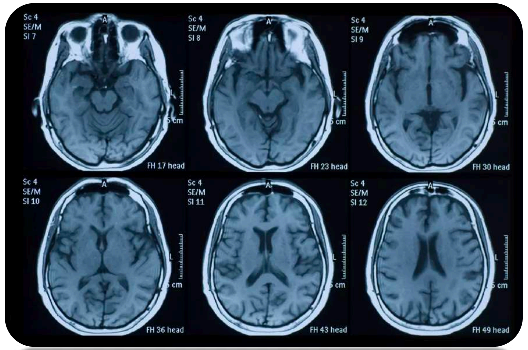
Clinical Data Modelling with Deep Neural Networks: Challenges and Solutions

Postdoctoral Associate, Xi Sheryl Zhang
Cornell University

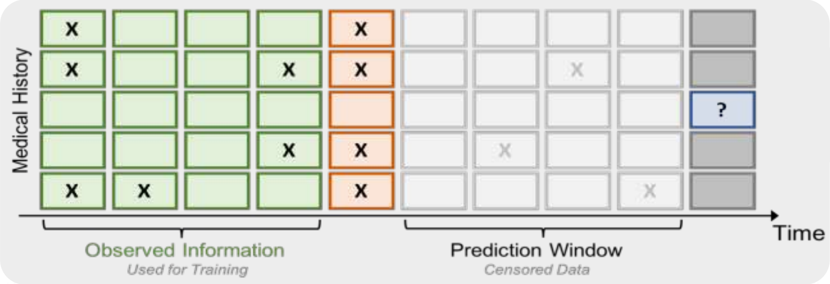
Why Deep Learning in Healthcare?

Data

Medical Images



Text

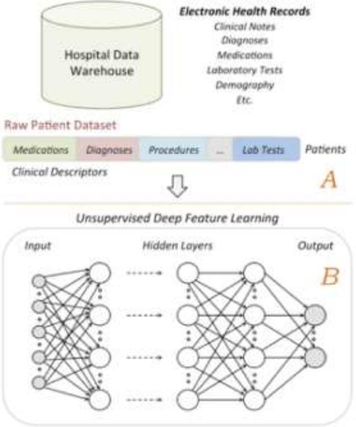


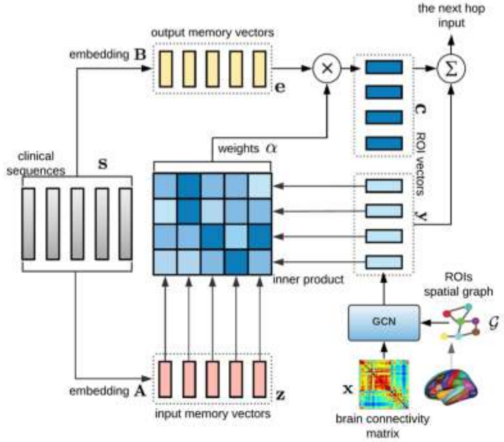
EHRs

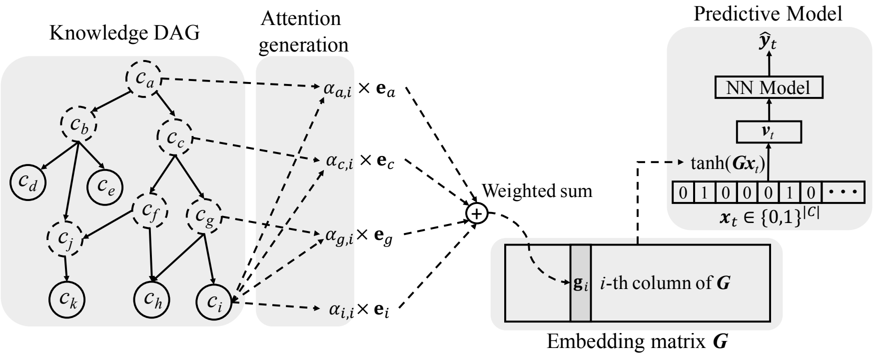
Genomics

… …

Algorithms







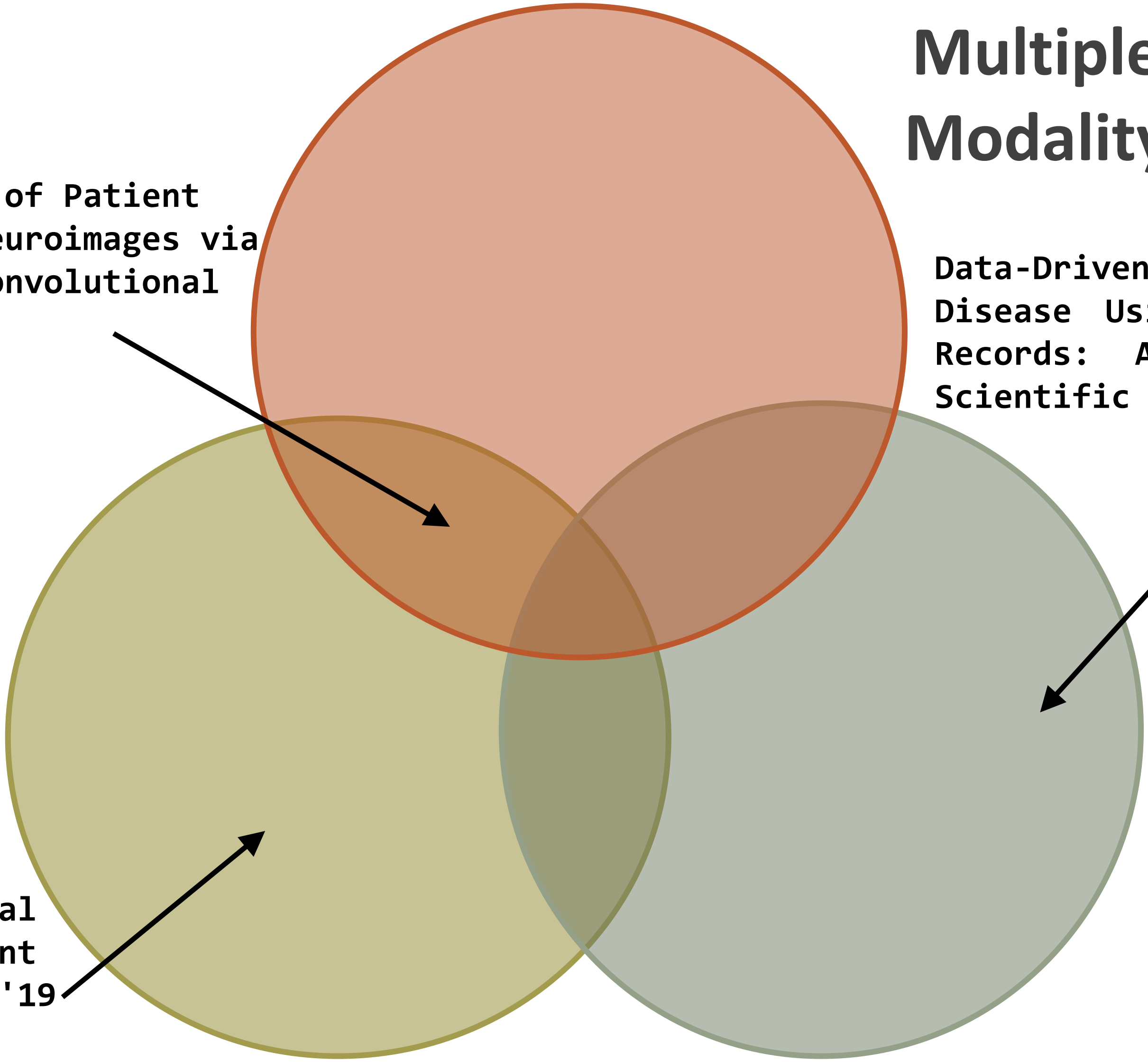
… …

Obstacles

Multiple Modality

Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM’18

Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Nature Scientific Reports, 2019



Data Scarcity

Heterogeneity

MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19

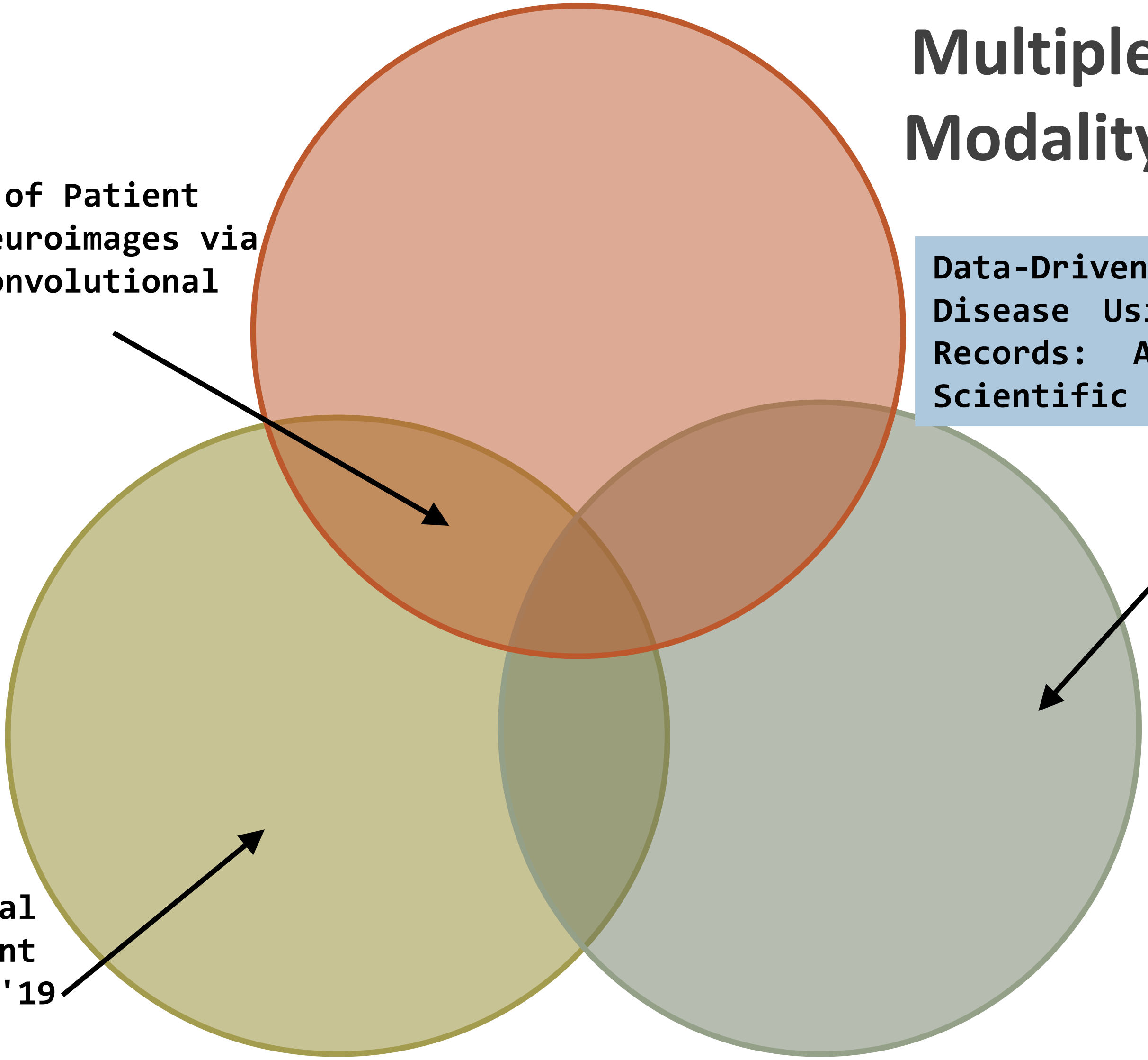
Outline

Part 1: Disease Subtyping on Clinical Times Series

Part 2: Integrative Disease Analysis via Multi-Modality

Part3: Meta-Learning on Limited Clinical Resources

Obstacles



Multiple Modality
Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Nature Scientific Reports, 2019
Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM’18
Data Scarcity
Heterogeneity
MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19

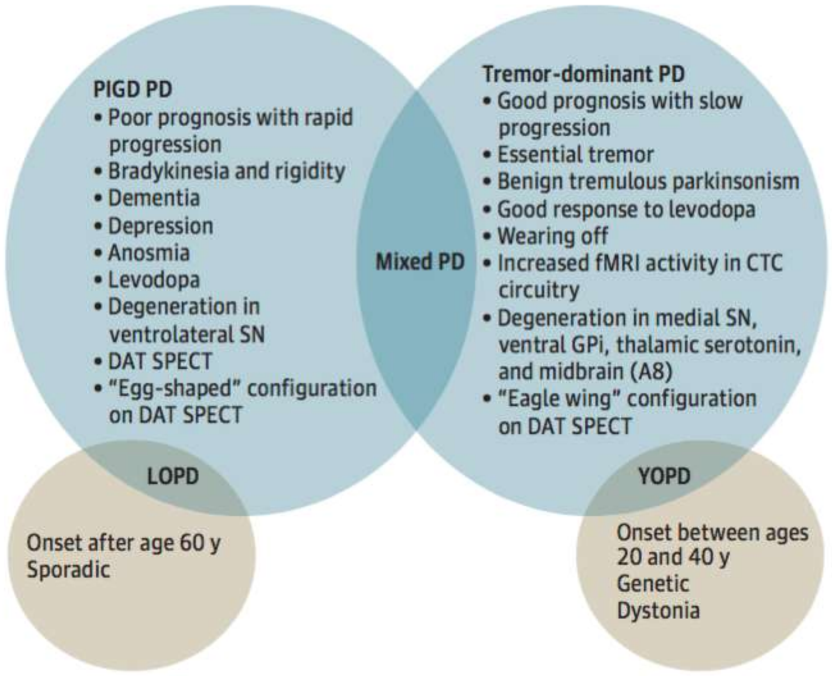
Disease Subtyping

Background: *Parkinson's Disease (PD) has been demonstrated heterogeneous in clinical representation and disease progression. Therefore, identifying subtypes with similar characteristics is an important task to study the disease.*

- The disease associates with clinical factors of motor, non-motor, and other variables, so that there is no widely accepted consensus on the criteria for patient groups.

- Data-driven approaches of clustering methodologies can identify subtypes without a priori hypothesis about disease knowledge.

- Recurrent neural networks are successful in many sequential learning tasks, and may allow us to find more PD progression patterns among clusters.



Disease Subtyping

Challenges

- Parkinson's Disease (PD) is clinically ***heterogeneous*** associated with a broad spectrum of clinical variable factors;
- How to identify ***disease progression*** biomarkers, so that we can provide a better population for modifying drug trials.

Solution

- We first concatenated the **multi-source records** according to their occurring timestamps to form a temporal sequence for each patient;
- A deep learning model **LSTM** is trained to encode the record sequences into a series of standardized embeddings.

Disease Subtyping

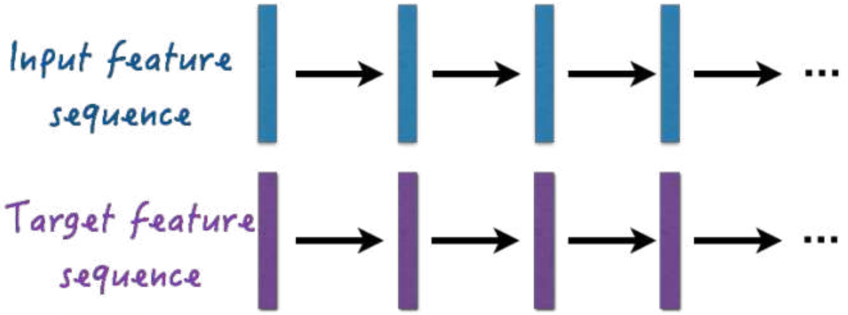
- Demographics
- Motor
- Non-motor
- Biospeci-men
- Imaging
- ... ...

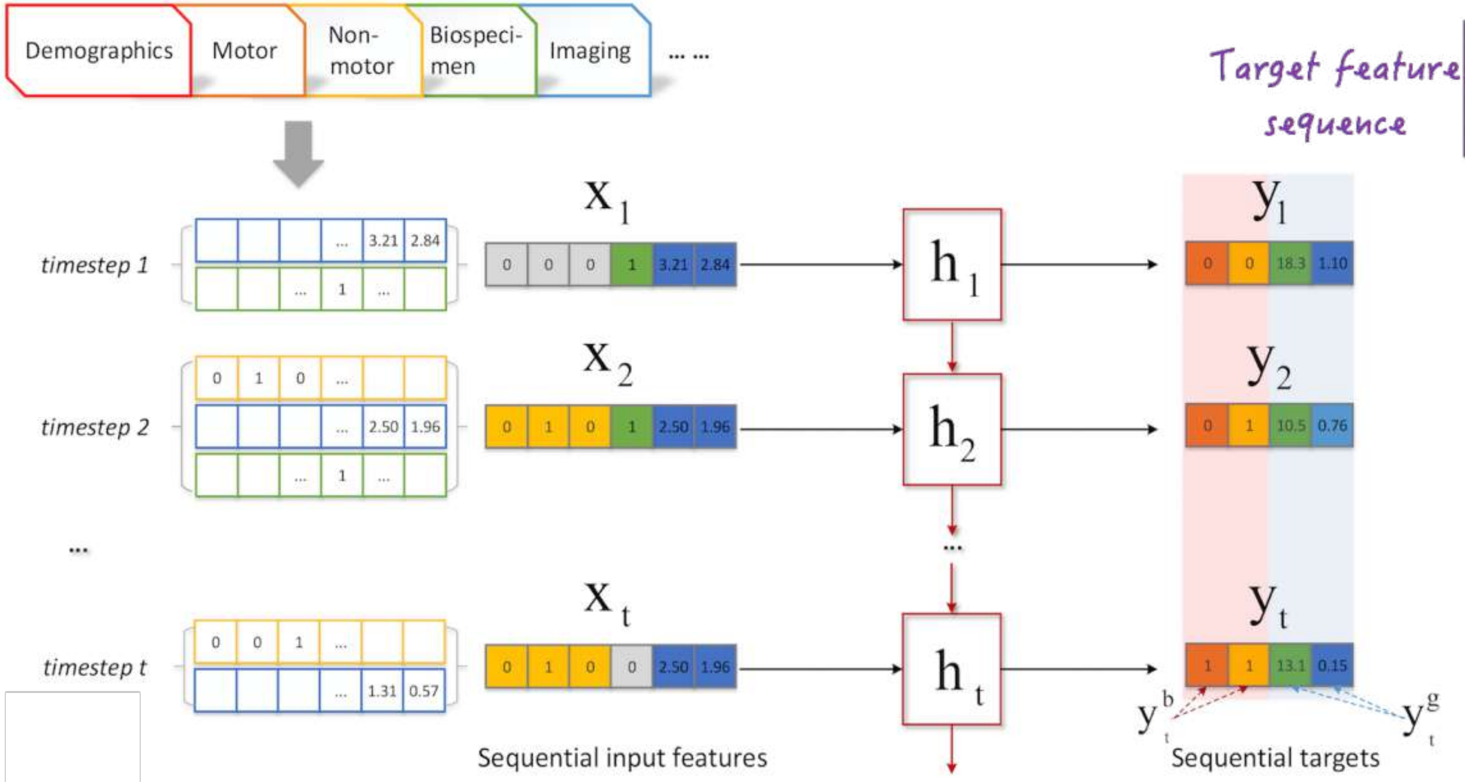
❖ Features

	Target Clinical Variables
1	Clinical Diagnosis
2	Demographics
3	Motor symptoms: MDS-UPDRS scores*
4	Cognitive Accessments: MoCA*
5	Cognitive Categorization: Normal Cognition; Mild Cognitive Impairment; Dementia
6	Other nonmotor variable: REM Sleep Disorder
7	Biospecimen: Lumber Puncture Sample Collection
8	Biospecimen: Laboratory Procedures containing DNA, RNA, Urine, Plasma, & Serum samples
9	Imaging Results: DaTScan Striatal Binding Ratio
10	Imaging Results: Magnetic Resonance Imaging

Ref: Fereshtehnejad, Seyed-Mohammad, Silvia Rios Romenets, Julius BM Anang, Véronique Latreille, Jean-François Gagnon, and Ronald B. Postuma. "New clinical subtypes of Parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes." JAMA neurology 72, no. 8 (2015): 863-873.

