



Intelligent Brain Analysis through Imaging and Networks

Daoqiang Zhang (张道强)

College of Computer Science and Technology

Nanjing University of Aeronautics and Astronautics

2018-5-15

Brain Projects



美国“脑活动图谱”计划

欧盟“人类脑计划”



中国脑计划



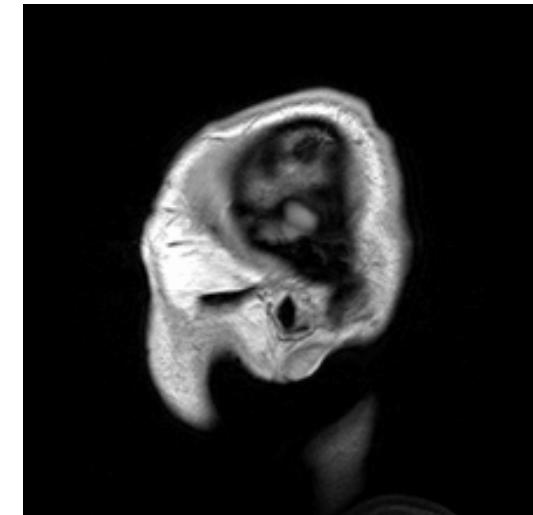
认识脑
保护脑
模拟脑



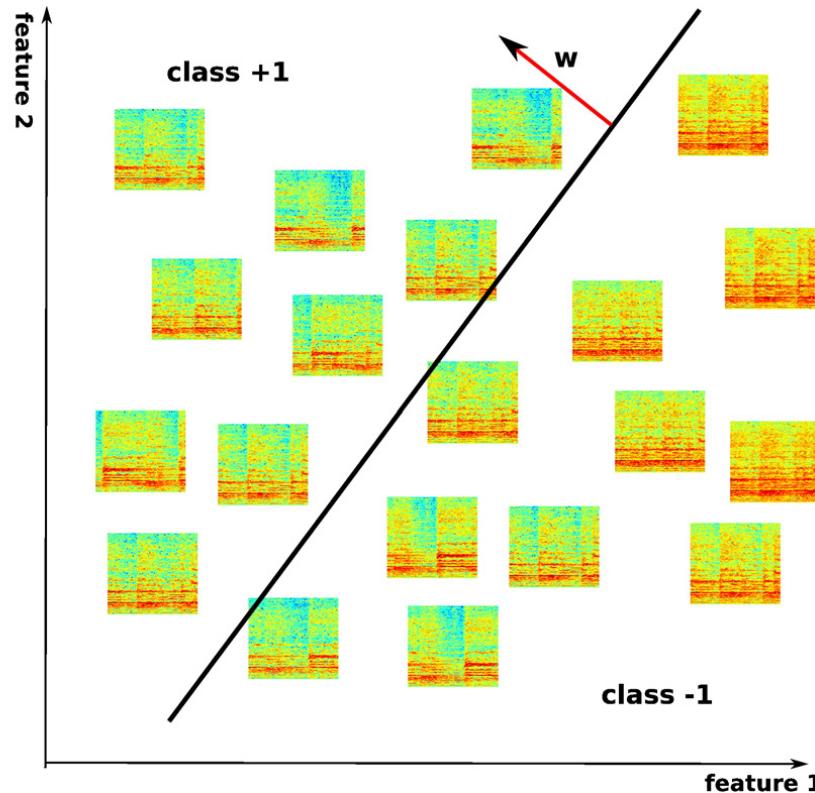
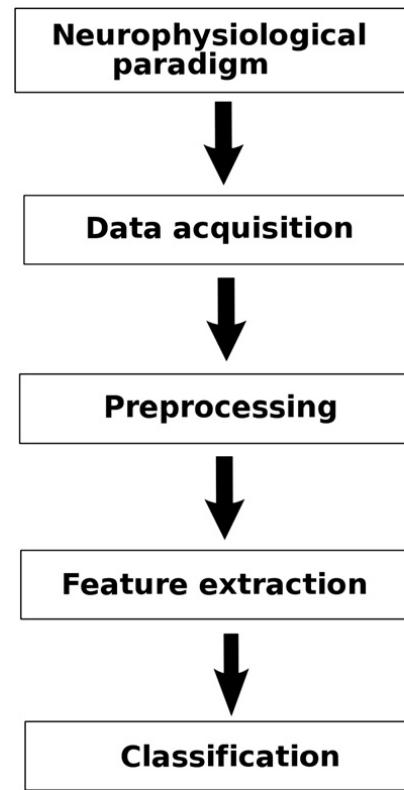
Brain Imaging (Neuroimaging)



- Neuroimaging includes the use of various techniques to either directly or indirectly image the structure or function of the brain
- Two broad categories
 - Structural neuroimaging deals with the structure of the brain
 - Functional neuroimaging is used to indirectly measure brain functions



Neuroimaging-based Classification

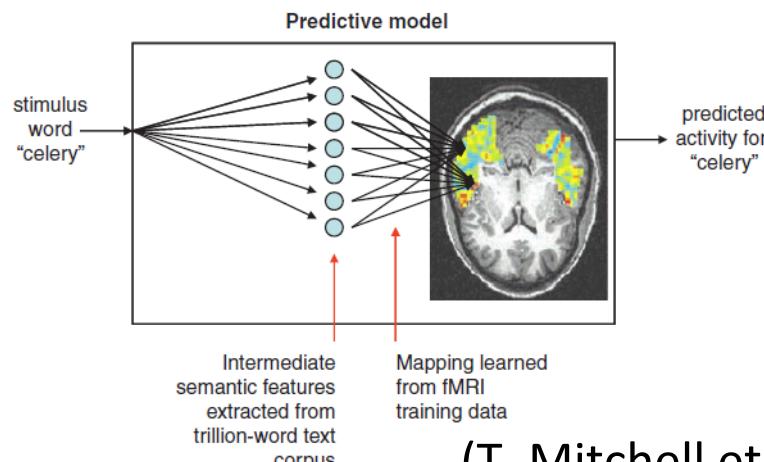
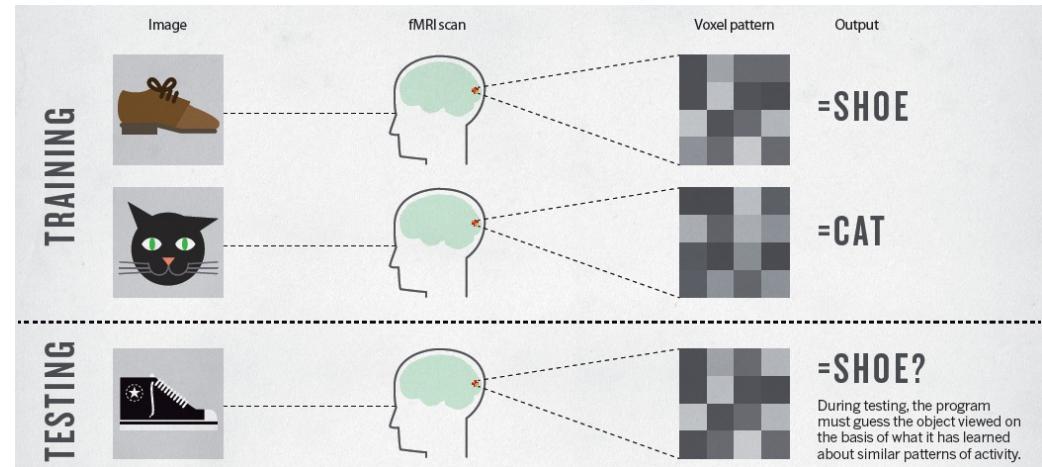


(S. Lemm, et al., Neuroimage, 2011)

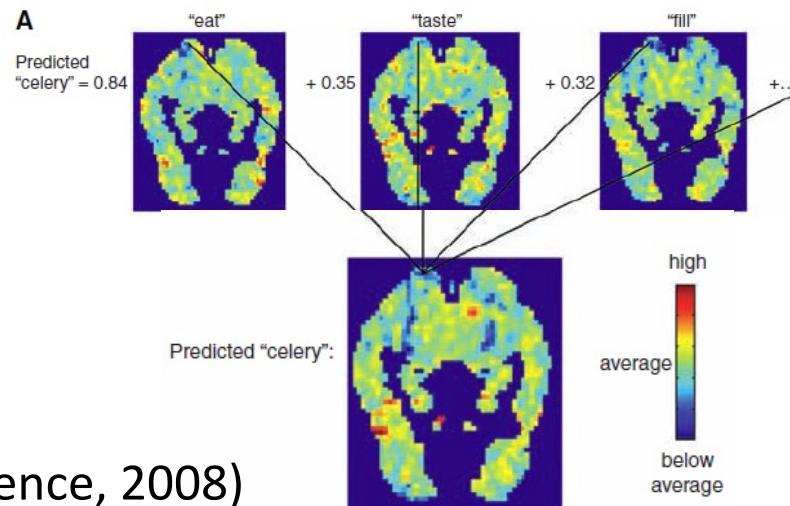
Example: Brain Decoding



(Nature Feature News, 2013)



(T. Mitchell et al., Science, 2008)



Recovery Movies



Outline



1

Backgrounds on Alzheimer's Disease

2

Brain-imaging based Analysis

3

Brain-network based Analysis

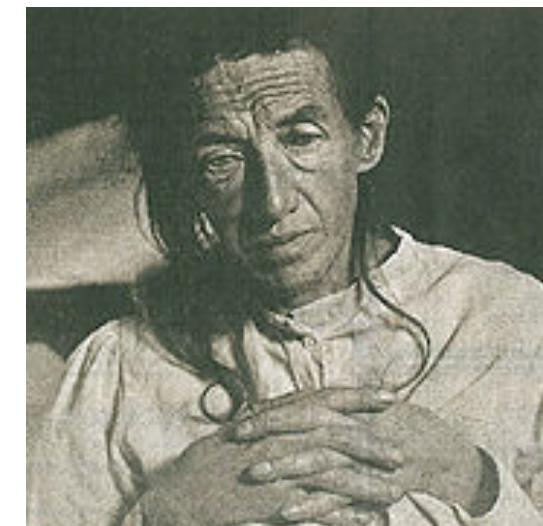
4

Summary

History of AD



- AD was first described by German psychiatrist and neuropathologist **Alois Alzheimer** in 1906 and was named after him
- The 51 y.o. woman (**Auguste Deter**) cared by Dr. Alzheimer until her death in 1906. He did an autopsy, examined her brain & described the typical abnormalities of what would be called later Alzheimer's Disease

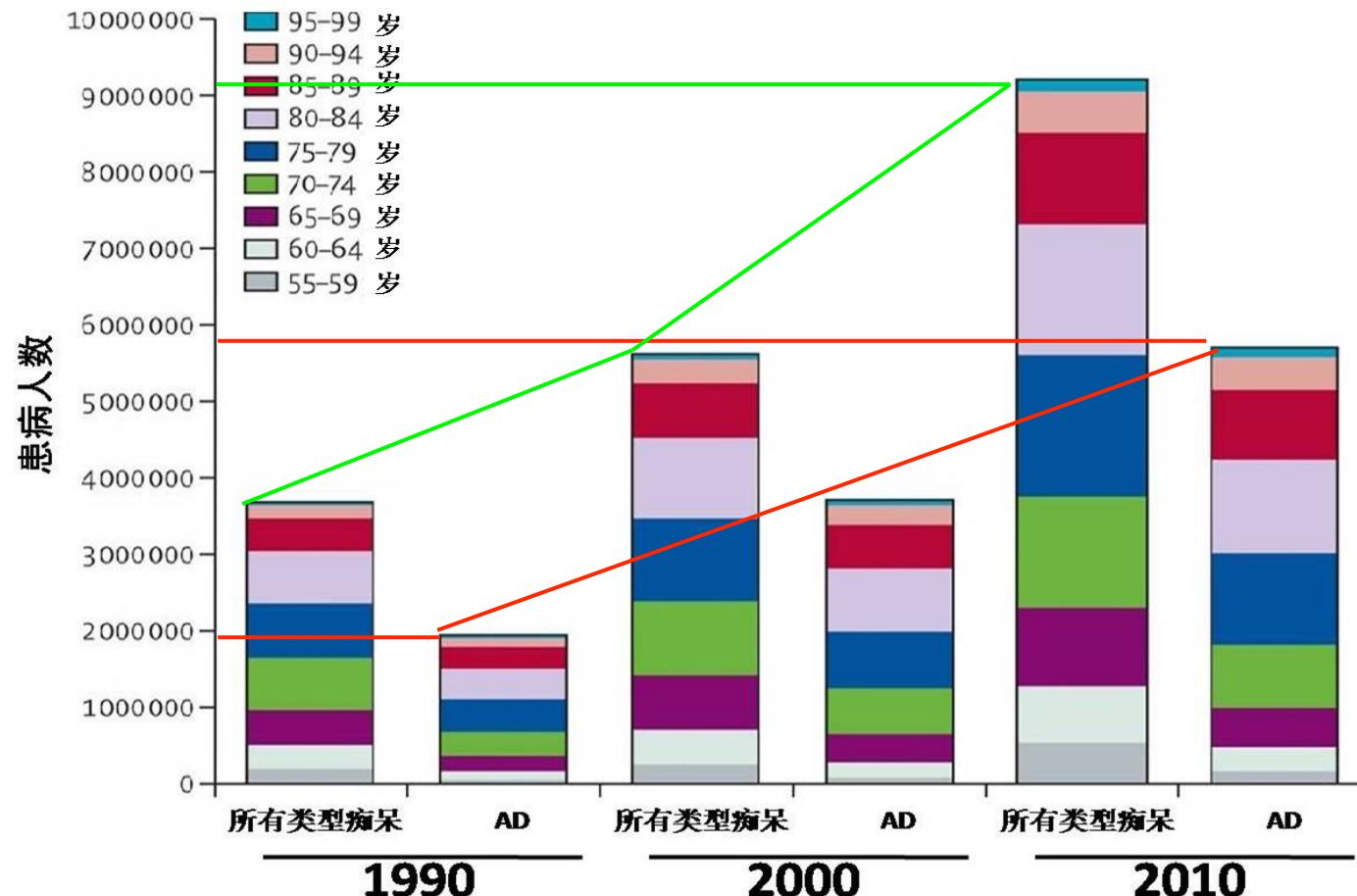


What Is AD?



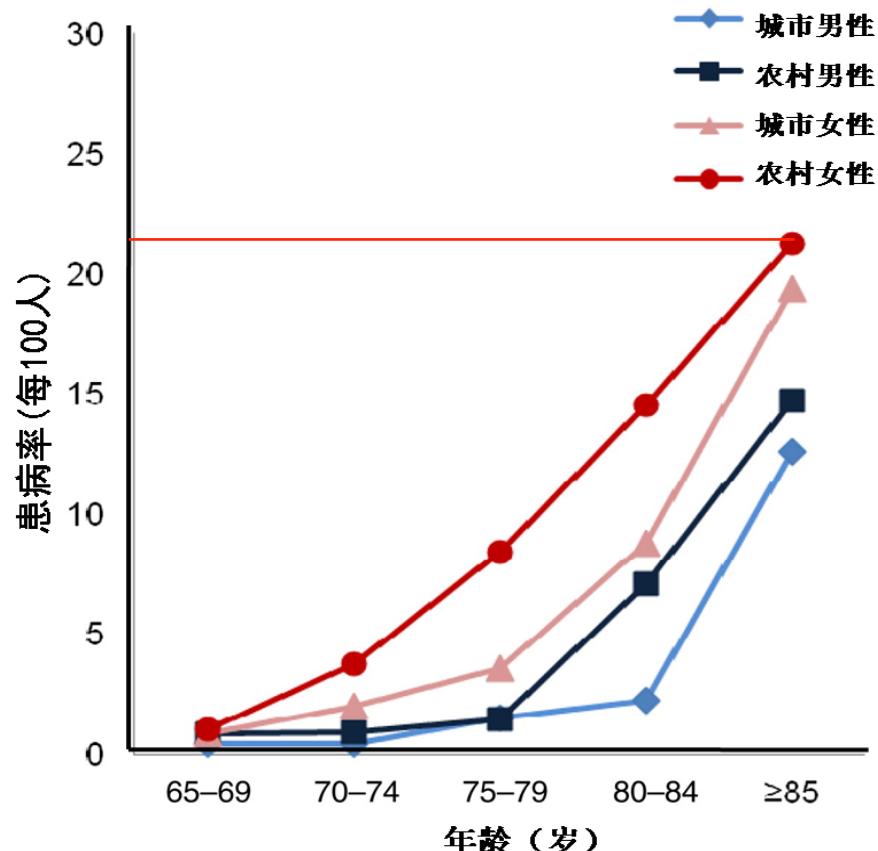
- It is the most common form of **dementia**
- There is **no cure** for the disease, which worsens as it progresses, and eventually leads to **death**
- Most often, AD is diagnosed in people over **65** years of age
- In 2006, there were **26.6 million** sufferers worldwide, and it is predicted to affect **1 in 85** people globally by 2050





不同年代我国痴呆和AD患者的人数

【Lancet. 2013】



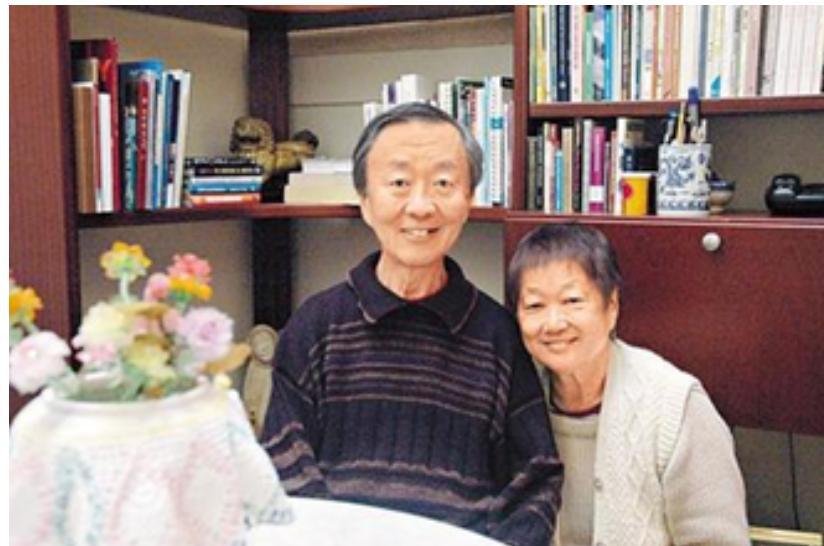
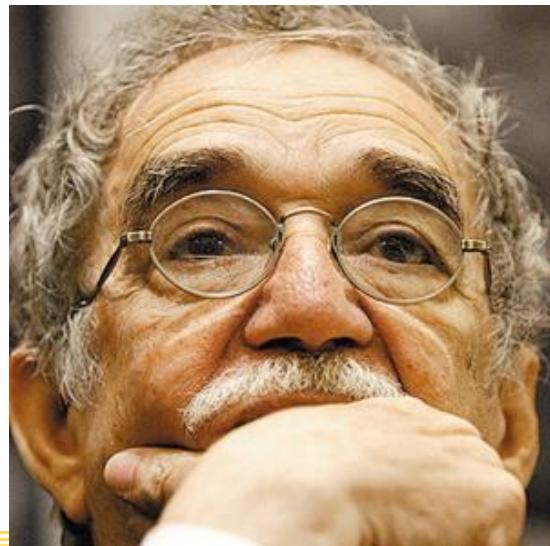
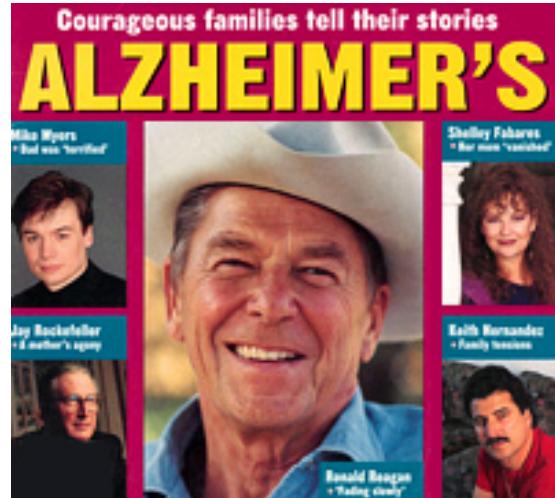
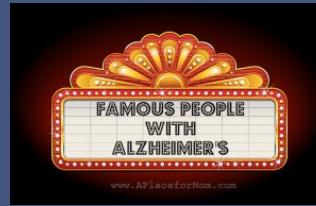
目前我国AD的患病率

【Alzheimer's & Dementia. 2013】

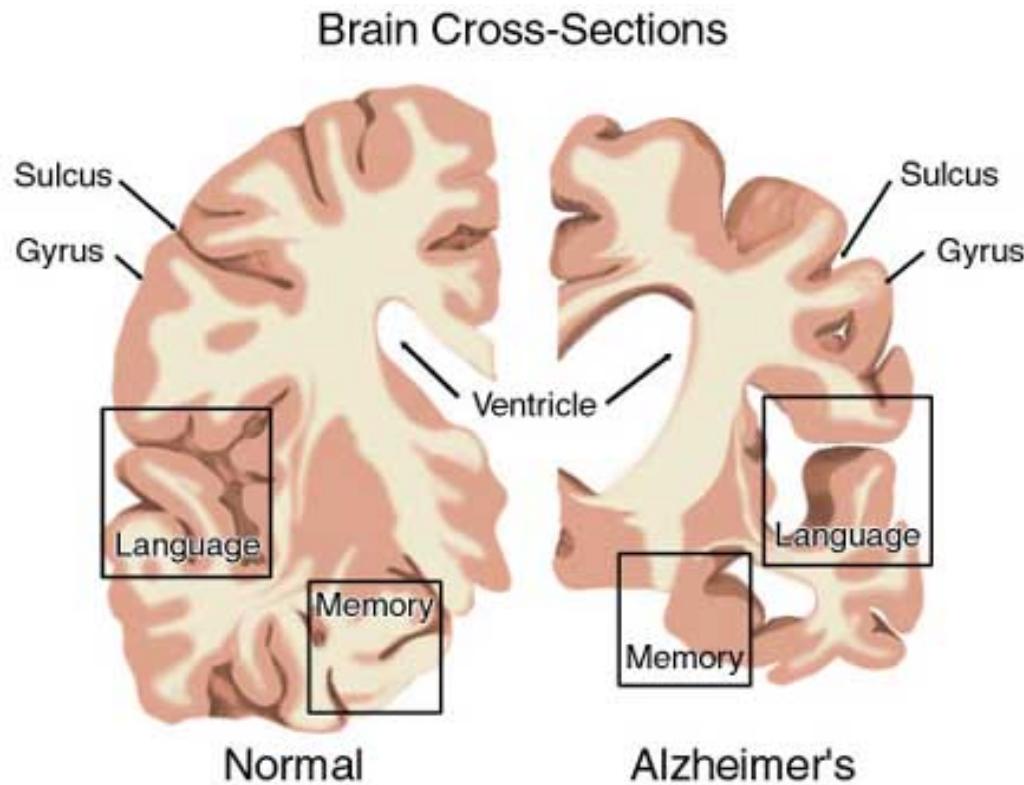
我国AD现状：

1. “三高”：患病率高、致残率高、负担重
2. “三低”：就诊率低、诊断率低、治疗率低

Celebrities with AD

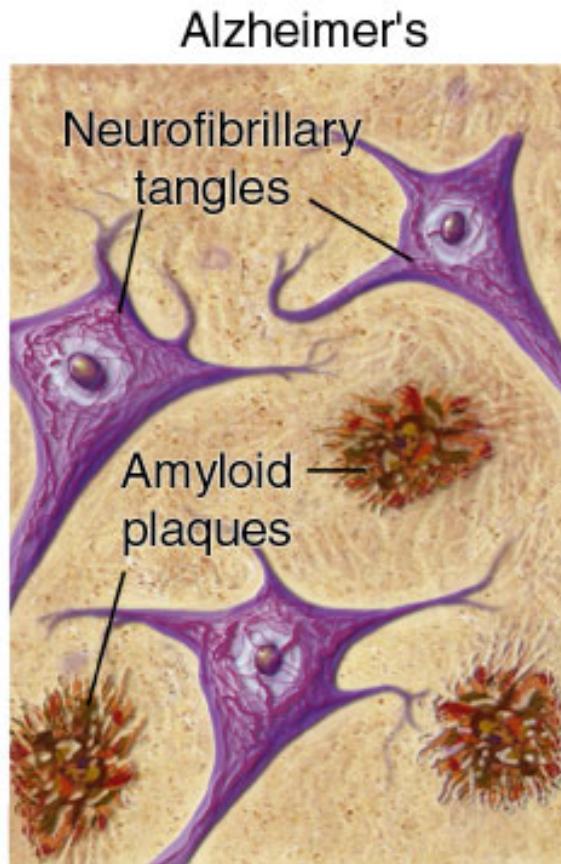
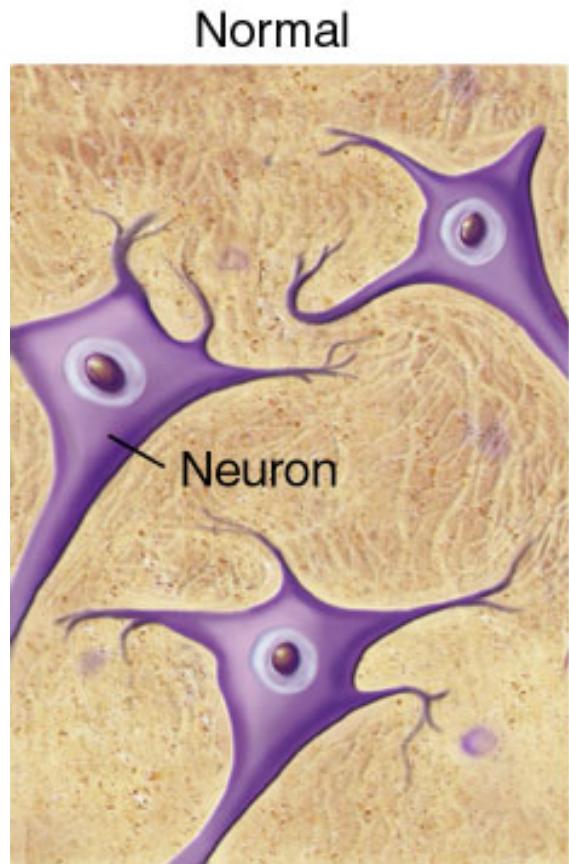


Normal vs. AD Brain



- In the normal brain there is a lot of healthy brain tissue in the **language** area. In the AD affected brain there is little in that area
- There are many differences between the two brains including the **memory**, **sulcus**, **gyrus**, **ventricle**, and **language** areas. In the AD brain, these are either shrunken or stretched out to unhealthy measures

Normal vs. AD Brain

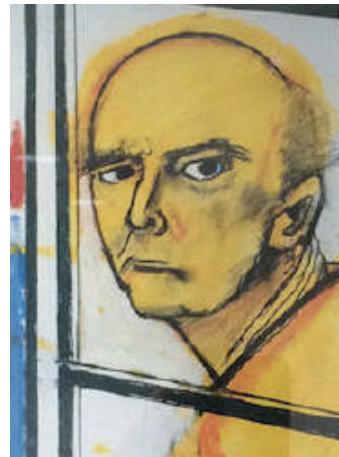


- Forms abnormal clumps called **amyloid plaques** and tangled bundles of fibers called **neurofibrillary tangles** in the brain

AD自画像



1967（早年）



1996（患病第2年）



1997（患病第3年）



1998



1999



2000



Normal or diseased?



(S. Crutch, et al., Lancet Neurology, 2012)



Normal or diseased?

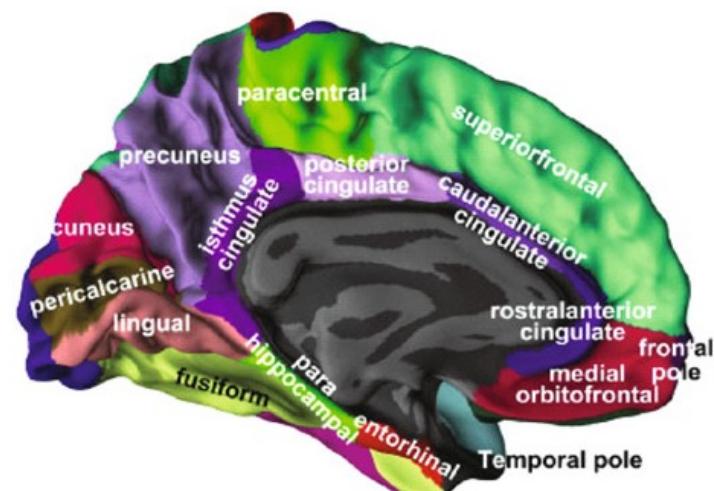
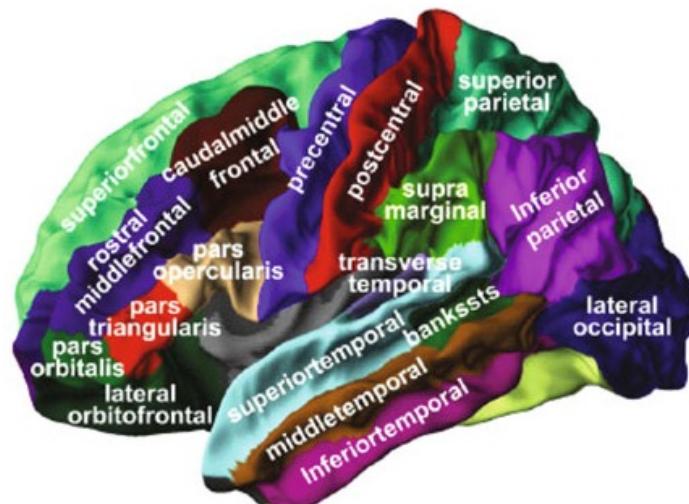


(S. Crutch, et al., Lancet Neurology, 2012)

AD Progression



- AD atrophy progresses
 - Starts in the medial temporal and limbic areas
 - Hippocampus and entorhinal cortex
 - Subsequently spreading to parietal association areas
 - Finally to frontal and primary cortices

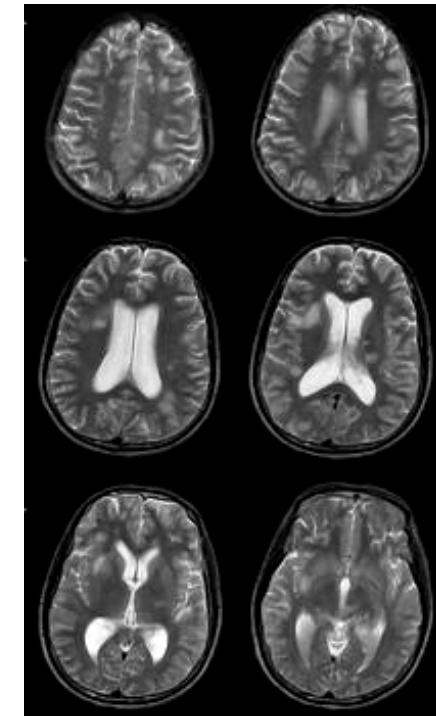


AD Biomarkers

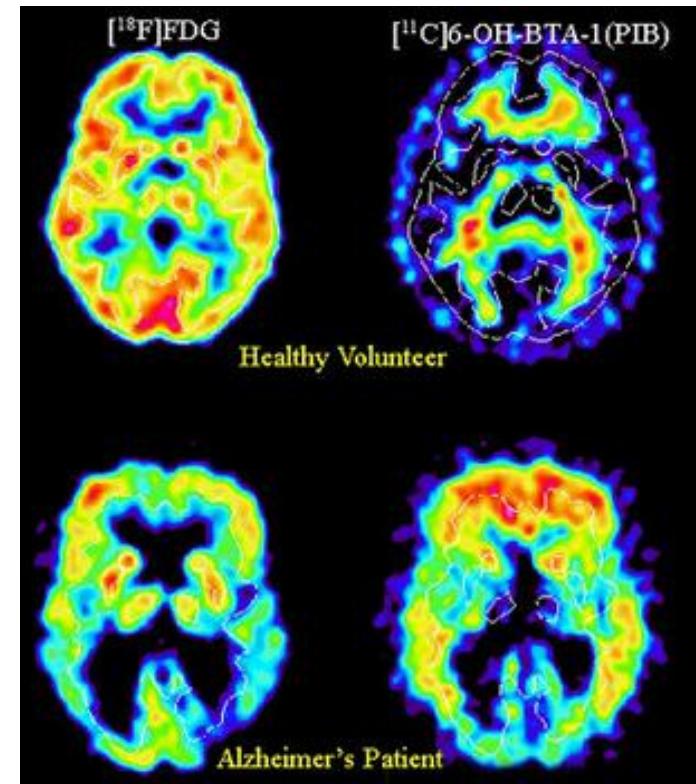


- Biomarkers for early diagnosis of AD
 - Magnetic resonance imaging (MRI)
 - Positron emission tomography (PET)
 - Cerebrospinal fluid (CSF)--- $\text{A}\beta_{42}$, t-tau and p-tau
 - ...

