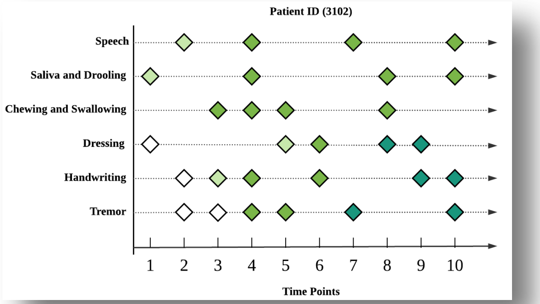
Outline

- Introduction
- Method
- Experiments
- Summary

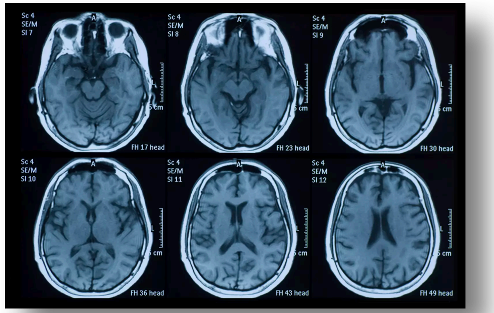
Why Integrative Analysis?

- **Background:** For complicated diseases such as Parkinson's and Alzheimer's, both ***<u>patients health records</u>*** and ***<u>neuroimaging</u>*** information are very important for disease understanding.

- **Goal:** Achieving superior classification performance on discriminating patients and controls, with an interpretable learning model based on ***<u>heterogeneous data structure</u>***.



