机器学习在脑影像分析及疾病诊断中的应用

张道强

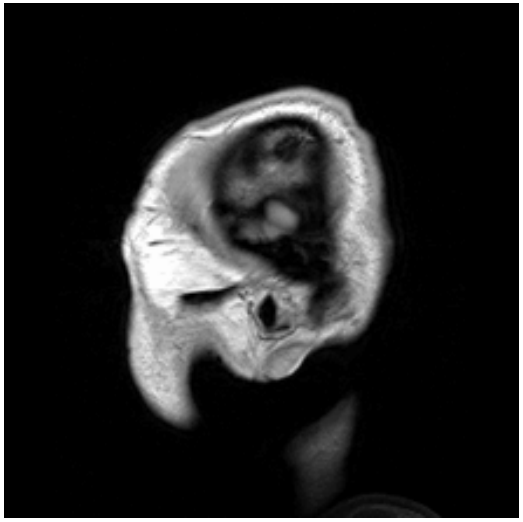
南京航空航天大学 计算机科学与技术学院

2013-11-02

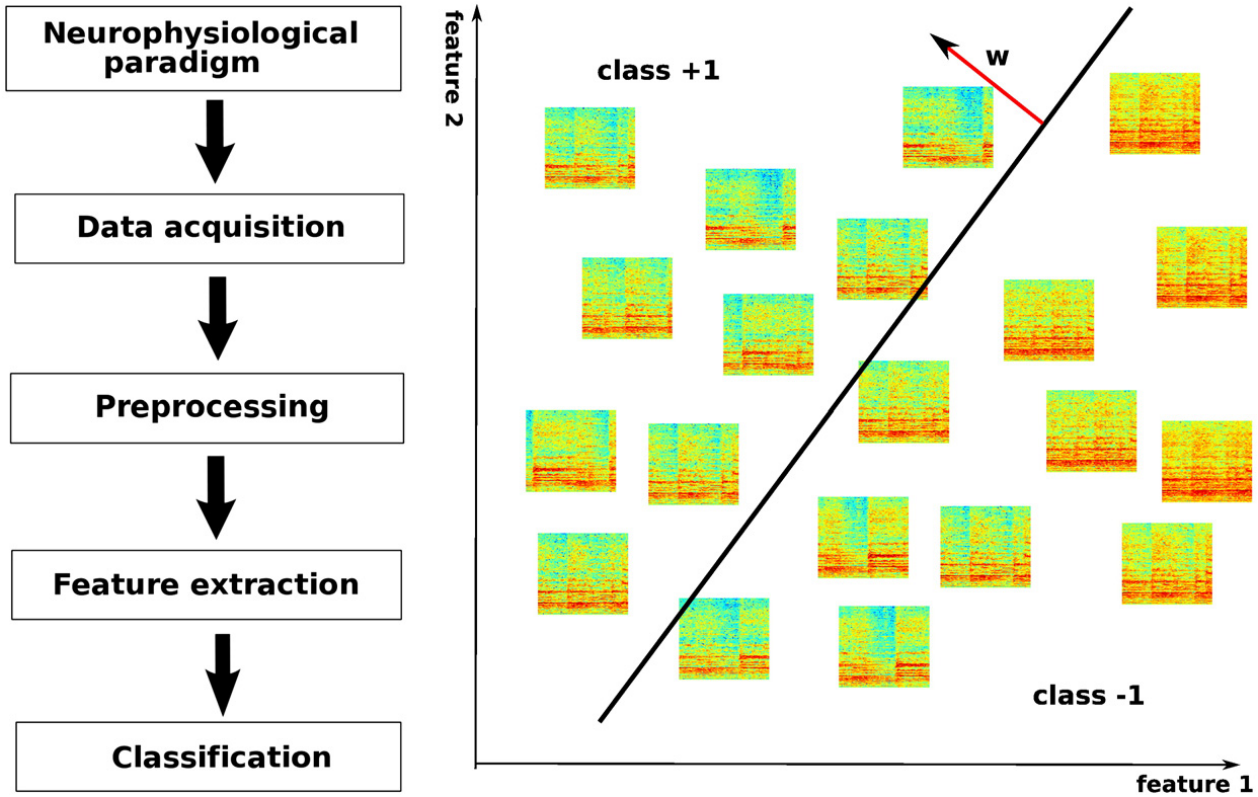
Brain Imaging (Neuroimaging)