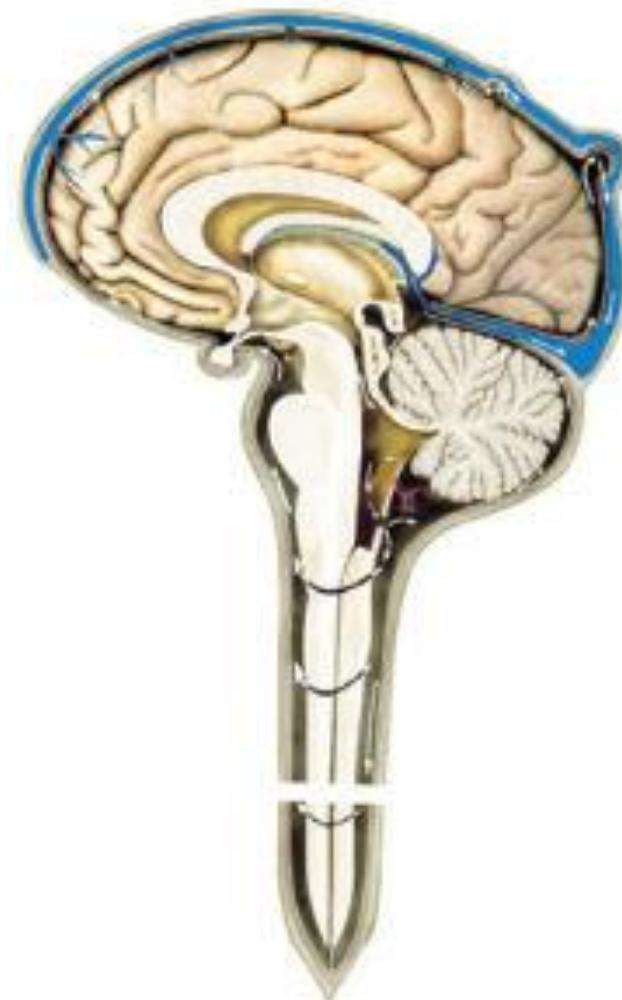
MRI



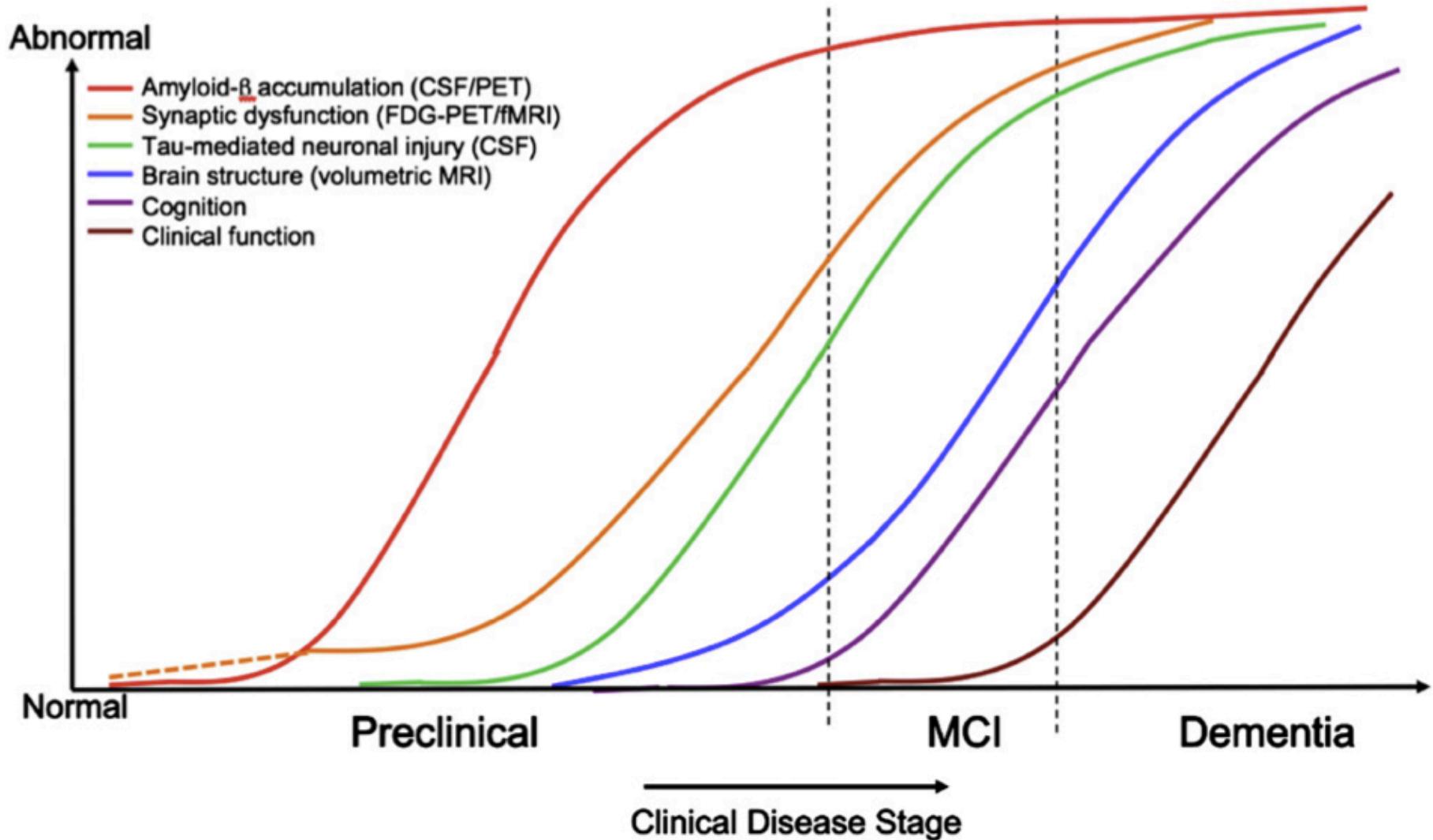
PET



CSF



Biomarkers



Outline



1

Backgrounds on Alzheimer's Disease

2

Brain-imaging based Analysis

3

Brain-network based Analysis

4

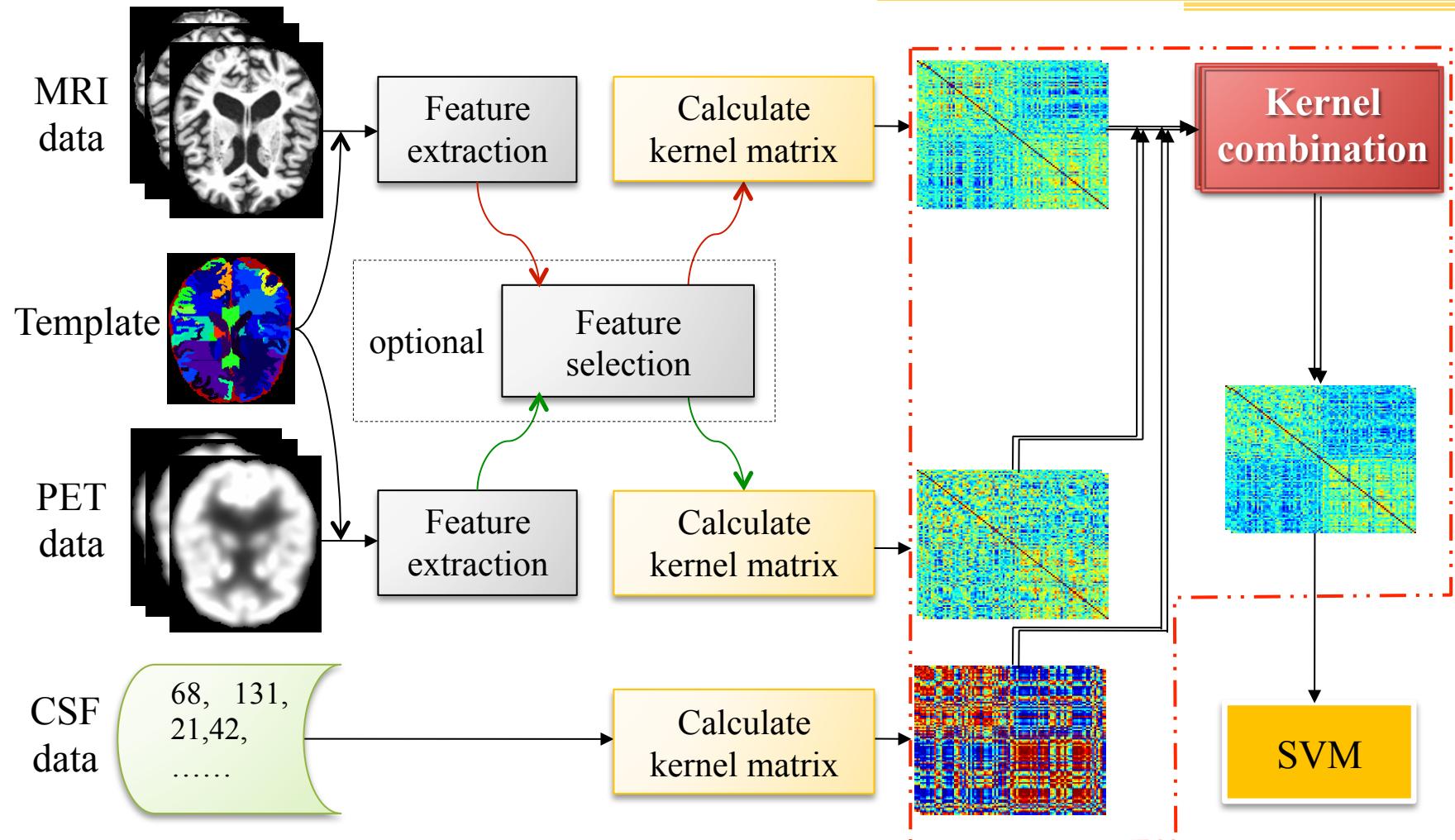
Summary

Multimodal Classification



- Motivation
 - Several modalities of biomarkers have been proved to be sensitive to AD, or its prodromal stage, i.e., mild cognitive impairment (MCI)
 - Different biomarkers provide **complementary** information, which may be useful for diagnosis of AD or MCI when used together
- **Question:** How can we effectively combine both imaging data (MRI and PET) and non-imaging data (CSF) for **multi-modality based classification?**

Flowchart



(D. Zhang, et al. Neuroimage, 2011)

Materials



Alzheimer's Disease Neuroimaging Initiative



- 202 subjects from ADNI, including 51 AD patients, 99 MCI and 52 healthy controls
 - 43 MCI converters who had converted to AD within 18 months and 56 MCI non-converters who had not converted
 - Only baseline data of MRI, CSF and PET are used

	AD (n=51; 18F/33M)			MCI (n=99; 32F/67M)			HC (n=52; 8F/34M)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age	75.2	7.4	59-88	75.3	7.0	55-89	75.3	5.2	62-85
Education	14.7	3.6	4-20	15.9	2.9	8-20	15.8	3.2	8-20
MMSE	23.8	2.0	20-26	27.1	1.7	24-30	29	1.2	25-30
CDR	0.7	0.3	0.5-1	0.5	0.0	0.5-0.5	0	0.0	0-0

Results

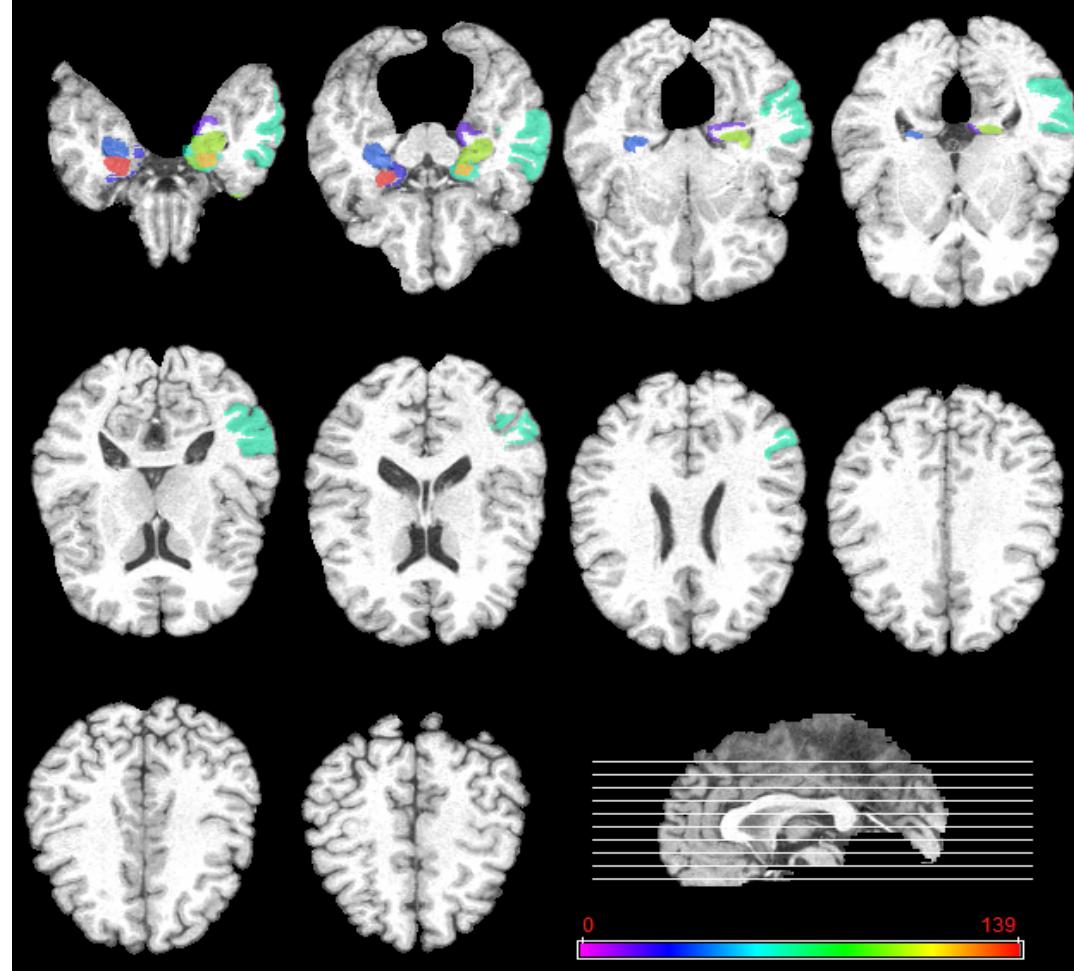


Comparison of performance of single-modal and multimodal classification methods

Methods	AD vs. HC			MCI vs. HC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
MRI	86.2 (82.9-89.0)	86 (82.7-88.7)	86.3 (83.1-89.1)	72.0 (68.4-74.7)	78.5 (75.6-80.6)	59.6 (55.1-63.7)
	82.1 (80-84.9)	81.9 (80-84.7)	82.3 (80-85.1)	71.4 (68.2-73.3)	78 (75.6-79.4)	58.8 (54.3-61.7)
	86.5 (82.9-90.5)	86.3 (82.7-90.3)	86.6 (83.1-90.6)	71.6 (67.4-74.7)	78.2 (75-80.6)	59.3 (52.9-63.7)
Combined	93.2 (89.0-96.5)	93 (88.7-96.3)	93.3 (89.1-96.6)	76.4 (73.5-79.7)	81.8 (79.4-84.4)	66.0 (62.6-70.3)
Baseline	91.5 (88.5-96.5)	91.4 (88.3-96.3)	91.6 (88.6-96.6)	74.5 (71.9-78.2)	80.4 (78.3-83.3)	63.3 (59.7-68.3)

(D. Zhang, et al. Neuroimage, 2011)

Results (cont'd)



Top 11 brain regions selected for MCI classification detected from MRI

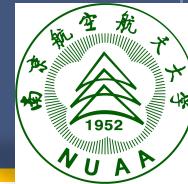
Results (cont'd)



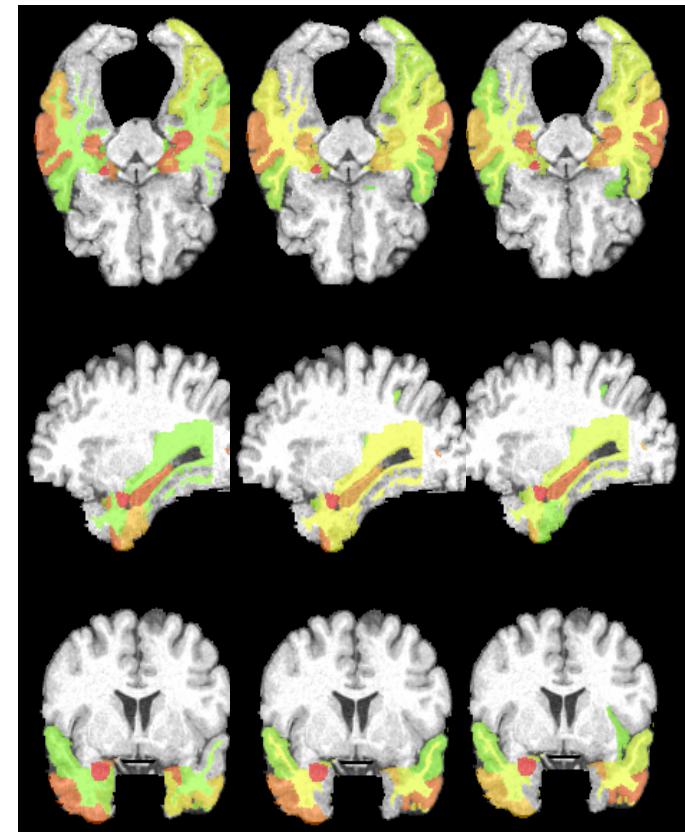
Top 11 brain regions detected from MRI and PET modalities for MCI classification

	MRI	PET
1	amygdala right (p<0.0001)	angular gyrus left (p=0.0003)
2	hippocampal formation left (p<0.0001)	precuneus left (p=0.0005)
3	hippocampal formation right (p<0.0001)	precuneus right (p=0.0021)
4	uncus left (p<0.0001)	inferior temporal gyrus left (p=0.0146)
5	entorhinal cortex left (p=0.0001)	anterior limb of internal capsule right (p=0.0154)
6	amygdala left (p=0.0001)	angular gyrus right (p=0.0189)
7	middle temporal gyrus left (p=0.0001)	anterior limb of internal capsule left (p=0.0204)
8	temporal pole left (p=0.0004)	globus palladus left (p=0.021)
9	perirhinal cortex left (p=0.0004)	globus palladus right (p=0.0259)
10	uncus right (p=0.0006)	posterior limb of internal capsule right (p=0.0272)
11	parahippocampal gyrus left (p=0.0009)	entorhinal cortex left (p=0.0286)

Multi-Modal Multi-Task Learning



- Motivation
 - Besides classification, there also exist **regression** tasks which estimate continuous clinical scores to **evaluate the stage of AD pathology** and predict **future progression**
 - Both regression and classification tasks are essentially related due to the **same underlying pathology**
- **Question: How can we jointly predict multiple regression and classification variables from multi-modality data?**



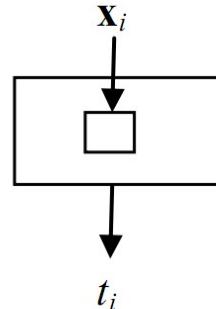
AD/MCI/HC MMSE ADAS-Cog

(D. Zhang, D. Shen. Neuroimage, 2012)

Four Learning Problems



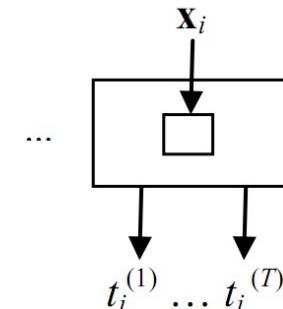
Input: Single-modality data



Output: Single task

(a) Single-model single task (SMST)

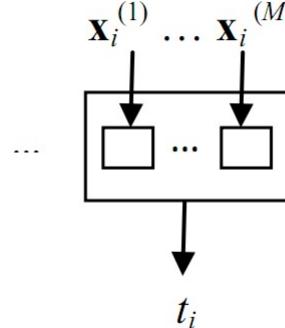
Input: Single-modality data



Output: Multiple tasks

(b) Multi-task learning

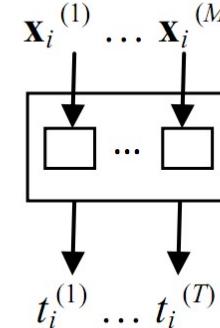
Input: Multi-modality data



Output: Single task

(c) Multi-modal learning

Input: Multi-modality data

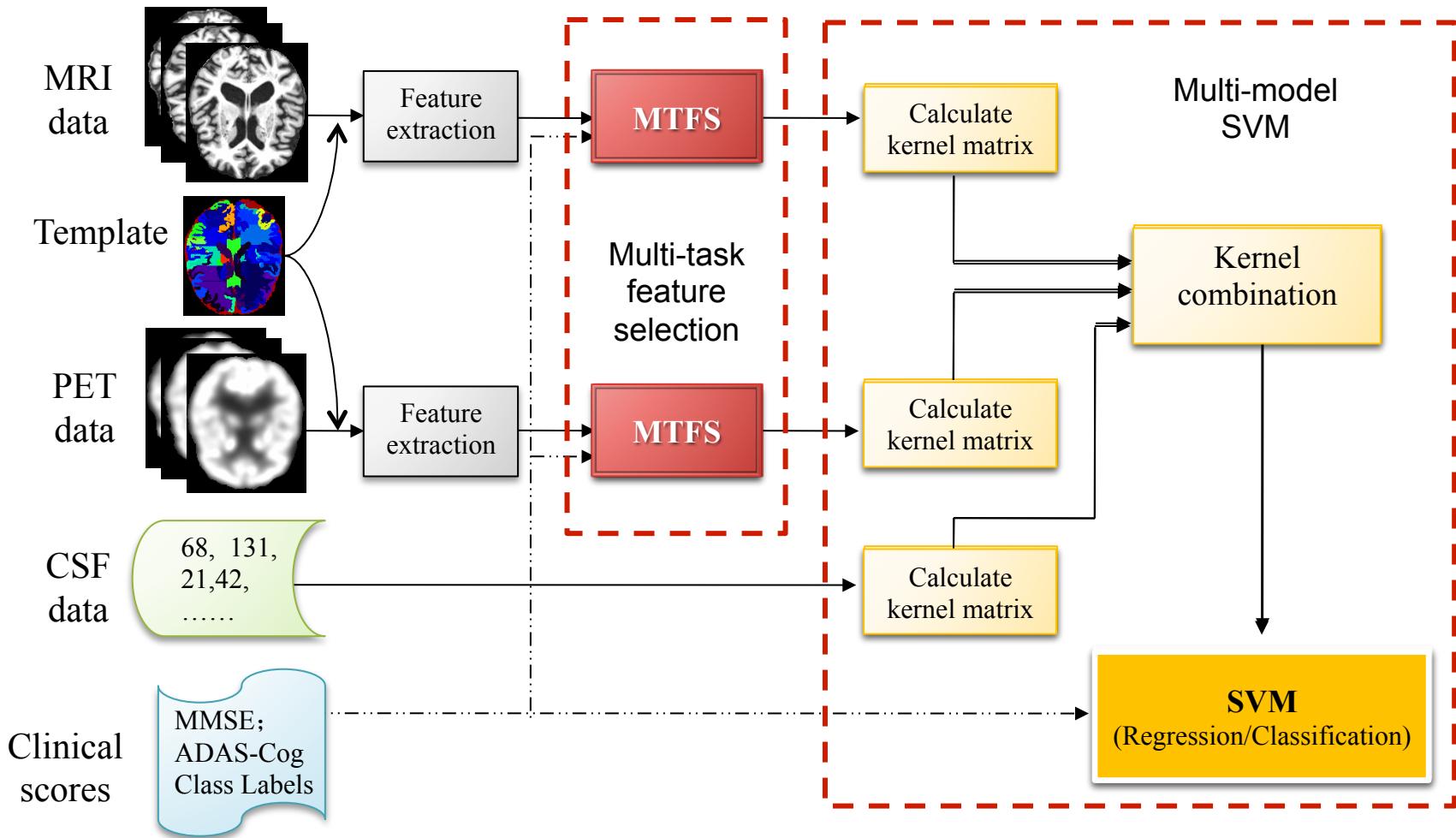


Output: Multiple tasks

(d) Multi-modal Multi-task (M3T)

(D. Zhang, D. Shen. Neuroimage, 2012)

Flowchart



(D. Zhang, D. Shen. Neuroimage, 2012)

Multi-Task Feature Selection



- Objective function

$$\begin{aligned}
 \min_{\mathbf{V}^{(m)}} \quad & \frac{1}{2} \sum_{j=1}^T \sum_{i=1}^N \left(t_i^{(j)} - \hat{t}^{(j)} (\mathbf{x}_i^{(m)}, \mathbf{v}_j^{(m)}) \right)^2 + \lambda \sum_{d=1}^{D^{(m)}} \|\mathbf{v}^{(m)}|_d\|_2 \\
 = \quad & \frac{1}{2} \sum_{j=1}^T \left\| \mathbf{y}^{(j)} - \mathbf{X}^{(m)} \mathbf{v}_j^{(m)} \right\|_2^2 + \lambda \sum_{d=1}^{D^{(m)}} \|\mathbf{v}^{(m)}|_d\|_2
 \end{aligned}$$

Group sparsity
 $\mathbf{V}^{(m)}$
 $\mathbf{v}^{(m)}|_1$
 $\mathbf{v}^{(m)}|_d$
 \vdots
 $\mathbf{v}^{(m)}|_D$
 $\mathbf{v}_1^{(m)} \quad \mathbf{v}_j^{(m)} \quad \mathbf{v}_T^{(m)}$
 $D \times T$

Weight matrix

$\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \dots, \mathbf{x}_i^{(m)}, \dots, \mathbf{x}_N^{(m)}]^T$
 $\mathbf{y}^{(j)} = [t_1^{(j)}, \dots, t_i^{(j)}, \dots, t_N^{(j)}]^T$
 $\mathbf{V}^{(m)} = [\mathbf{v}_1^{(m)}, \dots, \mathbf{v}_j^{(m)}, \dots, \mathbf{v}_T^{(m)}]$

Materials



- ADNI Subjects
 - 186 subjects (45AD, 91 MCI and 50 HCs), only baseline data, 3 modalities (MRI, CSF and PET)

	AD (n=45)	HC (n=50)	MCI-C (n=43)	MCI-NC (n=48)
Female/Male	16/29	18/32	15/28	16/32
Age	75.4 ± 7.1	75.3 ± 5.2	75.8 ± 6.8	74.7 ± 7.7
Education	14.9 ± 3.4	15.6 ± 3.2	16.1 ± 2.6	16.1 ± 3.0
MMSE (baseline)	23.8 ± 1.9	29.0 ± 1.2	26.6 ± 1.7	27.5 ± 1.6
MMSE (2 years)	19.3 ± 5.6	29.0 ± 1.3	23.8 ± 3.3	26.9 ± 2.6
ADAS-Cog (baseline)	18.3 ± 6.1	7.3 ± 3.3	12.9 ± 3.9	9.7 ± 4.0
ADAS-Cog (2 years)	27.3 ± 11.7	6.3 ± 3.5	16.1 ± 6.4	11.2 ± 5.7

Experiments



- Experiment 1
 - Estimating clinical stages
 - MMSE, ADAS-Cog, and class label (AD/MCI/HC)
- Experiment 2
 - Predicting 2-year MMSE and ADAS-Cog changes and MCI conversion

Results



- Comparison of performances of different methods on Experiment 1

Methods	Correlation coefficient		Classification accuracy	
	MMSE	ADAS-Cog	AD vs. HC	MCI vs. HC
MRI-based	0.504 ± 0.038	0.609 ± 0.014	0.848 ± 0.026	0.739 ± 0.028
PET-based	0.658 ± 0.027	0.670 ± 0.018	0.845 ± 0.035	0.797 ± 0.023
CSF-based	0.465 ± 0.019	0.474 ± 0.013	0.805 ± 0.022	0.536 ± 0.044
Baseline	0.658 ± 0.023	0.695 ± 0.011	0.920 ± 0.033	0.800 ± 0.024
Proposed M3T	0.697 ± 0.022	0.739 ± 0.012	0.933 ± 0.022	0.832 ± 0.015