Disease Subtyping





- Continuous target feature:

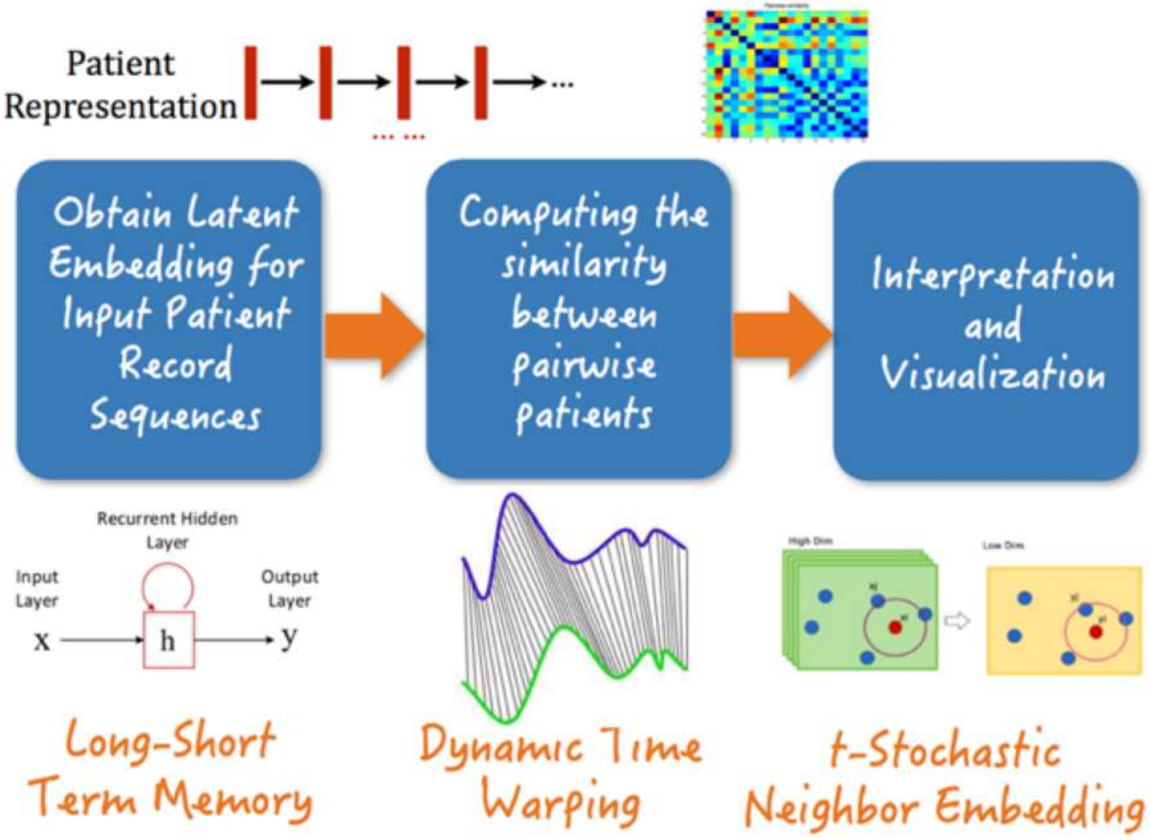
$$\frac{1}{2}\sum_t \left\|y_t^g - W_g h_t\right\|_2^2 + \lambda\left\|W_g\right\|_F^2$$

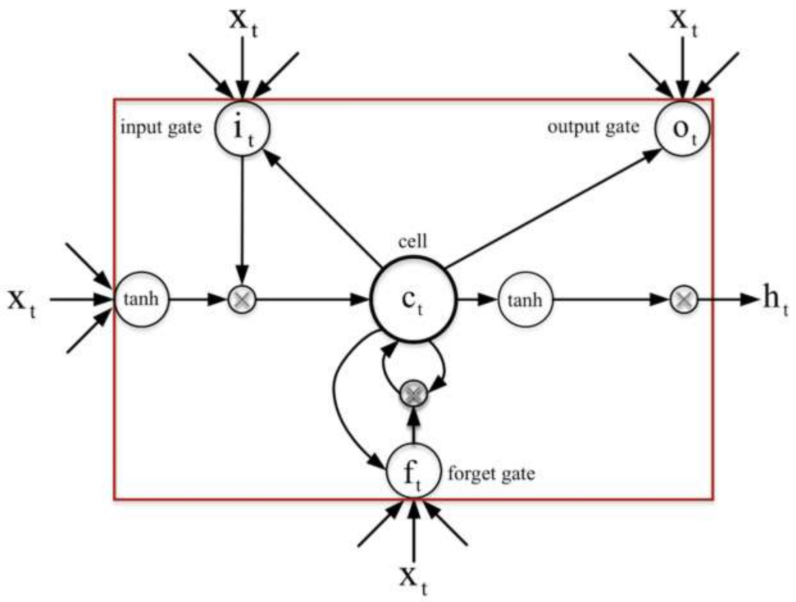
- Binary target feature:

$$\sum_t \sum_{j=1}^{m_b} \log\left(1 + \exp\left(-y_{t,j}^b (w_{b,j}^T h_t)\right)\right) + \lambda\|W_b\|_F^2$$

Disease Subtyping

❖ Method





$$i_t = \sigma(W_i x_t + W_i h_{t-1} + b_i)$$

$$f_t = \sigma\left(W_f x_t + W_f h_{t-1} + b_f\right)$$

$$o_t = \sigma(W_o x_t + W_o h_{t-1} + b_o)$$

$$c_t = f_t .* c_{t-1} + i_t .* tanh(W_c x_t + W_c h_{t-1} + b_c)$$

$$h_t = o_t .* tanh(c_t)$$

Disease Subtyping

❖ Dataset

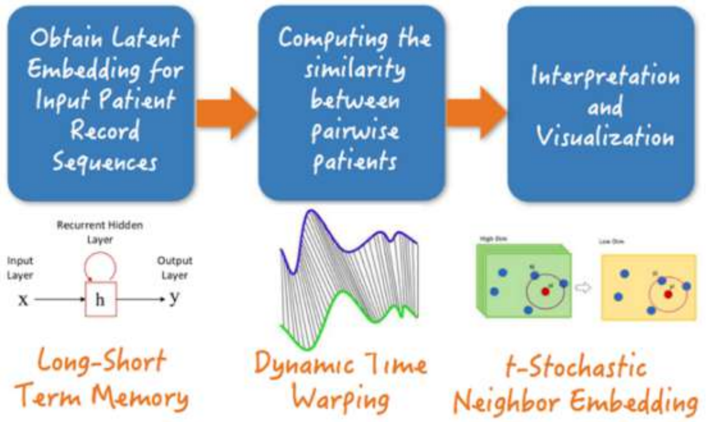
- The patient data were obtained from the Parkinson Progression Marker Initiative (PPMI) study. http://www.ppmi-info.org/

- The de-identified data contained archives of enrolled subjects from *June 1, 2010, to June 1, 2016*.

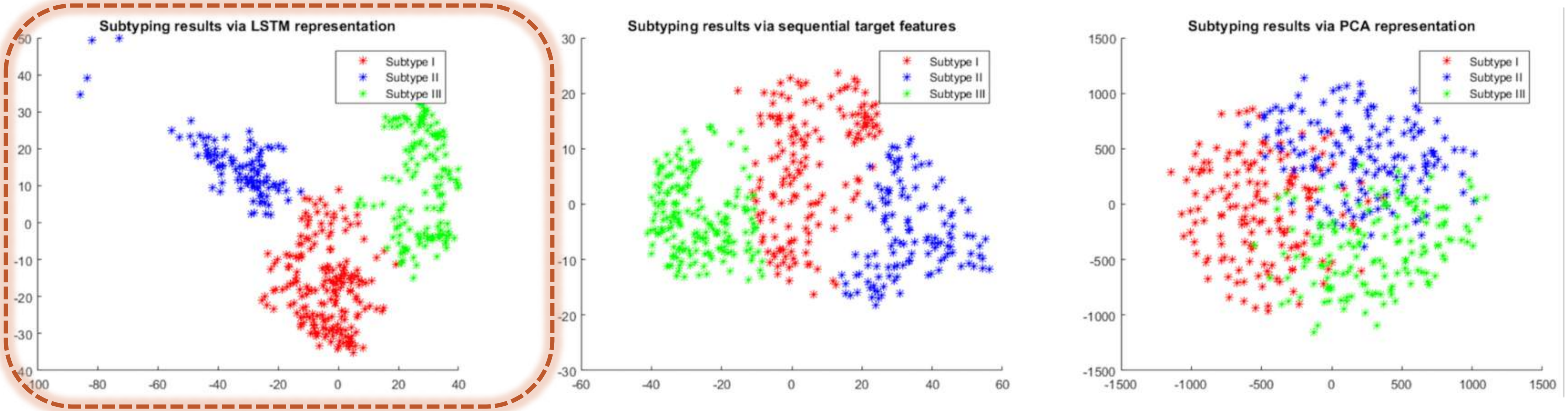
THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

Type	Code	# subjects
Idiopathic PD	1	466
Corticobasal degeneration	4	0
Dementia with Lewy bodies	5	1
Essential tremor	7	2
Multiple system atrophy	11	0
Psychogenic illness	15	1
No PD nor other neurological disorder	17	219
Other neurological disorder(s) (specify)	97	4
		759

Disease Subtyping: Visualization



❖ Results



Comparison with Traditional Clustering Methods

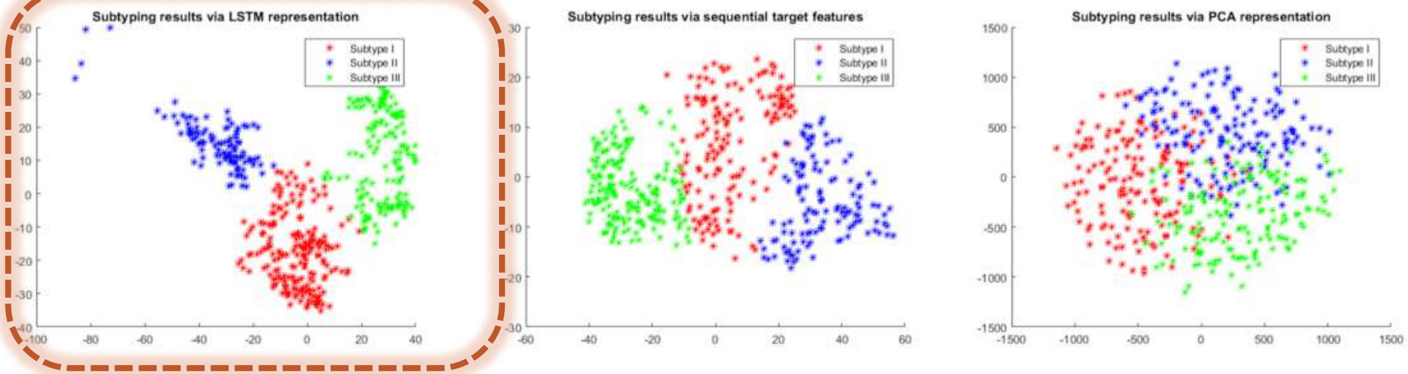
Disease Subtyping: Statistical Testing

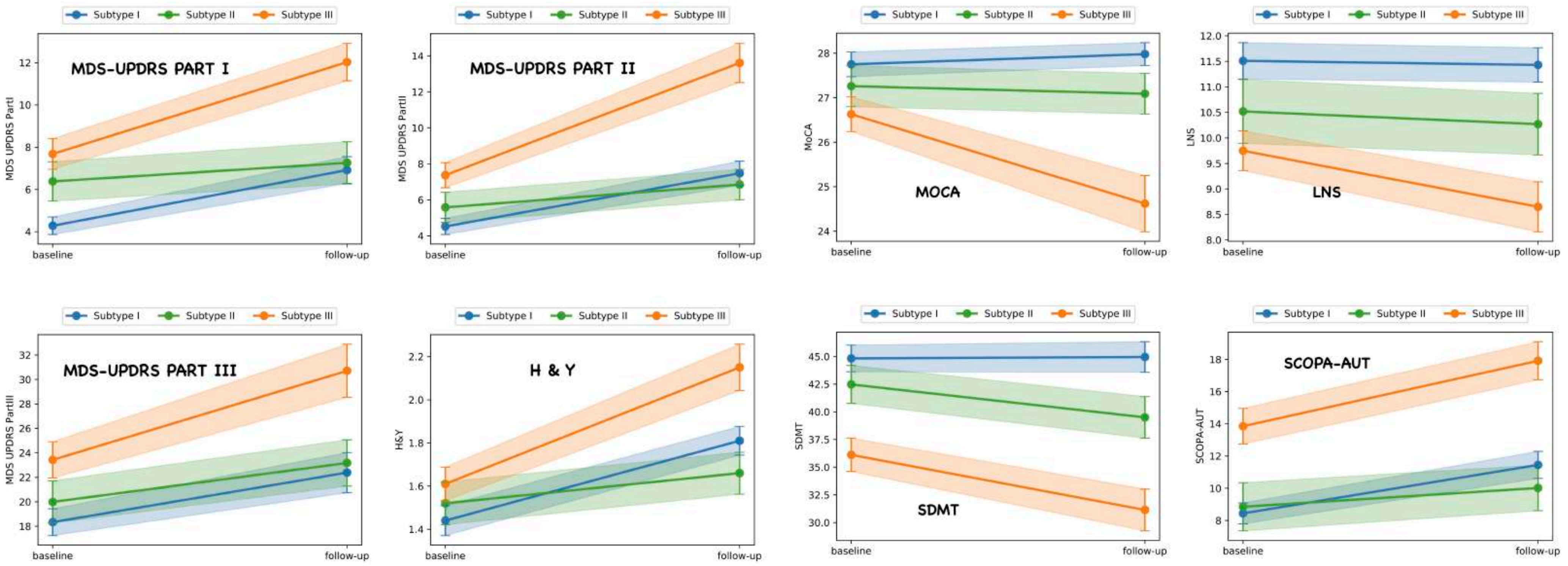
[a] *Chi-square test;* [b] *F-test;* [c] *One-way ANOVA test;* [d] *Kruskal-Wallis H-test*

Characteristics	Subtype I (N = 201)		Subtype II (N = 107)		Subtype III (N=158)		P-Value	
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Static	Progression
Age onset	58.79(9.5)		61.93(9.0)		65.32(8.8)		<0.0001[a]	
Hoehn and Yahr Stage	1.44(0.5)	1.81(0.4)	1.52(0.5)	1.66(0.5)	1.61(0.5)	2.15(0.6)	<0.0001[a]	<0.0001[a]
MDS-UPDRS Part I	4.28(2.9)	6.92(4.5)	6.38(4.8)	7.26(5.2)	7.68(4.6)	12.03(5.7)	<0.0001[a]	<0.0001[a]
MDS-UPDRS Part II	4.52(3.2)	7.48(4.8)	5.58(4.4)	6.85(4.4)	7.37(4.4)	13.61(7.0)	<0.0001[a]	<0.0001[a]
MDS-UPDRS Part III	18.34(7.9)	22.39(11.8)	19.99(9.0)	23.18(9.9)	23.43(9.5)	30.71(13.9)	0.1146[a]	<0.0001[a]
Montreal Cognitive Assessment	27.75(2.0)	27.98(1.8)	27.26(2.4)	27.09(2.4)	26.63(2.5)	24.62(4.0)	<0.0001[a]	<0.0001[a]
Geriatric Depression Scale	5.11(1.4)	5.2(1.3)	5.2(1.17)	5.31(1.2)	5.47(1.5)	5.96(1.8)	0.0017[a]	0.0010[a]
State Trait Anxiety Inventory	61.84(15.8)	59.52(16.0)	62.14(17.9)	61.89(18.1)	71.0(19.8)	74.25(20.1)	0.0053[a]	0.1717[a]
DaTScan	1.43(0.5)	1.23(0.5)	1.60 (0.6)	3.05 (0.6)	1.23 (0.5)	0.97 (0.5)	<0.0001[c]	<0.0001[c]