Patients Health Records

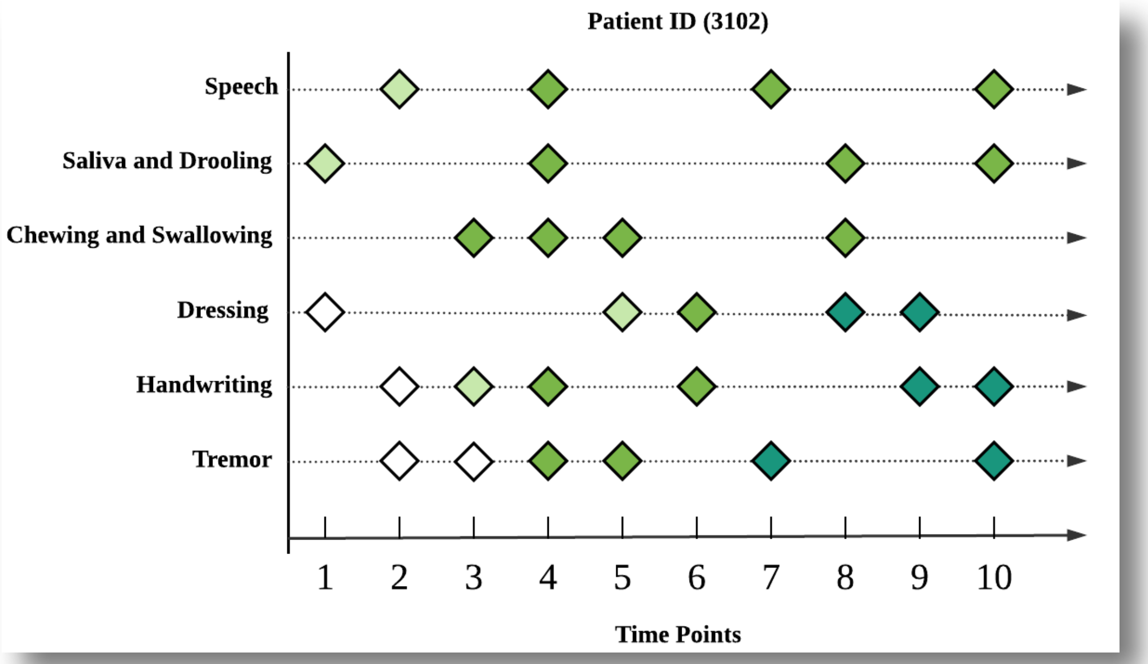


NeuroImages

Discriminating Patients and Health Controls

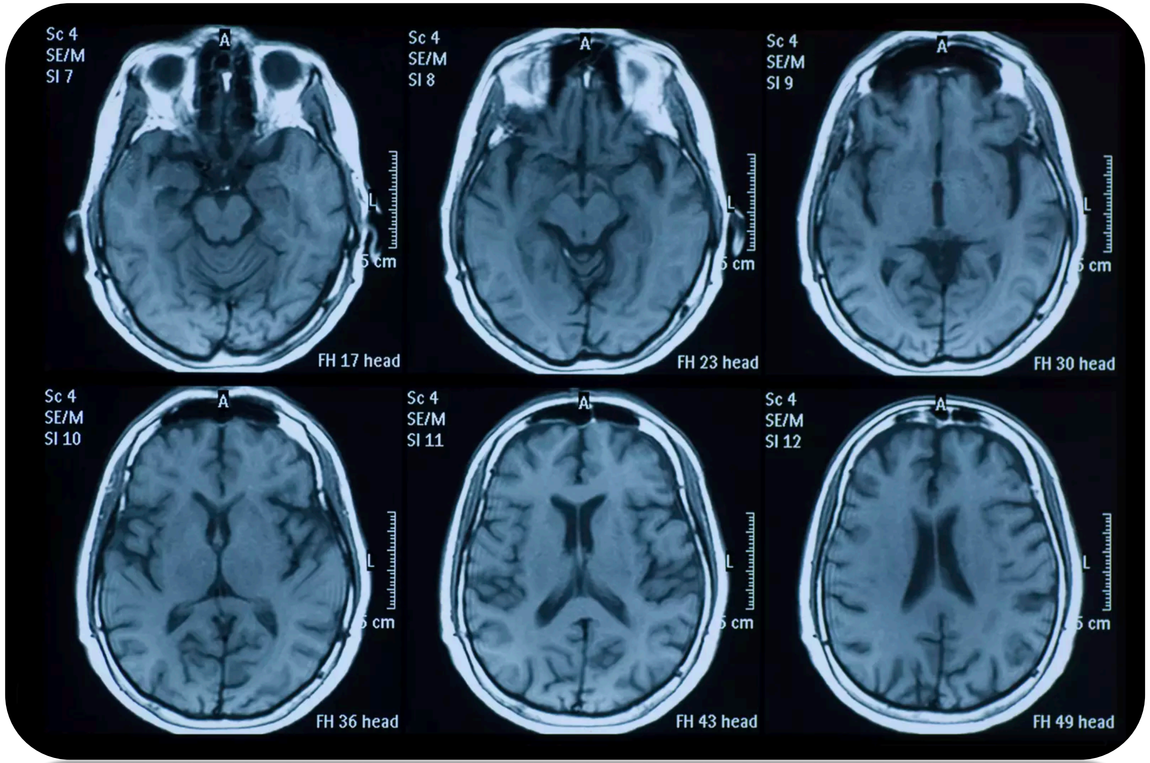
Modality-I: Patient Health Records

- A toy example of the patient health records for one patient

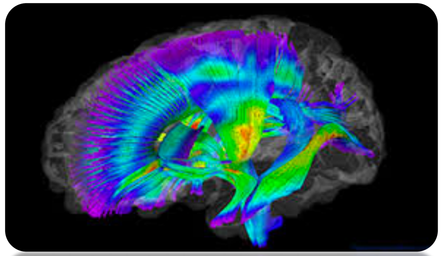


- ✓ Sequential structure
- ✓ Missing Values

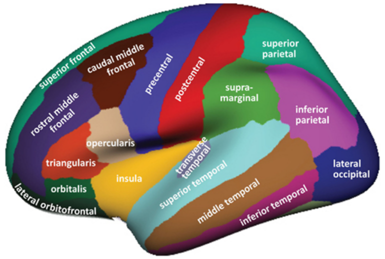
Modality-II: Neuroimages



http://time.com/2860630/mri-scans-can-detect-early-onset-of-parkinsons-study-finds/



DTI: Diffusion Tensor Imaging



ROI: Region of Interest

Desikan-Killiany 84

Challenges & Solution

Challenges

- ***Heterogeneity.*** The nature of EHR and neuroimage are completely different;
- ***Sequentiality.*** EHR data are sequential and a specific brain image is static.

Solution

We proposed a novel **Mem**ory-based **G**raph **C**onvolutional **N**etwork (MemGCN) to perform integrative analysis with both patient EHRs and neuroimages, using two major components: Graph Convolution and Memory array.

Outline

- Introduction
- Method
- Experiments
- Summary

Spatial Graph Construction

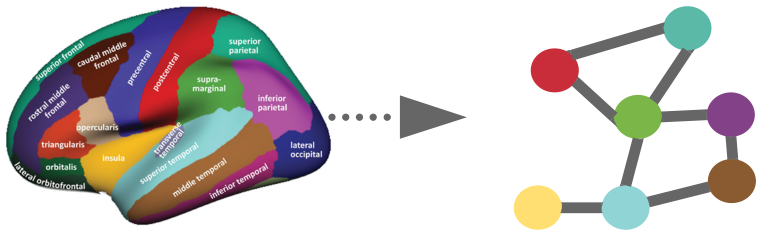
- Utilize 3-d brain coordinates of ROIs

Suppose we have a population of M acquisitions,

$$\bar{v}_i = \frac{1}{M}(\Sigma_m^M v_{i,m}^x, \Sigma_m^M v_{i,m}^y, \Sigma_m^M v_{i,m}^z), \forall i \in (1, \cdots, n).$$

the edges $\mathcal{E}$ can be constructed by

$$w_{ij} = \begin{cases} \exp(-\frac{\|\bar{v}_i - \bar{v}_j\|^2}{2\sigma^2}), & \text{if } i \in \mathcal{N}_j \text{ or } j \in \mathcal{N}_i \\ 0, & \text{otherwise.} \end{cases}$$