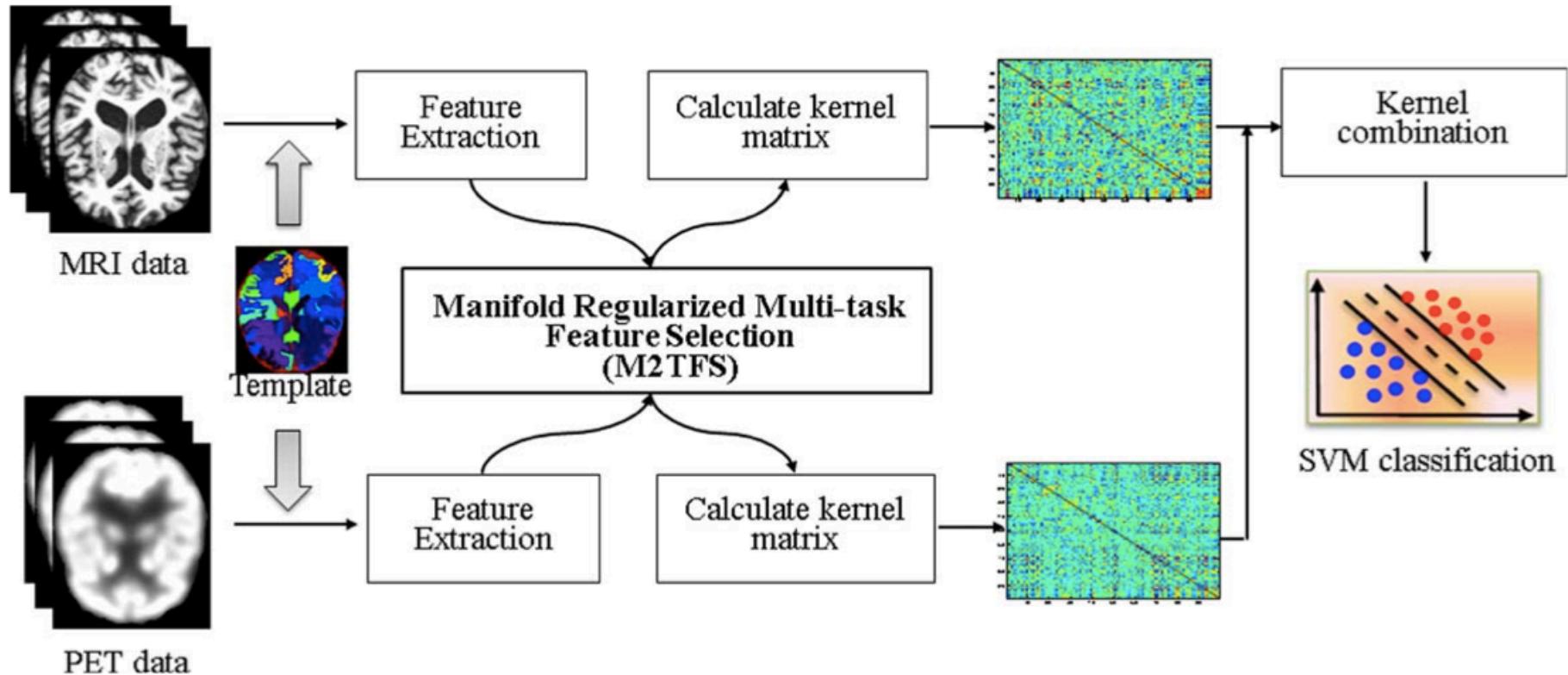
Results (cont'd)



- Comparison of performances of different methods on Experiment 2

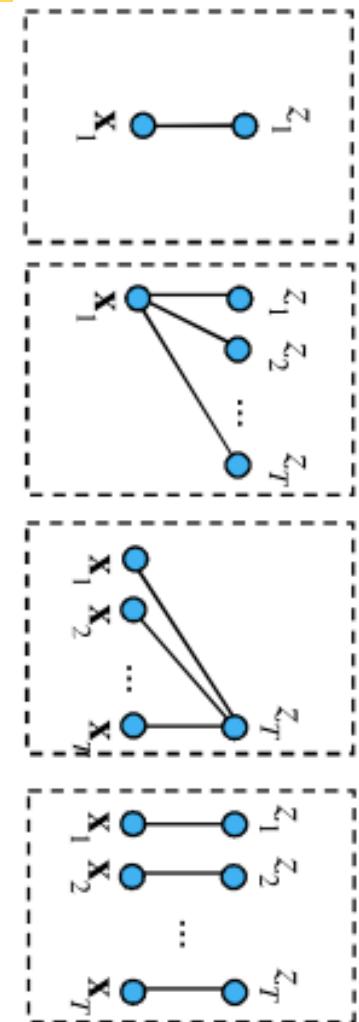
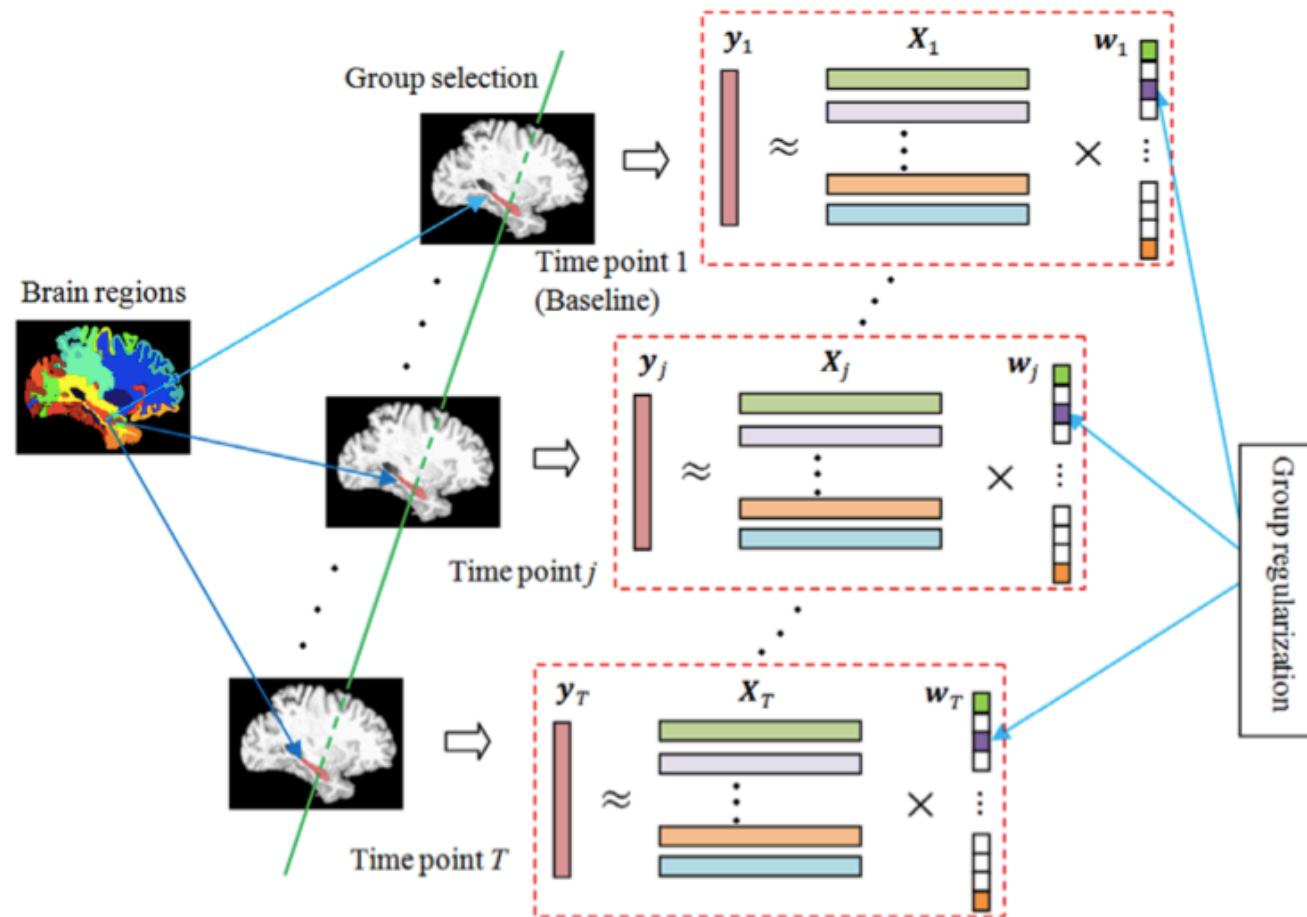
Methods	Correlation coefficient		Classification accuracy MCI-C vs. MCI-NC
	MMSE change	ADAS-Cog change	
MRI-based	0.419 ± 0.019	0.455 ± 0.037	0.620 ± 0.058
PET-based	0.434 ± 0.027	0.401 ± 0.046	0.639 ± 0.046
CSF-based	0.327 ± 0.018	0.425 ± 0.028	0.518 ± 0.086
Baseline	0.484 ± 0.009	0.475 ± 0.045	0.654 ± 0.050
Proposed M3T	0.511 ± 0.021	0.531 ± 0.032	0.739 ± 0.038

Manifold Regularized Multitask Learning



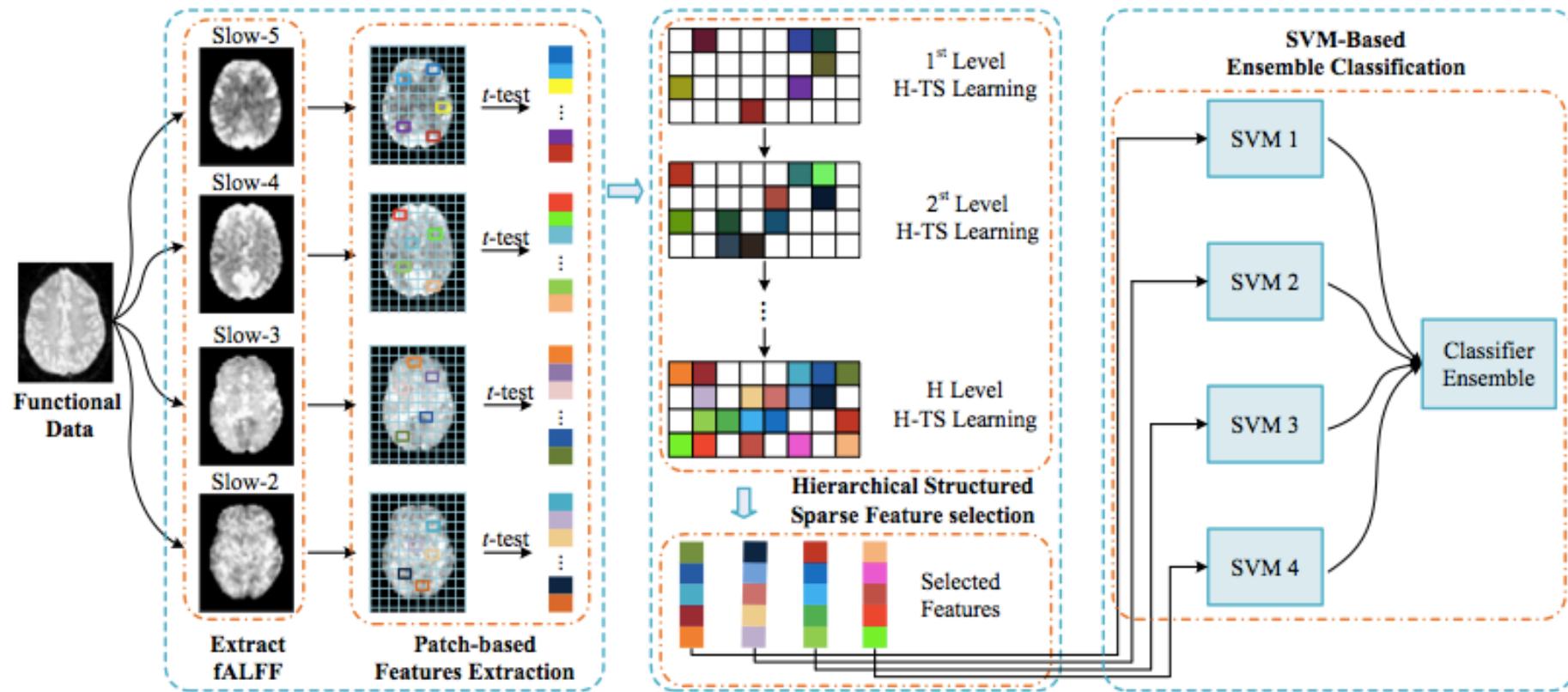
(B. Jie, D. Zhang, et al. Human Brain Mapping, 2015)

Longitudinal Multitask Learning



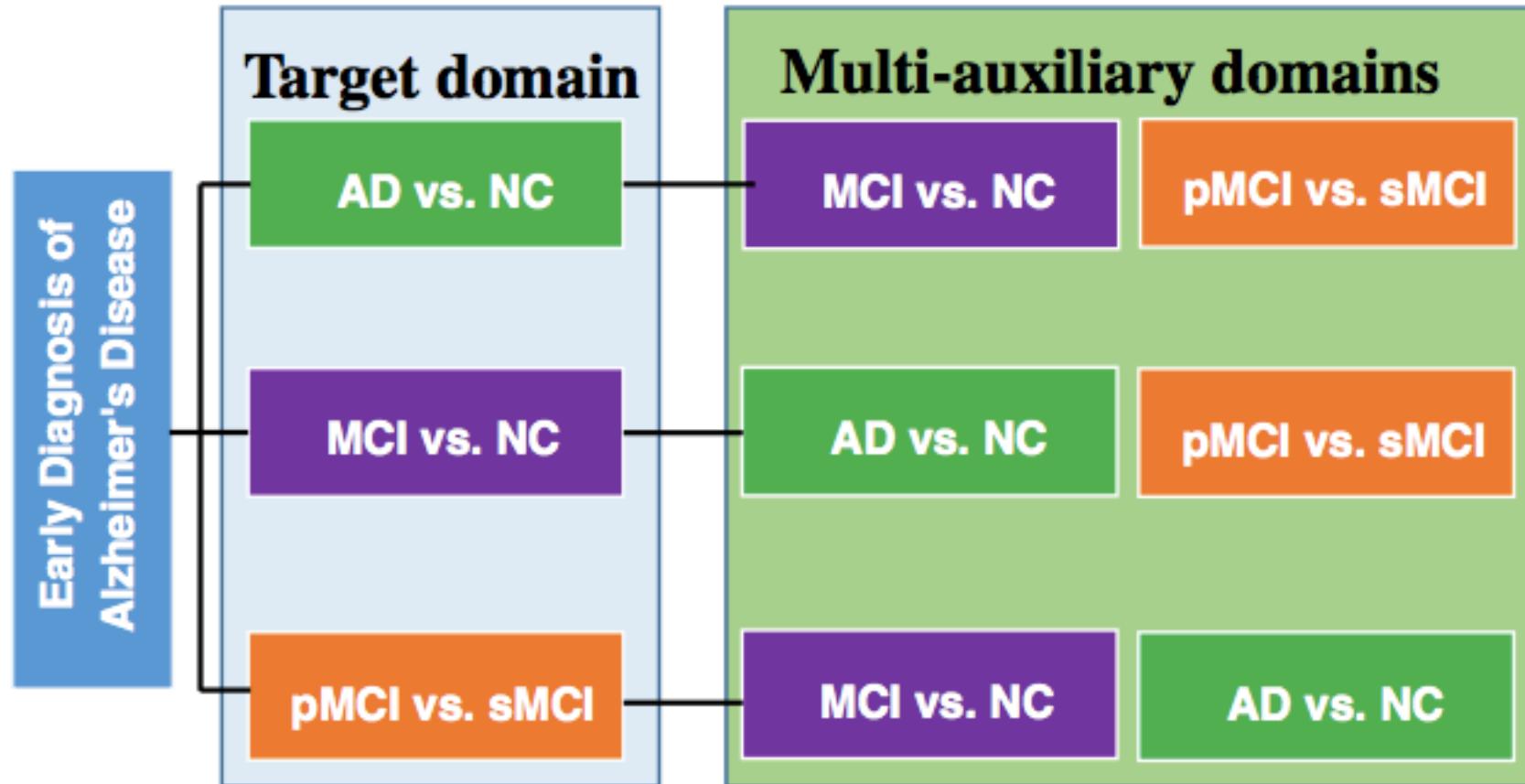
(D. Zhang, et al. PLOS ONE, 2012; B. Jie, et al. IEEE TBME 2017)

Multi-level Multitask Learning



(M. Wang, et al. MICCAI 2017)

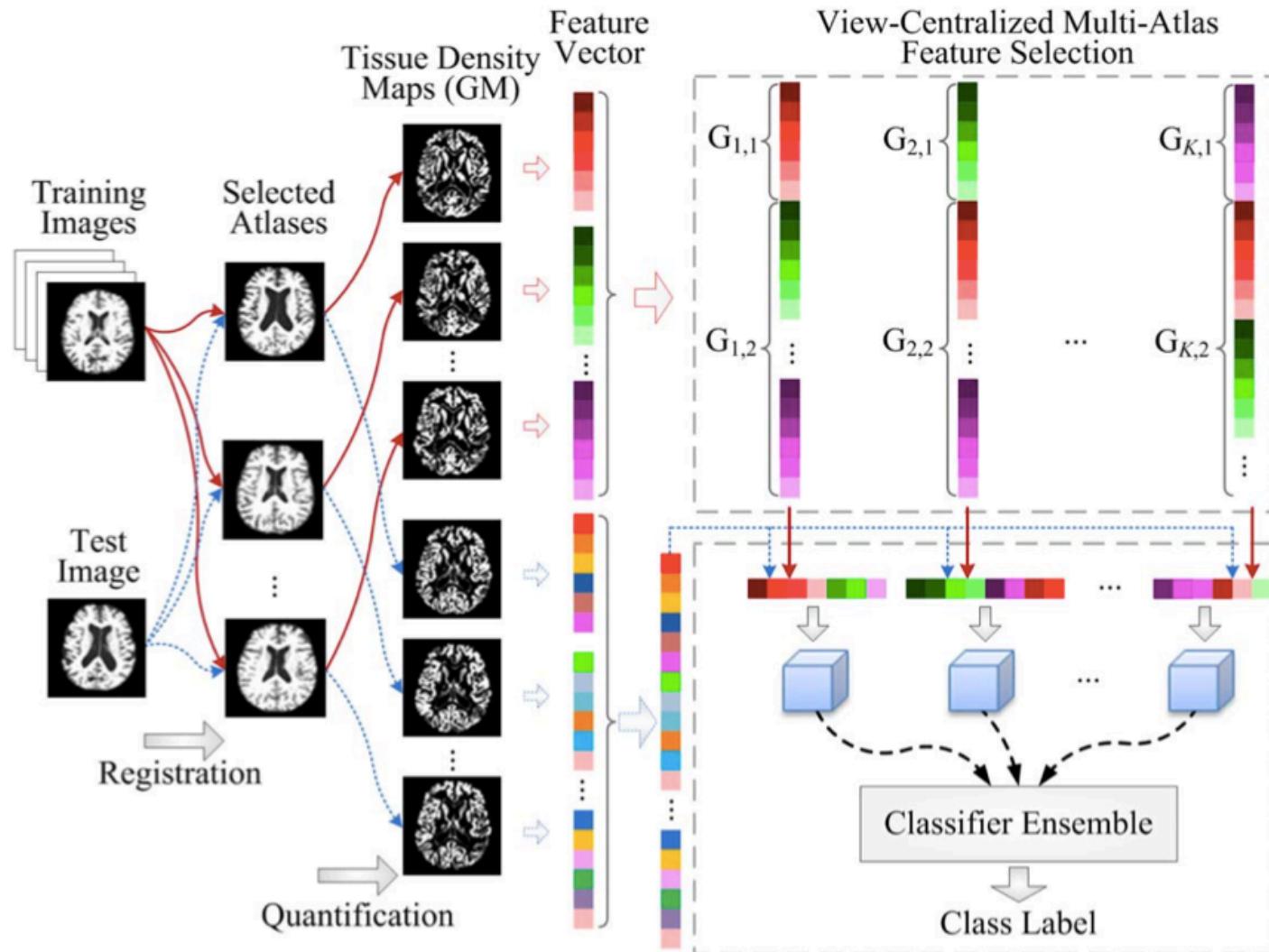
Multimodal Transfer Learning



(B. Cheng, et al. Brain Imaging & Behavior 2015; IEEE TBME 2015; Neuroinformatics 2017)



Multi-Atlas Classification

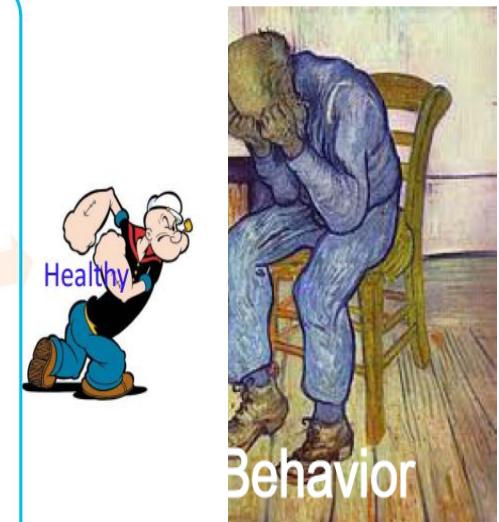
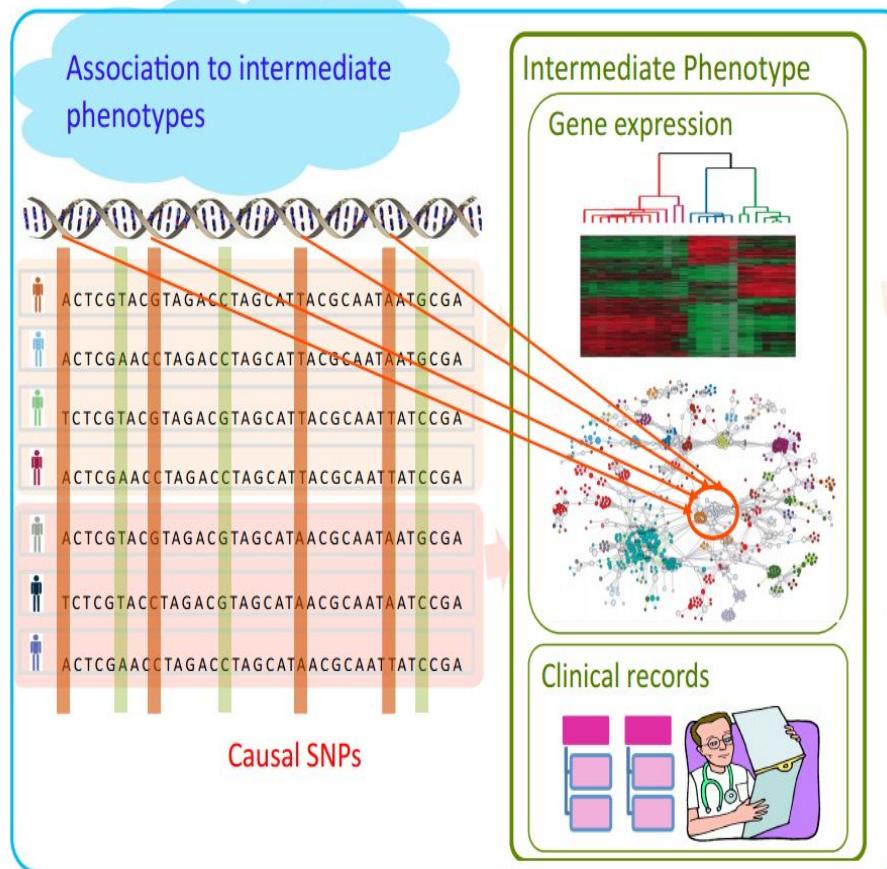
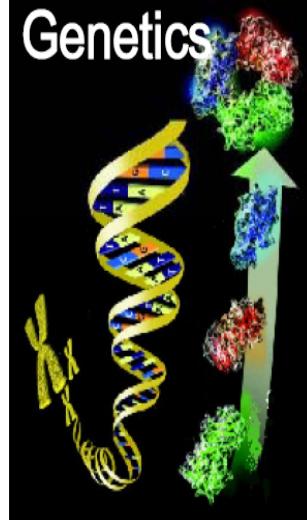


(M. Liu, D. Zhang, et al. Human Brain Mapping, 2015; IEEE TMI 2016; TBME 2016)