Disease Subtyping

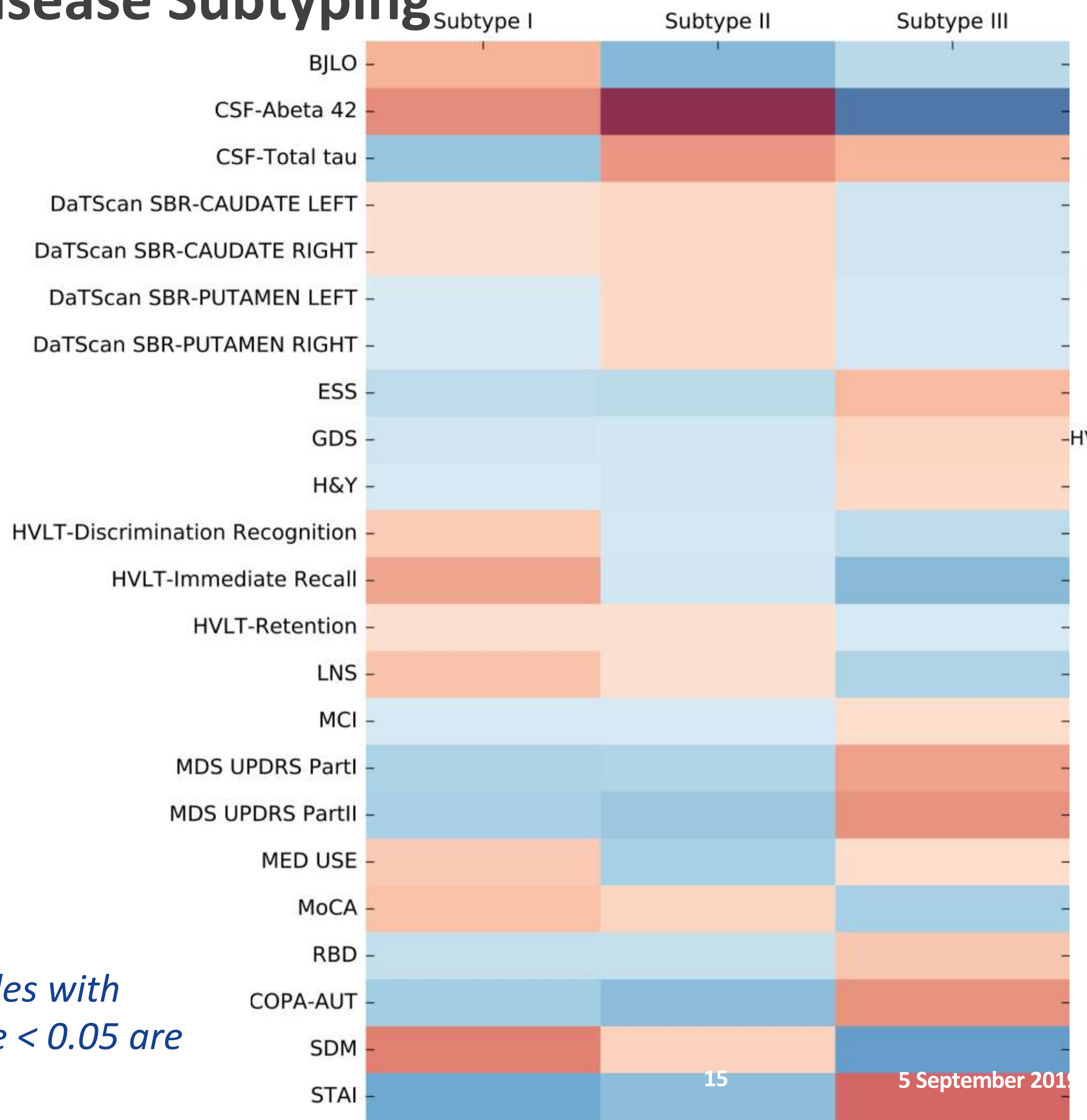
❖ Results

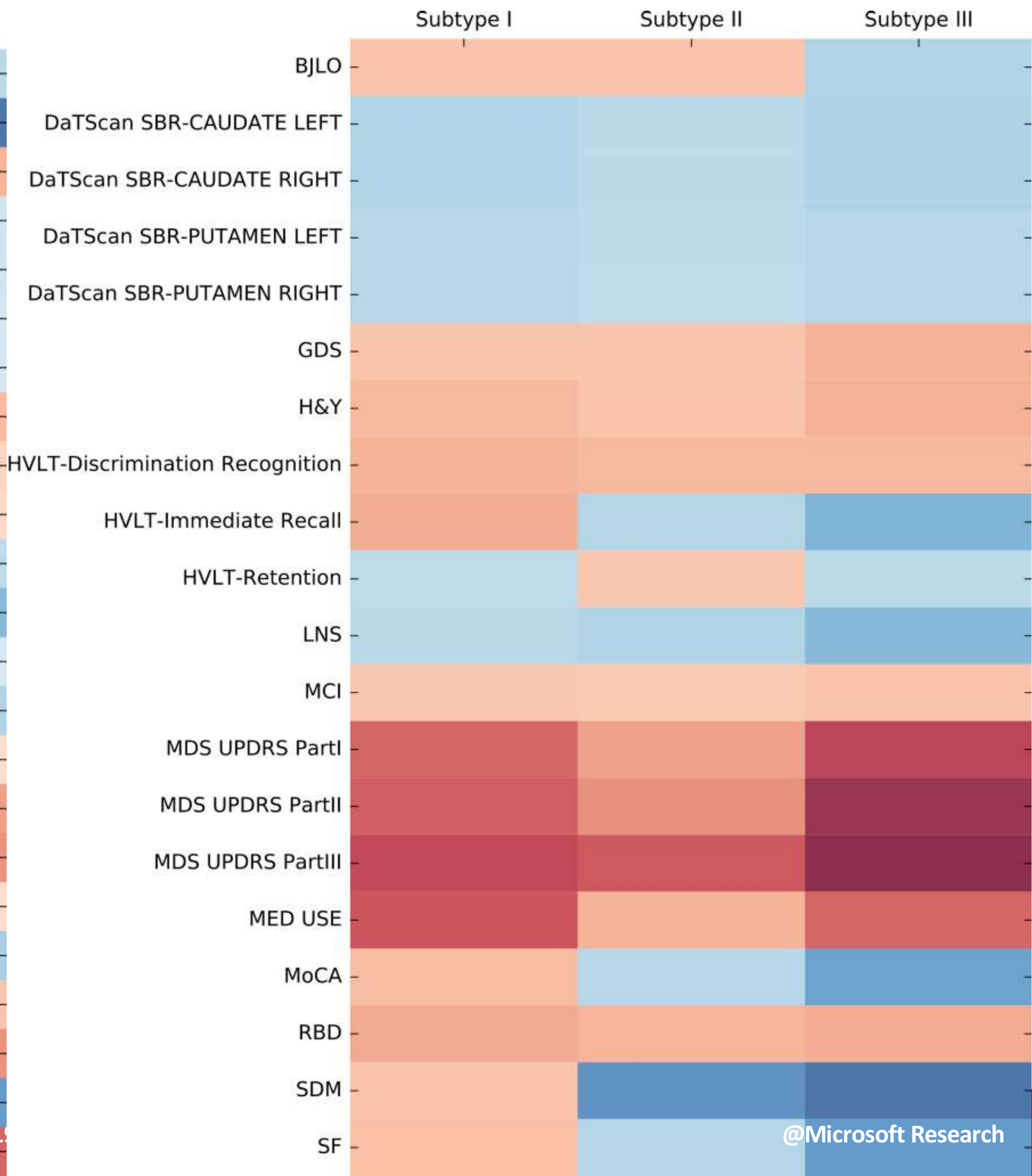




Progression in the Discriminative Variables

Disease Subtyping

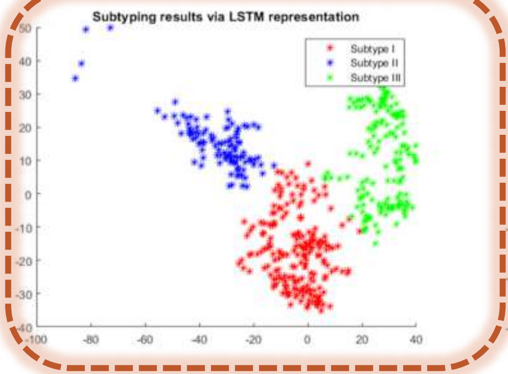


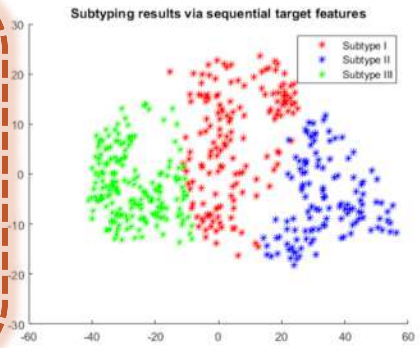


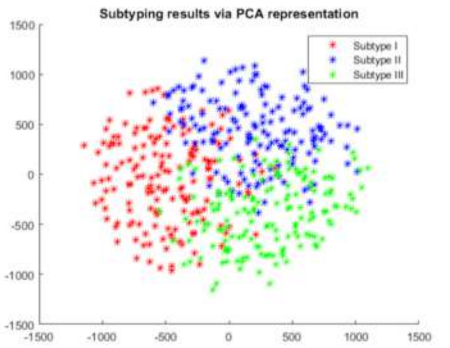
Variables with p-value < 0.05 are shown

Disease Subtyping

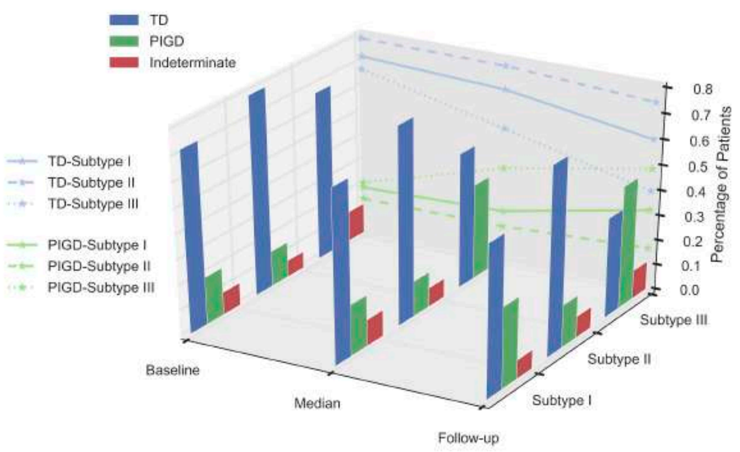
❖ Results





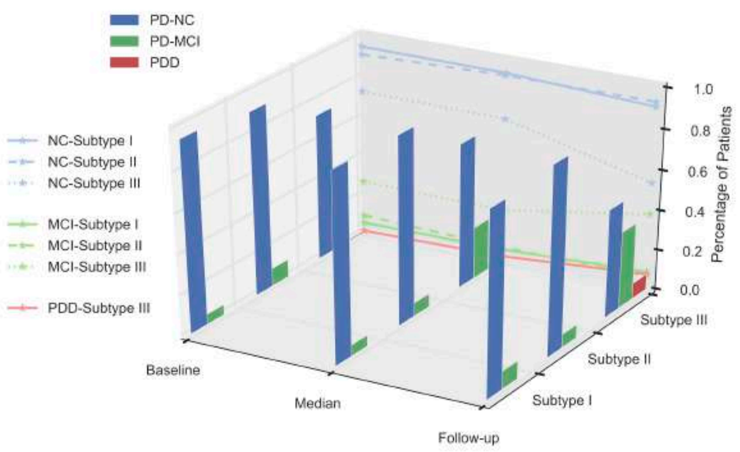


A



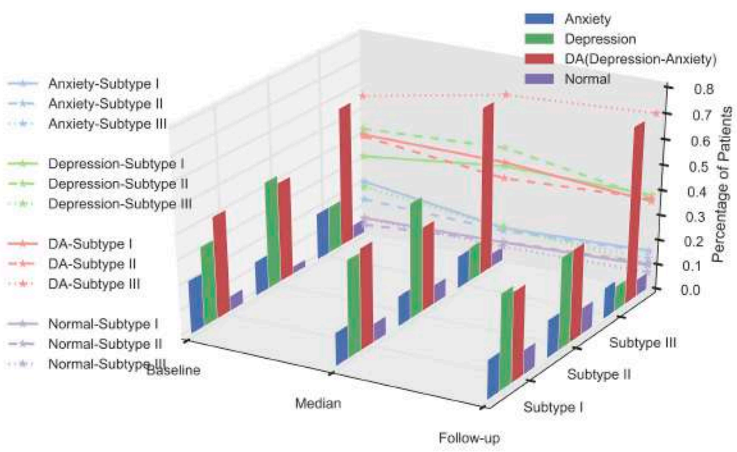
Correlation with Motor Subtypes

B



Correlation with Cognitive Subtypes

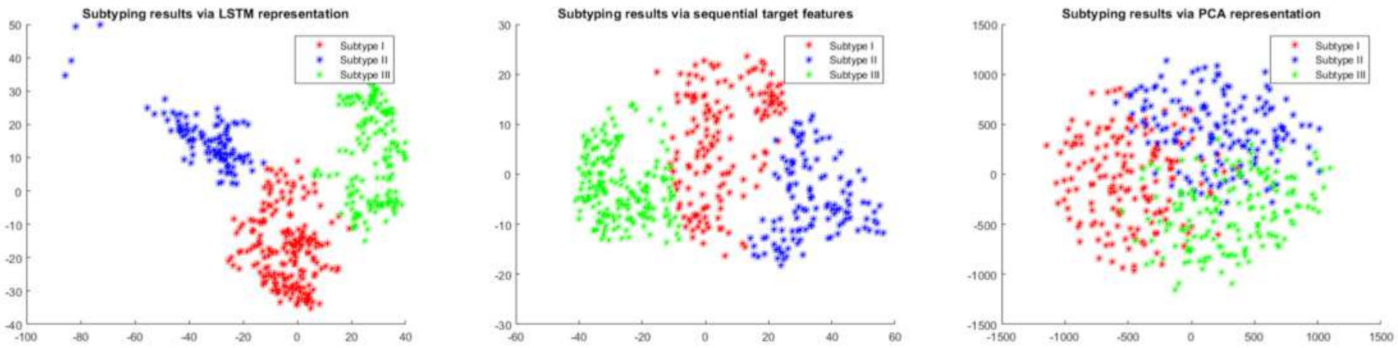
C



Correlation with Mood Subtypes

Relationship with Conventional PD Subtypes

Disease Subtyping

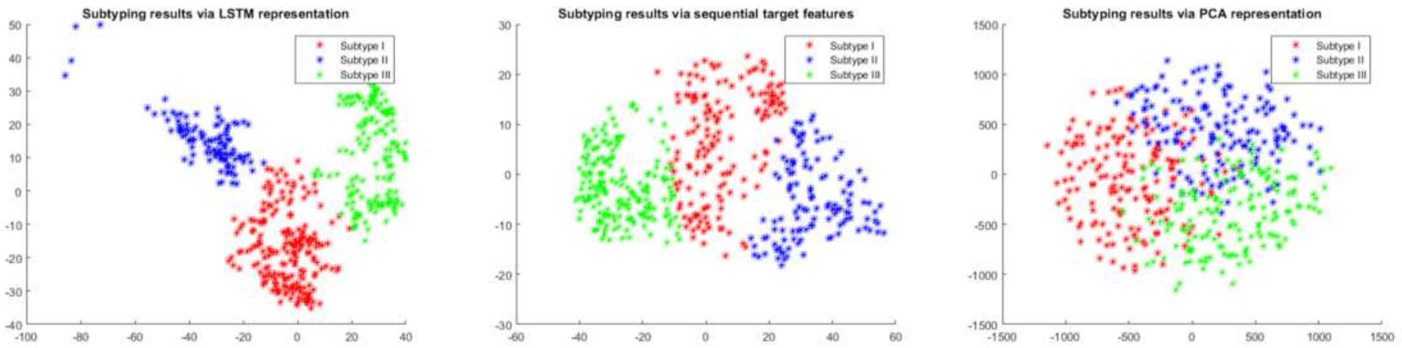


- Interpretation

Subtype I (43.1%)	Subtype II (22.9%)	Subtype III (33.9%)
58.79 years at baseline	61.93 years at baseline	65.32 years at baseline
Mild motor symptoms at baseline	Moderate motor symptoms at baseline	Poor motor symptoms at baseline
Mild non-motor symptoms at baseline	Moderate non-motor symptoms at baseline	Poor non-motor symptoms at baseline
Moderate motor decay	Mild motor decay	Severe motor decay

Xi Zhang, Jingyuan Chou, Jian Liang, Cao Xiao, Yize Zhao, Harini Sarva, Claire Henchcliffe, Fei Wang, Data-Driven Subtyping of Parkinson's Disease Using Longitudinal Clinical Records: A Cohort Study. Scientific Reports, Nature, 2018
.

Disease Subtyping



❖ Summarization

- ✓ This study is an initial attempt on leveraging advanced data analytics for identification of PD subtypes with longitudinal and heterogeneous clinical study data.
- ✓ Our approach has demonstrated strong potentials of identifcation of comprehensive progressive PD subtypes.

❖ Limitation

- ✓ the approach the deep learning (LSTM) procedure cannot be straightforwardly interpreted; Also, our study is only conducted on the PPMI cohort.

Source Code: https://github.com/sheryl-ai/Nature-Scientific-Reports

Xi Zhang, Jingyuan Chou, Jian Liang, Cao Xiao, Yize Zhao, Harini Sarva, Claire Henchcliffe, Fei Wang, Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Nature Scientific Reports, volume 9, Article number: 797 (2019).