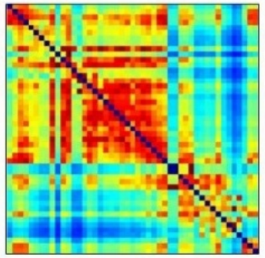


The Learning Problem

Input Data

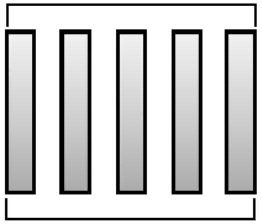
sample-level

connectivity matrix



$\mathbf{x} \in \mathbb{R}^{n \times n}$

sequential vectors



$\mathbf{s}_j, j = 1, \cdots, t$

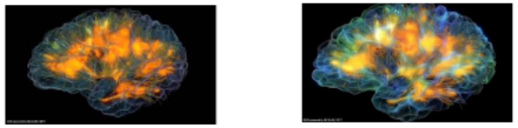
population-level

spatial graph

$\mathcal{G}$

Target

a sample pair



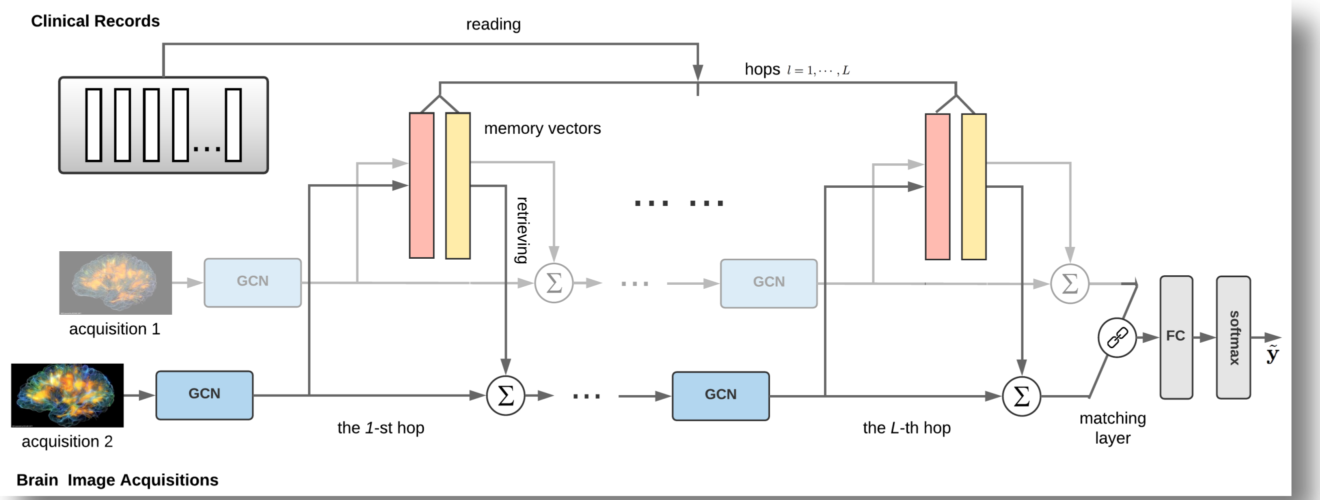
matching vs. non-matching $\tilde{\mathbf{y}}_{m,m'}$

x		**y**
PD	PD	*“same”*
PD	HC	*“different”*
HC	PD	*“different”*
HC	HC	*“same”*

Model Overview: MemGCN

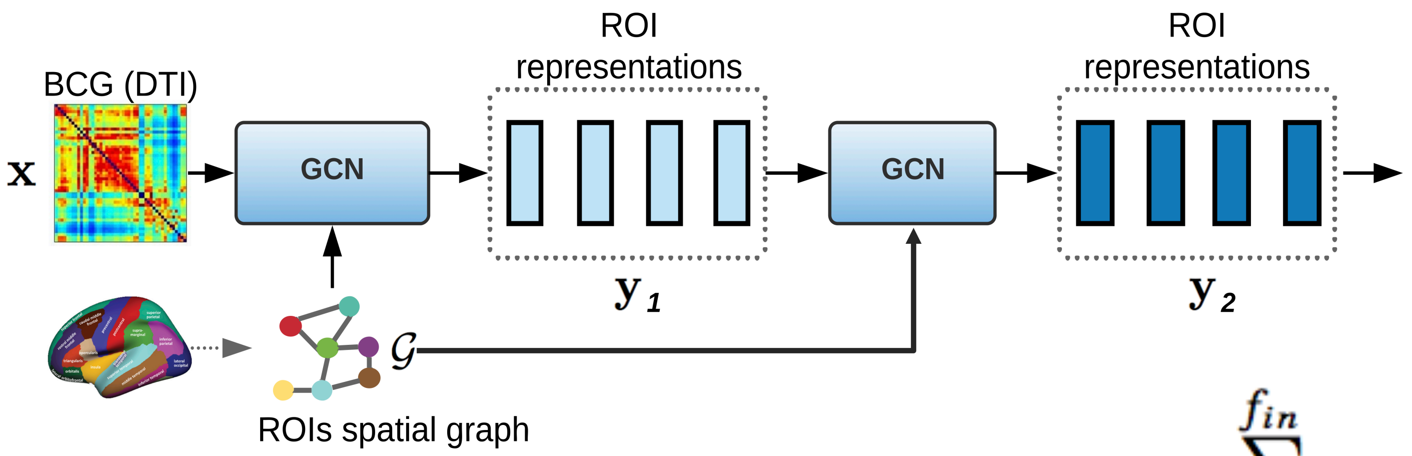
❖ Basic Blocks

- ✓ **B1:** Graph Convolutional Network
- ✓ **B2:** Memory Arrays
- ✓ **B3:** Multi-Hop Layer
- ✓ **B4:** Matching Layer



B1: Graph Convolutional Network (GCN)