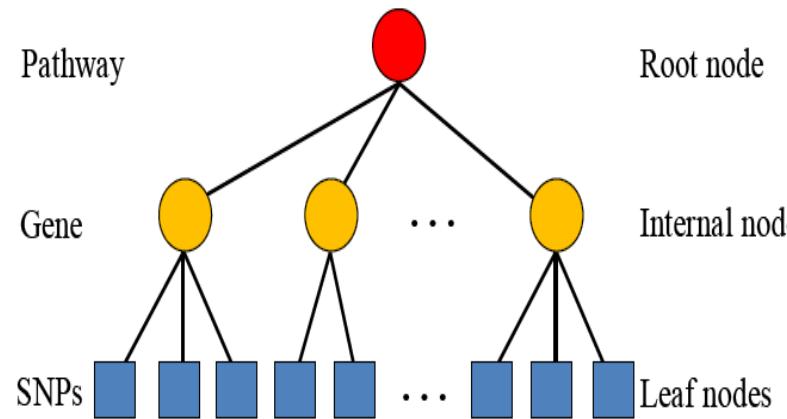
Imaging Genetics



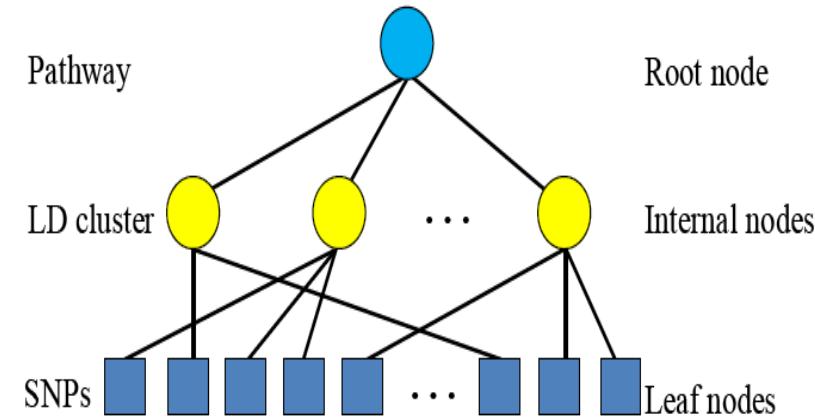
Genetic Basis of Complex Diseases



Tree Structure Among SNPs

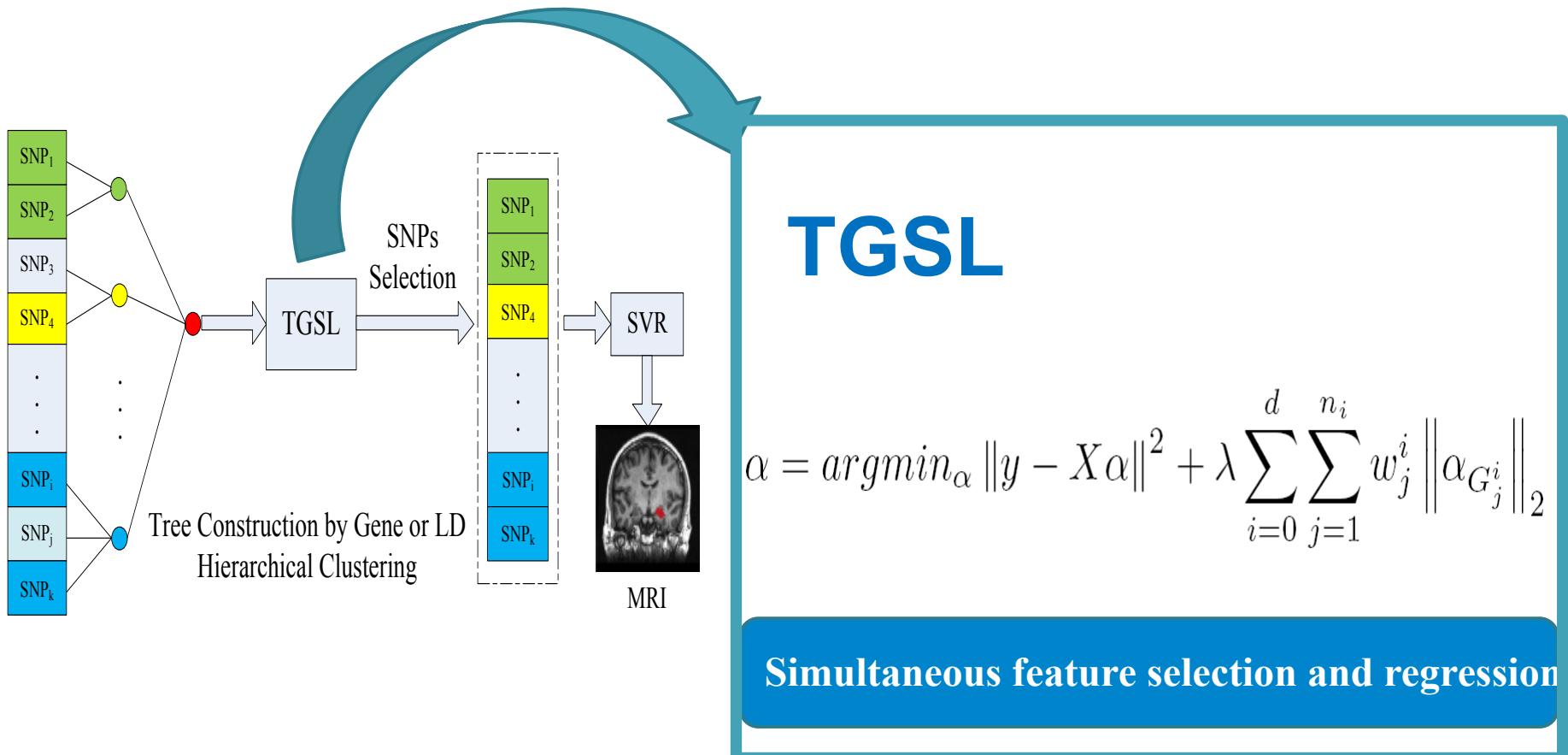


(a) group by gene

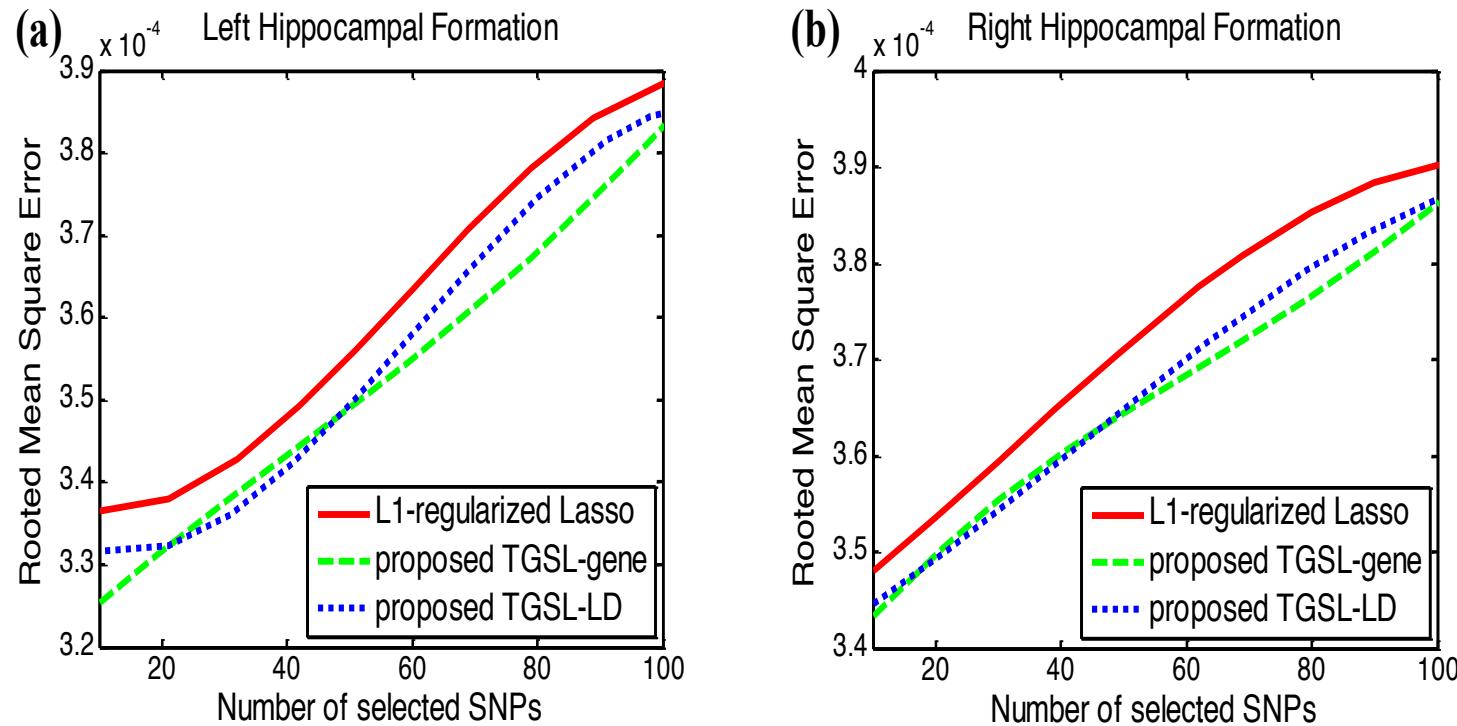


(b) group by linkage
disequilibrium (LD) cluster

Tree-Guided Sparse Learning

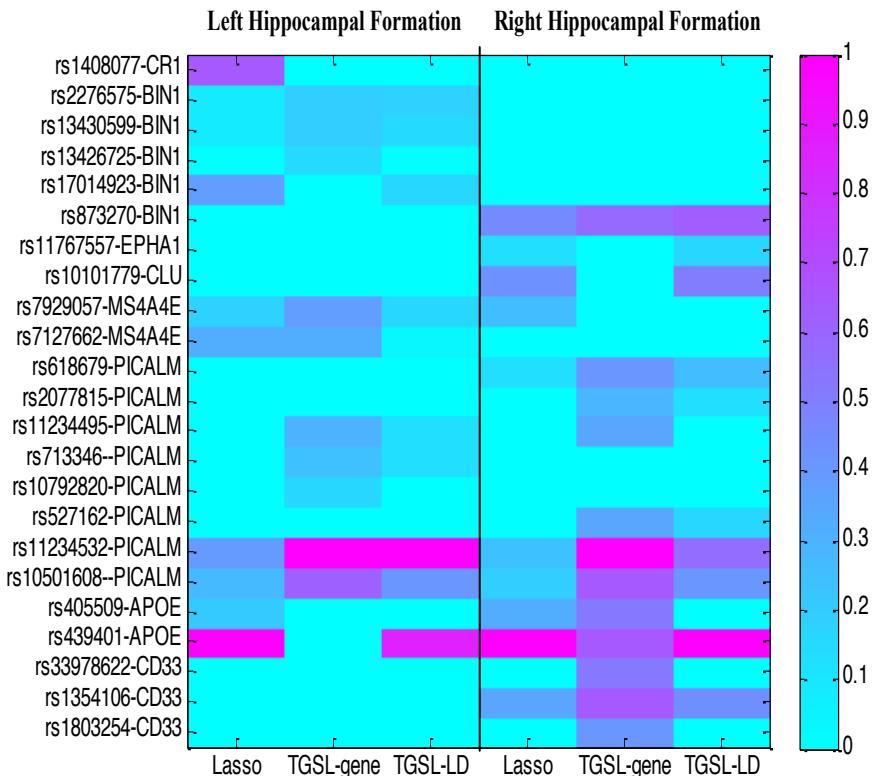


Results



Comparison of RMSE with respect to different number of selected SNPs

Genetic Marker Selection



Regression coefficients for the top 10 selected SNPs

The
Chromosome-Wide Association and Linkage Study in the Amish Detects a Novel Candidate Late-Onset Alzheimer Disease Gene

Biomarkers
we found are
consistent

nature.com > Journal home > Table of Contents

with the
original Article

Molecular Psychiatry (5 November 2013) | doi:10.1038/mp.2013.146

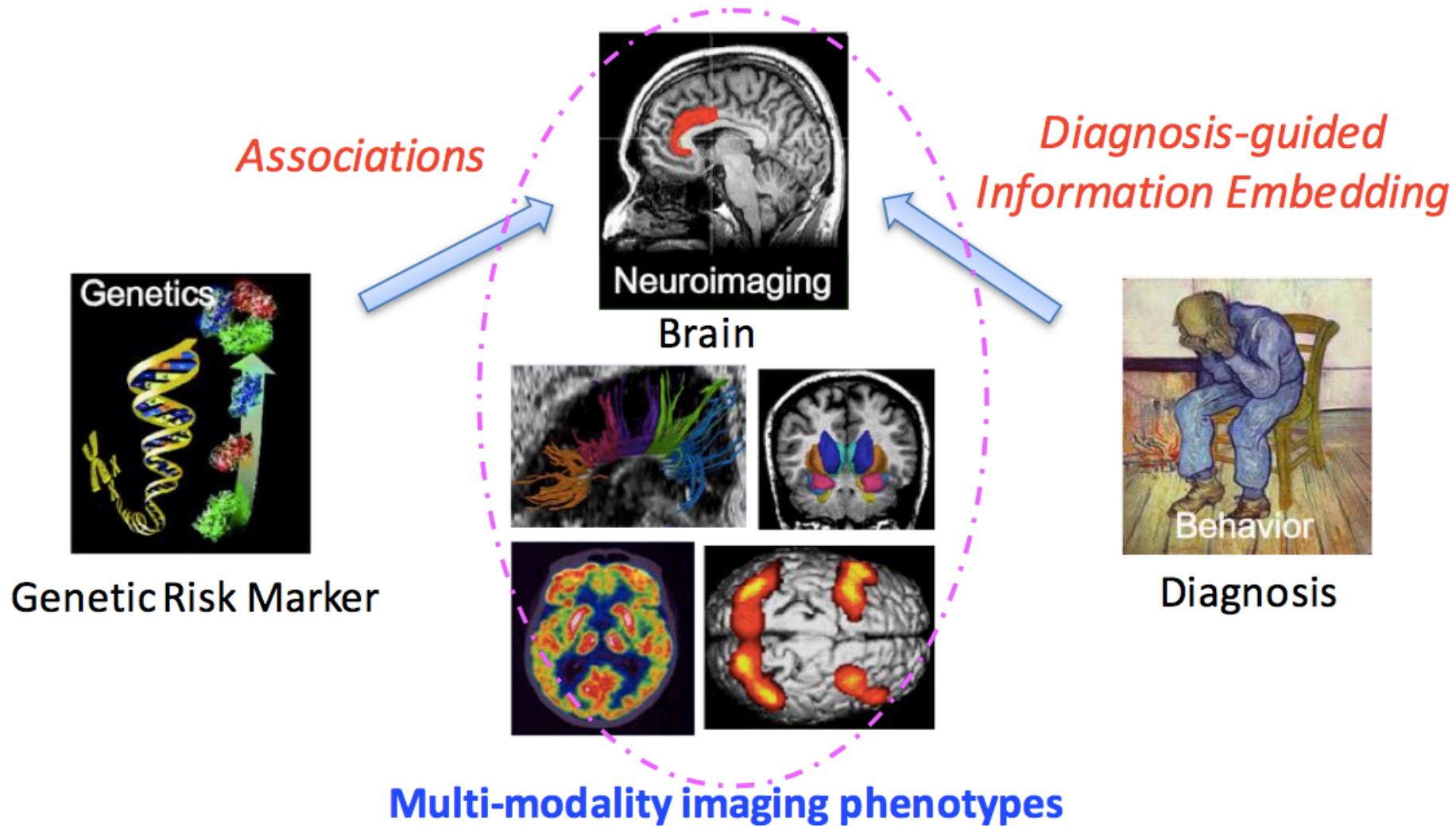
Common genetic variants on 1p13.2

associate with risk of autism

K Xia, H Guo, Z Hu, G Xun, L Zuo, Y Peng, K Wang, Y He, Z Xiong, L Sun, Q Pan, Z Long, X Zou, X Li, W Li, X Xu, L Lu, Y Liu, Y Hu, D Tian, L Long, J Ou, Y Liu, Y Li, L Zhang, Z Pan, J Chen, H Peng, Q Liu, X Luo, W Su, L Wu, D Hui, H Bai, Y Lin, Y Tang, Y Jiang, J Li, Z Miedzybrodzka, J Xia, Z Zhang, X Luo, X Zhang, D St Clair, J Zhao and F Zhang

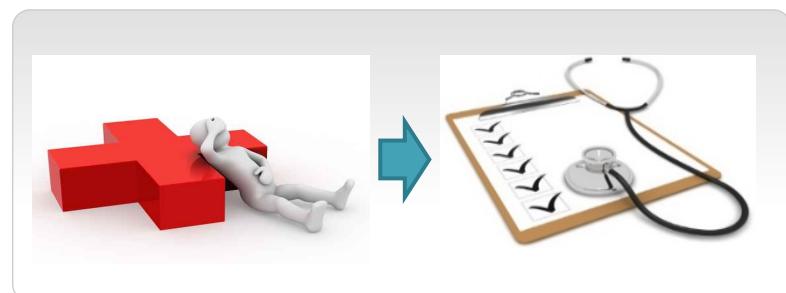
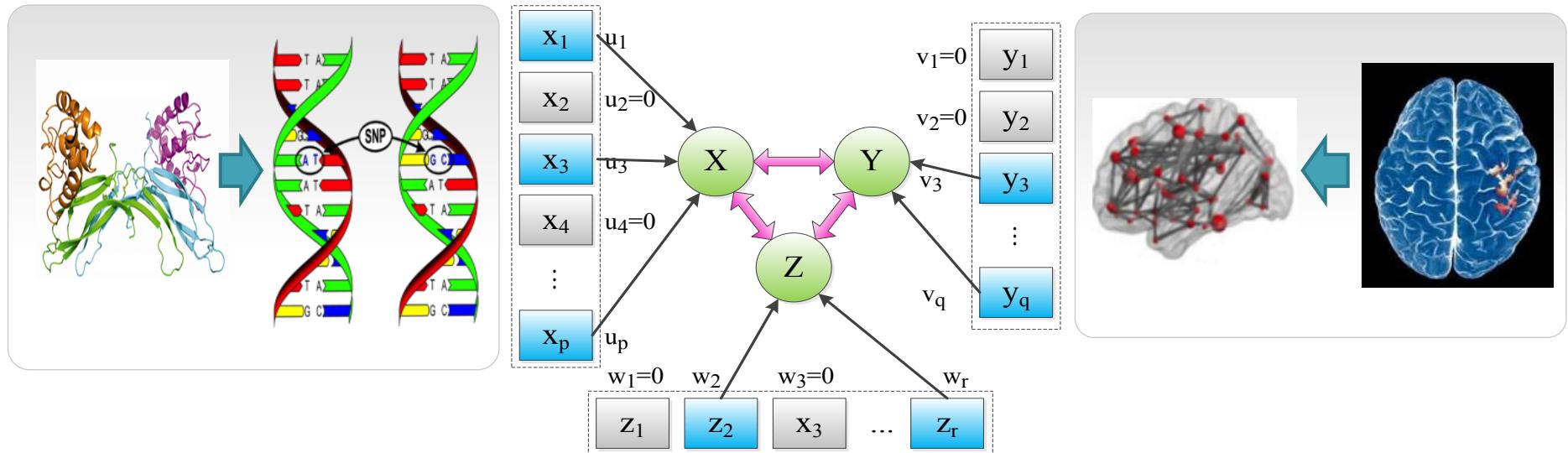
Molecular
Psychiatry

Multi-modality Phenotype Associations



(X. Hao, et al. PSB, 2016; Neuroinformatics 2016)

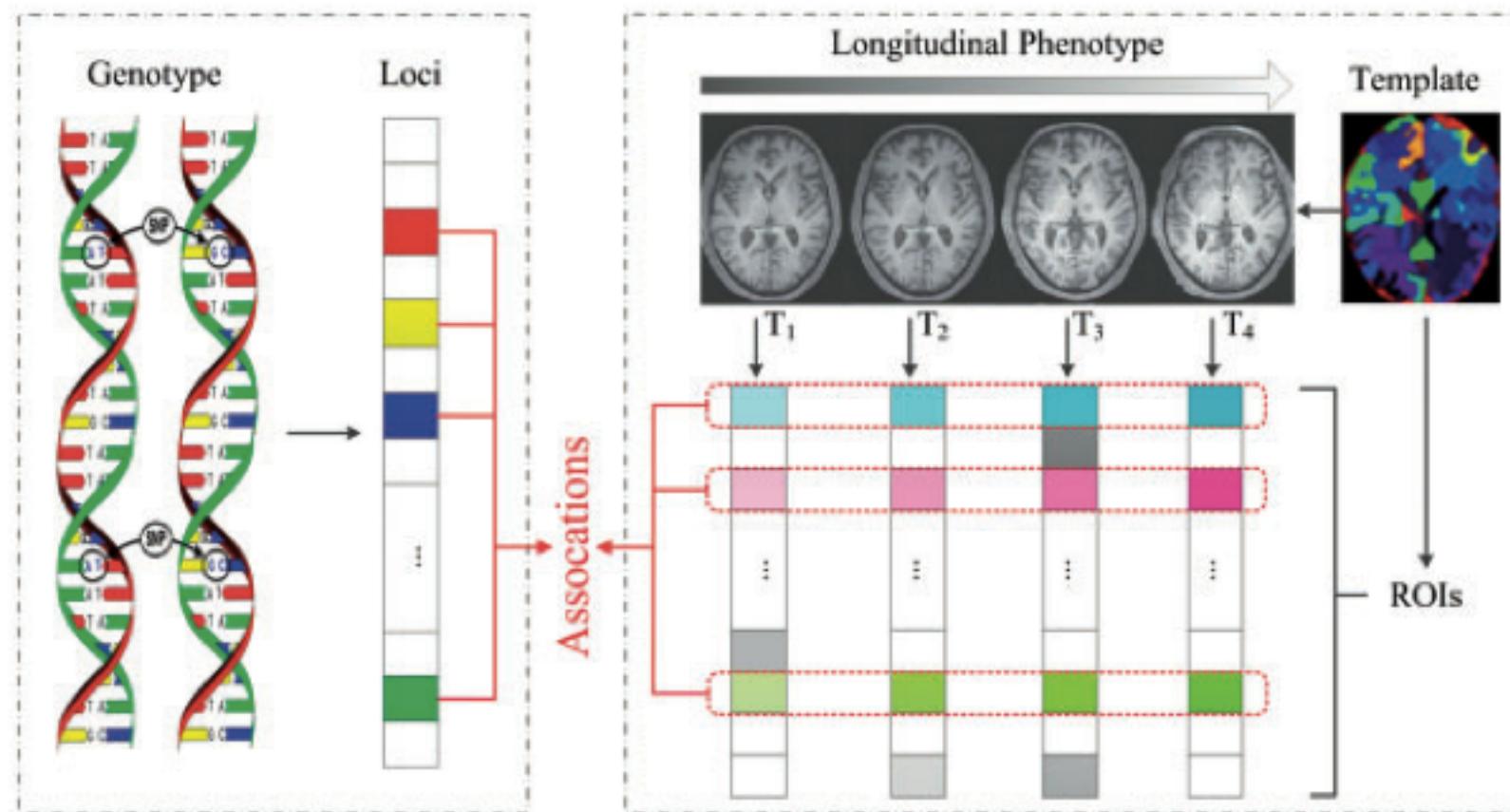
Multi-SNP-Multi-QT Associations



$$\begin{aligned} & \max_{w_1^T, w_2^T, w_3^T} w_1^T X^T Y w_2 + w_1^T X^T Z w_3 + w_2^T Y^T Z w_3 \\ \text{s. t. } & w_1^T X^T X w_1 \leq 1, \quad w_2^T Y^T Y w_2 \leq 1, \quad w_3^T Z^T Z w_3 \leq 1, \\ & \|w_1\|_1 \leq c_1, \quad \|w_2\|_1 \leq c_2, \quad \|w_3\|_1 \leq c_3 \end{aligned}$$

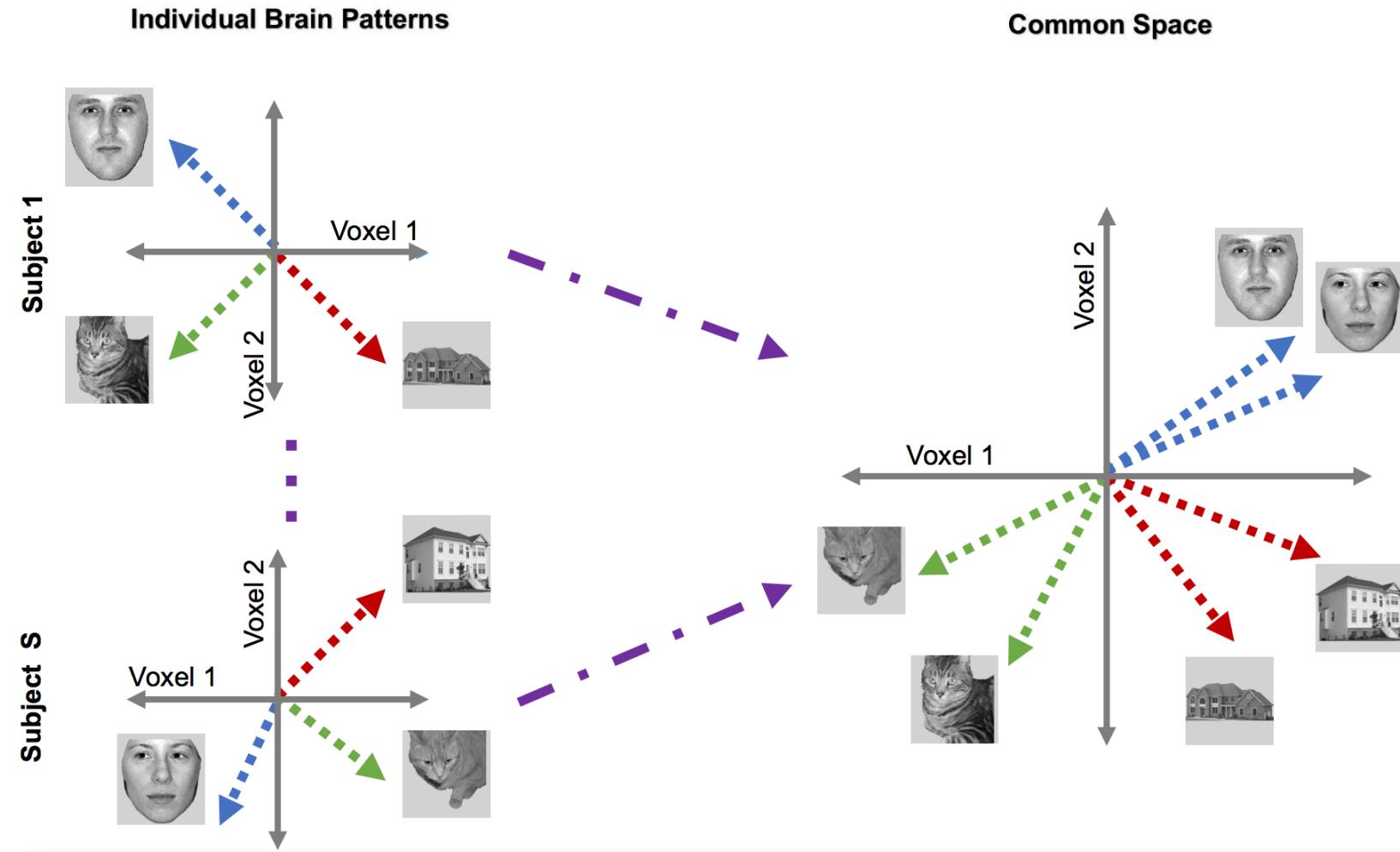
(X. Hao, et al. Scientific Reports 2017)

Longitudinal Phenotype Associations

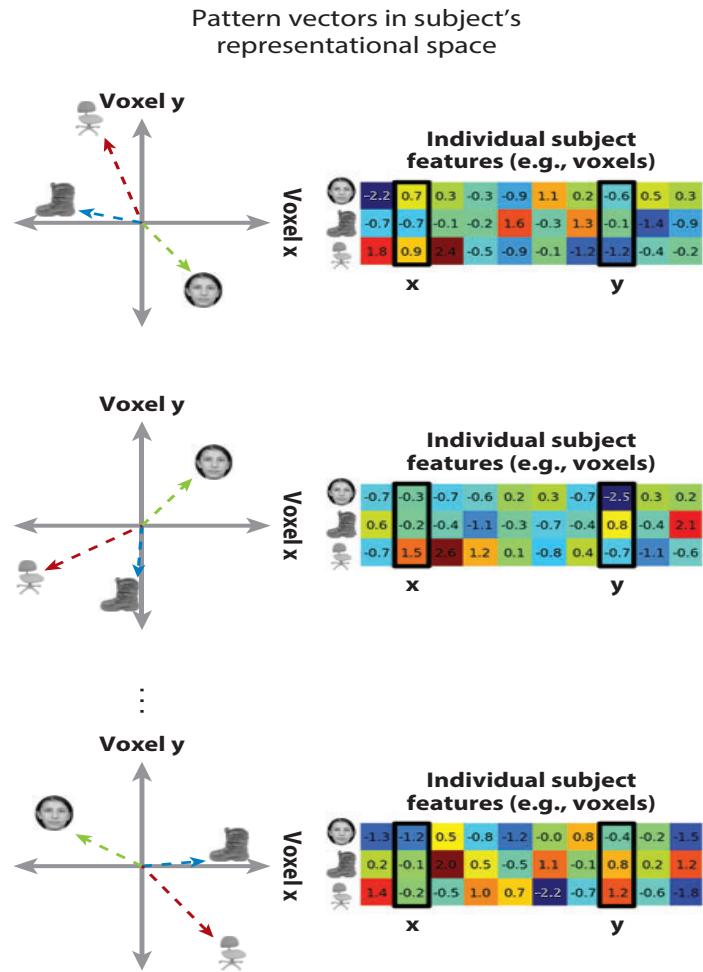


(X. Hao, et al. Bioinformatics 2017)

Brain Decoding: Hyperalignment



Brain Decoding: Hyperalignment



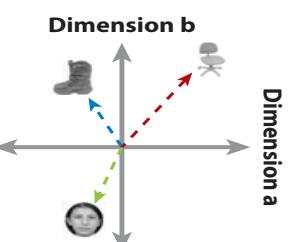
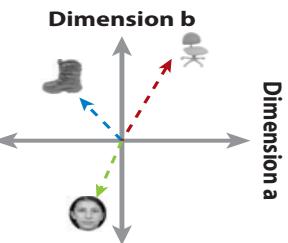
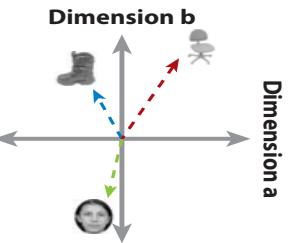
Hyperalignment of representational spaces

Transformation matrices (hyperalignment parameters)

Common model dimensions

Individual subject features	Common model dimensions
	0.4 0.3 0.3 -0.6 0.0 0.2 -0.1 0.4 0.3 -0.2 0.0 0.5 0.1 -0.5 0.2 0.2 0.2 -0.2 -0.5 -0.1 -0.4 0.3 0.1 -0.3 -0.2 -0.1 0.5 0.2 -0.2 0.4 -0.2 0.1 0.1 -0.3 -0.2 0.2 -0.5 -0.5 0.2 0.4 -0.0 0.6 -0.4 -0.3 -0.3 -0.2 0.0 0.3 0.4 0.2 -0.2 0.0 0.1 -0.2 0.5 -0.8 -0.3 -0.0 0.0 0.0 -0.5 -0.2 0.6 0.1 -0.2 0.0 0.0 0.0 0.2 -0.5 0.5 0.0 0.1 -0.1 -0.6 -0.5 0.2 -0.4 -0.1 -0.1 -0.3 0.3 -0.6 -0.2 -0.2 0.0 -0.1 -0.0 -0.6 0.0 0.1 -0.1 -0.1 0.3 0.0 0.5 -0.5 0.6 -0.1
	=
	Common model dimensions
	-1.6 0.3 0.5 0.9 1.3 -0.7 0.4 -0.9 1.0 0.0 -0.4 1.5 0.9 1.1 -0.8 -0.3 -0.2 1.2 0.8 0.7 0.1 1.6 0.8 -1.6 1.3 1.1 1.3 1.5 -0.9 1.1
a	
b	

Pattern vectors in common model space



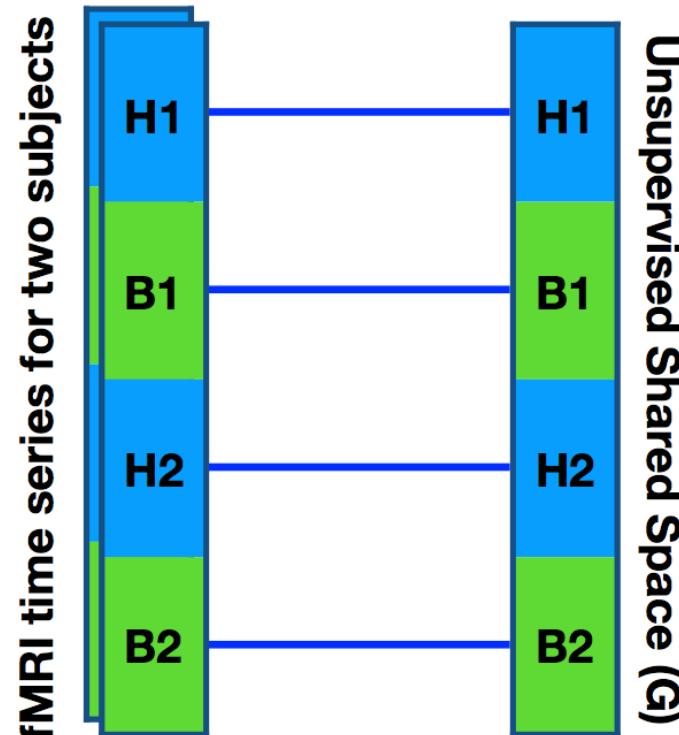
Individual subject features	Common model dimensions
	0.4 -0.3 -0.6 -0.1 -0.0 -0.3 -0.2 0.4 0.2 0.2 0.0 0.3 -0.1 -0.1 -0.5 -0.2 0.6 0.4 -0.4 0.0 -0.2 0.0 0.0 -0.5 0.4 0.6 0.1 0.5 0.0 -0.0 0.0 0.5 0.1 0.6 -0.0 -0.2 -0.1 -0.4 0.2 0.5 0.2 -0.2 0.4 -0.4 0.3 0.5 -0.0 0.1 -0.2 -0.5 0.1 0.4 -0.3 -0.3 -0.2 0.3 0.3 0.6 -0.2 -0.2 0.4 0.1 0.1 -0.0 -0.2 0.3 -0.6 0.1 0.6 0.1 0.4 0.0 0.3 -0.1 -0.5 0.2 -0.0 0.1 0.6 -0.2 0.5 0.3 -0.3 -0.2 -0.3 -0.1 -0.4 0.1 0.0 -0.5 -0.5 -0.5 0.3 -0.1 -0.4 -0.1 -0.1 0.1 0.0 0.4
	=
	Common model dimensions
	1.4 -0.3 0.4 0.9 1.2 -0.7 0.4 -0.9 1.2 -0.0 -0.5 0.4 0.9 1.0 -0.9 -0.2 -0.1 1.1 0.9 0.8 0.1 1.7 0.8 -1.7 1.3 1.1 1.3 1.5 1.0 1.2
a	
b	