Outline

Part 1: Disease Subtyping on Clinical Times Series

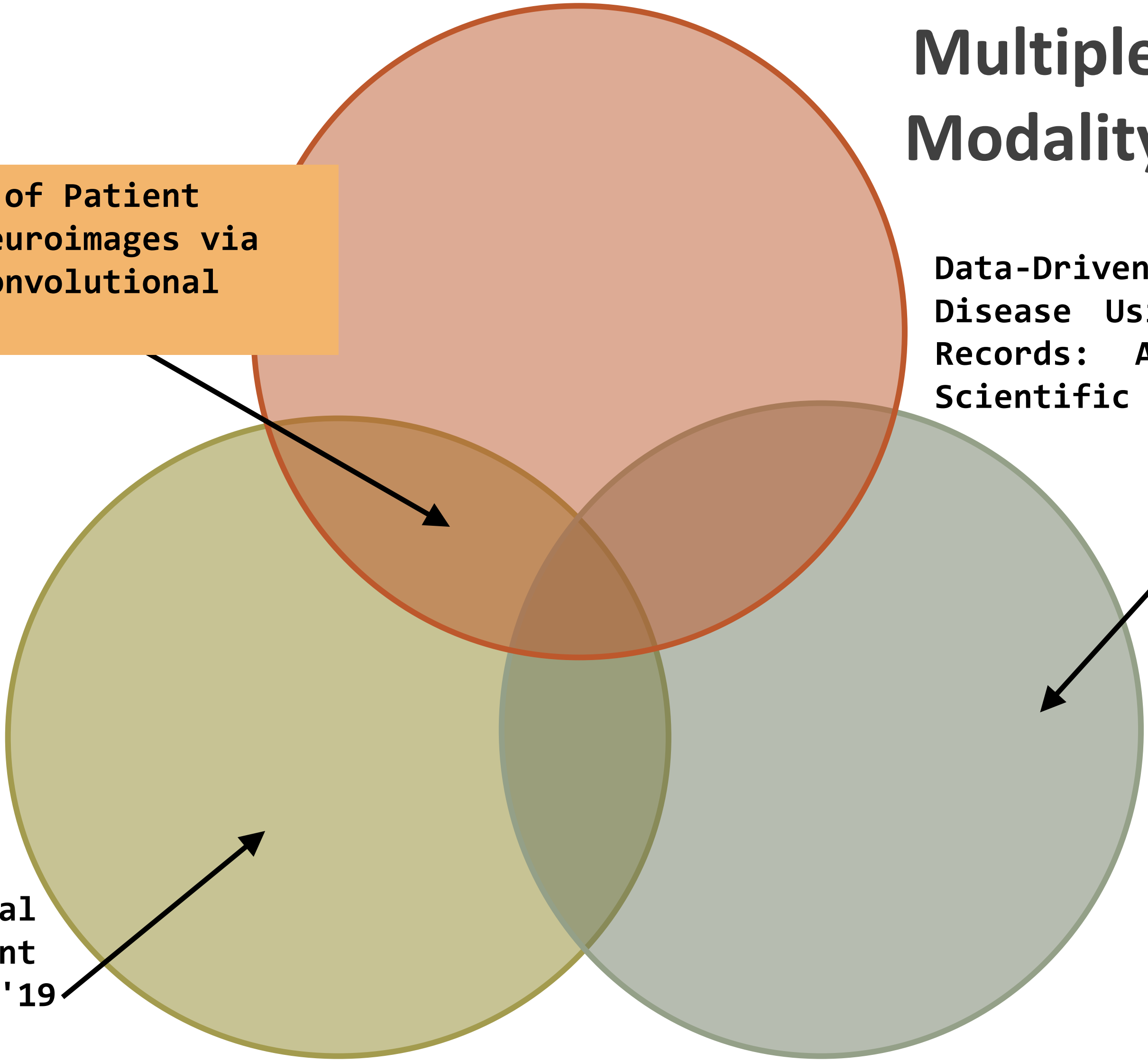
Part 2: Integrative Disease Analysis via Multi-Modality

Part3: Meta-Learning on Limited Clinical Resources

Obstacles

Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM’18

Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Nature Scientific Reports, 2019

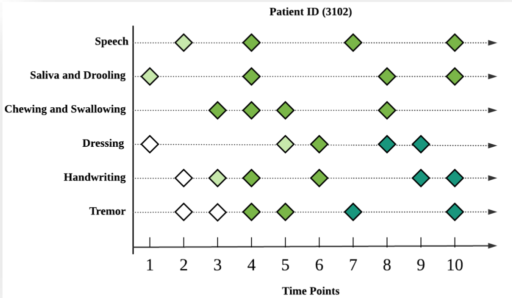


MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19

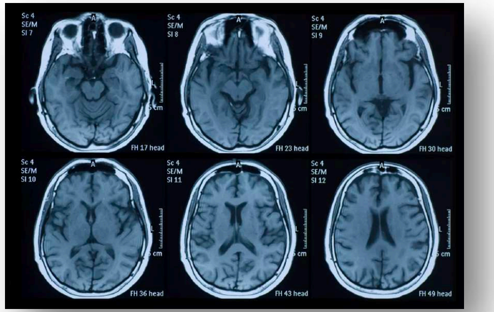
Multiple Modality Learning

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

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



Patients Health Records

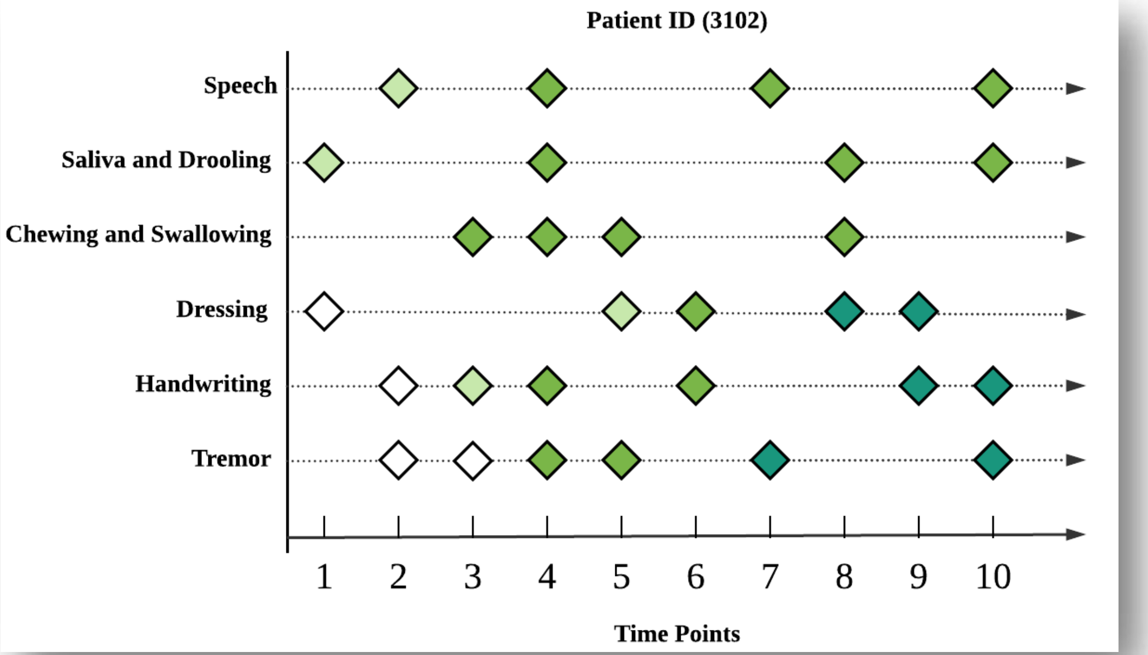


NeuroImages

Discriminating Patients and Health Controls

Multiple Modality Learning

❖ Modality I: Electronic Health Records (**Time Series**)

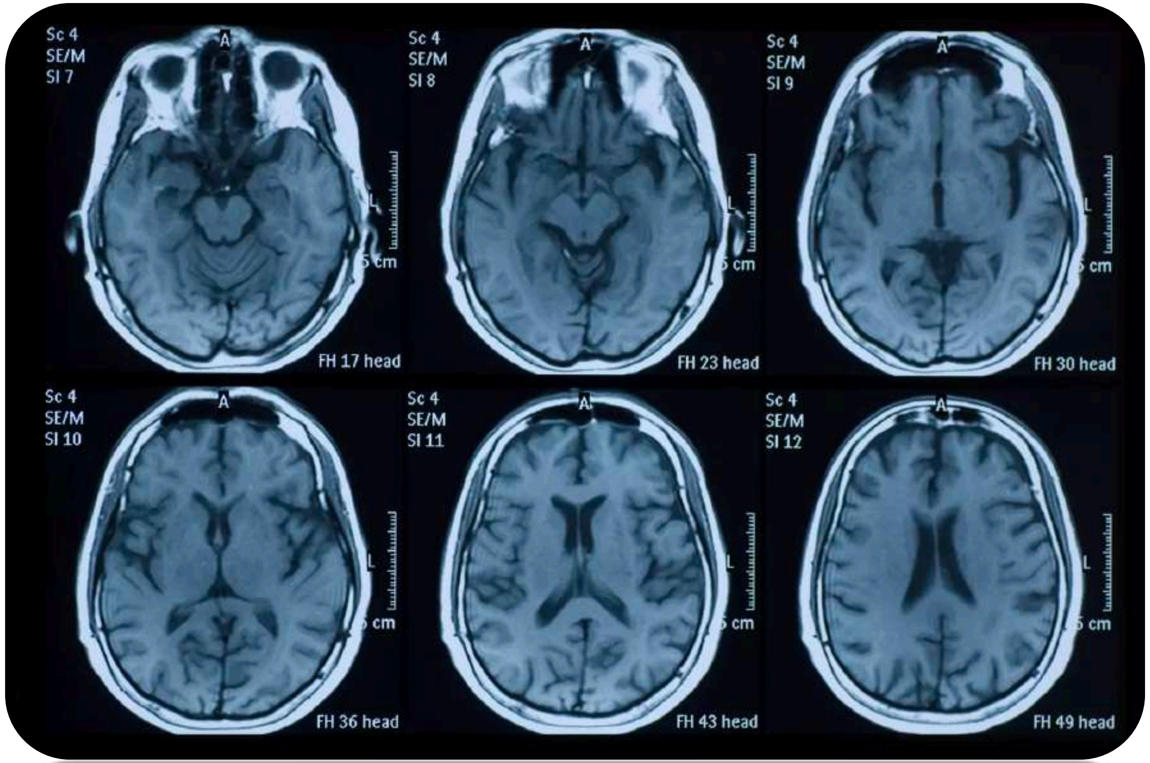


✓ **Sequential structure**

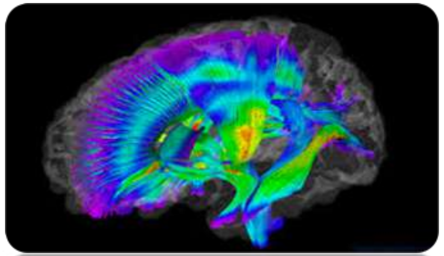
✓ **Missing Values**

Multiple Modality Learning

- Modality II: Neuroimages (**Graph Samples**)



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

Multiple Modality Learning

Challenges

- ***Multi-Modality.*** The nature of EHR and neuroimage are completely different where EHR data are sequential and a specific brain image is static, i.e., 3-dim tensor or graph.

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.

Multiple Modality Learning

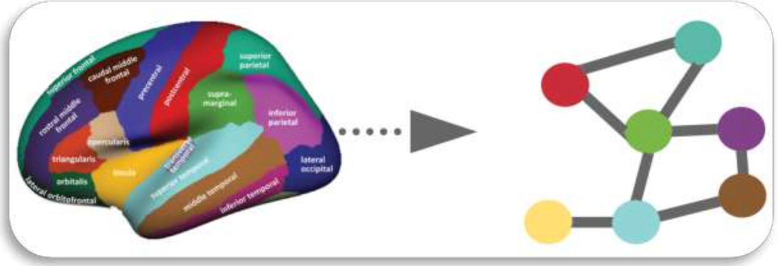
- Utilize 3-dimensional brain coordinates of ROIs

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

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

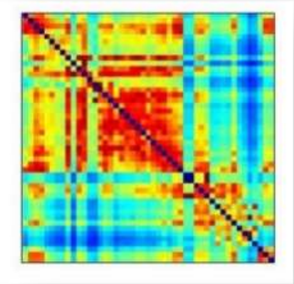


Multiple Modality Learning

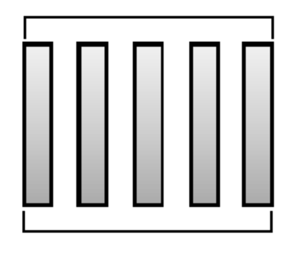
- Overview of the input data

sample-level

connectivity matrix



sequential vectors



population-level

spatial graph

Multiple Modality Learning

THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

- How about the amount of training data?

# of PD Subjects	# of HC Subjects
596	158

Small data modelling problem

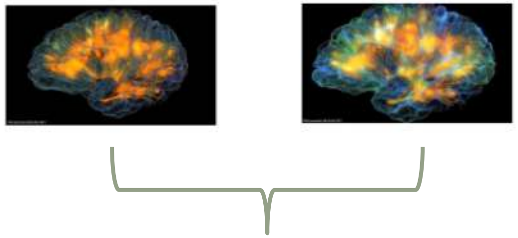
Multiple Modality Learning

- Pairwise-training strategy

# of Matching Samples	# of Non-Matching Samples
189,713	94,168

THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

a sample pair



matching vs.
non-matching

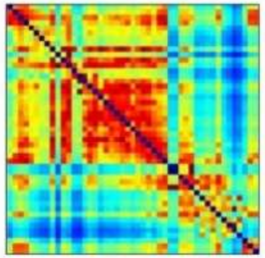
x		y
PD	PD	*"same"*
PD	HC	*"different"*
HC	PD	*"different"*
HC	HC	*"same"*

Multiple Modality Learning

❖ The Learning Problem

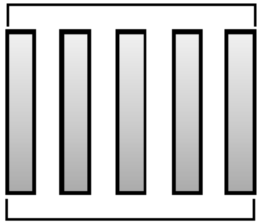
sample-level

- connectivity matrix



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

- sequential vectors



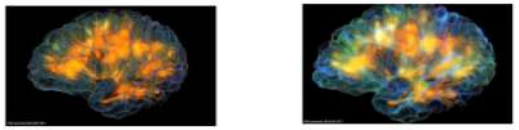
$\mathbf{s}_j, j = 1, \cdots, t$

population-level

- spatial graph

$\mathcal{G}$

a sample pair



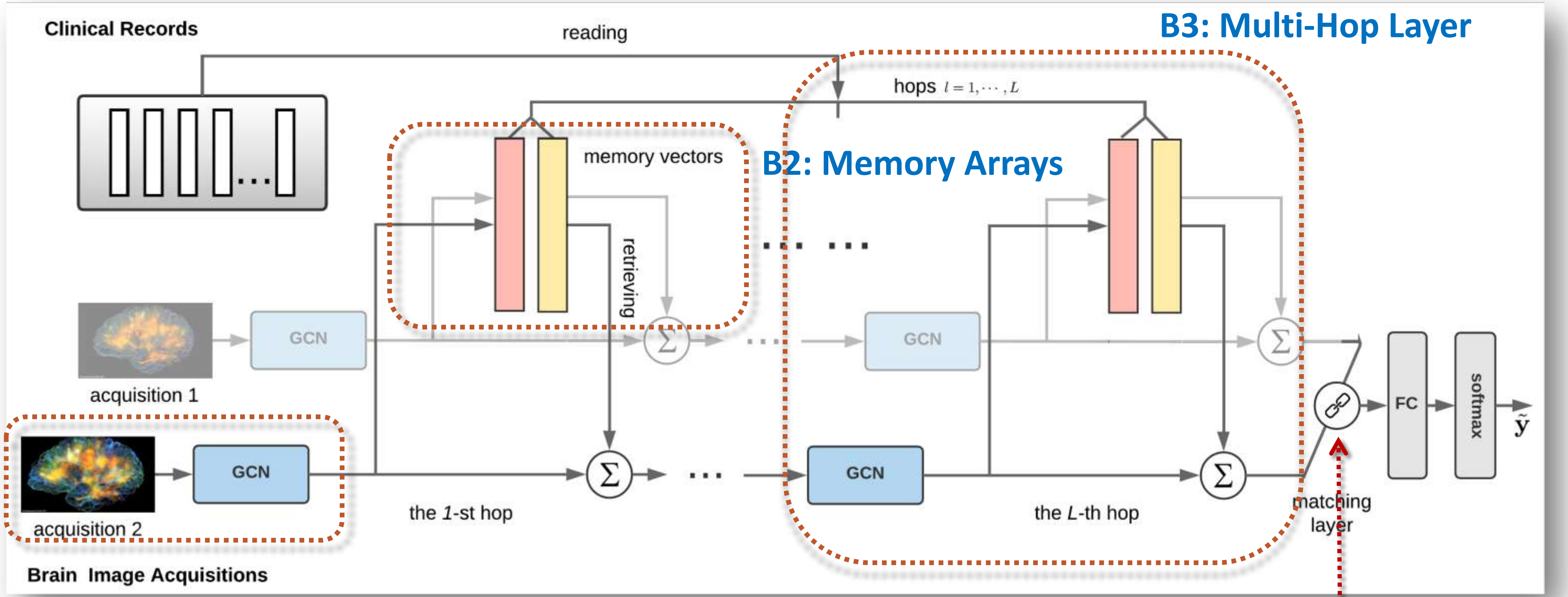
$\underbrace{\qquad\qquad}$

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

x		**y**
PD	PD	*"same"*
PD	HC	*"different"*
HC	PD	*"different"*
HC	HC	*"same"*

Multiple Modality Learning

- Network Architecture



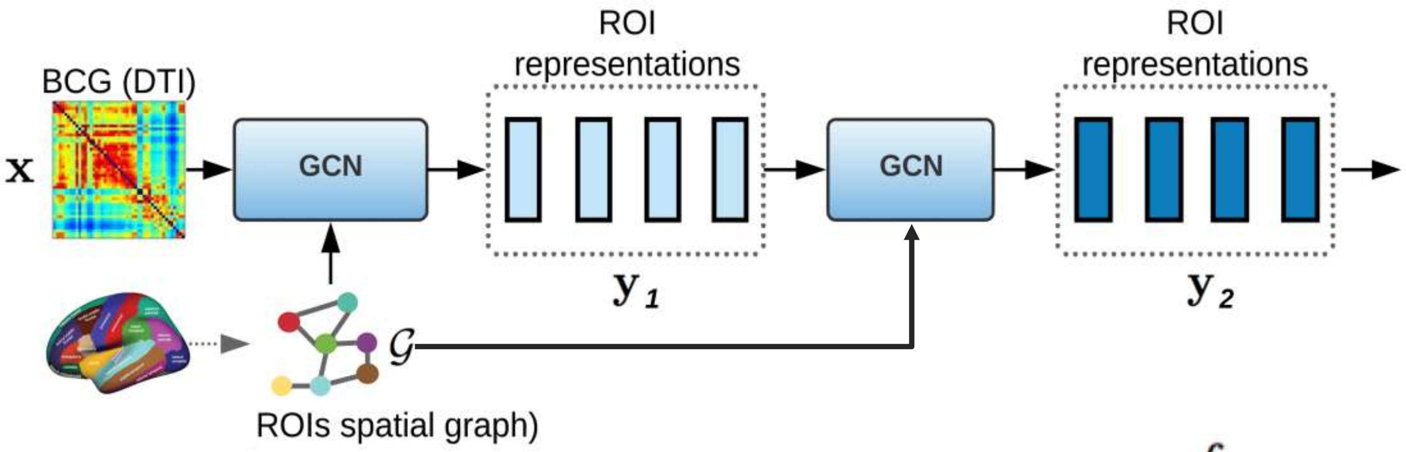
B1: Graph

Convolutional Network

B4: Matching Layer

Multiple Modality Learning

❖ B1: Graph Convolutional Network (GCN)