- An illustration of 2-layer GCN



Graph Fourier Transform

$$\mathbf{x} \star \mathbf{g} = \Phi(\Phi^{\mathrm{T}}\mathbf{x}) \odot (\Phi^{\mathrm{T}}\mathbf{g}) = \Phi g_{\theta}(\Lambda)\Phi^{\mathrm{T}}\mathbf{x}$$
$$= \Phi diag(\hat{g}_1, \cdots \hat{g}_n)\hat{\mathbf{x}}$$

Hadamard product

A general operator

Set as $\mathbf{x}_{m,i}$, l = 0

$$\mathbf{y}_{m,k^{l+1}} = \sum_{k^l=1}^{f_{in}} g_{\theta_{k^l,l+1}}(\mathbf{L})\mathbf{y}_{m,k^l} \in \mathbb{R}^n$$

$$g_{\theta}(\boldsymbol{\Delta}) = \sum_{p=0}^{r-1} \theta_p T_p(\tilde{\boldsymbol{\Delta}}) = \sum_{p=0}^{r-1} \theta_p \boldsymbol{\Phi} T_p(\tilde{\boldsymbol{\Lambda}})\boldsymbol{\Phi}^{\mathrm{T}}$$

ChebNet

Chebyshev polynomial

B2: Memory-Augmented GCN (MemGCN)

❖ Clinical Sequences Reading

To embed the sequential vectors $\mathbf{s}_1, \cdots \mathbf{s}_t$,

input memory $\quad \mathbf{z}_j = \mathbf{A}\mathbf{s}_j$

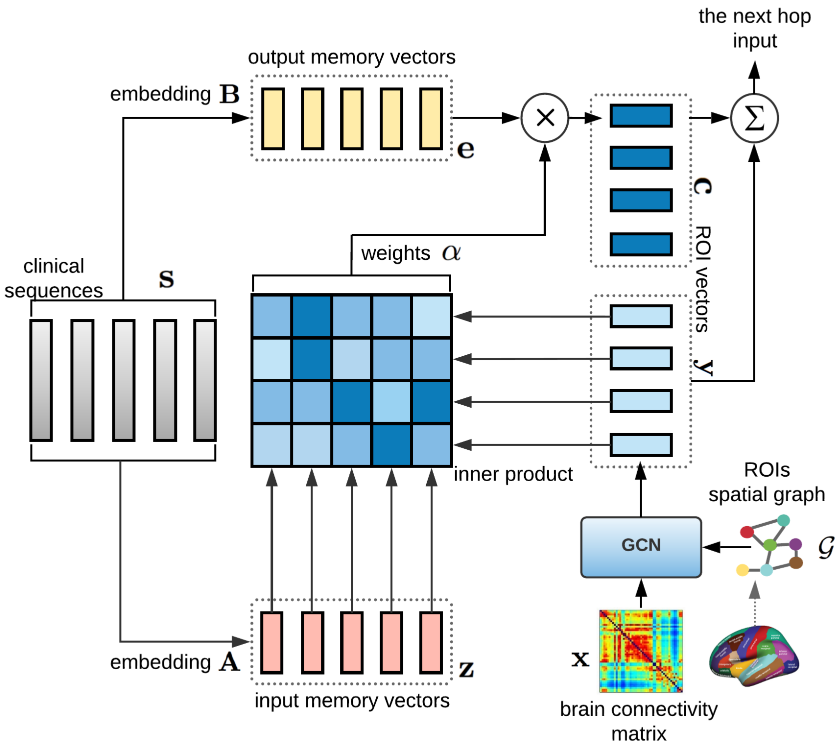
output memory $\quad \mathbf{e}_j = \mathbf{B}\mathbf{s}_j$

❖ Memory Representation Retrieving

To retrieve memory vectors from the embedding space,

$$\alpha_{ij} = softmax(\mathbf{y}_i\mathbf{z}_j) = \frac{\exp(\mathbf{y}_i\mathbf{z}_j)}{\sum_{j'=1}^{t}\exp(\mathbf{y}_i\mathbf{z}_{j'})}$$