Individual subject features	Common model dimensions
	0.4 0.4 0.4 -0.5 -0.2 0.1 -0.2 0.3 0.3 0.0 0.4 0.2 0.1 0.1 0.2 0.2 -0.4 -0.2 0.4 0.5 -0.7 0.3 0.0 0.2 -0.1 0.1 0.1 0.3 0.4 0.3 0.1 -0.1 -0.2 -0.4 -0.4 0.2 0.3 -0.2 0.6 0.3 -0.3 0.0 0.5 -0.3 0.6 0.2 -0.3 0.0 0.0 0.1 -0.2 0.3 0.2 -0.6 -0.1 -0.6 -0.3 0.2 -0.0 0.5 0.2 0.6 0.1 -0.3 0.2 0.1 -0.3 0.1 0.3 0.1 0.1 0.2 0.5 -0.1 0.6 0.1 0.5 -0.2 0.1 0.1 0.2 0.4 0.3 -0.3 -0.1 0.6 -0.3 0.3 -0.1 0.3 -0.3 -0.3 0.1 -0.3 -0.0 -0.1 0.4 0.6 -0.3
	=
	Common model dimensions
	-1.5 -0.2 0.5 1.0 1.2 -0.6 0.4 -0.8 -1.3 -0.1 -0.4 -1.6 0.8 1.1 -0.9 -0.3 -0.1 1.2 0.9 0.7 0.1 1.6 0.8 -1.7 1.4 1.1 1.3 1.5 -0.9 1.1
a	
b	

Local Discriminant Hyperalignment

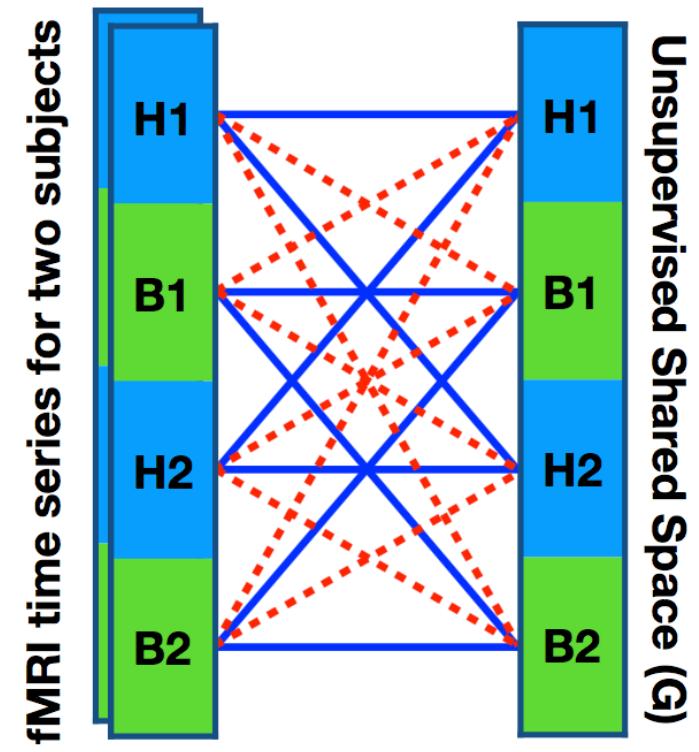


Unsupervised Hyperalignment (HA)



(a)

Local Discriminant HA (LDHA)



(b)

— Maximizing Correlation ----- Minimizing Correlation House Stimulus Bottle Stimulus

(Yousefnezhad & Zhang, AAAI 2017)

Results



Table 1: Accuracy of Classification Methods

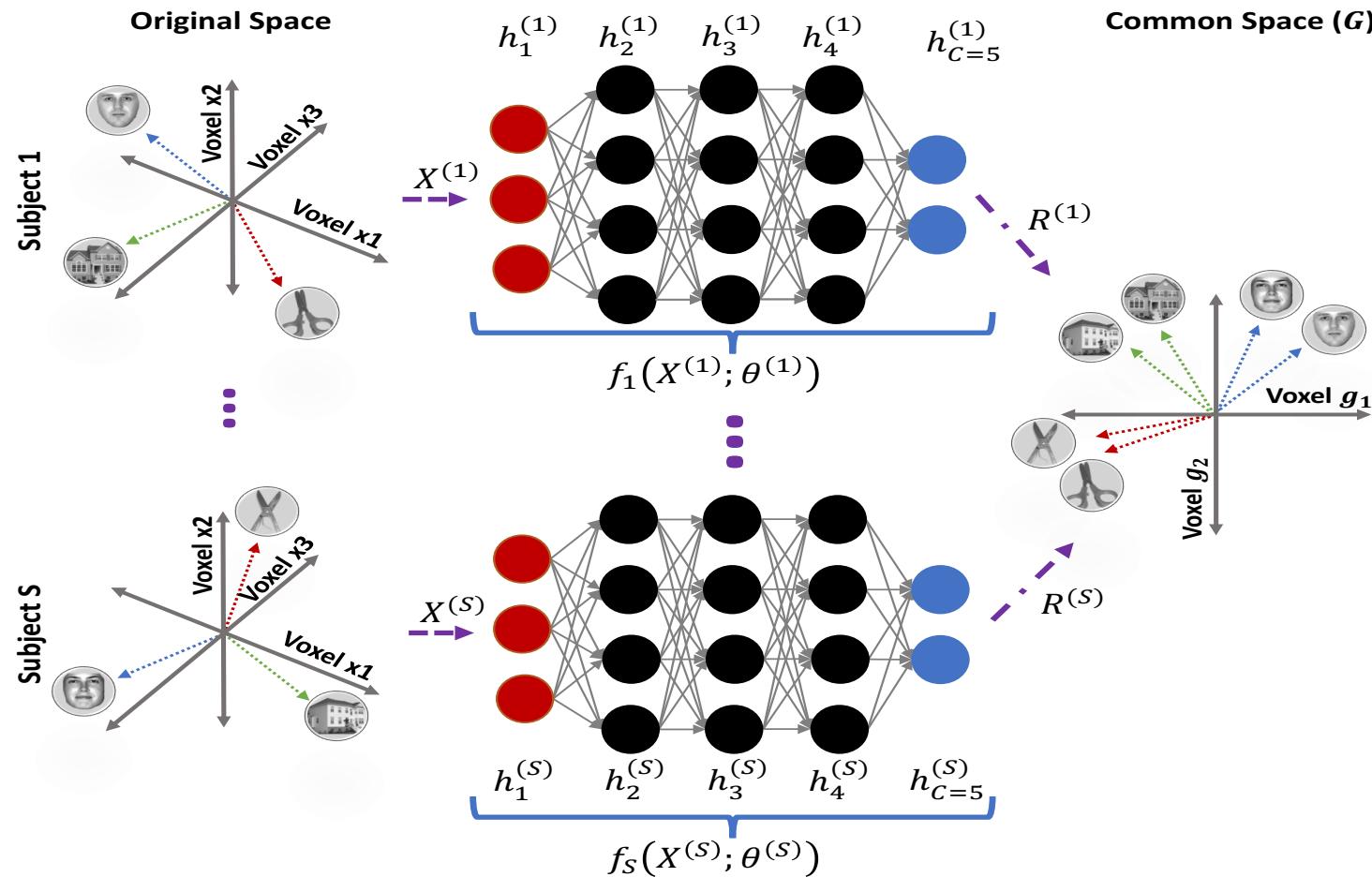
Data Sets	ν -SVM	HA	KHA	SCCA	SVD-HA	LDHA
DS005 (2 classes)	71.65±0.97	81.27±0.59	83.06±0.36	85.29±0.49	90.82±1.23	94.32±0.16
DS105 (8 classes)	22.89±1.02	30.03±0.87	32.62±0.52	37.14±0.91	40.21±0.83	54.04±0.09
DS107 (4 classes)	38.84±0.82	43.01±0.56	46.82±0.37	52.69±0.69	59.54±0.99	74.73±0.19
DS117 (2 classes)	73.32±1.67	77.93±0.29	84.22±0.44	83.32±0.41	95.62±0.83	95.07±0.27

Table 2: Area Under the ROC Curve (AUC) of Classification Methods

Data Sets	ν -SVM	HA	KHA	SCCA	SVD-HA	LDHA
DS005 (2 classes)	68.37±1.01	70.32±0.92	82.22±0.42	80.91±0.21	88.54±0.71	93.25±0.92
DS105 (8 classes)	21.76±0.91	28.91±1.03	30.35±0.39	36.23±0.57	37.61±0.62	53.86±0.17
DS107 (4 classes)	36.84±1.45	40.21±0.33	43.63±0.61	50.41±0.92	57.54±0.31	72.03±0.37
DS117 (2 classes)	70.17±0.59	76.14±0.49	81.54±0.92	80.92±0.28	92.14±0.42	94.23±0.94

(Yousefnezhad & Zhang, AAAI 2017)

Deep Hyperalignment



(Yousefnezhad & Zhang, NIPS 2017)

Results



Table 1: Accuracy of HA methods in post-alignment classification by using simple task datasets

↓Algorithms, Datasets→	DS005	DS105	DS107	DS116	DS117
ν -SVM [17]	71.65±0.97	22.89±1.02	38.84±0.82	67.26±1.99	73.32±1.67
HA [1]	81.27±0.59	30.03±0.87	43.01±0.56	74.23±1.40	77.93±0.29
RHA [2]	83.06±0.36	32.62±0.52	46.82±0.37	78.71±0.76	84.22±0.44
KHA [3]	85.29±0.49	37.14±0.91	52.69±0.69	78.03±0.89	83.32±0.41
SVD-HA [4]	90.82±1.23	40.21±0.83	59.54±0.99	81.56±0.54	95.62±0.83
SRM [5]	91.26±0.34	48.77±0.94	64.11±0.37	83.31±0.73	95.01±0.64
SL [9]	90.21±0.61	49.86±0.4	64.07±0.98	82.32±0.28	94.96±0.24
CAE [6]	94.25±0.76	54.52±0.80	72.16±0.43	91.49±0.67	95.92±0.67
DHA	97.92±0.82	60.39±0.68	73.05±0.63	90.28±0.71	97.99±0.94

Outline



1

Backgrounds on Alzheimer's Disease

2

Brain-imaging based Analysis

3

Brain-network based Analysis

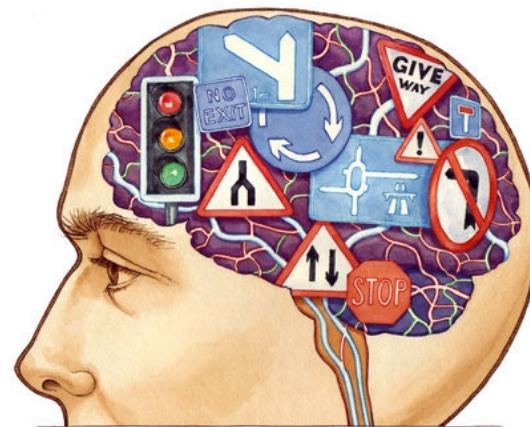
4

Summary

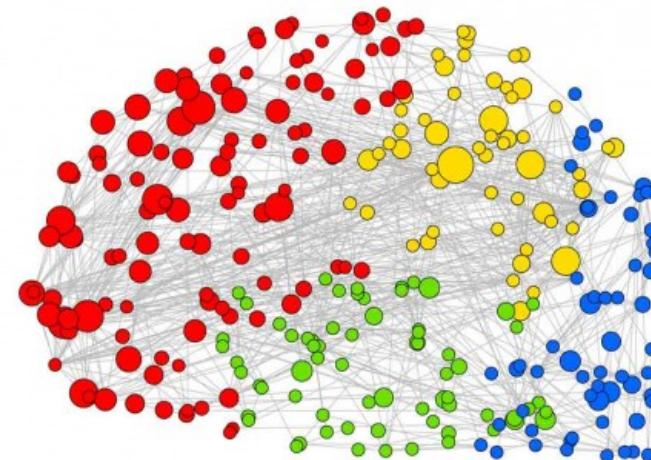
Brain Connectomics



- *Studies the interaction of brain functional regions at systems level*

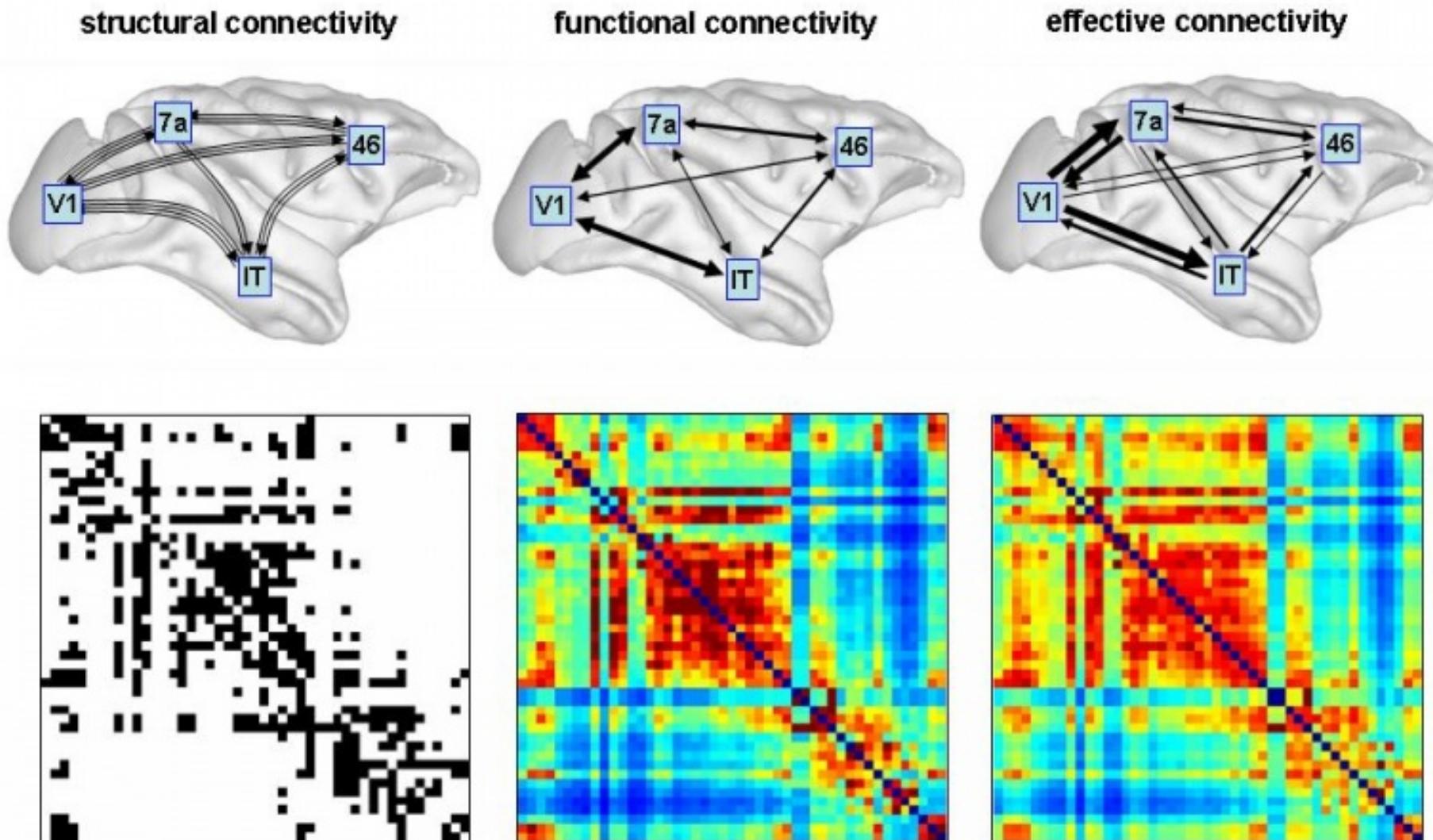


Steve Carroll



(Petra E. Vértes, et al., PNAS, 2012)

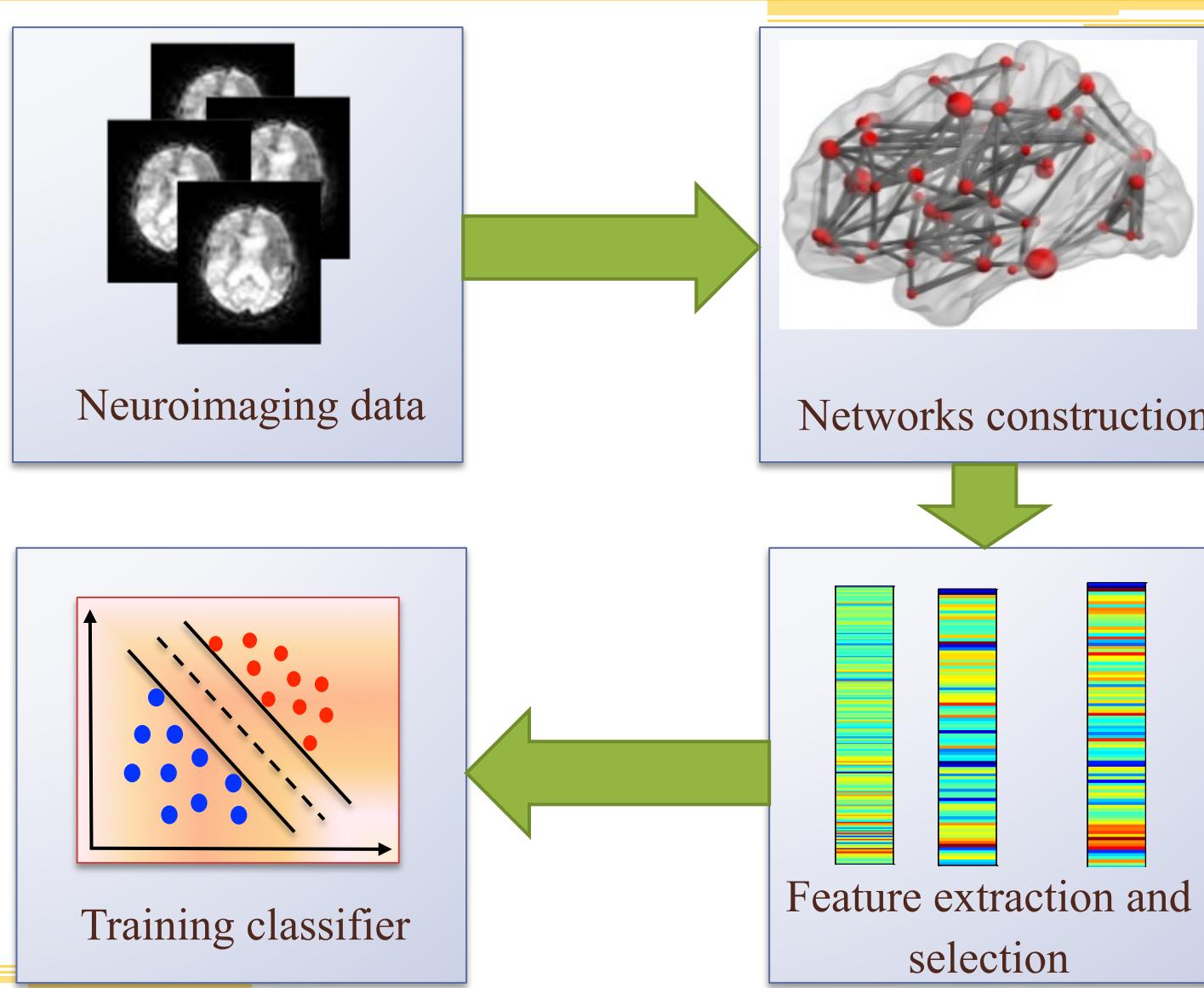
- *Mapping the human brain is one of the great scientific challenges of the 21st century*



(Honey et al., PNAS, 2007)

A decorative element consisting of several horizontal bars at the bottom of the slide. From left to right, there is a yellow bar with a thin white horizontal line across it, a dark blue bar, and a dark blue bar with a thick white horizontal line across it.

Network-based Classification



Network-based Classification

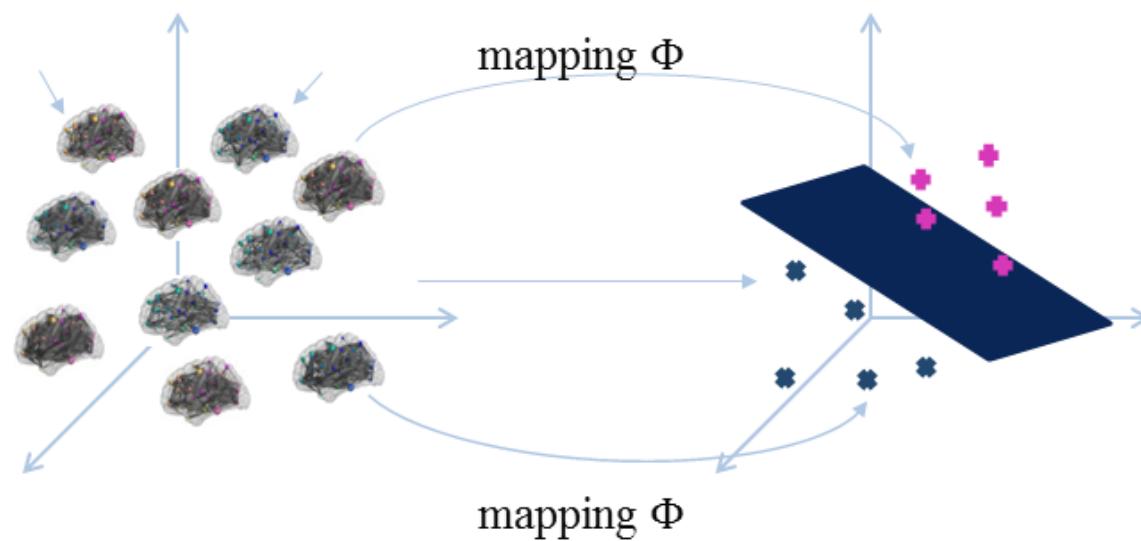


- Motivation
 - Brain connectivity networks have been used for classification of AD/MCI from normal controls (NC)
 - In conventional methods, **local measures of connectivity networks** are first extracted from each ROI as network features, and then concatenated into a long vector
 - Some useful structural information of network, especially **global topological information**, may be lost
- **Question:** How can we better preserve the network topological information for more effective brain network based classification?

Topological Graph Kernel



- Topology-based graph kernel
 - The kernel is defined on graphs, which can be used to compute the similarity of a pair of graphs



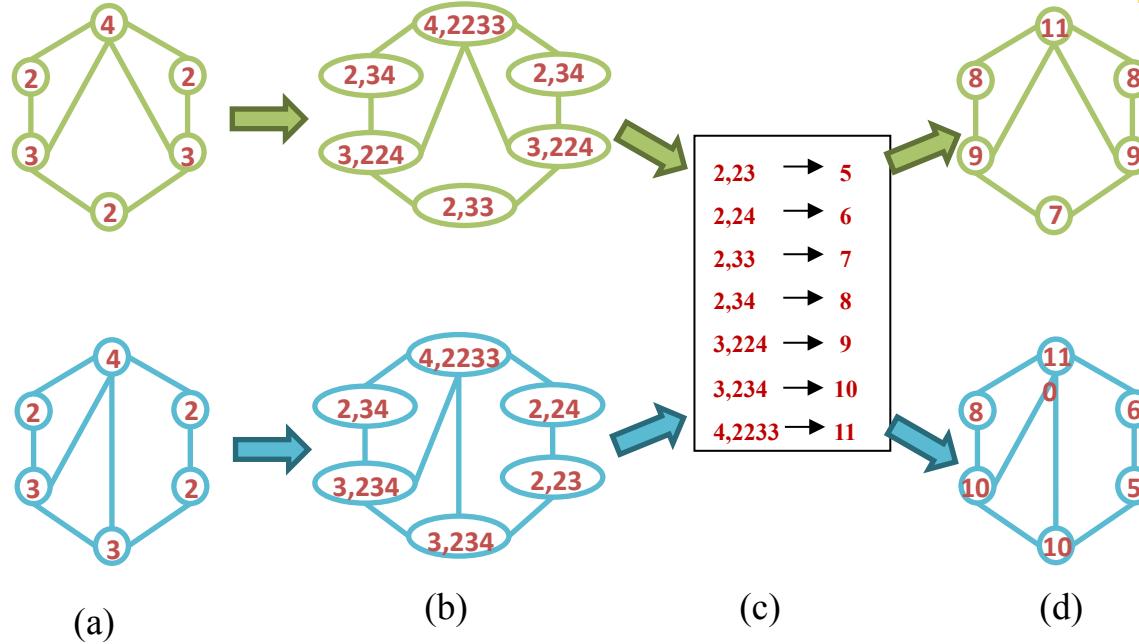
Subtree-based Graph Kernel



- Weisfeiler-Lehman test
 - First, label every vertex of a graph with degree of that vertex
 - Then, at each iteration, augment the label of each vertex in graph by the sorted set of node labels of neighboring nodes, and compress these augmented labels into new short labels
 - This process proceeds **iteratively** until the node label sets of two graphs differ, or the number of iteration reaches the maximum

(N. Shervashidze, et al., JMLR, 2011)

Example



$$L = \{L_0, L_1\} = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$$

$$\emptyset(G) = (3, 2, 1, 0, 0, 1, 2, 2, 0, 1)$$

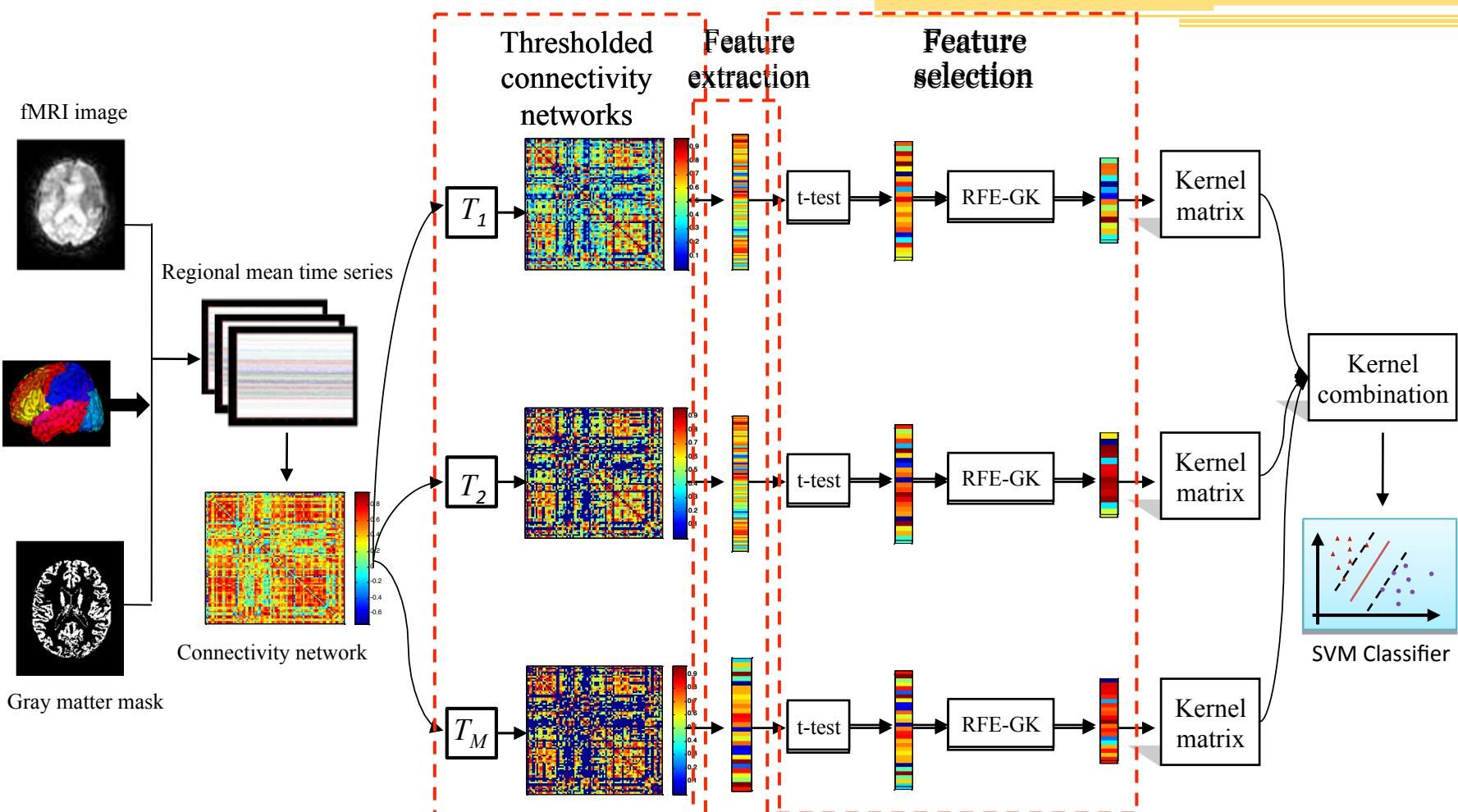
$$\emptyset(H) = (3, 2, 1, 1, 1, 0, 1, 0, 2, 1)$$

$$k(G, H) = \langle \emptyset(G), \emptyset(H) \rangle = 17$$

(e)

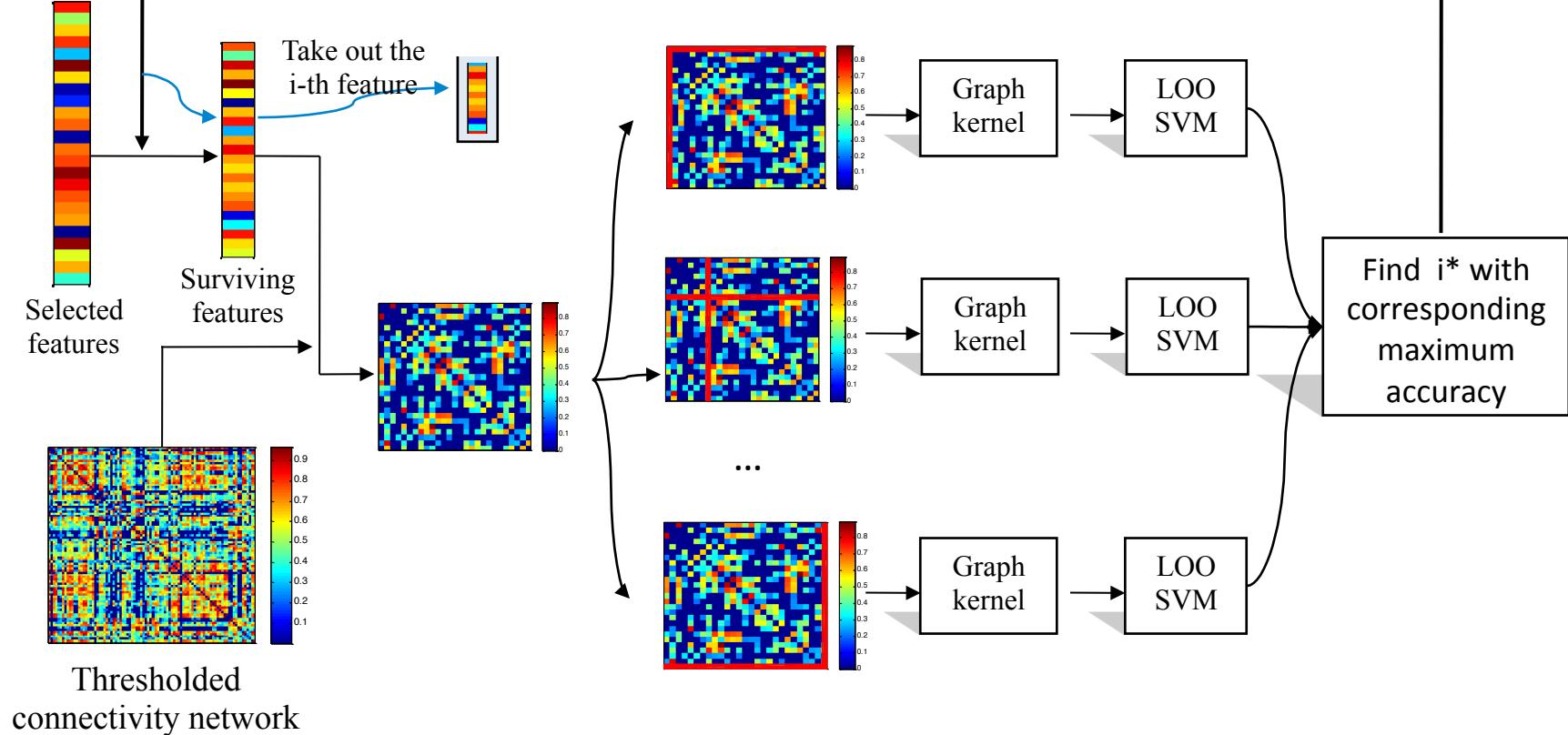
Illustration of the construction process of the Weisfeiler-Lehman subtree kernel with for two graphs G and H . Here, (a) the initial labeled graph G and H , (b) augmented labels on graph G and H , (c) label compression, (d) relabeled graph G and H , (e) computation of the kernel on Graph G and H