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

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

Graph Fourier Transform

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

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

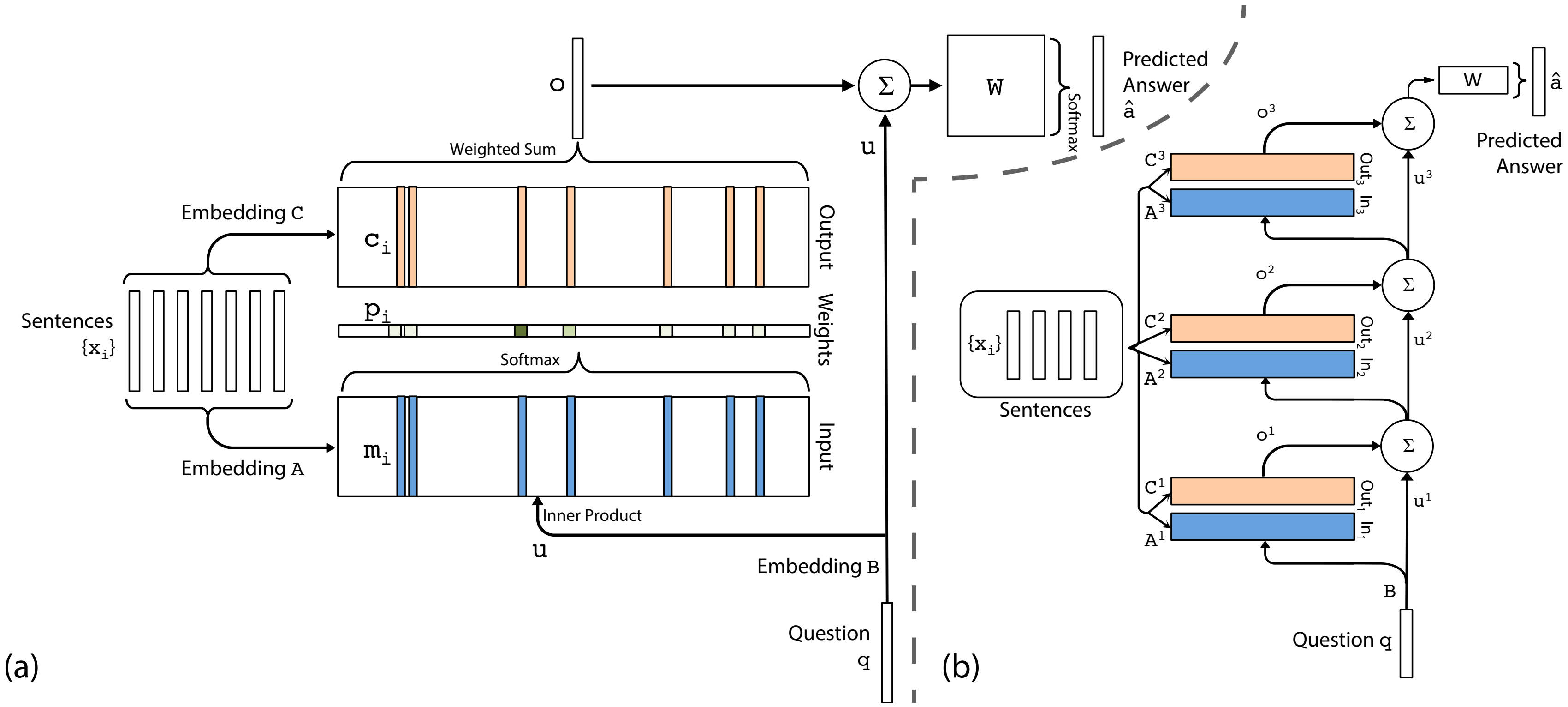
A general operator

ChebNet

Chebyshev polynomial

Multiple Modality Learning

❖ Preliminary: End-to-End Memory Network



Sukhbaatar et al. '15

Multiple Modality Learning

- B2: Memory-Augmented GCN (MemGCN)
  - ✓ Clinical Sequences Reading

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

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

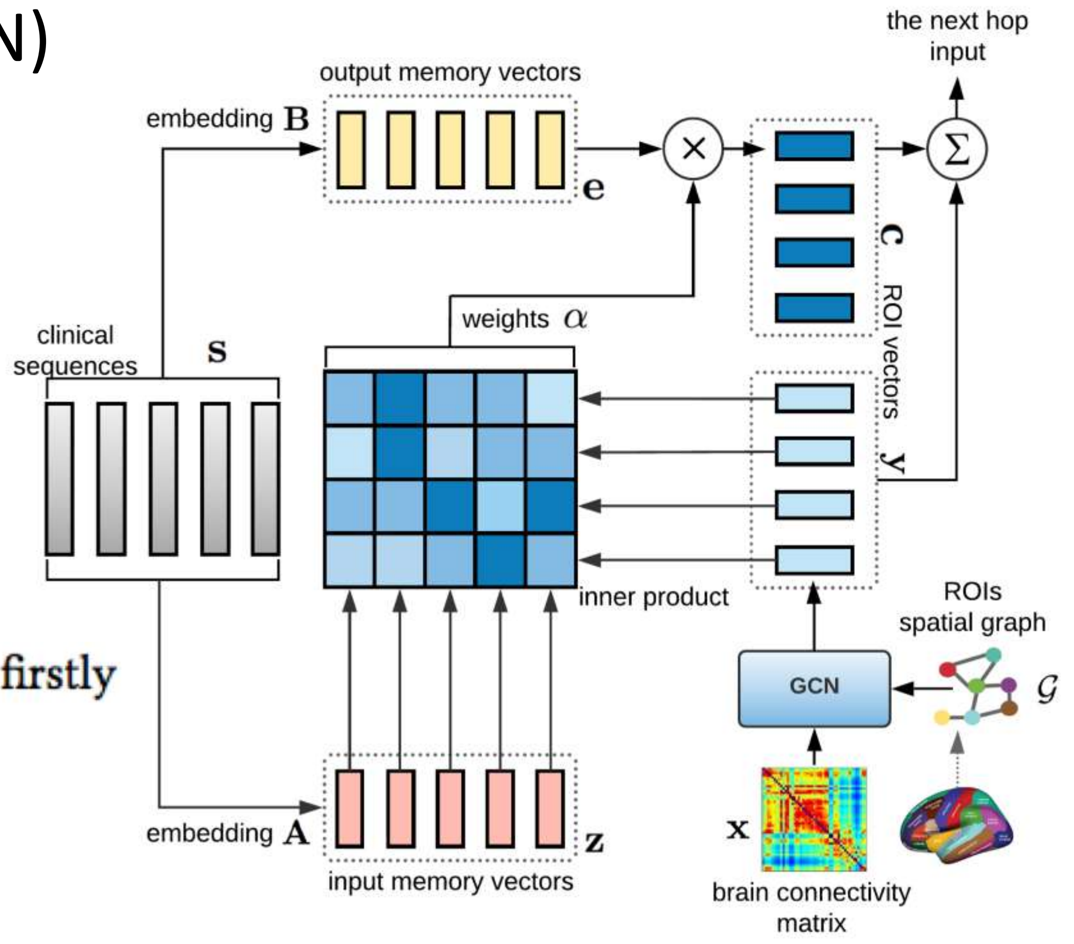
    $$\text{output memory} \quad \mathbf{e}_j = \mathbf{B}\mathbf{s}_j$$

  - ✓ Memory Representation Retrieving

    To retrieve memory vectors from the embedding space, we firstly

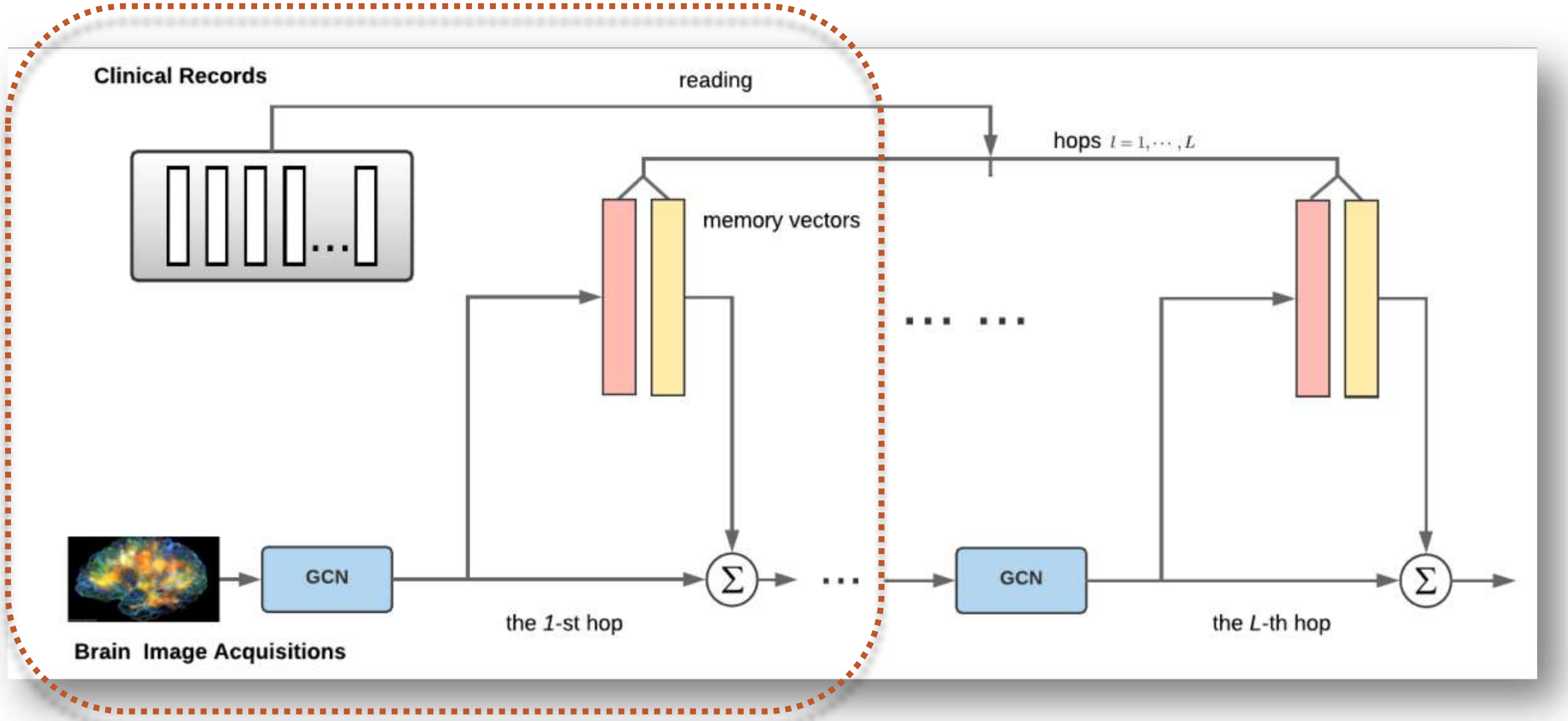
    $$\alpha_{ij} = softmax(\mathbf{y}_i\mathbf{z}_j) = \frac{\exp(\mathbf{y}_i\mathbf{z}_j)}{\sum_{j'=1}^{t}\exp(\mathbf{y}_i\mathbf{z}_{j'})}$$

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



Multiple Modality Learning

- B3: Extend to multiple hop architecture



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

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

Multiple Modality Learning

- B4: Matching Layer
  - ✓ 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

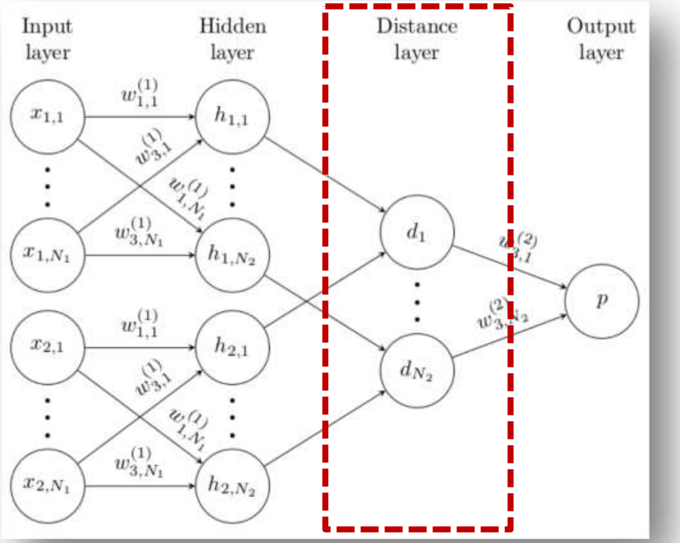
$$sim_{i,j}(\mathbf{x}_m, \mathbf{x}_{m'}) = (\mathbf{y}^L_{m,i})^{\mathrm{T}} \mathbf{M} \mathbf{y}^L_{m',j}, \; i, j = 1, \cdots, n.$$

parameter matrix

Siamese-like Network

Impose structure➔

Learning a metric space



Koch et al. ’15

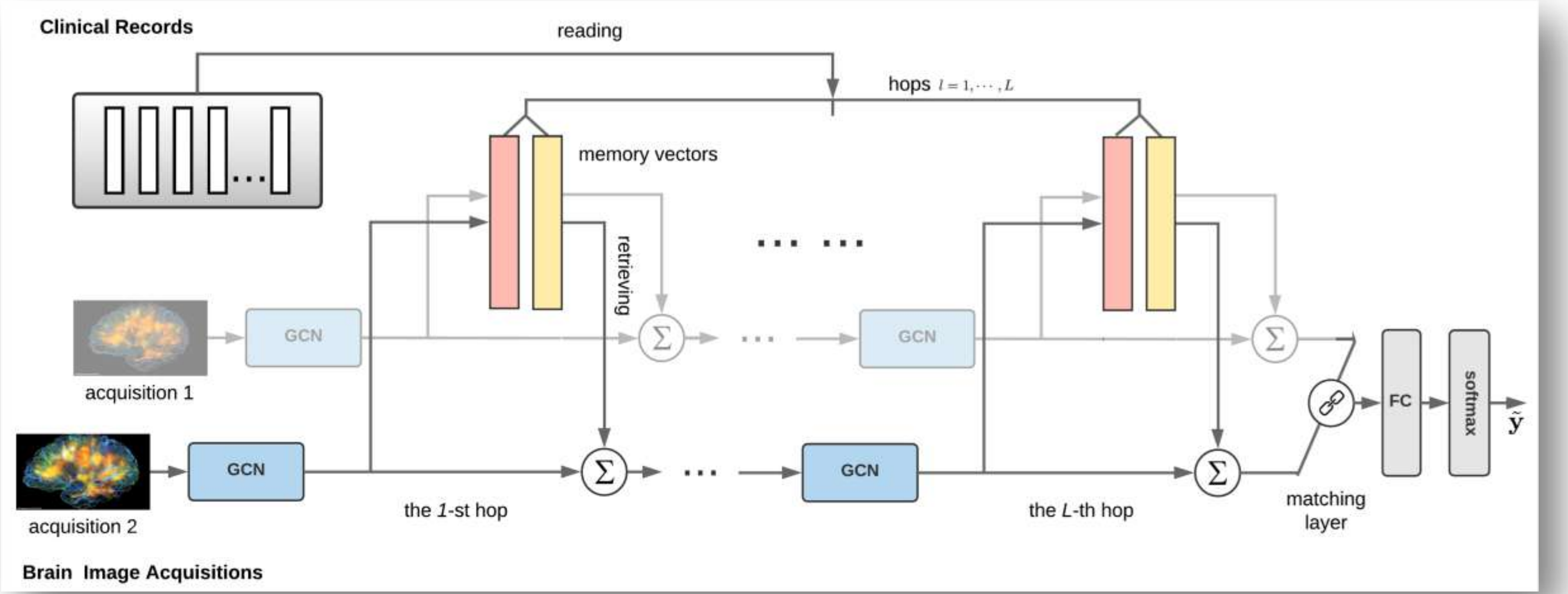
Multiple Modality Learning

❖ Objective Function (cross entropy, pairwise strategy)

$$\mathcal{L} = \sum_{m,m'}^{N} \tilde{\mathbf{y}}_{m,m'} \log \mathbf{p}_{m,m'} + (1 - \tilde{\mathbf{y}}_{m,m'}) \log(1 - \mathbf{p}_{m,m'}) + \gamma \|\mathbf{\Theta}\|_2$$

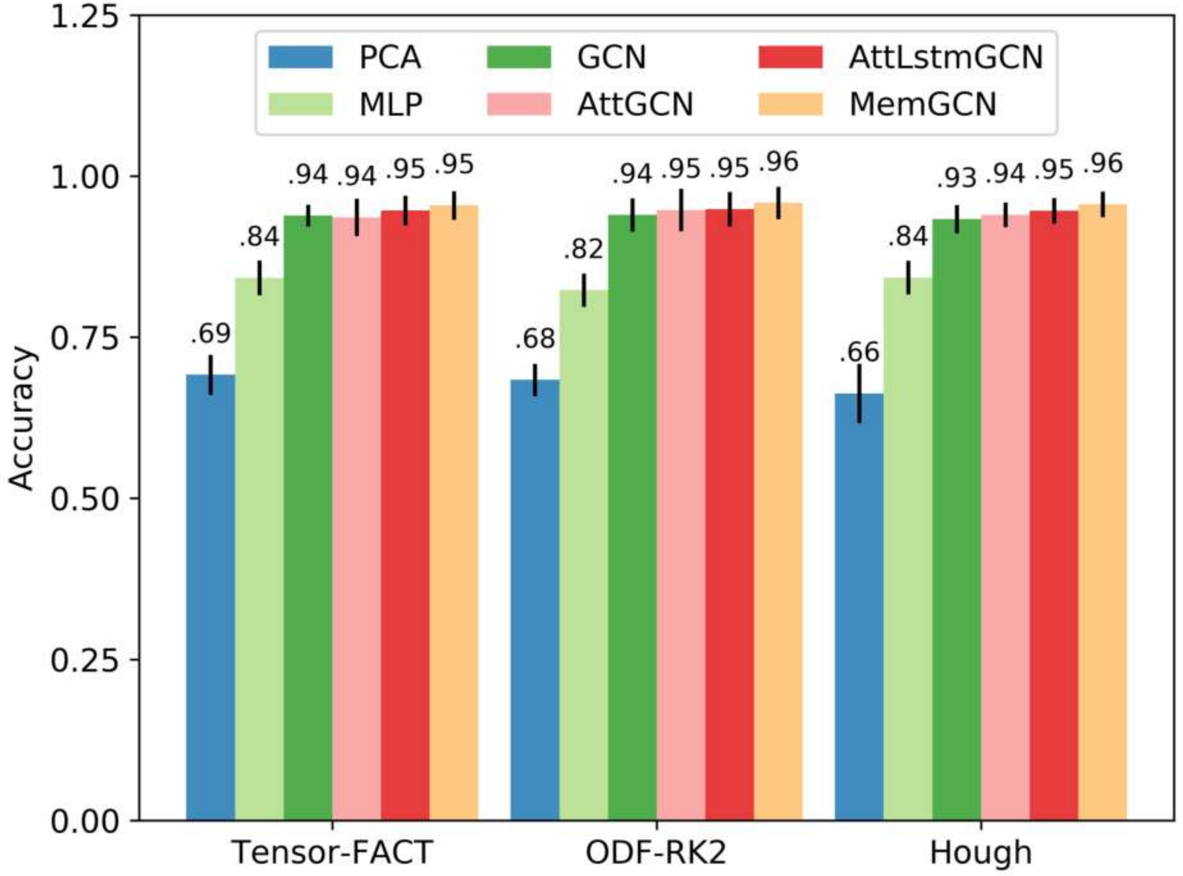
where $\tilde{\mathbf{y}}_{m,m'}$ denotes the label for sample pair $(\mathbf{x}_m, \mathbf{x}_{m'})$