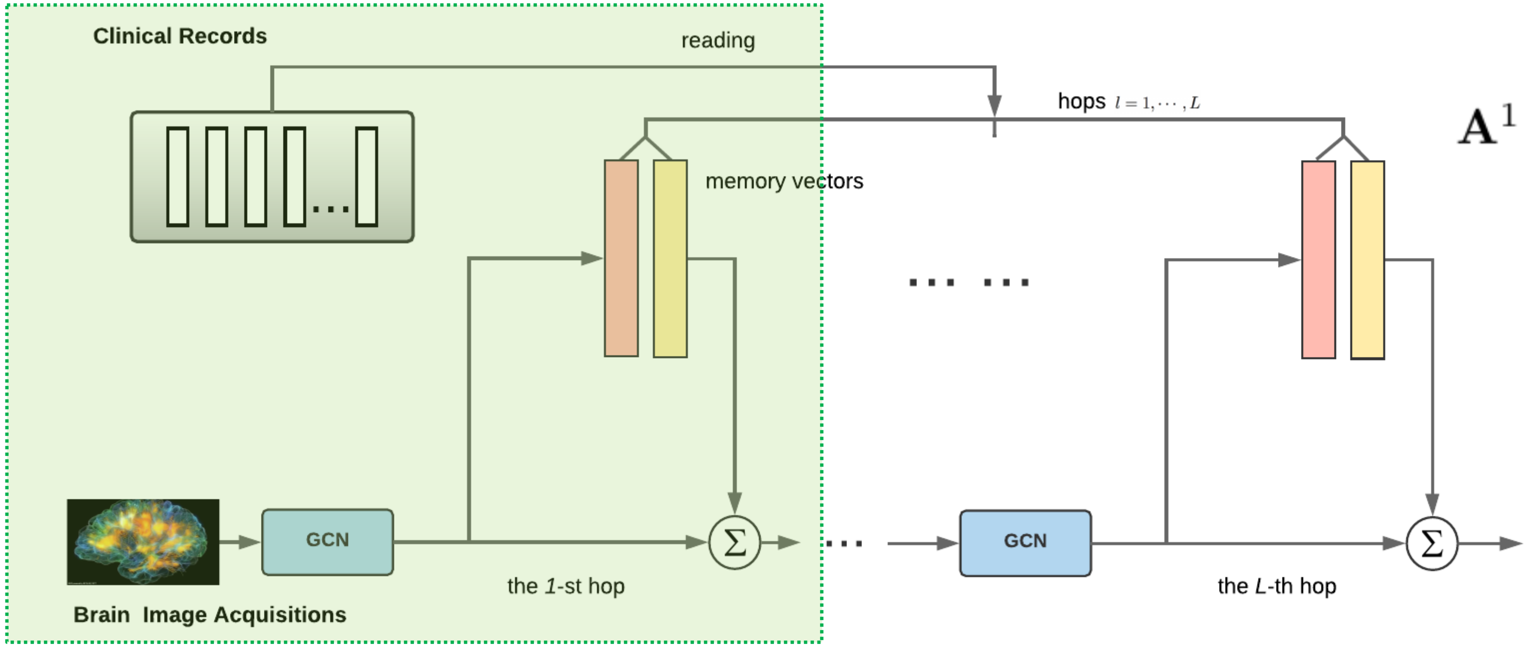
$$\mathbf{c}_i = \sum_{j=1}^{t}\alpha_{ij}\mathbf{e}_i \qquad \hat{\mathbf{y}}_i = \mathbf{y}_i + \mathbf{c}_i$$



1-hop MemGCN

B3: Multi-Hop Layer

- Extend to multiple hop architecture (B1 ⊂ B2 ⊂ B3)



1-hop MemGCN

$$\mathbf{A}^1 = \cdots = \mathbf{A}^L \text{ and } \mathbf{B}^1 = \cdots = \mathbf{B}^L$$

$$\alpha_{ij}^l = \frac{\exp(\mathbf{y}_i^l \mathbf{z}_j^l)}{\sum_{j'=1}^{t} \exp(\mathbf{y}_i^l \mathbf{z}_{j'}^l)}$$

$$\mathbf{c}_i^l = \sum_{j=1}^{t} \alpha_{ij}^l \mathbf{e}_i^l$$

the output feature map $\hat{\mathbf{y}}$ at the l-th hop can be rewritten as

$$\mathbf{y}^{l+1} = \mathbf{H}\mathbf{y}^l + \mathbf{c}^l, l = 1, \cdots, L$$

B4: Matching Layer

- What if the size of training data is small?
  - Siamese-like Network
    - ✓ Inner Product Matching

$$sim_i(\mathbf{x}_m, \mathbf{x}_{m'}) = (\mathbf{y}^L_{m,i})^{\mathrm{T}} \mathbf{y}^L_{m',i}, \ i = 1, \cdots, n.$$

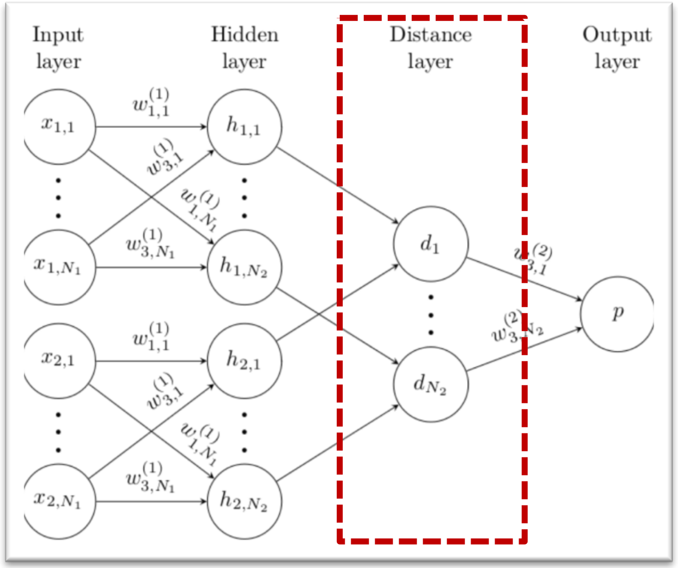
    - ✓ Bilinear Matching

$$sim_{i,j}(\mathbf{x}_m, \mathbf{x}_{m'}) = (\mathbf{y}^L_{m,i})^{\mathrm{T}} \mathbf{M} \mathbf{y}^L_{m',j}, \ i, j = 1, \cdots, n.$$