Flowchart



(B. Jie, D. Zhang, et al., Human Brain Mapping, 2014)

RFE-GK Flowchart



(B. Jie, D. Zhang, et al., Human Brain Mapping, 2014)

Experiment & Results



- Dataset: 12 MCI patients and 25 healthy controls

Group	MCI	Normal
No. of subjects (male/female)	6/6	9/16
Age (mean \pm SD)	75.0 ± 8.0	72.9 ± 7.9
Years of edu (mean \pm SD)	18.0 ± 4.1	15.8 ± 2.4
MMSE (mean \pm SD)	28.5 ± 1.5	29.3 ± 1.1

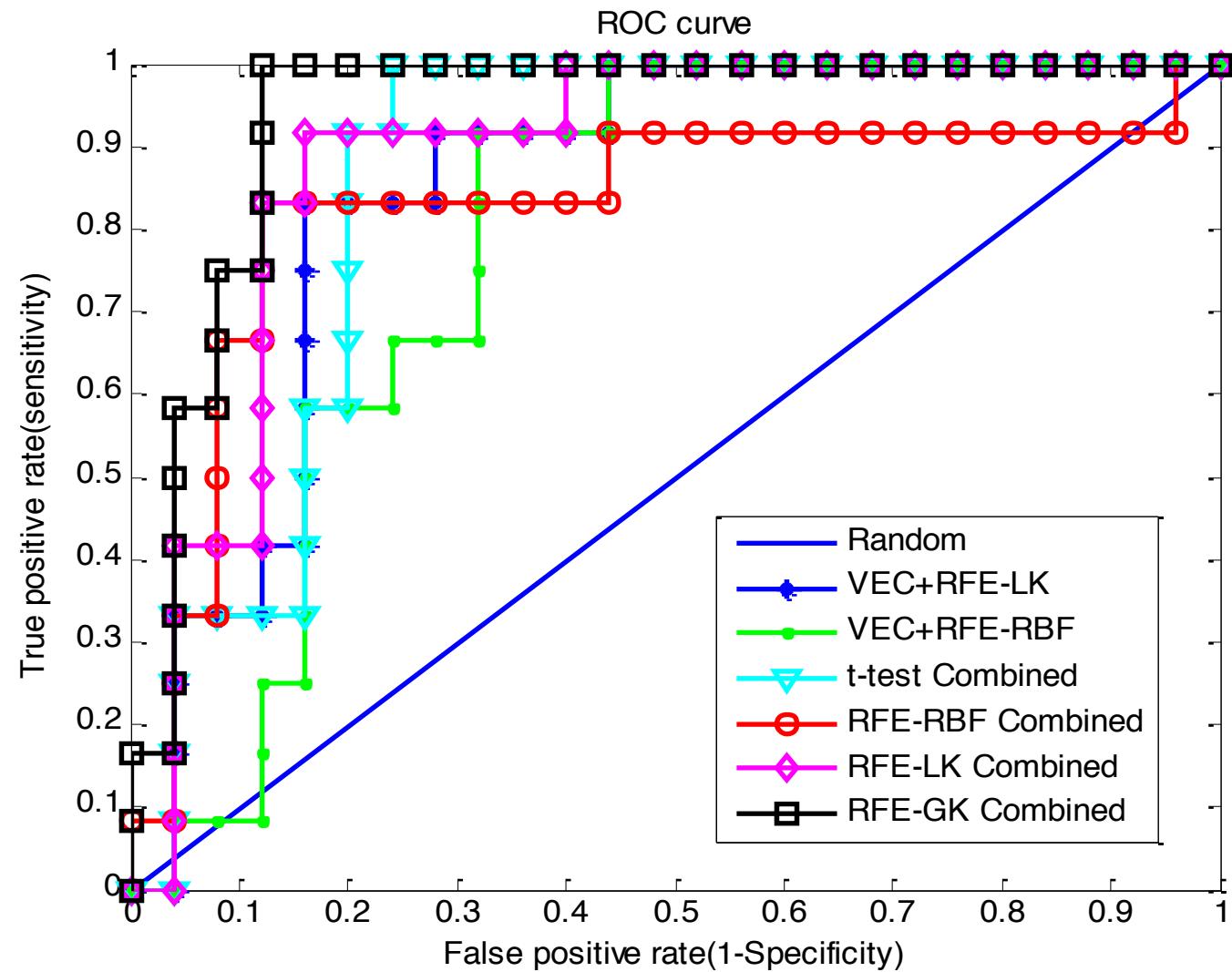
- A leave-one-out (LOO) cross-validation strategy was used to evaluate the classification performance.

Classification results

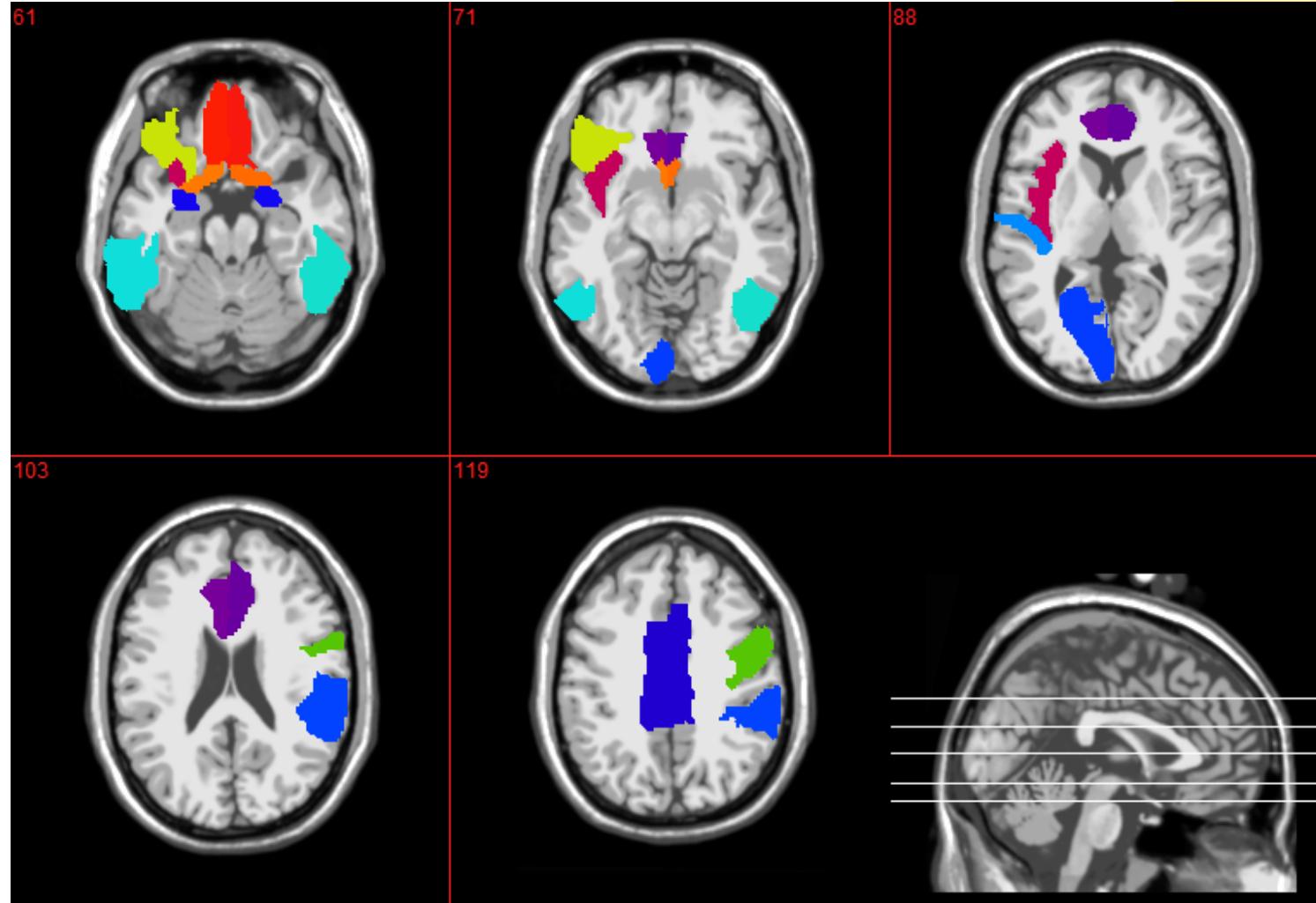


	Methods	T ₁	T ₂	T ₃	T ₄	T ₅	combined
ACC(%)	VEC-RFE-LK	-	-	-	-	-	83.8
	VEC-RFE-RBF	-	-	-	-	-	73.0
	t-test	75.7	78.4	64.9	64.9	64.9	81.1
	RFE-RBF	78.4	73.0	67.6	73.0	78.4	86.5
	RFE-LK	83.8	70.3	64.9	78.4	64.9	86.5
	RFE-GK	86.5	83.8	75.7	75.7	64.9	91.9
SEN(%)	VEC-RFE-LK	-	-	-	-	-	83.3
	VEC-RFE-RBF	-	-	-	-	-	66.7
	t-test	75.0	75.0	50.0	50.0	50.0	83.3
	RFE-RBF	58.3	66.7	25.0	33.3	50.0	83.3
	RFE-LK	91.7	58.3	41.7	66.7	50.0	91.7
	RFE-GK	91.7	75.0	58.3	66.7	50.0	100.0
SPE(%)	VEC-RFE-LK	-	-	-	-	-	84.0
	VEC-RFE-RBF	-	-	-	-	-	76.0
	t-test	76.0	80.0	72.0	72.0	72.0	80.0
	RFE-RBF	88.0	76.0	88.0	92.0	92.0	88.0
	RFE-LK	80.0	76.0	76.0	84.0	72.0	84.0
	RFE-GK	84.0	88.0	84.0	80.0	72.0	88.0
AUC	VEC-RFE-LK	-	-	-	-	-	0.85
	VEC-RFE-RBF	-	-	-	-	-	0.79
	t-test	0.84	0.86	0.74	0.71	0.68	0.86
	RFE-RBF	0.68	0.77	0.75	0.65	0.76	0.83
	RFE-LK	0.87	0.82	0.70	0.79	0.72	0.89
	RFE-GK	0.85	0.86	0.77	0.78	0.60	0.94

ROC curve



Top Selected ROIs



(B. Jie, D. Zhang, et al., Human Brain Mapping, 2014)

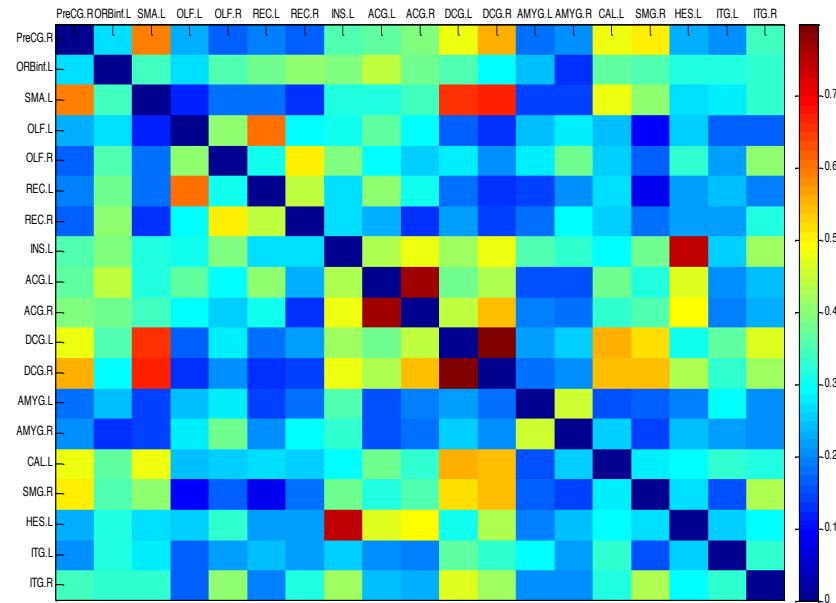
Selected ROIs



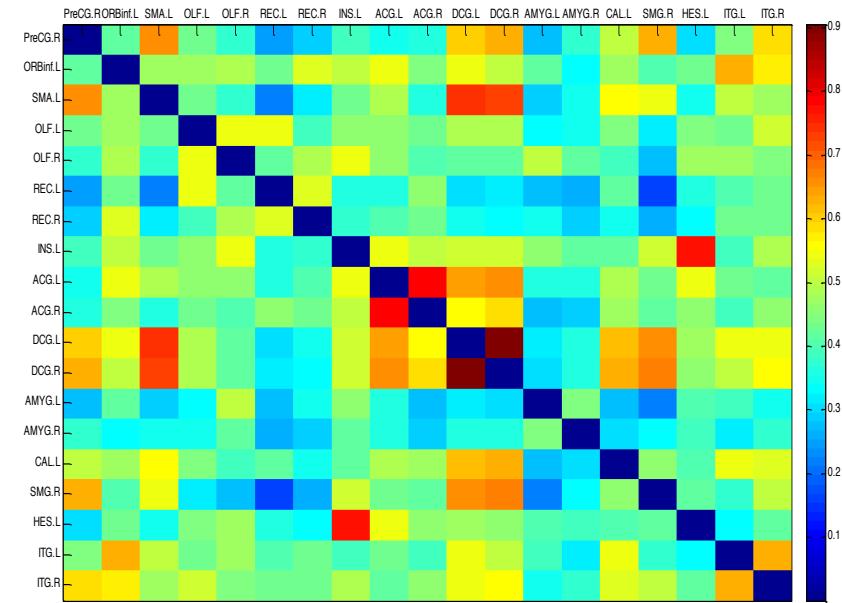
T₁	T₂	T₃
L olfactory cortex	L inferior temporal gyrus	L anterior cingulate gyrus
R inferior temporal gyrus	L olfactory cortex	L olfactory cortex
L inferior temporal gyrus	R inferior temporal gyrus	R middle cingulate
L anterior cingulate gyrus	R gyrus rectus	L amygdala
R supramarginal gyrus	R amygdala	L calcarine sulcus
L supplementary motor area	R precentral gyrus	R olfactory cortex
L orbital part of inferior frontal gyrus	L gyrus rectus	L middle cingulate
L gyrus rectus	L orbital part of inferior frontal gyrus	R inferior temporal gyrus
R gyrus rectus	L supplementary motor area	L orbital part of inferior frontal gyrus
R amygdala	R supramarginal gyrus	R amygdala
R precentral gyrus	L anterior cingulate gyrus	L heschl gyrus
R anterior cingulate gyrus	R anterior cingulate gyrus	L gyrus rectus

T₄	T₅	All
L olfactory cortex	L amygdala	R precentral gyrus
R middle cingulate	R middle cingulate	L orbital part of inferior frontal gyrus
R olfactory cortex	L orbital part of inferior frontal gyrus	L supplementary motor area
L gyrus rectus	L olfactory cortex	L olfactory cortex
L anterior cingulate gyrus	R olfactory cortex	R olfactory cortex
L amygdala	L gyrus rectus	L gyrus rectus
R inferior temporal gyrus	L anterior cingulate gyrus	R gyrus rectus
L orbital part of inferior frontal gyrus	R inferior temporal gyrus	L insula
R amygdala	R anterior cingulate gyrus	L anterior cingulate gyrus
L inferior temporal gyrus	L insula	R anterior cingulate gyrus
L calcarine sulcus	L inferior temporal gyrus	L middle cingulate
R anterior cingulate gyrus	L calcarine sulcus	R middle cingulate

Connectivity Sub-networks



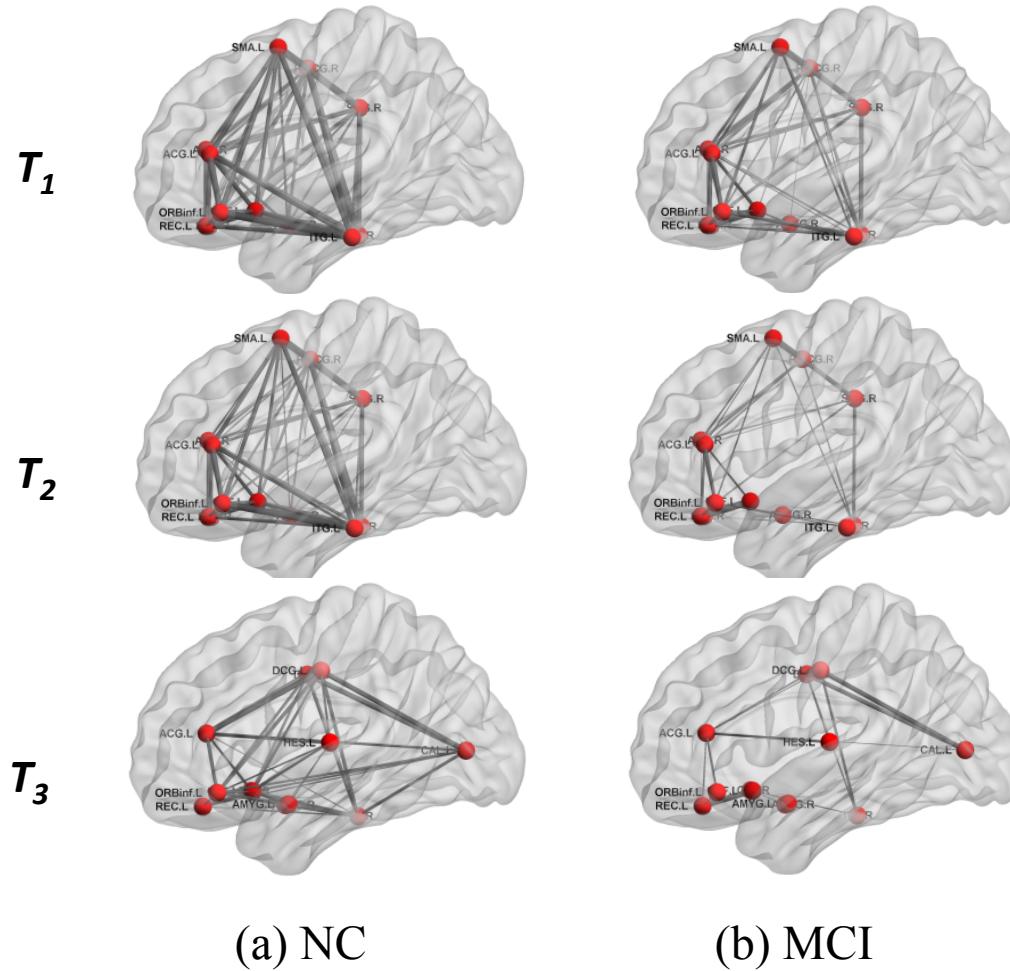
(a) MCI group



(a) NC group

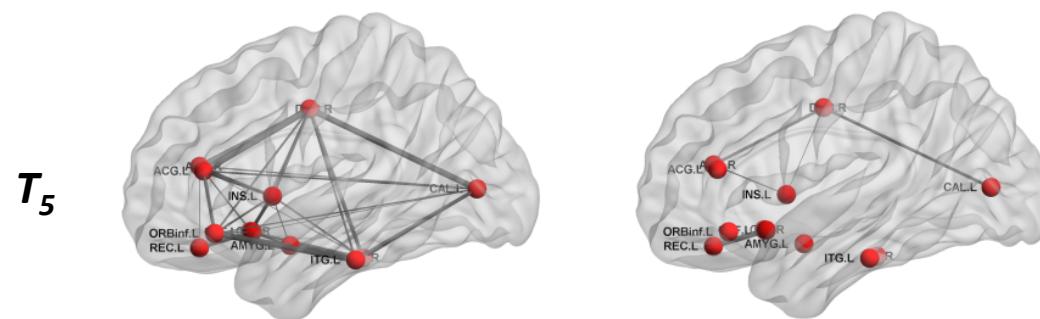
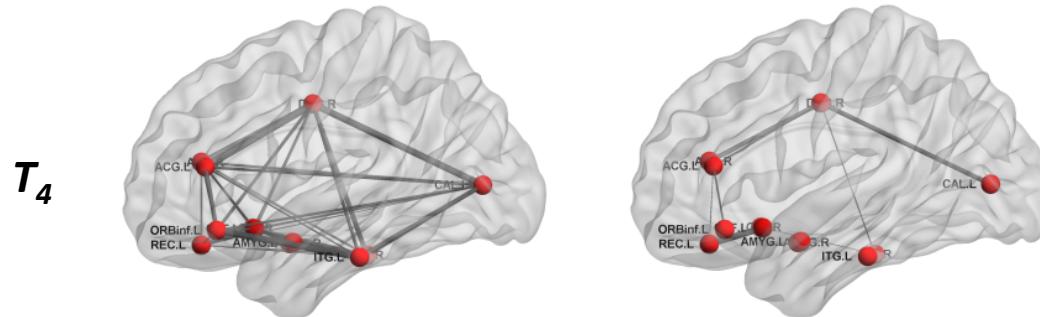
Visualization on average connectivity networks (matrices)
constructed using top selected ROIs

Brain Sub-networks



Thresholded average connectivity sub-network based
on top selected ROIs

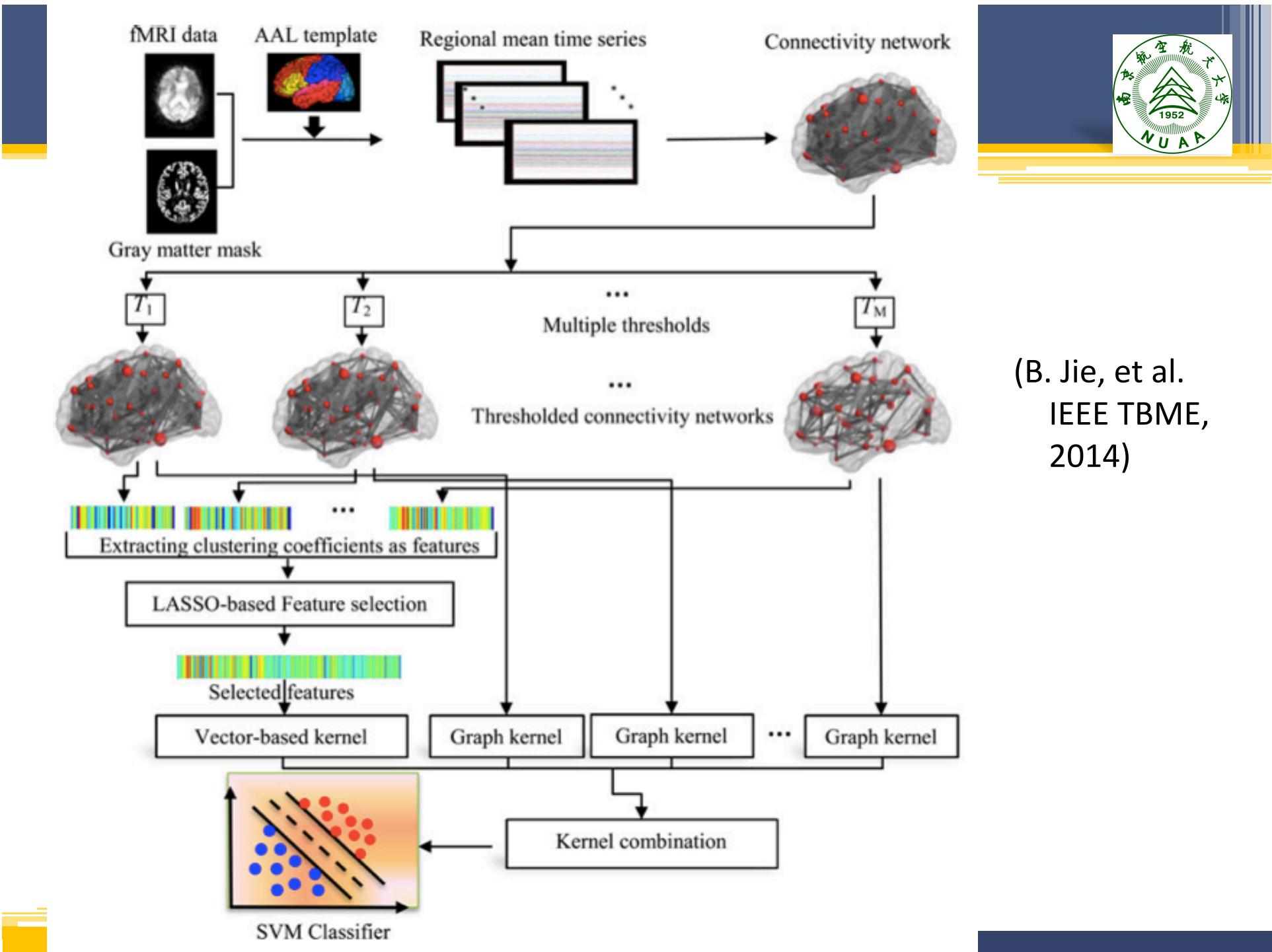
Brain Sub-networks



(a) NC

(b) MCI

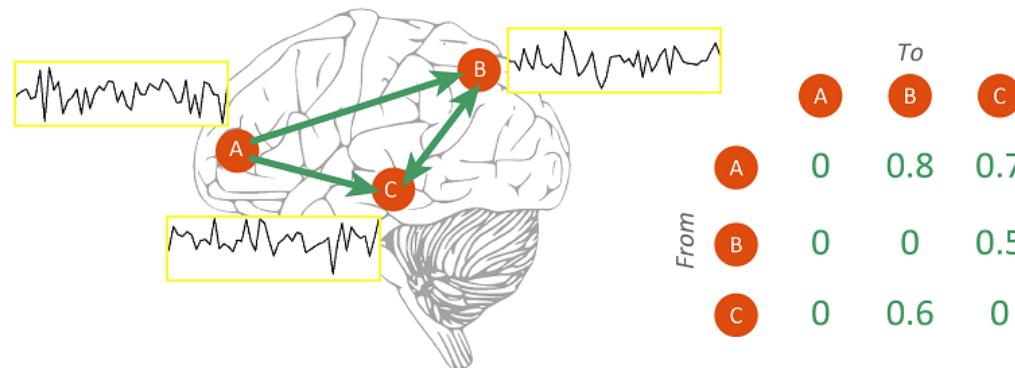
Thresholded average connectivity sub-network based on
top selected ROIs



Hyper-network based Method

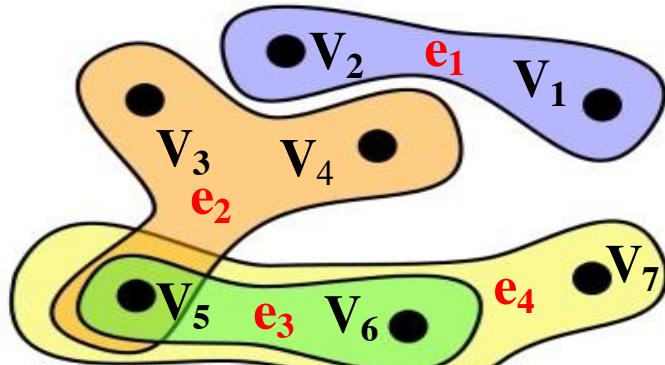
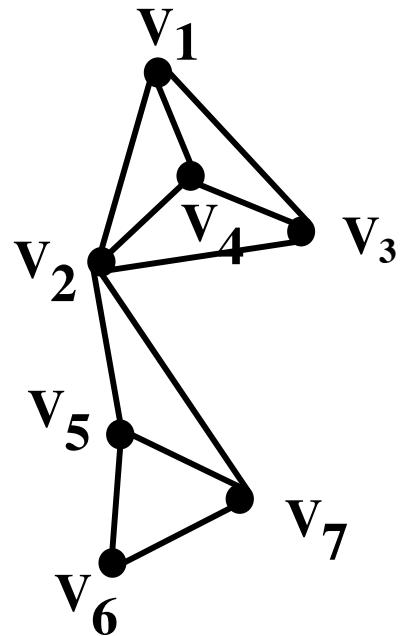


- Motivation
 - Conventional connectivity network is usually constructed based on the **pairwise correlation** among brain regions
 - Cannot reflect the useful **higher-order** relationship among brain regions



- **Question:** how to characterize the higher-order relationship among brain regions?