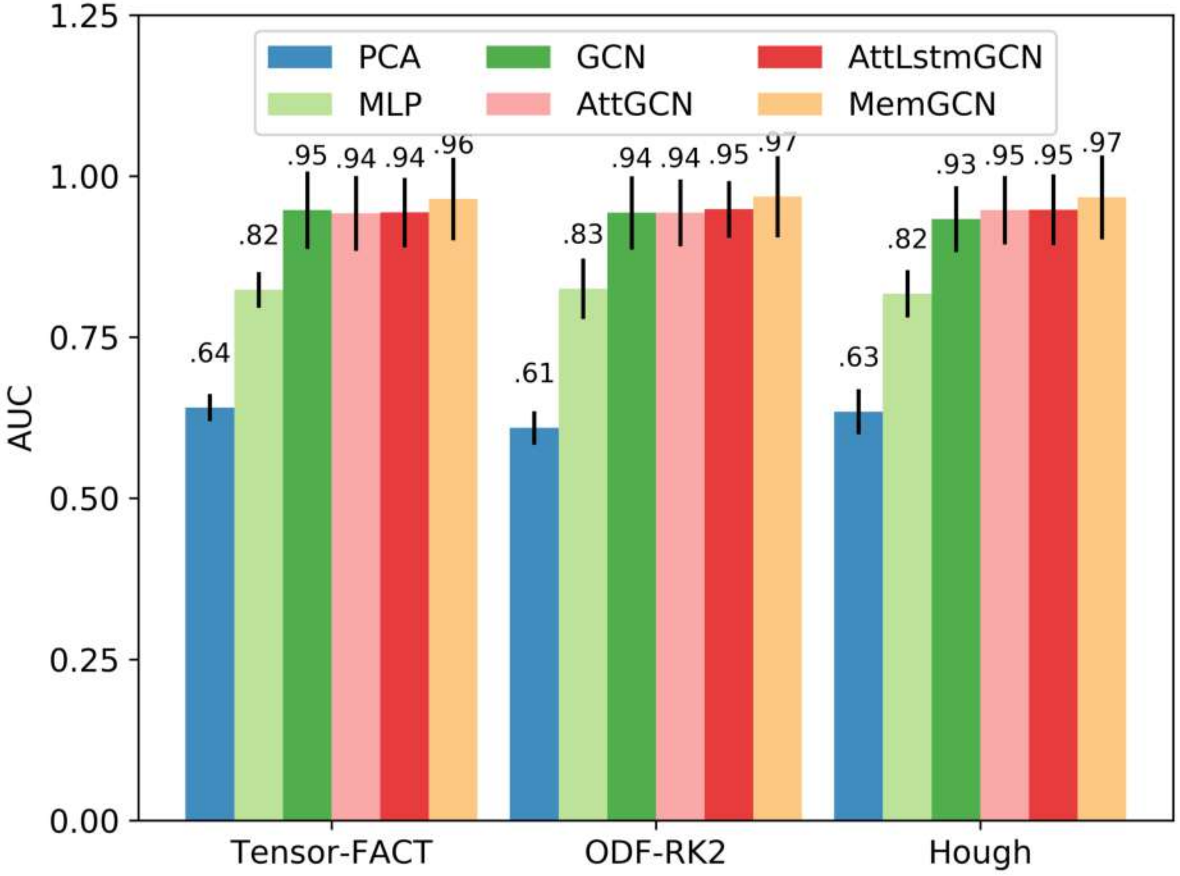
$$\mathbf{p} = softmax(\mathbf{w}_c^{\mathrm{T}} \mathbf{r})$$



Multiple Modality Learning

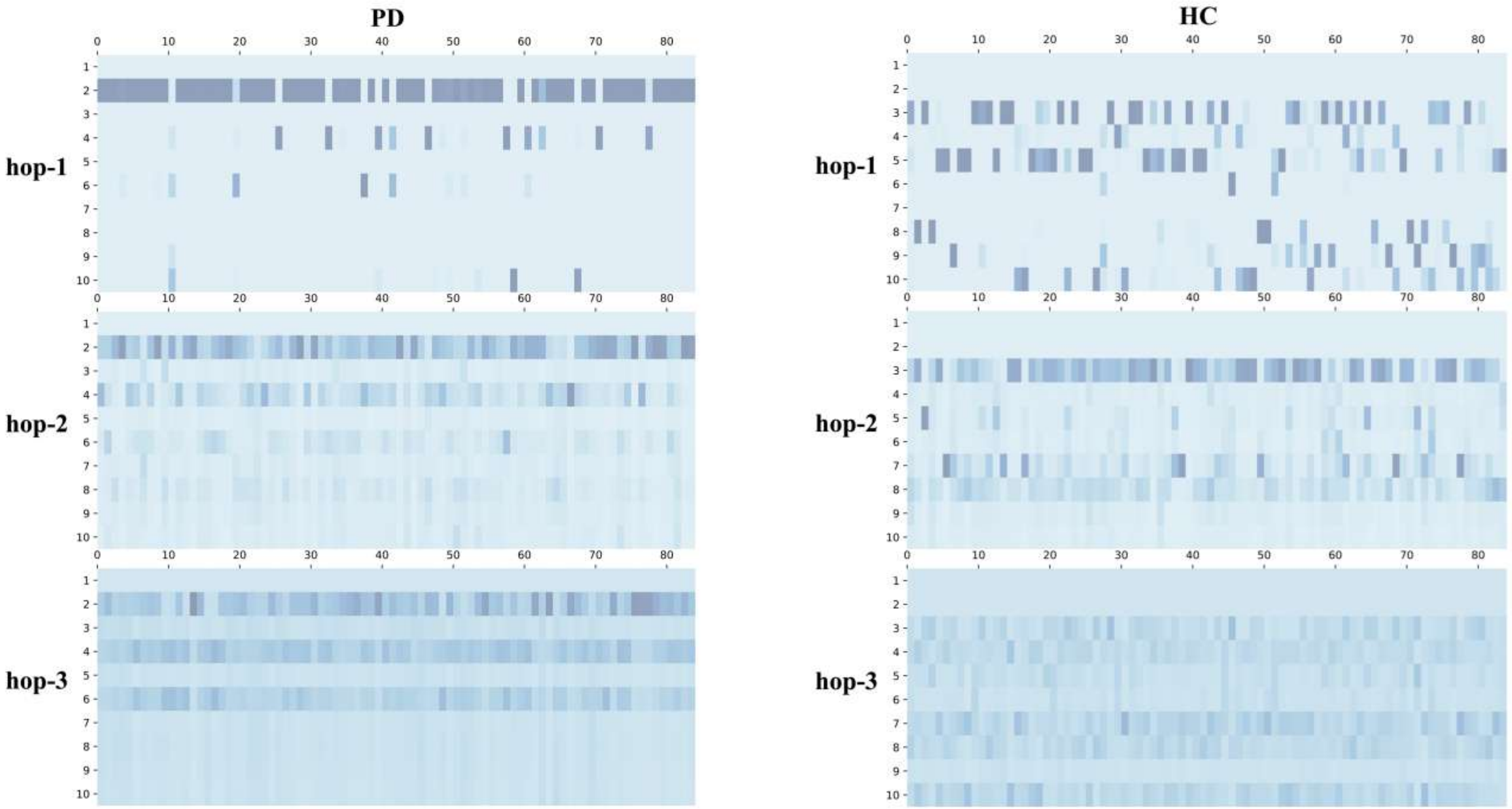
- Matching vs Non-Matching Results

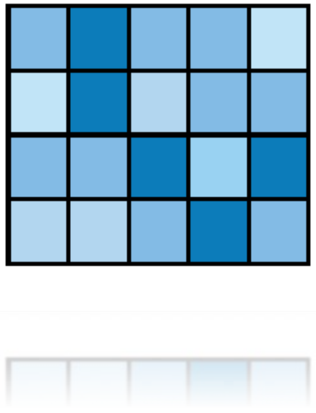




Multiple Modality Learning

❖ Longitudinal Alignment: Case Study





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

Multiple Modality Learning

Interpretation: Learned Similarity (Region Scores)

- 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.

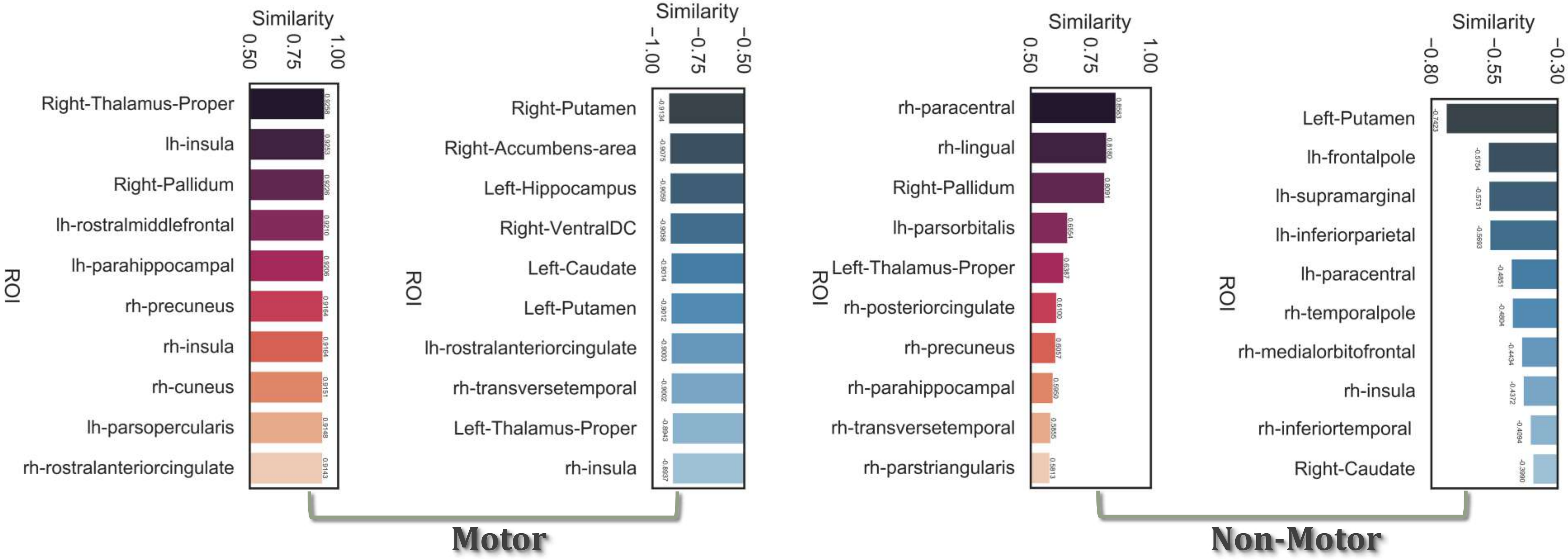
Ref: 1. Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects, *Brain,* Volume 118, Issue 4, August 1995, Pages 913–933;

2. The functions of the basal ganglia and the paradox of stereotaxic surgery in Parkinson's disease, *Brain*, Volume 117, Issue 4, August 1994, Pages 877–897;

3. Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls, *Brain*, Volume 127, Issue 4, April 2004, Pages 791–800.

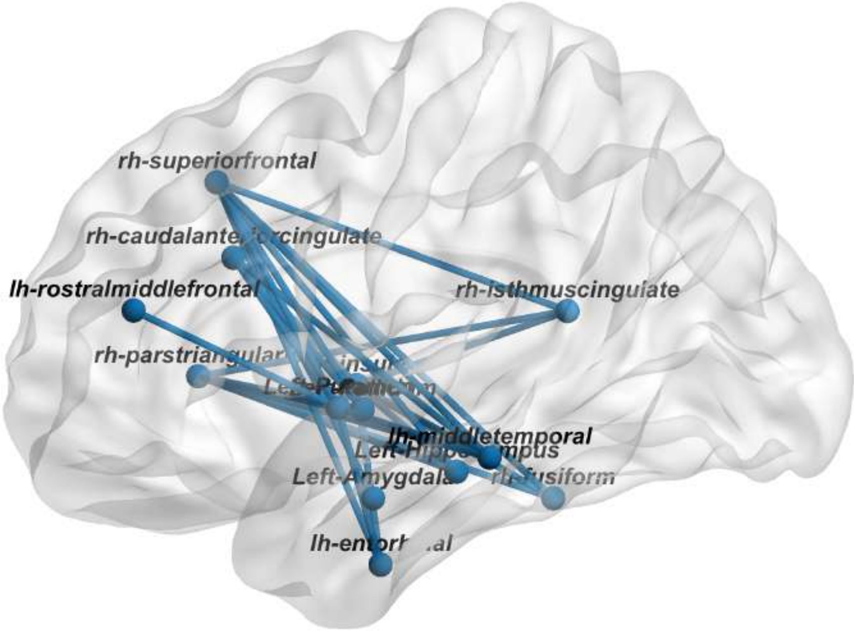
Multiple Modality Learning

- Interpretation: Learned Similarity (Region Scores)



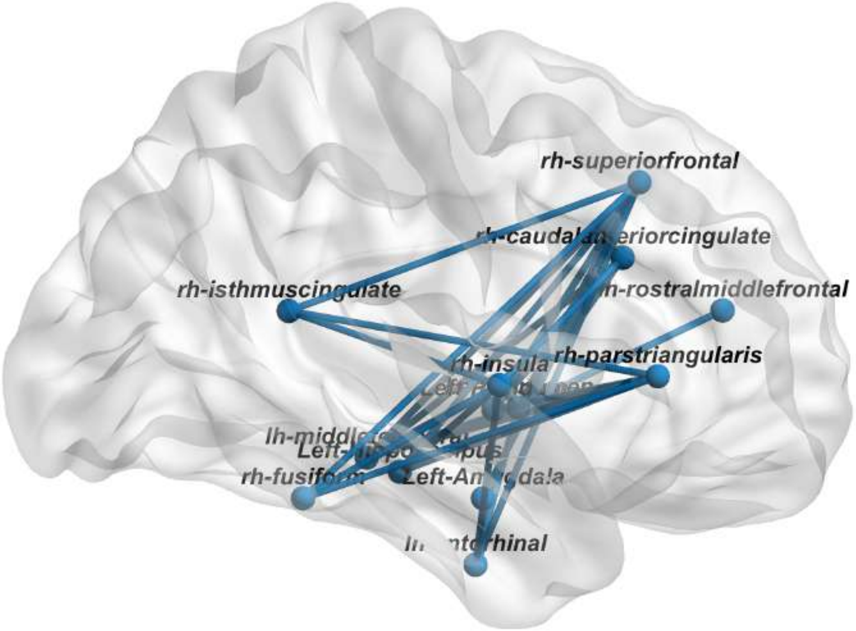
Multiple Modality Learning

❖ Interpretation: Learned Similarity (Identical Connection)



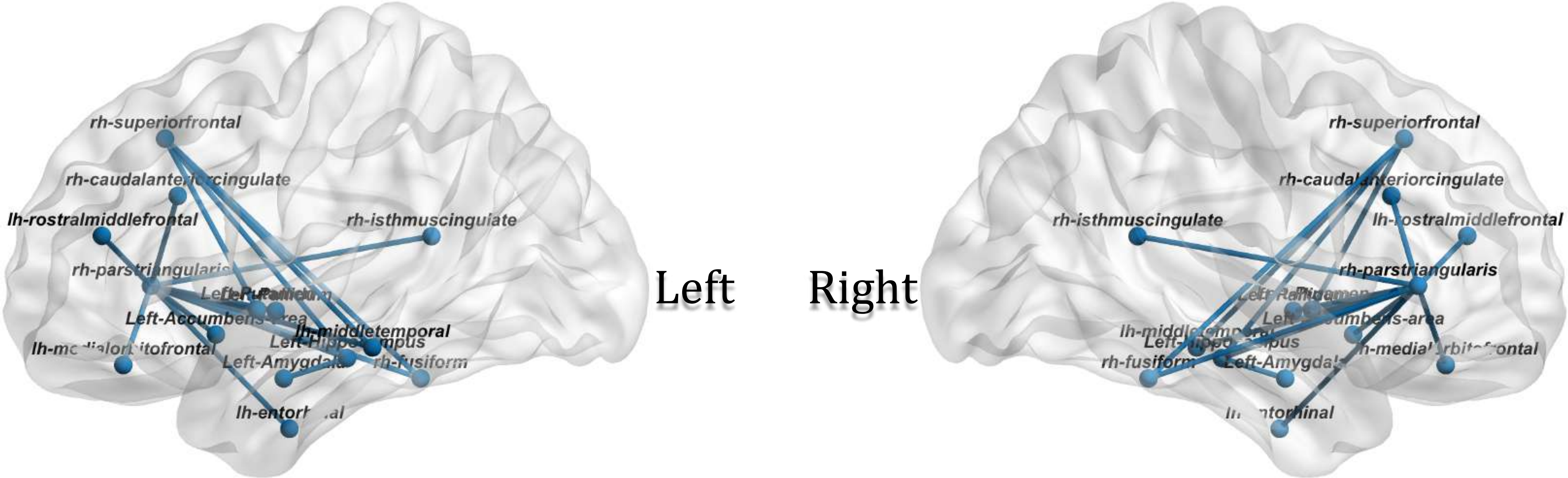
Left

Right



Multiple Modality Learning

❖ Interpretation: Learned Similarity (Discriminative Connection)

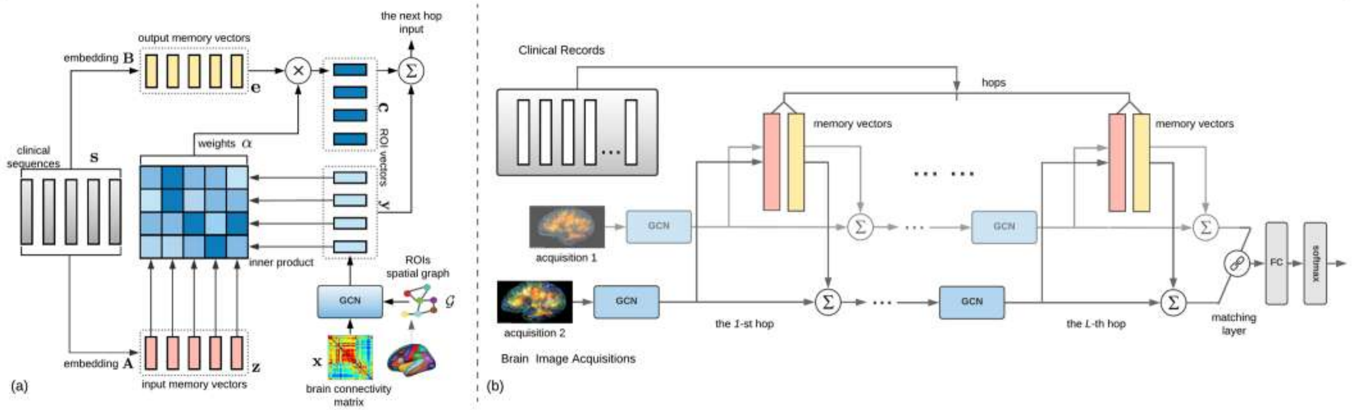


Source Code: https://github.com/sheryl-ai/MemGCN

Xi Zhang, Jingyuan Chou, Fei Wang, Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM’18: IEEE International Conference on Data Mining, 2018.