parameter matrix

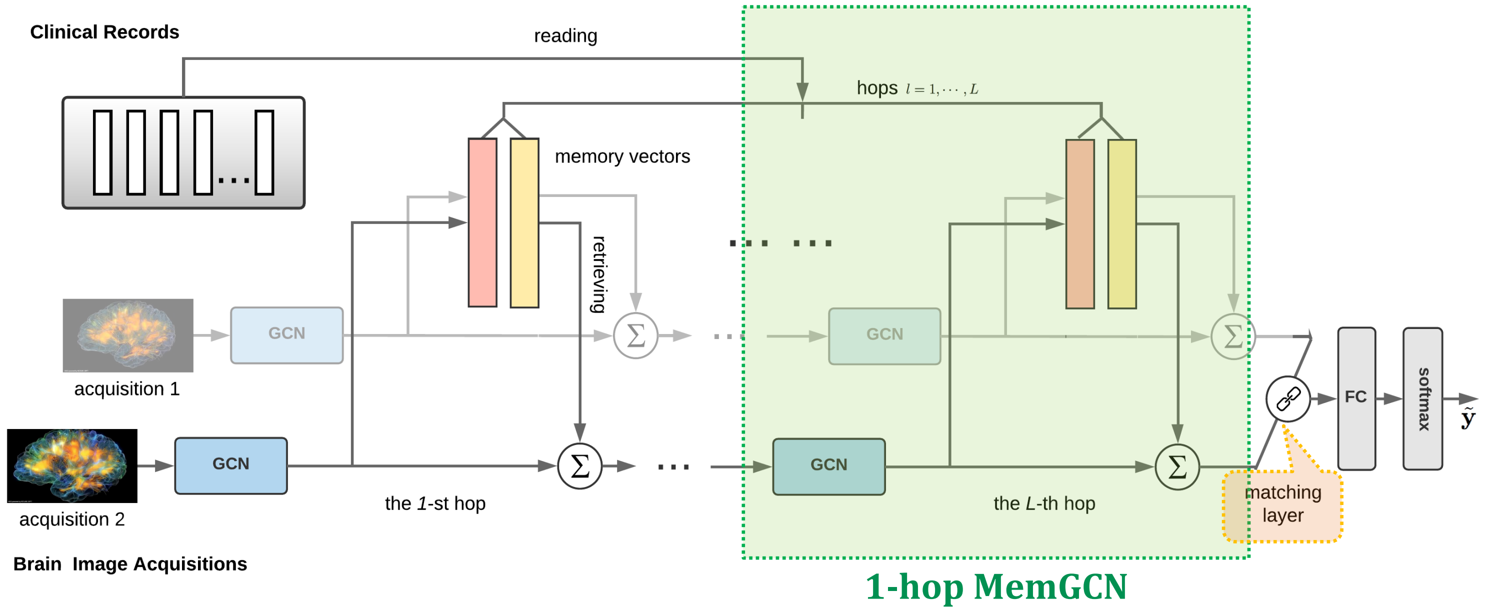
Impose structure➔

Learning a metric space



Koch et al. '15

Model Training



- Objective Function: (pairwise training)

$$\mathcal{L} = \sum_{m,m'}^{N} \tilde{\mathbf{y}}_{m,m'} \log \mathbf{p}_{m,m'} + (1 - \tilde{\mathbf{y}}_{m,m'}) \log(1 - \mathbf{p}_{m,m'}) + \gamma \|\Theta\|_2$$ where $\tilde{\mathbf{y}}_{m,m'}$ denotes the label for sample pair $(\mathbf{x}_m, \mathbf{x}_{m'})$

$$\mathbf{p} = softmax(\mathbf{w}_c^{\mathrm{T}} \mathbf{r})$$ where $\mathbf{w}_c \in \mathbb{R}^2$

Outline

- Introduction
- Method
- Experiments
- Summary

Dataset: PPMI

- NeuroImages (DTI acquisitions)

# of Matching Samples	# of Non-Matching Samples	# of PD Subjects	# of HC Subjects
189, 713	94, 168	596	158

- Patient Health Records

Extra Modality	Clinical Assessments	
Motor	MDS-UPDRS Part II	MDS-UPDRS Part III
Non-Motor	MDS-UPDRS Part I	MoCA

Relationship Classification

- 5-fold cross validation (held-out subjects)
- Evaluation(binary classification)
  - ✓ AUC
  - ✓ Accuracy
- Baselines:

  1. Raw edges; 2. PCA; 3. FCN; 4. GCN; 5. AttGCN; 6. AttLstmGCN.

Relationship Classification

- Results for classifying on test sets of tensor-FACT, ODF-RK2, and Hough.

Extra Modalities	Methods	tensor-FACT		ODF-RK2		Hough	
		Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
None	Raw Edges	65.94±3.78	58.47±4.05	67.56±4.12	60.93±5.60	67.90±4.09	64.49±3.56
	PCA	69.19±3.13	64.10±2.10	68.38±2.50	60.93±2.63	66.28±4.60	63.46±3.52
	FCN	71.65±3.58	66.17±2.00	70.66±3.79	68.80±2.80	70.01±3.28	61.91±3.42
	FCN-*2layer*	84.22±2.76	82.36±2.87	82.31±2.68	82.53±4.74	84.27±2.63	81.77±3.74
	GCN-*inner*	93.69±2.15	92.67±4.94	93.23±2.63	93.04±5.26	92.80±2.51	93.90±5.48
	GCN-*bilinear*	93.89±1.76	94.77±6.08	94.00±2.65	94.32±5.72	93.34±2.26	93.35±5.14
Fusion	AttGCN	93.62±2.99	94.25±5.88	94.76±3.31	94.33±5.23	94.01±1.94	94.74±5.35
	AttLstmGCN	94.70±2.35	94.38±5.41	94.89±2.71	94.87±4.49	94.64±2.02	94.80±5.51
	MemGCN-*inner*	95.43±2.22	96.42±6.36	95.54±2.98	96.59±6.44	95.48±2.34	96.49±6.41
	MemGCN-*bilinear*	**95.47±2.25**	**96.48±6.40**	**95.87±2.56**	**96.84±6.36**	**95.64±2.00**	**96.74±6.51**

proposed methods

Interpretation: learned similarity

❖ Identical ROIs vs. Discriminative ROIs

- Average the learned representations for pairwise sample groups (by inner product)