Solution: Hyper-graph



$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

$$\mathcal{V} = \{v_1, v_2, \dots, v_7\}, \quad \mathcal{E} = \{e_1, e_2, e_3, e_4\}$$

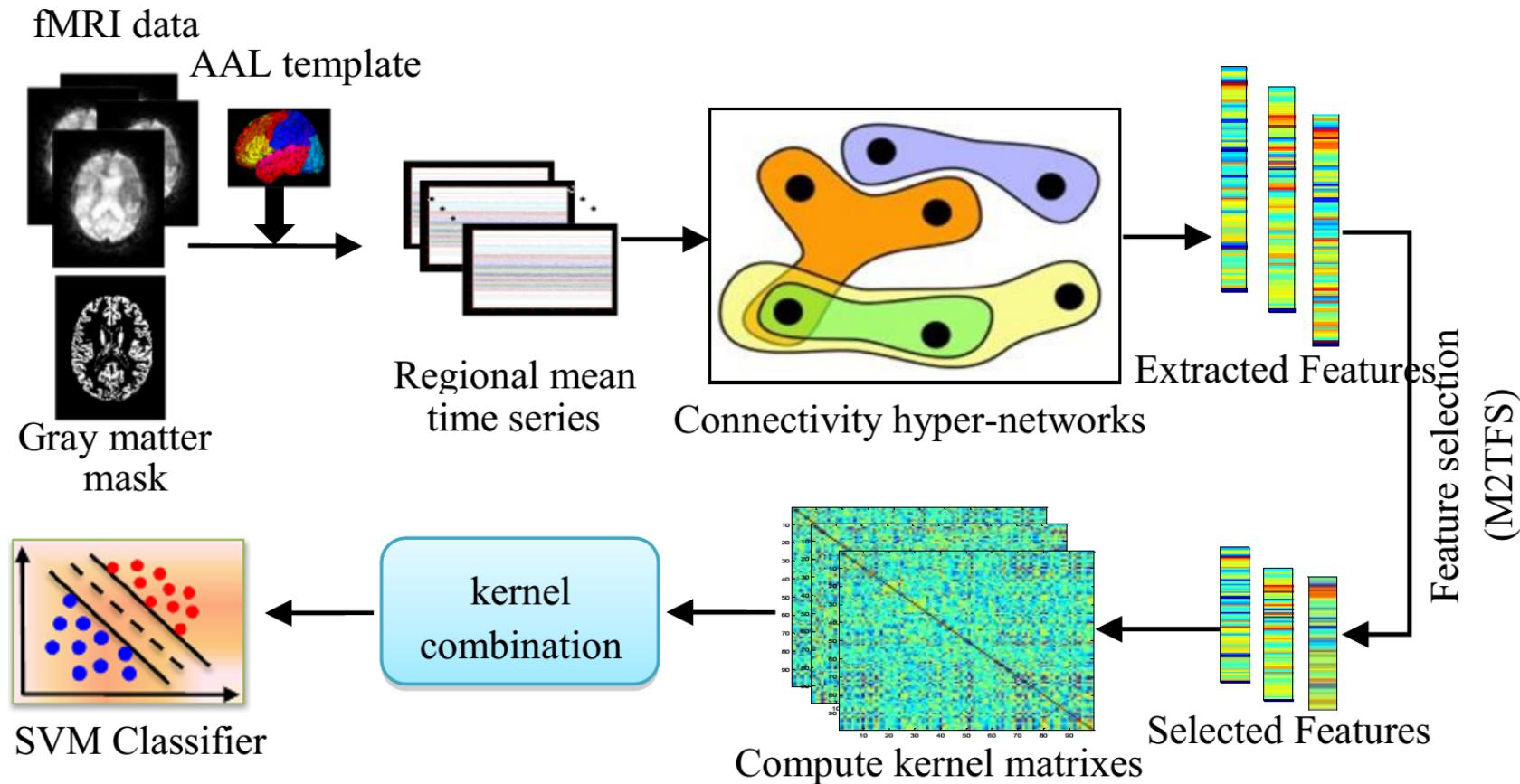
$$e_1 = \{v_1, v_2\}, \quad e_2 = \{v_3, v_4, v_5\}$$

$$e_3 = \{v_5, v_6\}, \quad e_4 = \{v_5, v_6, v_7\}$$

	e_1	e_2	e_3	e_4
v_1	1	0	0	0
v_2	1	0	0	0
v_3	0	1	0	0
v_4	0	1	0	0
v_5	0	1	1	1
v_6	0	0	1	1
v_7	0	0	0	1

Hyper-graph vs. graph. Left: a conventional graph in which two nodes are connected together by an edge. Middle: a hyper-graph in which each hyper-edge can connect more than two nodes. Right: the incidence matrix for the hyper-graph in the middle

Flowchart

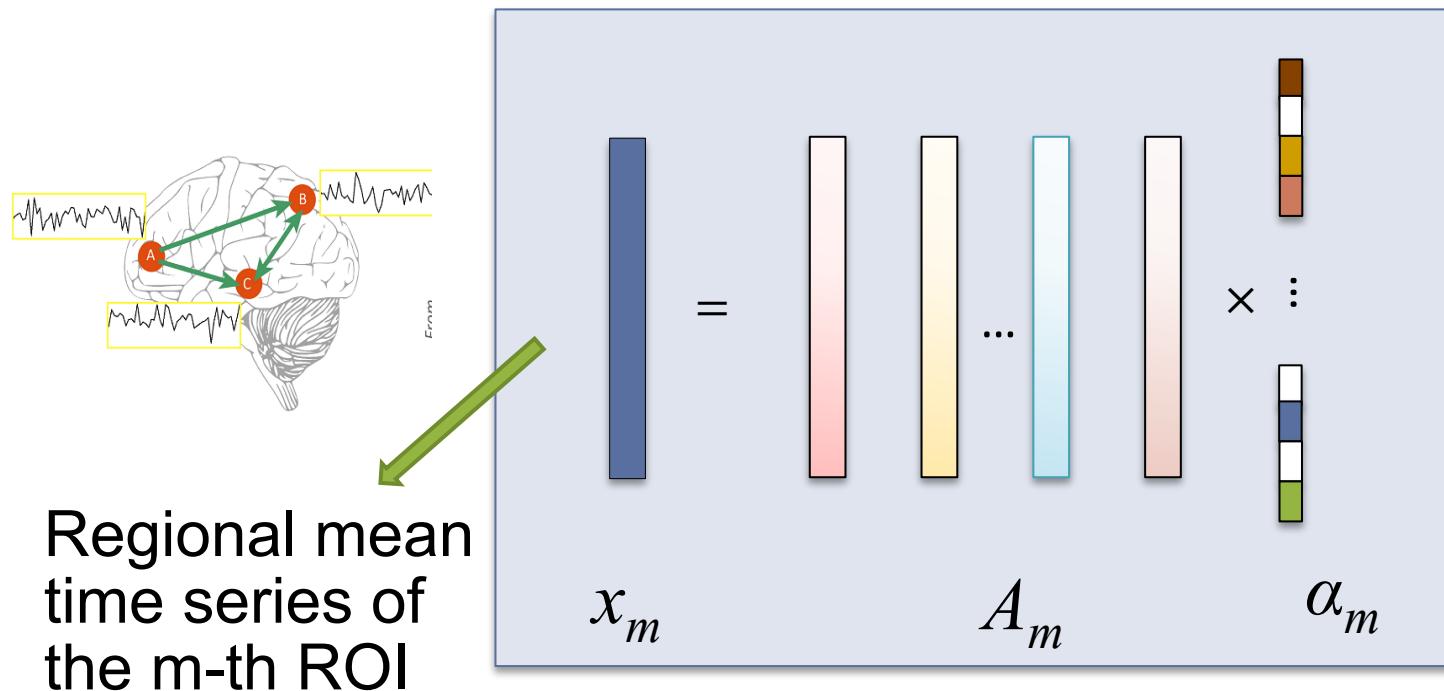


(B. Jie, D. Shen, D. Zhang, MICCAI'14; Medical Image Analysis 2016)

Hyper-network Construction



- Sparse representation



$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda \|\alpha_m\|_1$$



Feature Extraction

- Hyper-graph Clustering coefficients
 - hyper-network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $S(v) = \{e_i \in \mathcal{E}: v \in e_i\}$ represent the hyper-edges adjacent to the node v . Let $N(v) = \{u \in \mathcal{V}: \exists e \in \mathcal{E}, u, v \in e\}$ be the nodes that are neighbors of the node v . Then three different clustering coefficients on the node v can be, respectively, defined as:

$$HCC^1(v) = \frac{2 \sum_{u,t \in N(v)} I(u, t, \neg v)}{|N(v)|(|N(v)| - 1)}$$

$$HCC^2(v) = \frac{2 \sum_{u,t \in N(v)} I'(u, t, v)}{|N(v)|(|N(v)| - 1)}$$

$$HCC^3(v) = \frac{2 \sum_{e \in S(v)} (|e| - 1) - |N(v)|}{|N(v)|(|S(v)| - 1)}$$

- where $I(u, t, \neg v)=1$ if there exists $e_i \in \mathcal{E}$ such that $u, t \in e_i$ but $v \notin e_i$, and 0 otherwise; $I'(u, t, v)=1$ if there exists $e_i \in \mathcal{E}$ such that $u, t, v \in e_i$, and 0 otherwise.



Feature selection

- The manifold regularized multi-task feature selection (M2TFS)*

$$\min_W \frac{1}{2} \sum_{k=1}^3 \|Y - Z^k w^k\|_2^2 + \beta \sum_{k=1}^3 (Z^k w^k)^T L^k (Z^k w^k) + \gamma \|W\|_{2,1}$$

- where
 - $Z^k = [z_1^k, \dots, z_n^k, \dots, z_N^k]^T \in R^{N \times M}$ represent three sets of features,
 - $z_n^k = [HCC^k(v_i)]_{i=1:M}$ represents the vector of clustering coefficients
 - $Y = [y_1, y_2, \dots, y_N]^T \in R^N$ be the response vector for those N training subjects,

(B.Jie, D. Zhang, et al. MICCAI'13; Human Brain Mapping 2015)

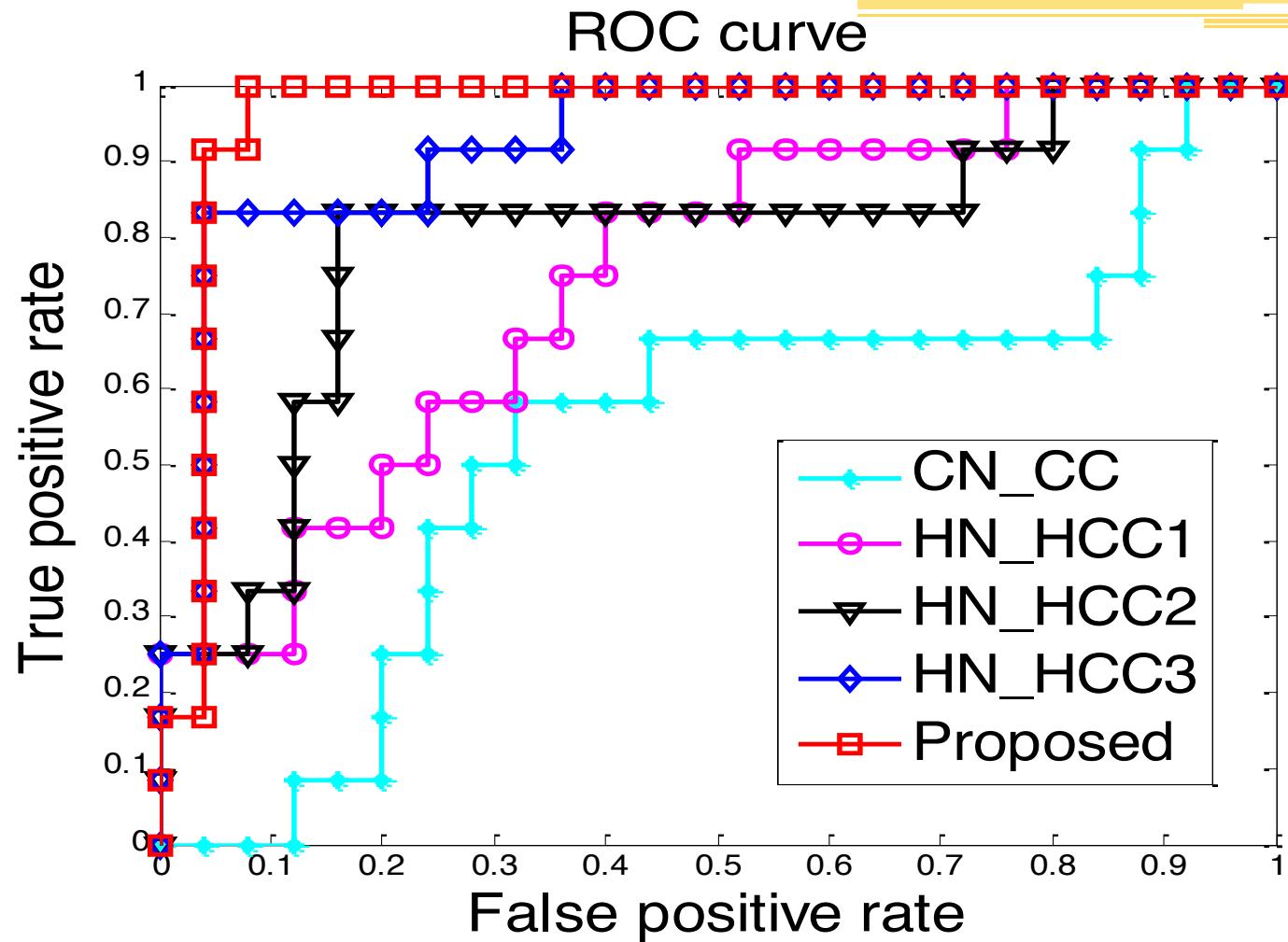
Experimental Results



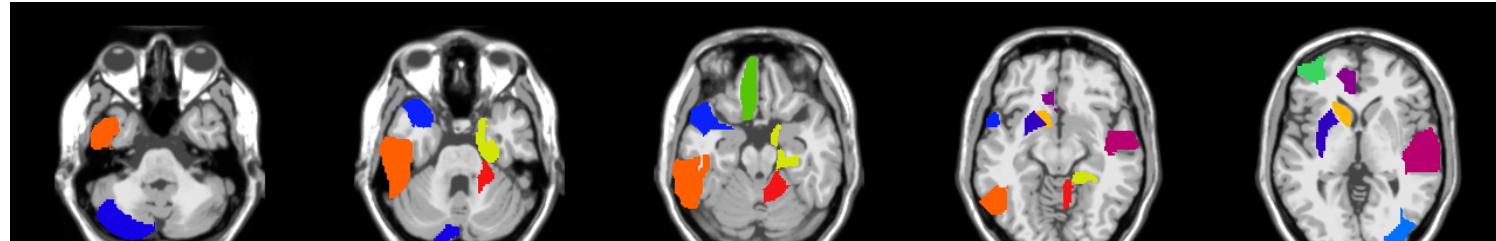
- Classification performances of different methods

Method	Accuracy	Sensitivity	Specificity	AUC
CN_CC	62.2	41.7	72.0	0.54
HN_HCC ¹	75.7	41.7	92.0	0.75
HN_HCC ²	81.1	75.0	84.0	0.80
HN_HCC ³	89.2	83.3	92.0	0.93
Proposed	94.6	91.7	96.0	0.96

Results

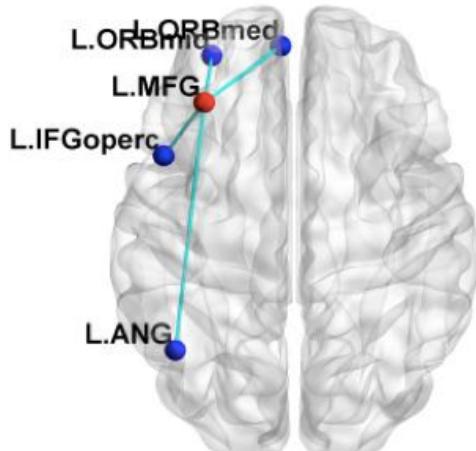


Brain Regions

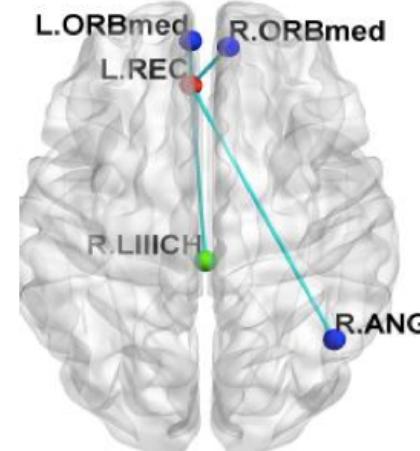
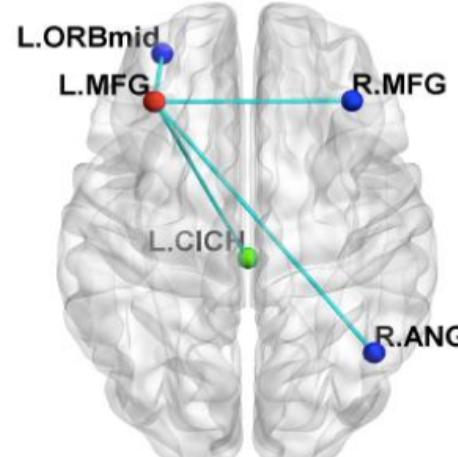


ID	ROIs	p-value (HCC ¹)	p-value (HCC ²)	p-value (HCC ³)	p-value (CN)
7	L. middle frontal gyrus	0.057	0.045	0.026	0.582
27	L. rectus gyrus	0.642	0.010	0.056	0.038
31	L. anterior cingulate gyrus	0.076	0.467	0.005	0.225
33	L. middle cingulate gyrus	0.869	0.521	0.044	0.283
40	R. paraHippocampal gyrus	0.018	0.078	0.014	0.380
52	R. middle occipital gyrus	0.025	0.001	0.010	0.454
71	L. caudate	0.002	0.002	0.001	0.925
73	L. putamen	0.128	0.055	0.017	0.409
82	R. superior temporal gyrus	0.056	0.037	0.439	0.645
83	R. temporal pole (superior)	0.022	0.011	0.002	0.909
89	L. inferior temporal left	0.036	0.009	0.062	0.943
93	L. crus II of cerebellar hemisphere	0.295	0.033	0.071	0.911
98	R. lobule IV, V of cerebellar hemisphere	0.048	0.154	0.877	0.703

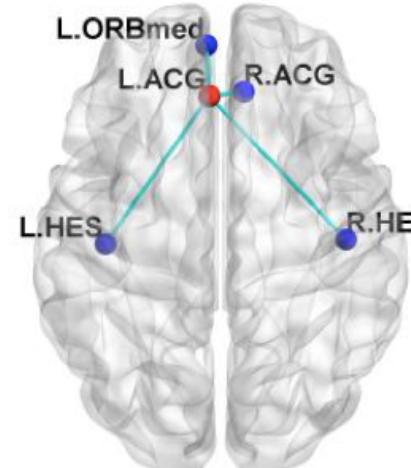
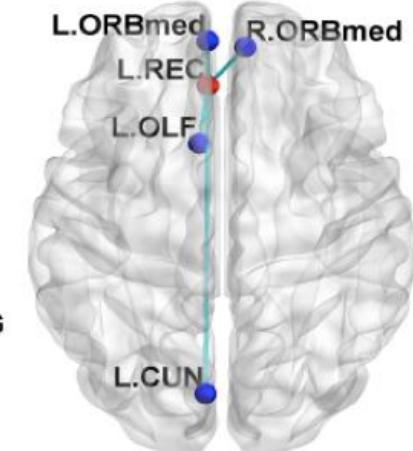
Hyperedges



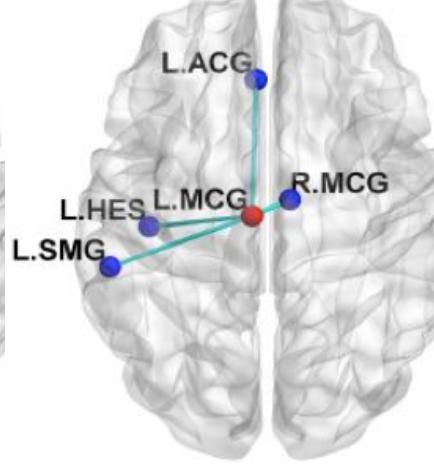
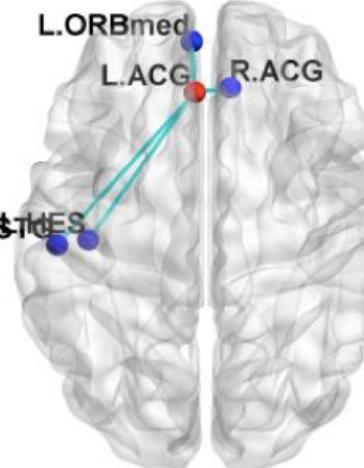
(a) L. middle frontal gyrus (L.MFG)



(b) L. rectus gyrus (L.REC)



(c) L. anterior cingulate gyrus (L.ACG)



(d) L. middle cingulate gyrus (L.MCG)

NC

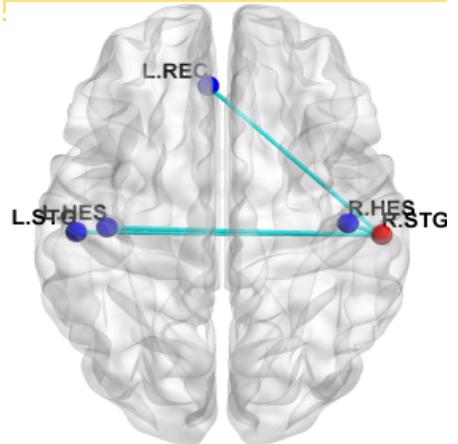
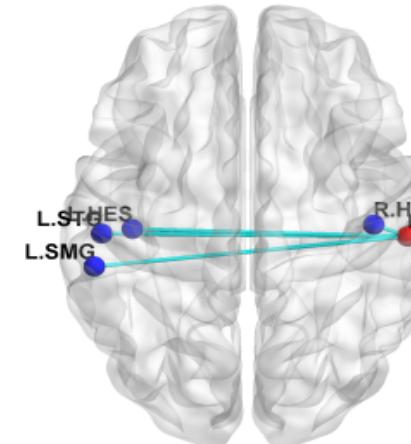
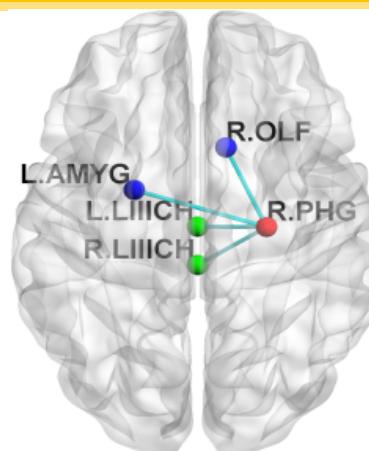
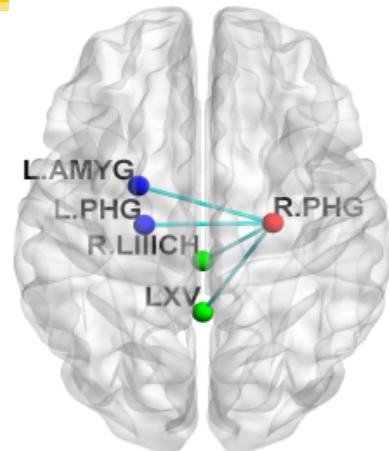
MCI

NC

MCI

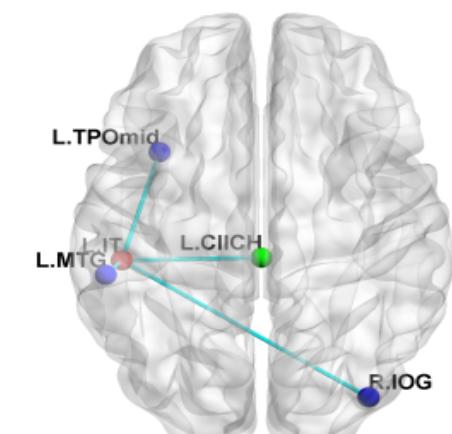
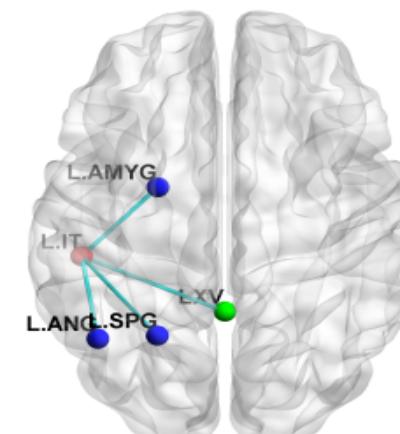
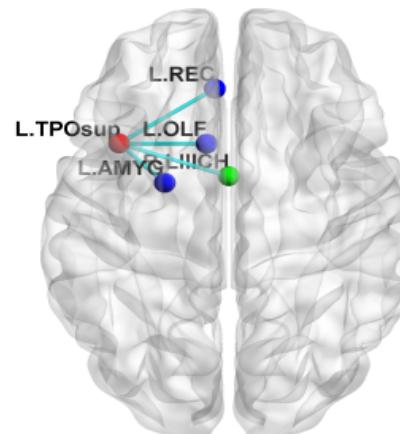
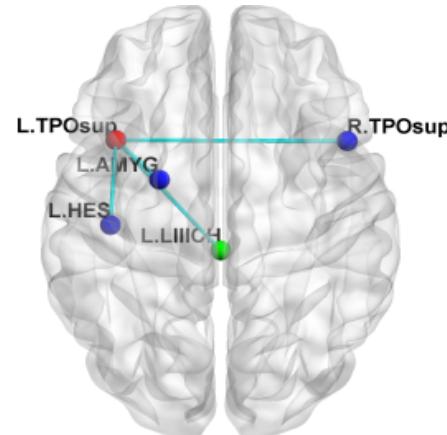


Hyperedges (cont'd)



(e) R. paraHippocampal gyrus (R.PHG)

(f) R. superior temporal gyrus (R.STG)



(g) R. temporal pole (superior) (R.TPosup)

(h) L. inferior temporal (L.IT)

NC

MCI

NC

MCI

Sub-network Mining based Method

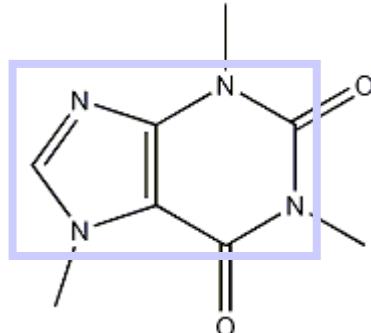


- Motivation
 - Our hypothesis is that there exist different frequent and discriminative subnetwork patterns between MCI group and NC group
 - **Main idea:** directly mine the discriminative subnetworks from connectivity networks and then use them for subsequent classification between MCI patients and normal controls

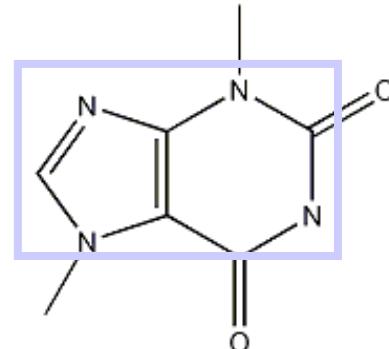
Frequent Subgraph Mining



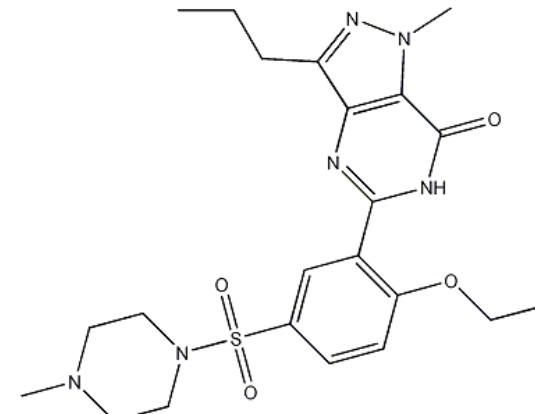
CHEMICAL COMPOUNDS



(a) caffeine



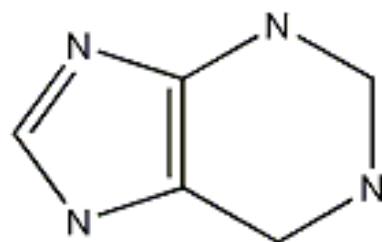
(b) diurobromine



(c) viagra

...

FREQUENT SUBGRAPH





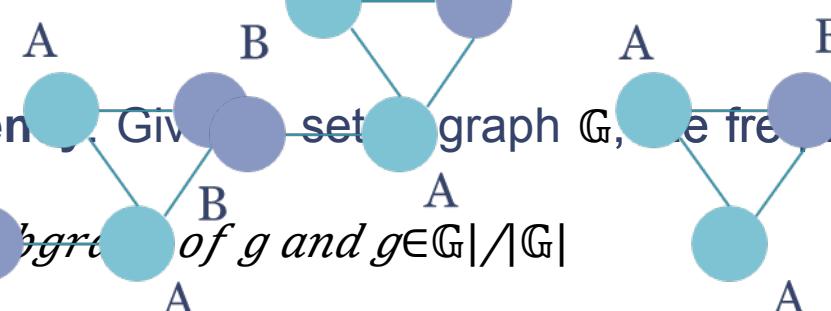
Frequent subgraph mining

➤ **Labeled Undirected Graph.** Let $G=(V,E,L,l)$ be a labeled undirected graph, where V is a set of nodes and $E \subseteq V \times V$ is a set of edges. $e=\{u,v\}$ indicates an edge between the nodes u and v . L is a set of labels, and l is a mapping function that assigns labels to vertices V and edges E .

➤ **Subgraph.** For two labeled undirected graphs $G_{ls}=(V_{ls},E_{ls},L_{ls},l_{ls})$ and $G=(V,E,L,l)$, we say G_{ls} is a subgraph of G if $V_{ls} \subseteq V$, $E_{ls} \subseteq E$, $L_{ls} \subseteq L$ and $l_{ls} = l$.

➤ **Subgraph Frequency.** Given a set of graph \mathbb{G} , the frequency of a subgraph g_{ls} is defined as:

$$f_{qg_{ls}} \mathbb{G} = |g_{ls} \text{ is a subgraph of } g \text{ and } g \in \mathbb{G}| / |\mathbb{G}|$$

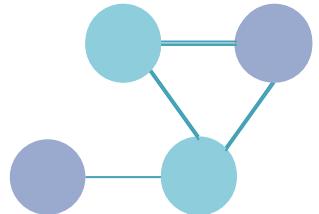
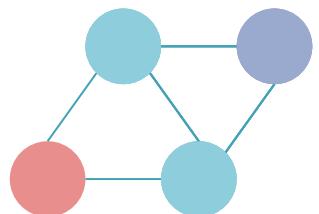


➤ **Frequent Subgraph Mining.** Given a set of labeled undirected graph \mathbb{G} and a support parameter s where $0 < s \leq 1$, find all undirected graphs that are subgraphs in at least $s|\mathbb{G}|$ of the input graphs.