Multiple Modality Learning

❖Summary



- ✓ Making a progress in modelling a small cohort data such as PPMI.
- ✓ Interpretable high-level representations extracted from MemGCN are explored.
- ✓ Experiments on classification of Parkinson's Disease demonstrate the superiority of MemGCN.

Source Code: https://github.com/sheryl-ai/MemGCN

Xi Zhang, Jingyuan Chou, Fei Wang, Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM'18: IEEE International Conference on Data Mining, 2018.

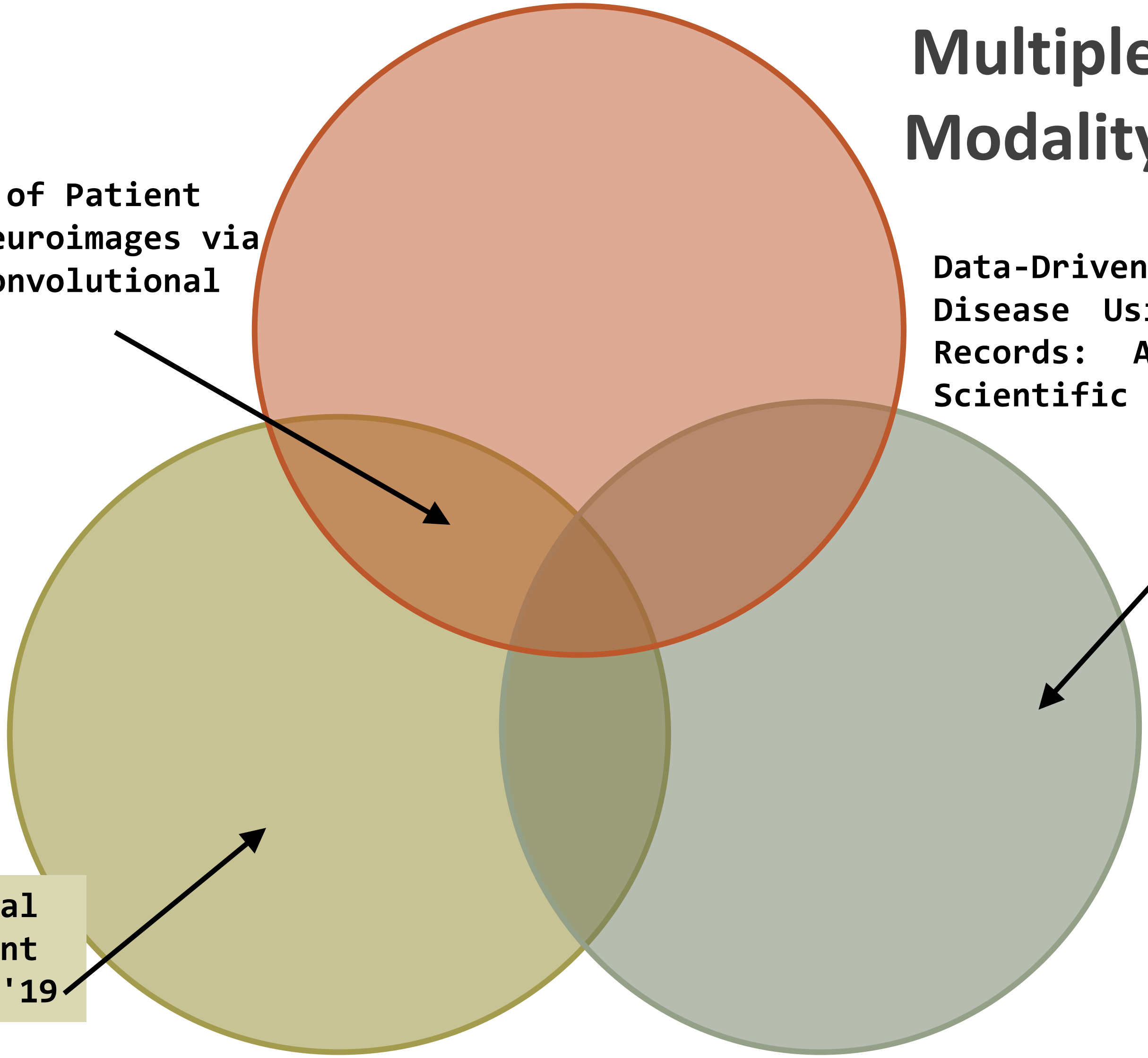
Outline

Part 1: Disease Subtyping on Clinical Times Series

Part 2: Integrative Disease Analysis via Multi-Modality

Part3: Meta-Learning on Limited Clinical Resources

Obstacles



Multiple Modality

Integrative Analysis of Patient Health Records and Neuroimages via Memory-based Graph Convolutional Network. ICDM’18

Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Nature Scientific Reports, 2019

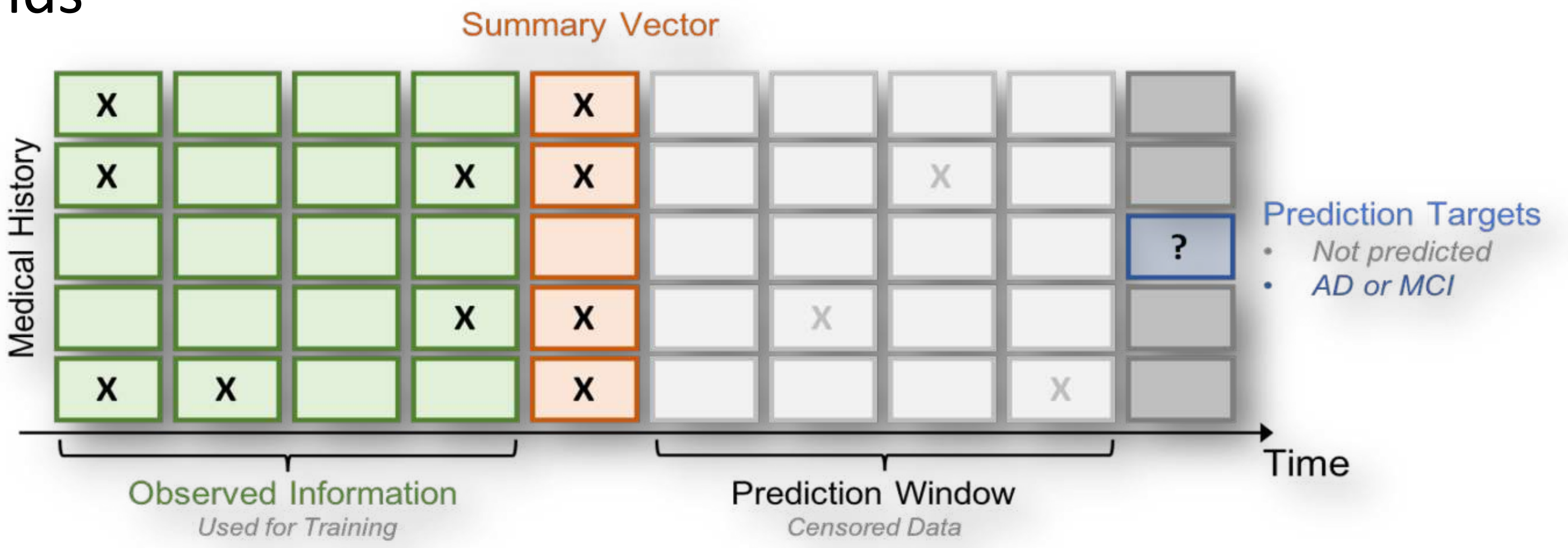
Data Scarcity

Heterogeneity

MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19

Meta-Learning on Limited Clinical Resource

❖ Backgrounds



- Patient EHRs: each patient has a sequence of vectors;
- Predictive models: build for clinical risks, such as in-hospital mortality, hospital readmission, chronic disease onset, condition exacerbation, etc.
  - LR, SVM, k-Nearest Neighbor, Random Forest, MLP;
  - RNN, CNN.

Meta-Learning on Limited Clinical Resource

- How about patient samples that are insufficient?
  - it is expensive and sometimes even impossible for obtaining labelled new samples
  - reusing data on other domain/tasks becomes a feasible strategy

    - transfer learning

    - meta-learning (learning to transfer)

Using the learning experiences from a set of relevant tasks ...

Meta-Learning on Limited Clinical Resource

Challenges

- ***Data Scarcity:*** EHRs are suffering sparsity, irregularity, temporality;
- ***Label Insufficient:*** labelled samples in medicine (patients) are relatively limited, and creates troubles for building an effective predictive model.

Solution

- We proposed a **MetaPred**, a model agnostic meta-learning framework for low-resource predictive modelling with patient EHRs.

Meta-Learning on Limited Clinical Resource

❖ Motivation

Goal: is to predict the risks of target disease with few labeled patients, which give rise to a low-resource classification.

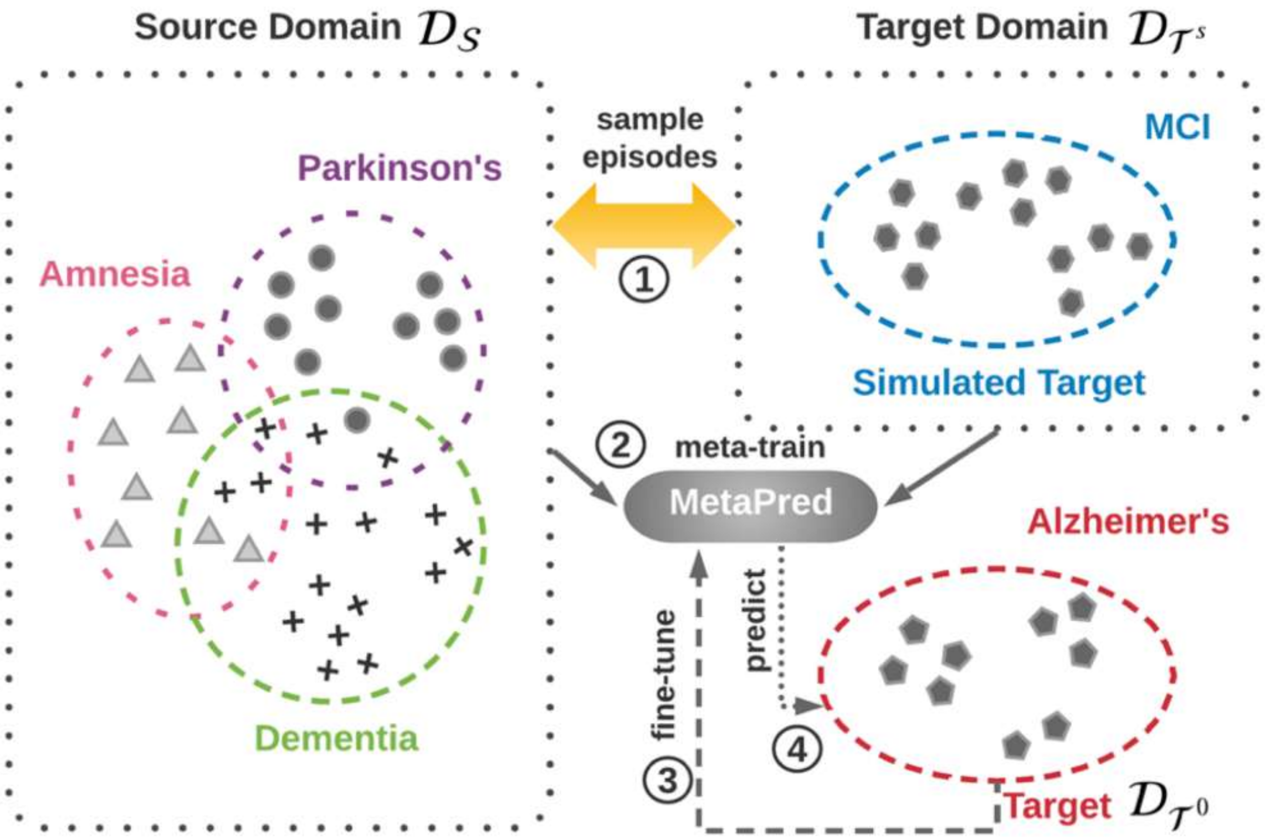
The idea: is to take advantage of labeled patients from other relevant high-resource domains and design the learning to transfer framework with sources and a simulated target.

Meta-Learning for Clinical Risk Prediction

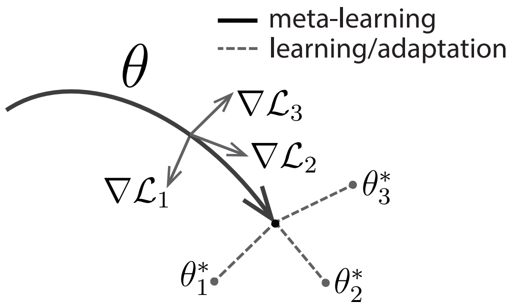
❖ Problem Setup

4 steps:

- ✓ sample episode
- ✓ meta-train
- ✓ fine-tune
- ✓ predict



Meta-Learning on Limited Clinical Resource



- Preliminary: MAML

Meta-learning, also known as *learning to learn*, aims to optimize the objective over a variety of learning tasks $\mathcal{T}$ which are associated with the corresponding datasets $\mathcal{D}_{\mathcal{T}}$.

$$\Theta^* = \arg\min_{\Theta} \mathbb{E}_m \mathbb{E}_{\mathcal{D}_{epi}^m \sim p(\mathcal{D}_{\mathcal{T}})} \mathcal{L}_{\Theta}(\mathcal{D}_{\mathcal{T}})$$

Loss function: $$\mathcal{L}_{\Theta} = \frac{1}{|\mathcal{D}_{epi}^{te}|} \sum_{(\mathrm{X}_i, \mathrm{y}_i) \in \mathcal{D}_{epi}^{te}} \mathcal{L}_{\Theta}\left((\mathrm{X}_i, \mathrm{y}_i); \mathcal{D}_{epi}^{tr}\right)$$

Finn et al. '17

Meta-Learning on Limited Clinical Resource

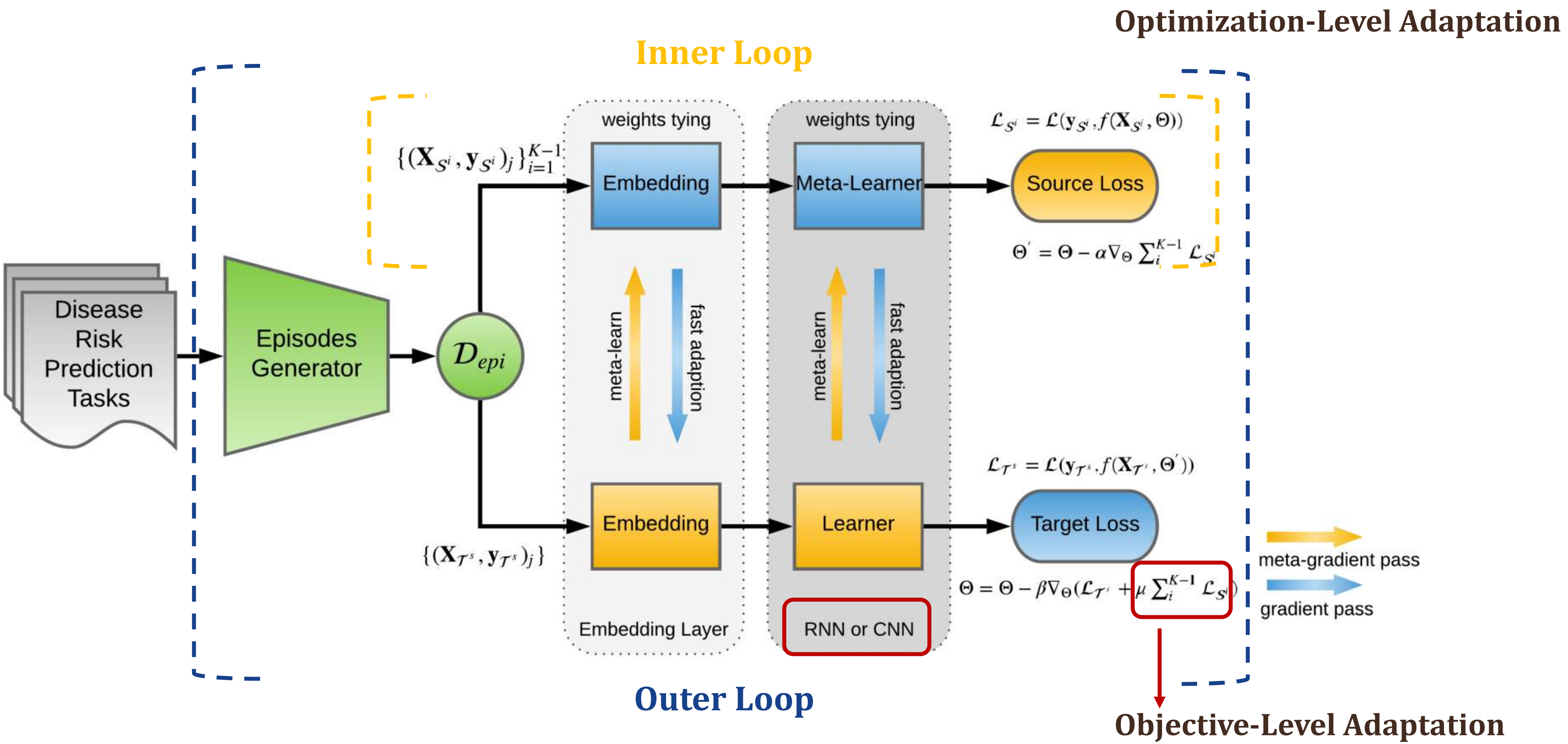
- Two-level adaptation

$$\Theta^* = \textbf{Learner}(\mathcal{T}^s; \textbf{MetaLearner}(\mathcal{S}^1, \cdots, \mathcal{S}^{K-1}))$$

$$\Theta^* = \textbf{Learner}\left(\mathcal{T}^s, \{\mathcal{S}^i\}_i^{K-1}; \textbf{MetaLearner}(\{\mathcal{S}^i\}_i^{K-1})\right)$$

- ✓ ***Parameter-level adaptation***: Model-agnostic, typing parameters for source/target domain.
- ✓ ***Objective-level adaptation***: bounded target risk via empirical risk of source domains.