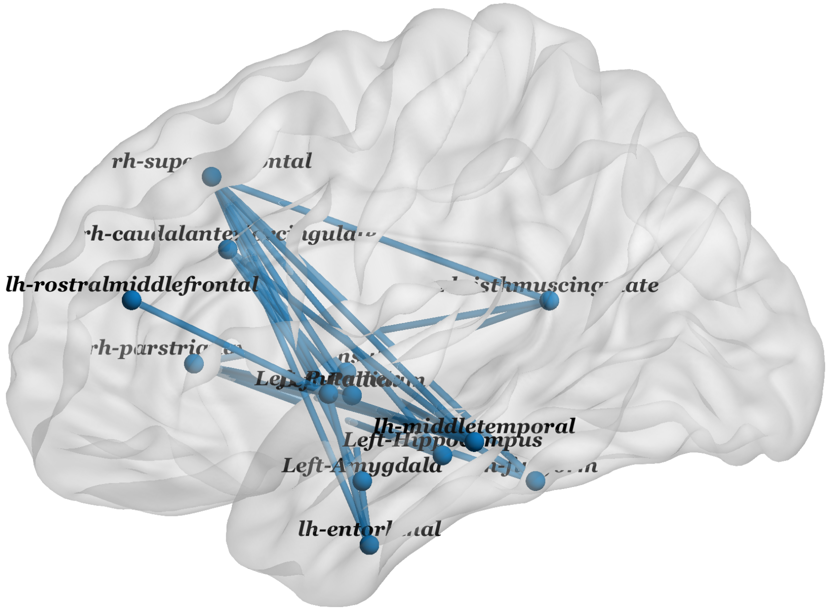
	Motor		Non-motor		Fusion	
	ROI Name	Score	ROI Name	Score	ROI Name	Score
Identical ROIs (PD Group)	Right Thalamus Proper	0.9258	Rh Paracentral	0.8563	Rh Pars Opercularis	0.9344
	Lh Insula	0.9253	Rh Lingual	0.8180	Rh Lateral Occipital	0.8372
	Right Pallidum	0.9226	Right Pallidum	0.8091	Left Accumbens Area	0.7887
	Lh Rostral Middle Frontal	0.9210	Lh Parsorbitalis	0.6554	Rh Parahippocampal	0.7827
	Parahippocampal	0.9206	Left Thalamus Proper	0.6387	Rh Frontalpole	0.7742
Discriminative ROIs (PD vs. HC Group)	Right Putamen	-0.9134	Left Putamen	-0.7423	Right Thalamus Proper	-0.8960
	Right Accumbens Area	-0.9075	Lh Frontal Pole	-0.5754	Left Caudate	-0.8439
	Left Hippocampus	-0.9059	Lh Supramarginal	-0.5731	Lh Paracentral	-0.8227
	Right VentralDC	-0.9058	Lh Inferior Parietal	-0.5693	Lh Middle Temporal	-0.7865
	Left Caudate	-0.9014	Lh Paracentral	-0.4851	Lh Cuneus	-0.7528

* Lh and Rh are the abbreviations of Left Hemisphere and Right Hemisphere respectively.

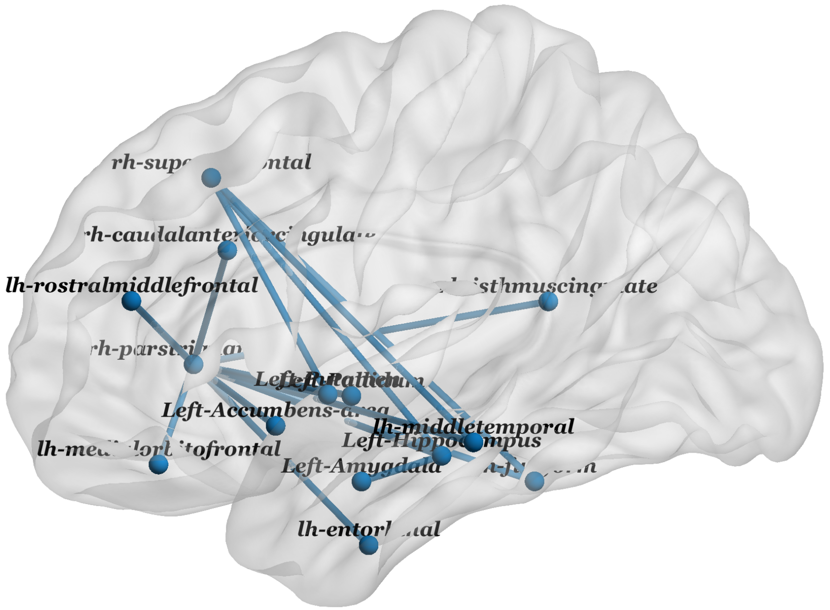
***Ref**: 1. Morphological alterations in the caudate, putamen, pallidum, and thalamus in Parkinson's disease;*
2. Resting state functional connectivity of the striatum in Parkinson's disease.