- 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

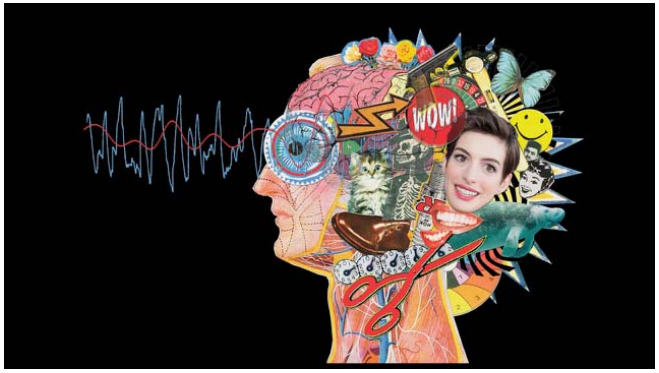


Neuroimaging-based Classification

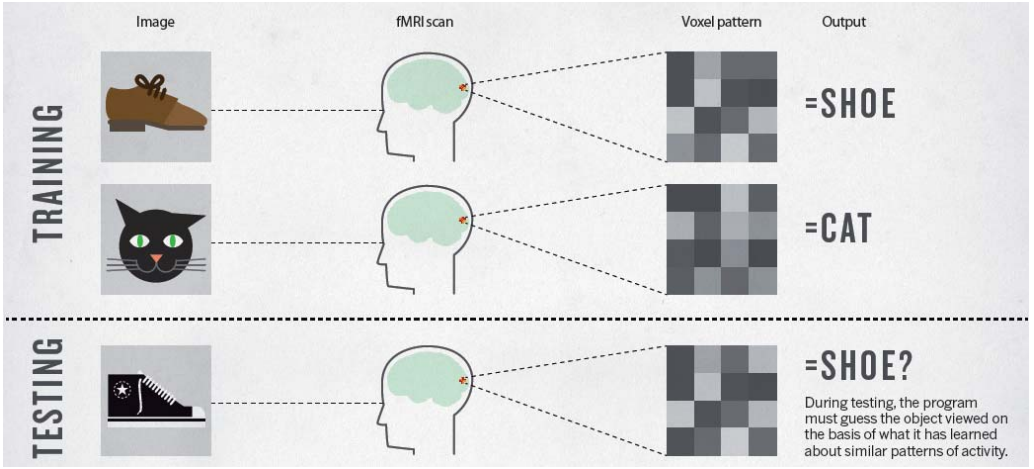


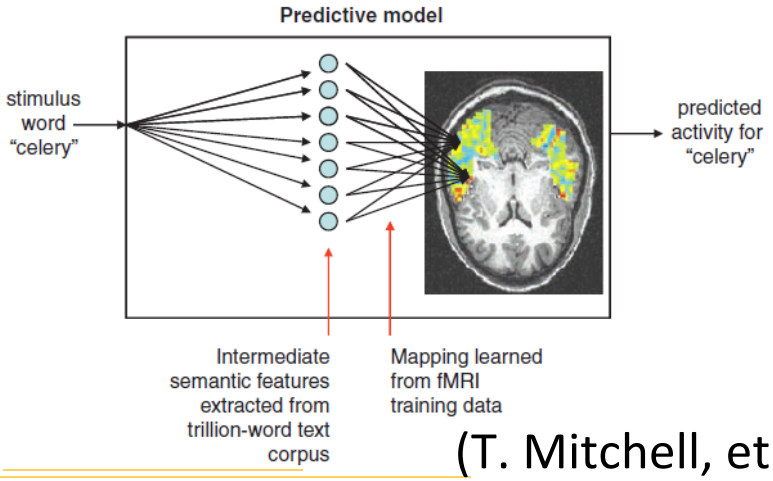
(S. Lemm, et al., Neuroimage, 2011)