Meta-Learning on Limited Clinical Resource: Framework



Meta-Learning on Limited Clinical Resource

- Objective adaptation

Domain Adaptation Bound Ben-David et al. '10

Theorem 1 *For a hypothesis h,*

$$\epsilon_T(h) \le \epsilon_S(h) + d_1(\mathcal{D}_S, \mathcal{D}_T) + \min\left\{\mathrm{E}_{\mathcal{D}_S}\left[|f_S(\mathbf{x}) - f_T(\mathbf{x})|\right], \mathrm{E}_{\mathcal{D}_T}\left[|f_S(\mathbf{x}) - f_T(\mathbf{x})|\right]\right\}.$$

$$\mathcal{L}_{\mathcal{T}}(f_{\Theta'}) = \mathcal{L}_{\mathcal{T}^s}(f_{\Theta'}) + \mu \sum_{i}^{K-1} \mathcal{L}_{\mathcal{S}^i}(f_{\Theta})$$

$$= \sum_{\mathcal{D}_{epi}^{\mathcal{T}^s}} \mathcal{L}\left(\mathbf{y}_{\mathcal{T}^s}, f(\mathbf{X}_{\mathcal{T}^s}, \Theta')\right) + \boxed{\mu \sum_{i}^{K-1} \sum_{\mathcal{D}_{epi}^{\mathcal{S}^i}} \mathcal{L}\left(\mathbf{y}_{\mathcal{S}^i}, f(\mathbf{X}_{\mathcal{S}^i}, \Theta)\right)}$$

all the domains share the same feature space

Meta-Learning on Limited Clinical Resource

❖ Parameter adaptation

Feedforward 1:
Embedding → LSTM → MLP

Feedforward 2:
Embedding → 1d-CNN → MLP

Obj: Source Domains

Obj: Simulated Target Domain

Algorithm 1 MetaPred Training

Require: Source domains $\mathcal{S}^i$; Simulated target domain $\mathcal{T}^s$;
Require: Hyperparameters α, β, μ;

Initialize model parameter Θ randomly
while Outer-Loop not done **do**
  Sample batch of episodes $\{\mathcal{D}_{epi}\}$ from $\mathcal{D}_{\mathcal{S}^i}$ and $\mathcal{D}_{\mathcal{T}^s}$
  while Inner-Loop not done **do**
    $\{(\mathbf{X}_{\mathcal{S}^i}, \mathbf{y}_{\mathcal{S}^i})\}_{i=1}^{K-1}, \{(\mathbf{X}_{\mathcal{T}^s}, \mathbf{y}_{\mathcal{T}^s})\} = \{\mathcal{D}_{epi}\}$
    Compute $\mathcal{L}_{\mathcal{S}^i} = \mathcal{L}(\mathbf{y}_{\mathcal{S}^i}, f(\mathbf{X}_{\mathcal{S}^i}, \Theta)), i = 1, \cdots, K-1$
    Parameter fast adaption with gradient descent:
    $\Theta' = \Theta - \alpha \nabla_\Theta \sum_i^{K-1} \mathcal{L}_{\mathcal{S}^i}$
  end while
  Compute $\mathcal{L}_{\mathcal{T}^s} = \mathcal{L}(\mathbf{y}_{\mathcal{T}^s}, f(\mathbf{X}_{\mathcal{T}^s}, \Theta'))$
  Update $\Theta = \Theta - \beta \nabla_\Theta (\mathcal{L}_{\mathcal{T}^s} + \mu \sum_i^{K-1} \mathcal{L}_{\mathcal{S}^i})$ using Adam
end while

Meta-Learning on Limited Clinical Resource

OREGON HEALTH & SCIENCE UNIVERSITY

❖ Datasets

Disease	ICD-9 Codes
Mild Cognitive Impairment	331.83, 331.89, 331.9
Alzheimer's Disease	331.0, 331.2, 331.6, 331.7
Parkinson's Disease	332.*
Dementia	290.*, 291.*, 294.*, 331.82
Amnesia	780.93
Huntington's Disease	333.4
Mechanical Obstructions	331.3, 331.4, 331.5
Frontotemporal Dementia	331.1, 331.11, 331.19

* means that all the codes in this diagnosis group are included.

Targets

Domain	Case	Control	# of visit	Ave. # of visit
MCI	1,965	4,388	161,773	22.24
Alzheimer's	1,165	4,628	136,197	20.73
Parkinson's	1,348	3,588	105,053	20.01
Dementia	3,438	1,591	98,187	18.06
Amnesia	2,974	4,215	180,091	21.60

Meta-Learning on Limited Clinical Resource

OREGON HEALTH & SCIENCE UNIVERSITY

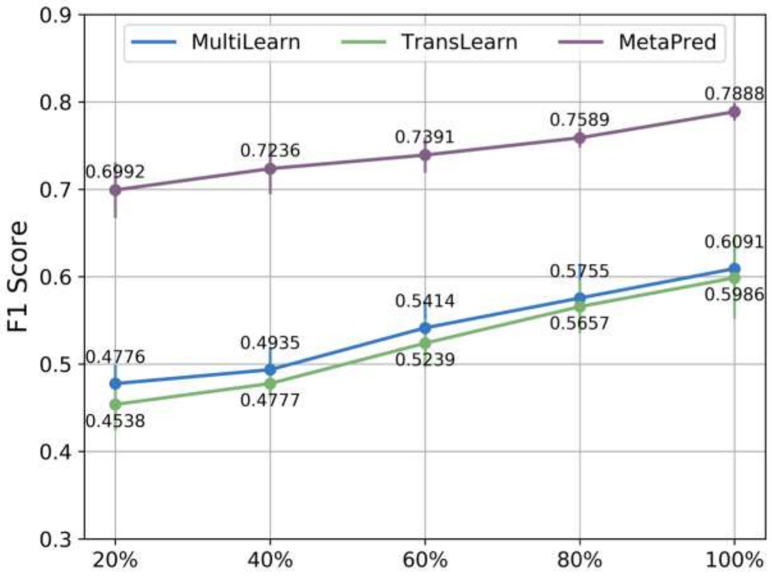
- Prediction Accuracy

Training Data	Model	MCI		Alzheimer's Disease	
		AUCROC	F1 Score	AUCROC	F1 Score
Fully Supervised	LR	0.5861 (.01)	0.3813 (.02)	0.5369 (.01)	0.2216 (.02)
	*k*NN	0.6106 (.01)	0.4540 (.01)	0.6713 (.02)	0.4686 (.03)
	RF	0.6564 (.01)	0.4998 (.01)	0.6300 (.02)	0.4111 (.04)
	MLP	0.6515 (.01)	0.5077 (.01)	0.6639 (.02)	0.4901 (.03)
	CNN	0.6999 (.01)	0.5816 (.02)	0.6755 (.03)	0.4935 (.04)
	LSTM	0.6874 (.01)	0.5666 (.02)	0.6902 (.01)	0.5316 (.02)
Low-Resource	Meta-CNN	0.7624 (.02)	0.6992 (.02)	0.7682 (.01)	0.6434 (.03)
	Meta-LSTM	0.7876 (.02)	0.7225 (.02)	0.7464 (.02)	0.6170 (.03)
Fully Fine-Tuned	Meta-CNN	0.8470 (.01)	0.7888 (.02)	**0.8461 (.01)**	**0.7375 (.01)**
	Meta-LSTM	**0.8477 (.01)**	**0.7963 (.02)**	0.8232 (.01)	0.7364 (.01)

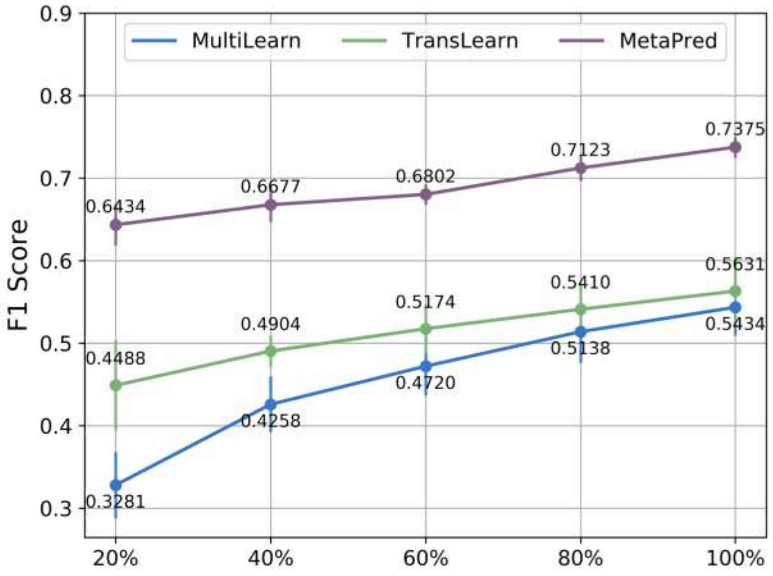
Outperform fully supervised model

Meta-Learning on Limited Clinical Resource

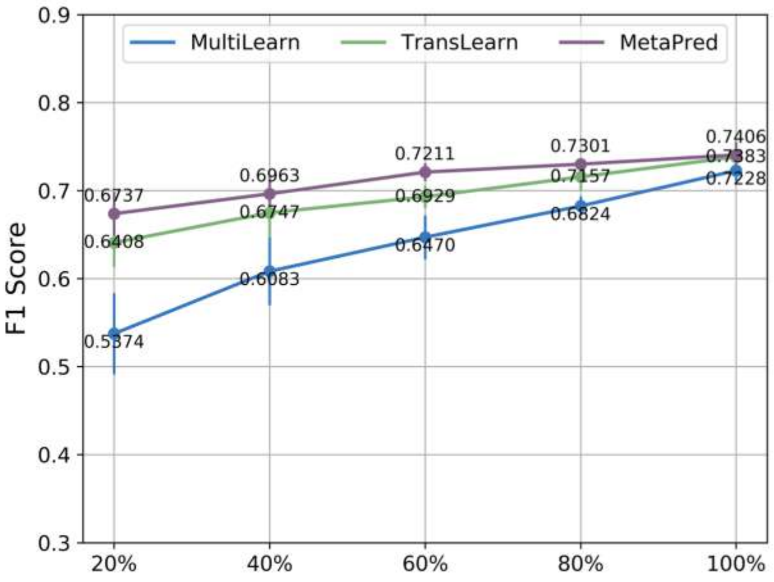
❖ Fine-tuning on Target Domain



(a) MCI



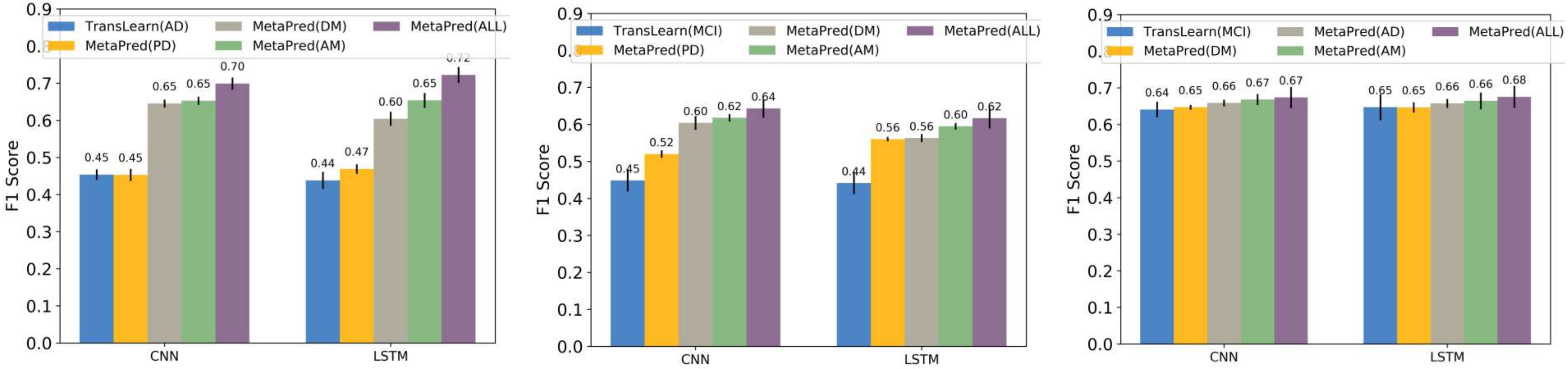
(b) Alzheimer's



(c) Parkinson's

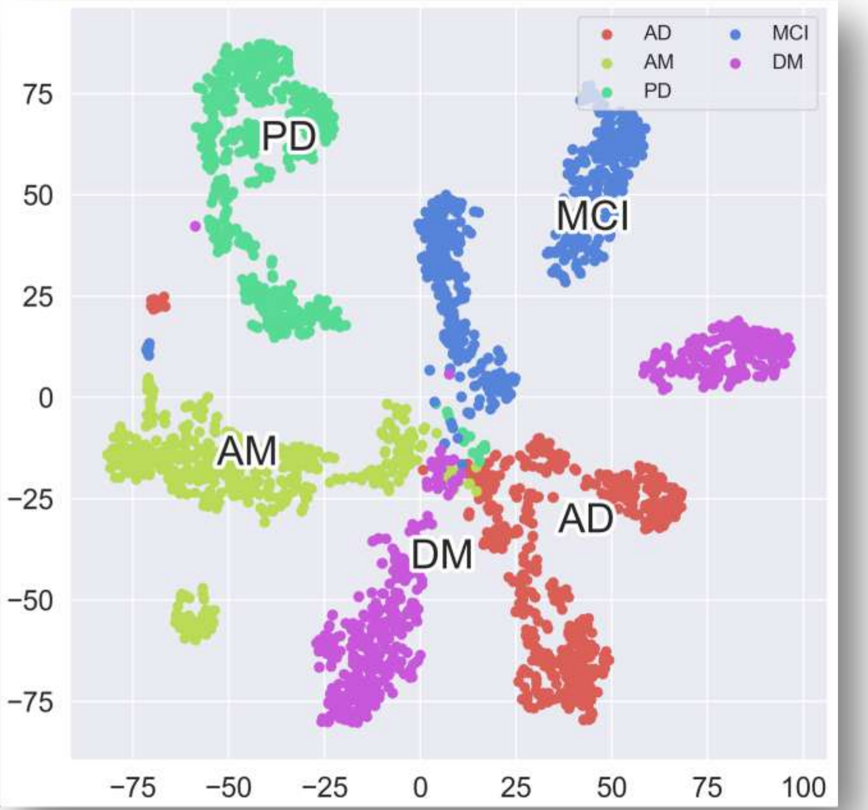
Meta-Learning on Limited Clinical Resource

- Different combinations of source disease domains



(a) MCI (b) Alzheimer's (c) Parkinson's

Meta-Learning on Limited Clinical Resource

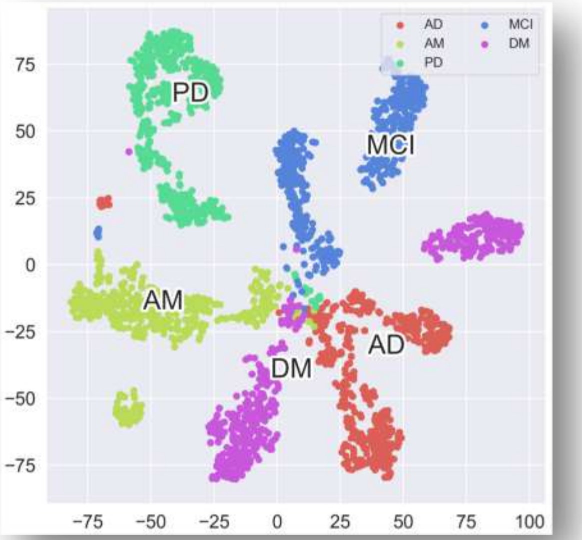


t-SNE Visualization

Source Code: https://github.com/sheryl-ai/MetaPred

Xi Zhang, Fengyi Tang, Hiroko Dodge, Jiayu Zhou, Fei Wang, MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.

Meta-Learning on Limited Clinical Resource



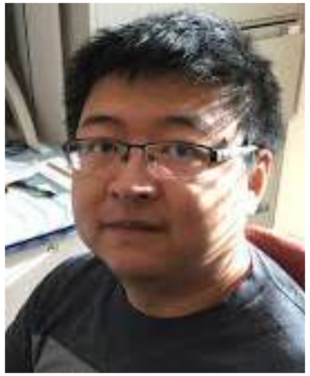
❖ Summary

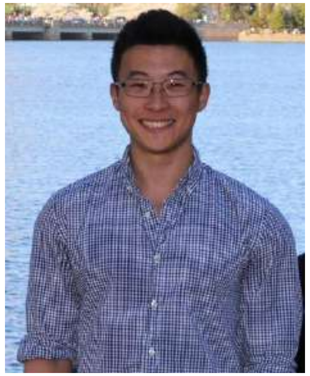
- ✓ Leverages deep predictive modeling with the model agnostic meta-learning to exploit the medical records from high-resource domain.
- ✓ Introduce two different kinds of adaptation, which are parameter-level adaptation, objective-level adaptation.
- ✓ Extensive evaluation involving 5 cognitive diseases is conducted on real-world EHR data for risk prediction tasks.

Source Code: https://github.com/sheryl-ai/MetaPred

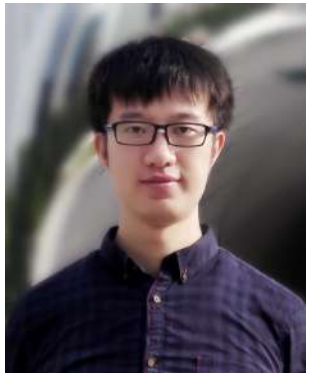
Xi Zhang, Fengyi Tang, Hiroko Dodge, Jiayu Zhou, Fei Wang, MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. SIGKDD'19: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019.


Hiroko H. Dodge
UMich

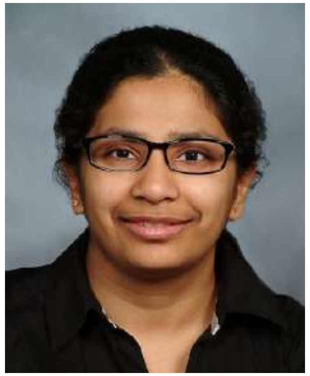

Fei Wang
Cornell


Andy Tang
MSU


Jiayu Zhou
MSU


Jian Liang
Tencent


Claire Henchcliffe
Cornell


Harini Sarva
Cornell


Yize Zhao
Yale


Jingyuan Chou
UVA


Cao (Danica) Xiao
IQVIA

Thank You!

Q & A