Interpretation: Learned Similarity

❖ Identical Connections vs. Discriminative Connections

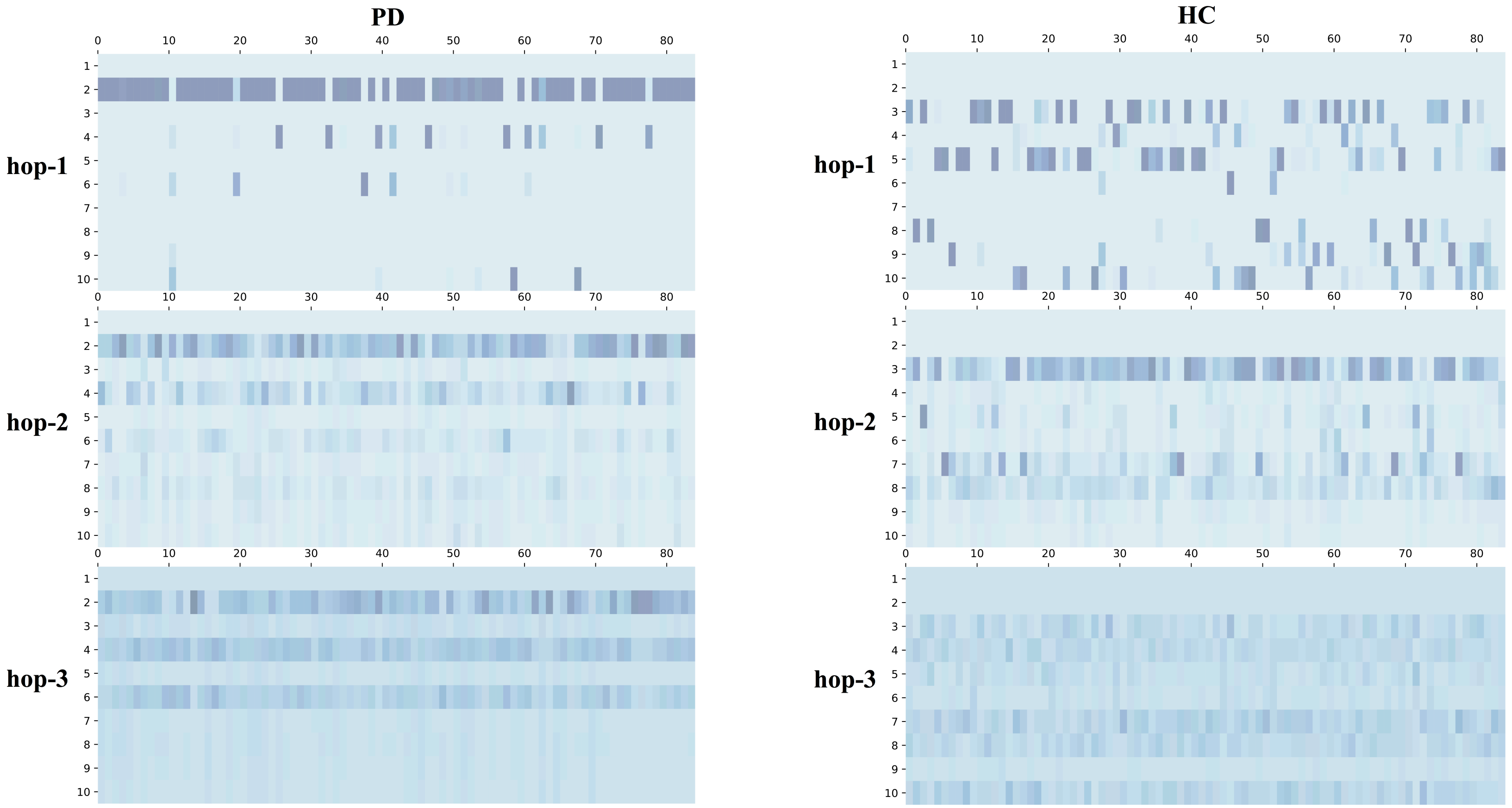


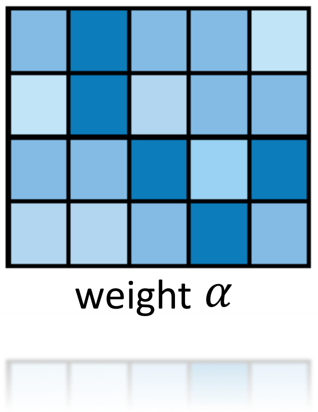
(a) The top identical connections for PD group



(b) The top discriminative connections between PD and HC groups

Longitudinal Alignment: Case Study





Visualizations of **attention interaction** matrices for one PD case and one HC case during 3 memory hops.

Outline

- Introduction
- Method
- Experiments
- Summary

Summary

- Proposed an **integrative analysis** architecture MemGCN to deal with **patient health records** and **neuroimages** for discriminating patients and healthy controls;
- **Interpretable** high-level representations extracted from MemGCN are explored;
- Our framework makes a progress in modeling a **small** cohort data such as PPMI.

Thank You!
Q&A

xi-zhang.net

@sheryl_zx

sheryl.zhangxi@gmail.com