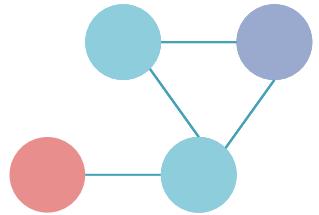
Frequent subgraph mining



➤ Example



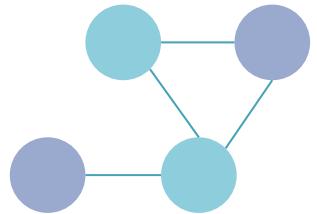
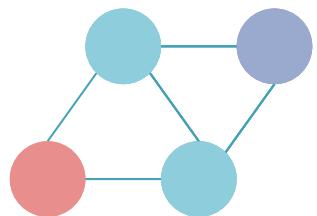
Support=3/3



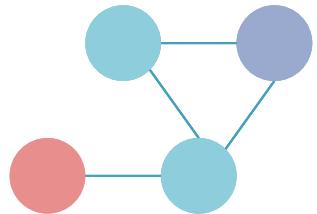
Frequent subgraph mining



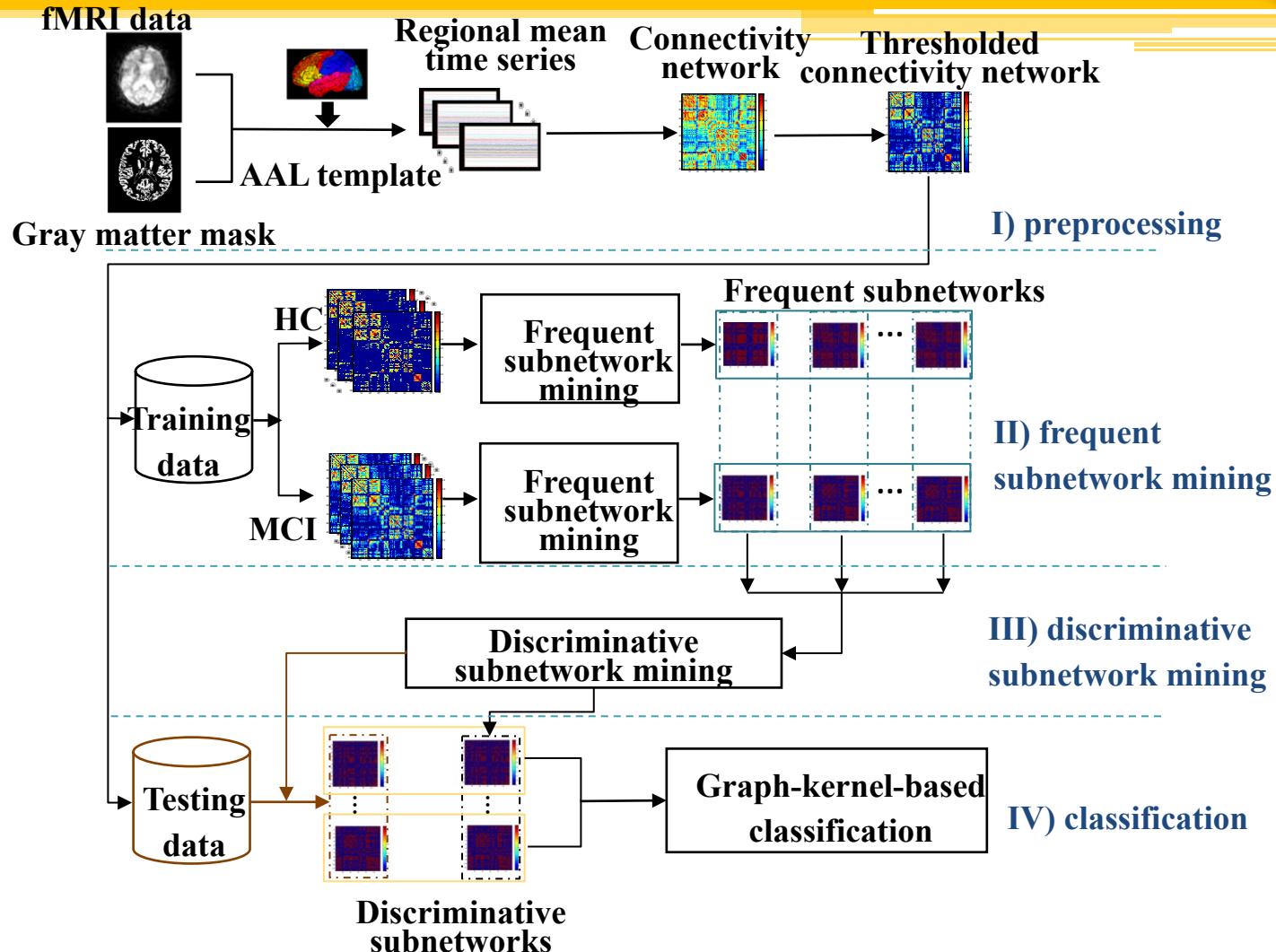
➤ Example



Support=2/3



Flowchart



(F. Fei, B. Jie, D., Zhang, Brain Connectivity, 2014)

Experimental Results



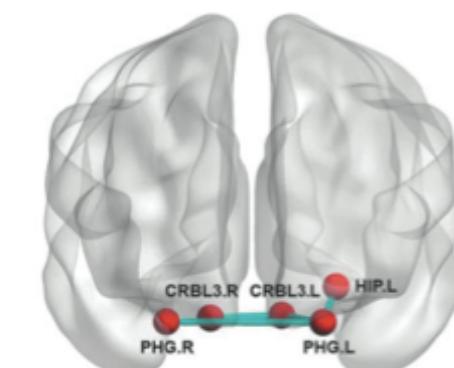
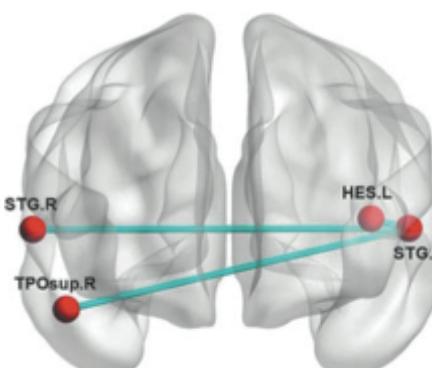
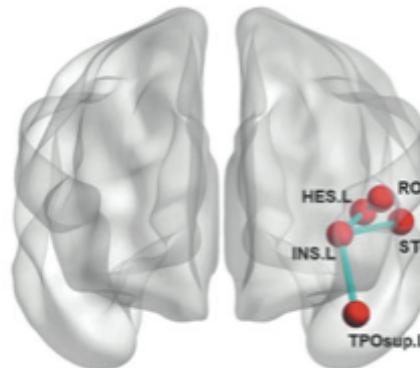
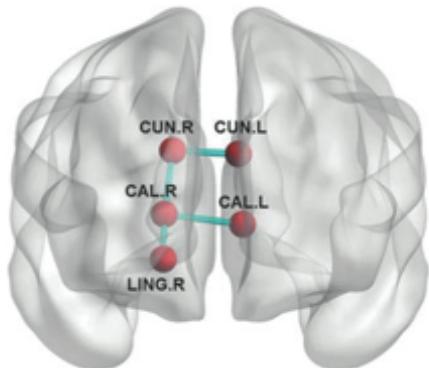
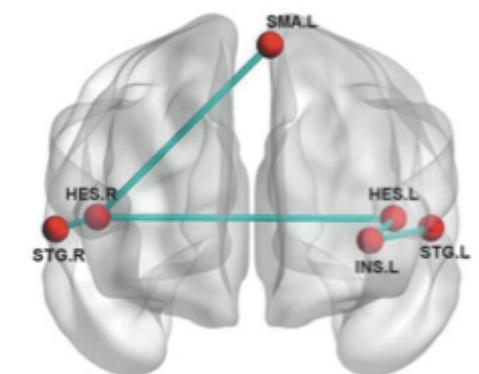
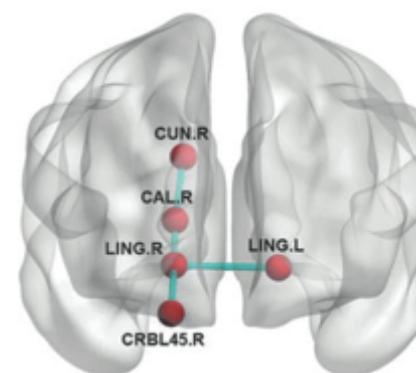
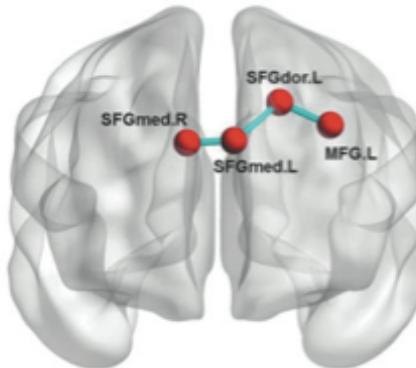
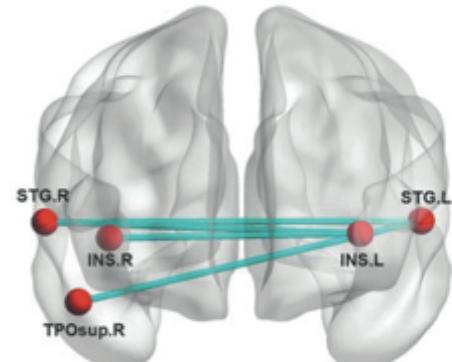
➤ Classification performance

Classification performance of different methods

	Methods	T1	T2	T3	T4	T5
Accuracy (%)	Baseline	73.0	73.0	70.3	73.0	75.7
	RFE-GK	86.5	83.8	75.7	75.7	64.9
	Proposed	89.2	97.3	97.3	91.9	91.9
AUC	Baseline	0.51	0.79	0.63	0.83	0.71
	RFE-GK	0.85	0.86	0.77	0.78	0.60
	Proposed	0.90	0.96	0.96	0.86	0.94



Mined Sub-networks



NC

MCI

NC

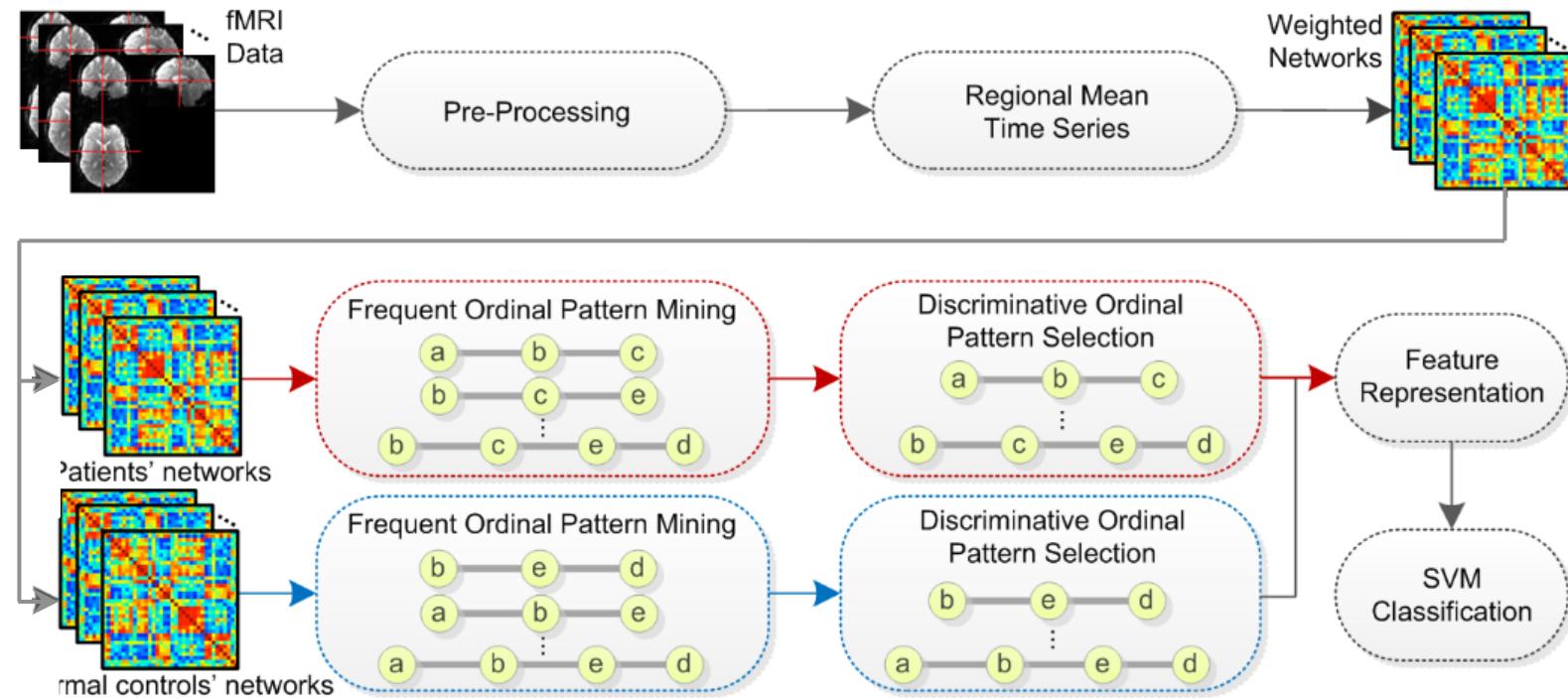
MCI

Ordinal Patterns Mining



- Existing network descriptors
 - Node degrees
 - Clustering coefficients
 - Sub-networks ...
- Limitations of previous work
 - Designed on un-weighted brain connectivity networks
 - Focus on individual brain regions other than local structures of brain networks

Flowchart



An overview of ordinal pattern based learning for brain disease diagnosis

(X. Liu, et al., MICCAI 2016; D. Zhang, et al., IEEE TMI 2018)

Illustration

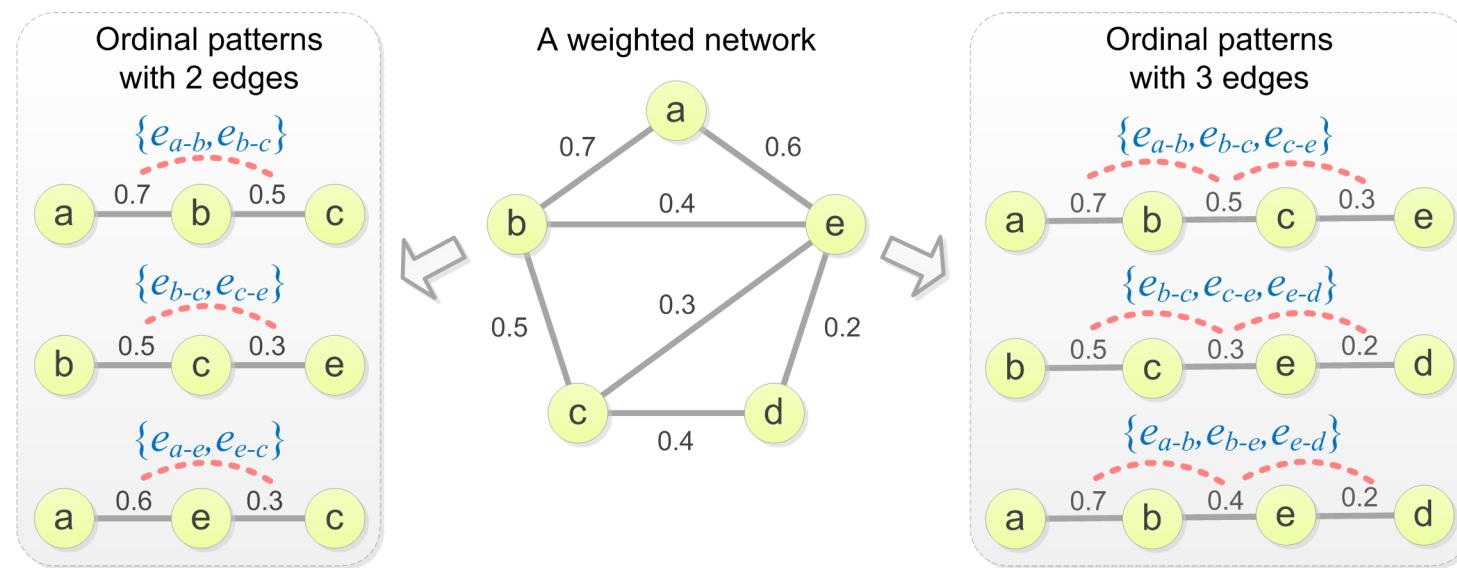


Illustration of the proposed ordinal patterns

Illustration

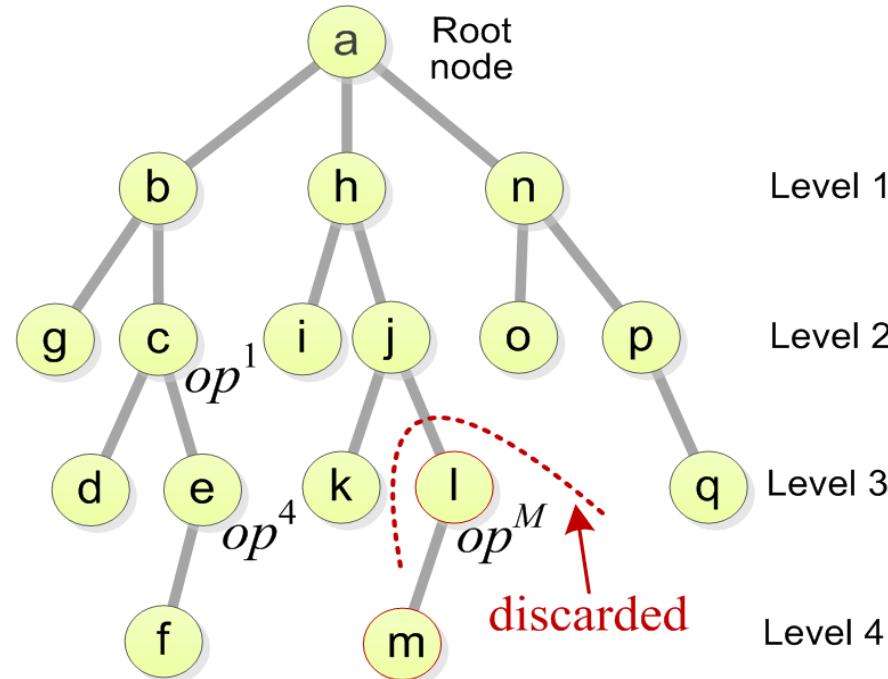


Illustration of the proposed frequent ordinal pattern mining algorithm

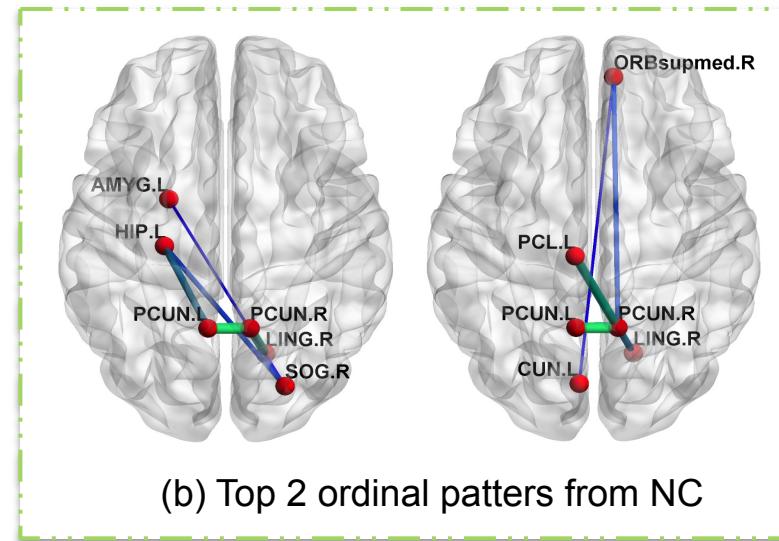
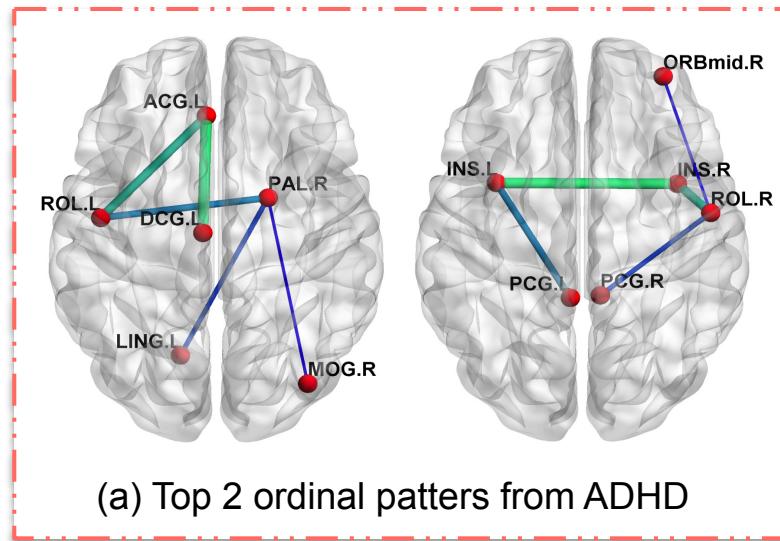
Experimental Results



Comparison of different methods in three classification tasks

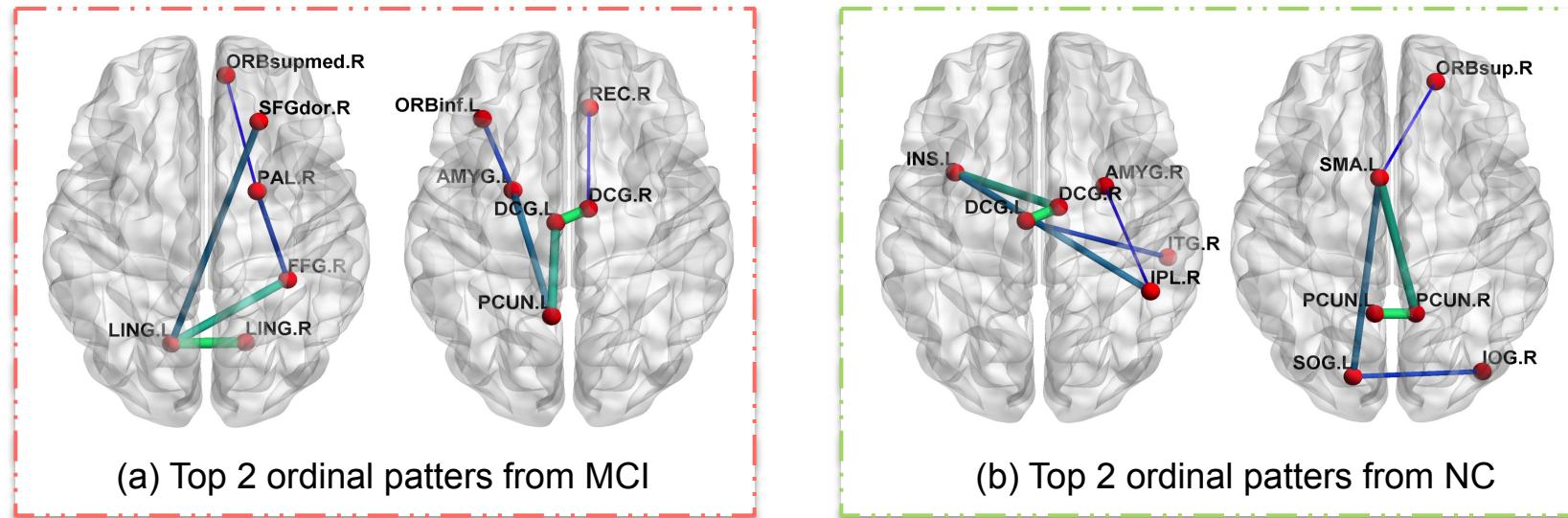
Method	AD vs. NC			MCI vs. NC			ADHD vs. NC		
	ACC	SEN	AUC	ACC	SEN	AUC	ACC	SEN	AUC
CC	72.62	73.53	70.94	71.14	72.73	68.69	71.29	72.03	70.51
CCMT	80.95	82.35	76.35	74.50	75.76	74.79	74.53	75.43	77.64
DS	76.19	76.47	75.59	77.18	78.79	74.89	81.01	81.36	80.82
DSMT	85.71	85.29	87.59	79.19	80.81	76.99	83.79	84.74	84.63
Proposed	94.05	96.77	96.35	88.59	87.27	84.57	87.50	88.89	87.37

Results (Cont'd)



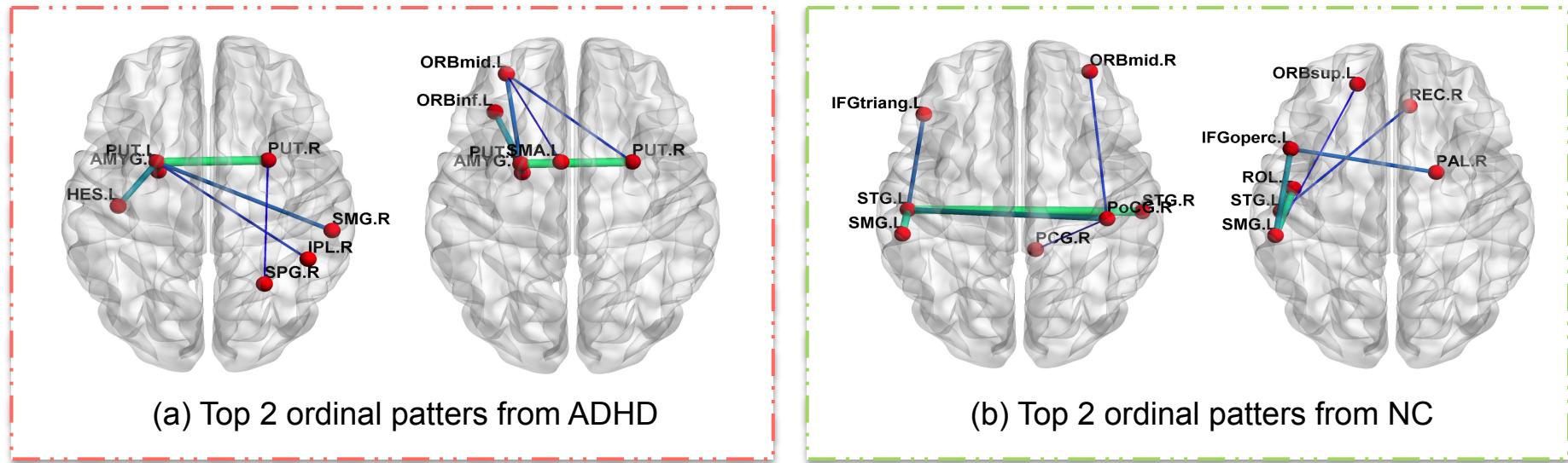
Ordinal patterns identified in **AD vs. NC** classification

Results (Cont'd)



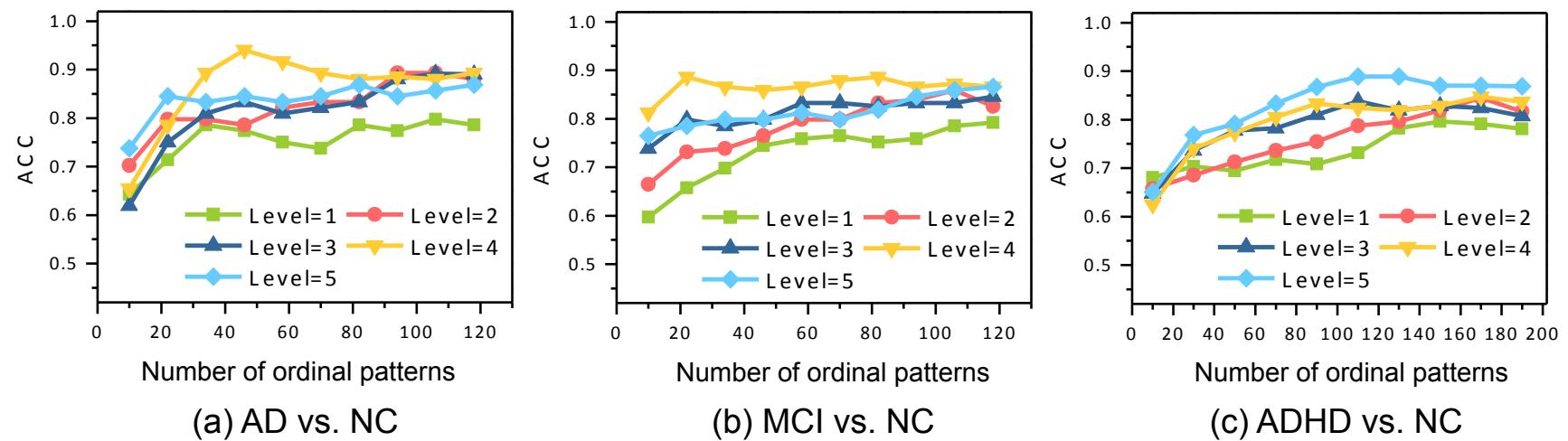
Ordinal patterns identified in **MCI vs. NC** classification

Results (Cont'd)



Ordinal patterns identified in **ADHD vs. NC** classification

Results (Cont'd)



Influence of the level number in frequent ordinal pattern mining method and the number of discriminative ordinal patterns in three classification tasks

Outline



1

Backgrounds on Alzheimer's Disease

2

Brain-imaging based Analysis

3

Brain-network based Analysis

4

Summary

Summary



- Brain imaging and brain network are two basic tools for analyzing brain
- Intelligent machine learning methods play important roles in neuroimaging-based analysis and brain disease diagnosis
- More opportunities for applying machine learning techniques in neuroimaging

Acknowledgement



- Collaborators
 - Prof. Songcan Chen (NUAA)
 - Prof. Zhi-Hua Zhou (NJU)
 - Prof. Dinggang Shen (UNC)
 - Dr. Guorong Wu (UNC)
 - Dr. Chong-Yaw Wee (UNC)
 - Prof. Manhua Liu (SJTU)
 - Dr. Yinghuan Shi (NJU)
 - Dr. Jun Liu (SAS)
 - Dr. Shu Liao (Siemens)
- Students
 - Biao Jie
 - Mingxia Liu
 - Xiaoke Hao
 - Bo Cheng
 - Tony
 - Fei Fei
 - Junqiang Du
 - ...

Thanks for your attention!



For more details, contact:

dqzhang@nuaa.edu.cn

<http://ibrain.nuaa.edu.cn>