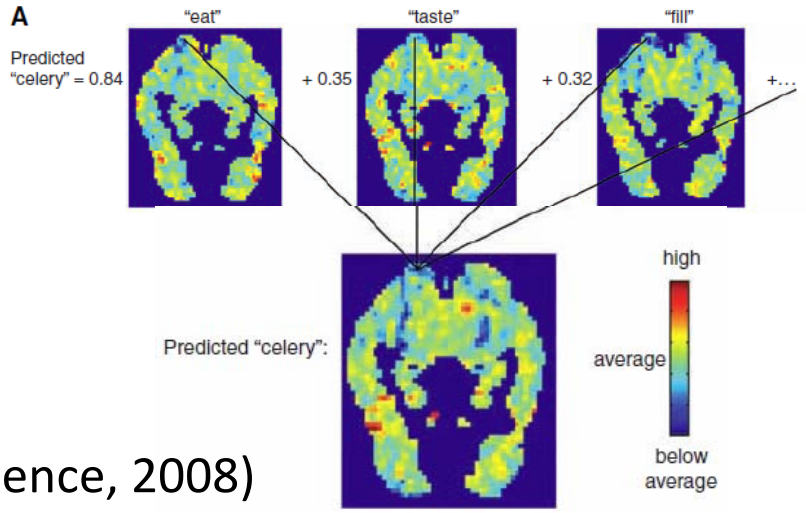
Example: Brain Decoding



(Nature Feature News, 2013)







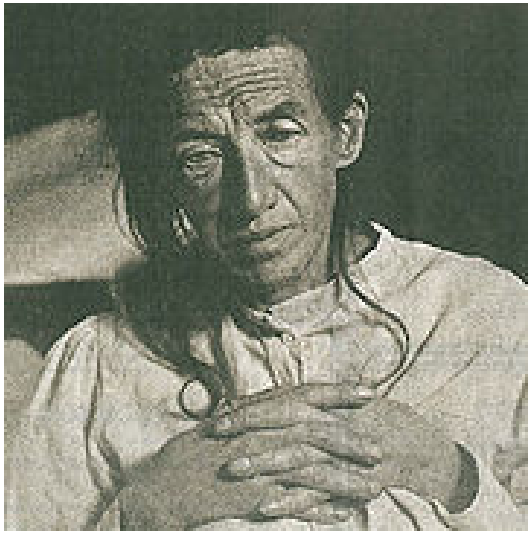
(T. Mitchell, et al., Science, 2008)

Outline

History of AD

- AD was first described by German psychiatrist and neuropathologist Alois Alzheimer in 1906 and was named after him

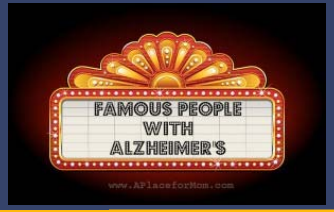
- The 51 y.o. woman (Auguste Deter) cared by Dr. Alzheimer until her death in 1906. He did an autopsy, examined her brain & described the typical abnormalities of what would be called later Alzheimer's Disease

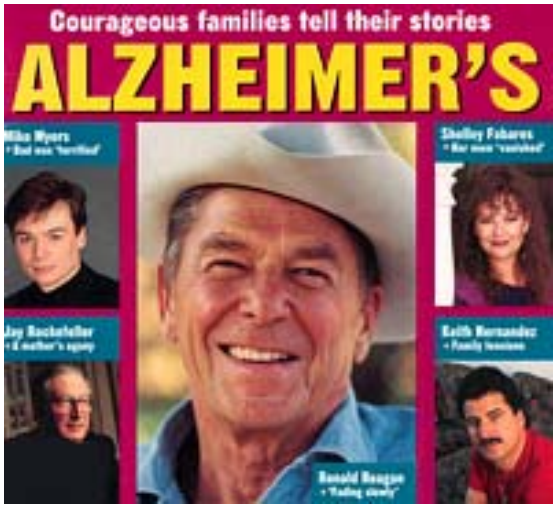


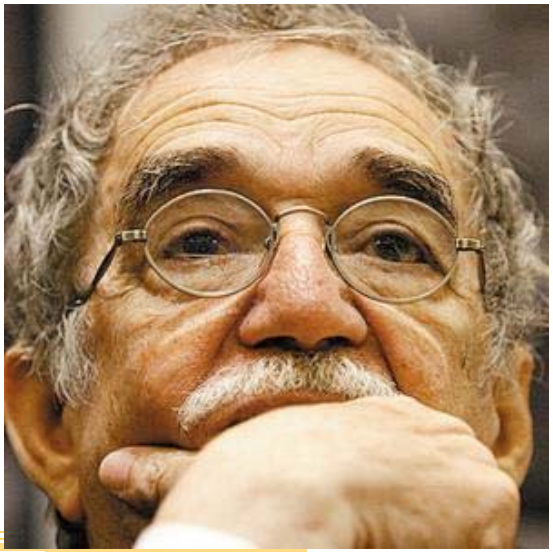
What Is AD?

- 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

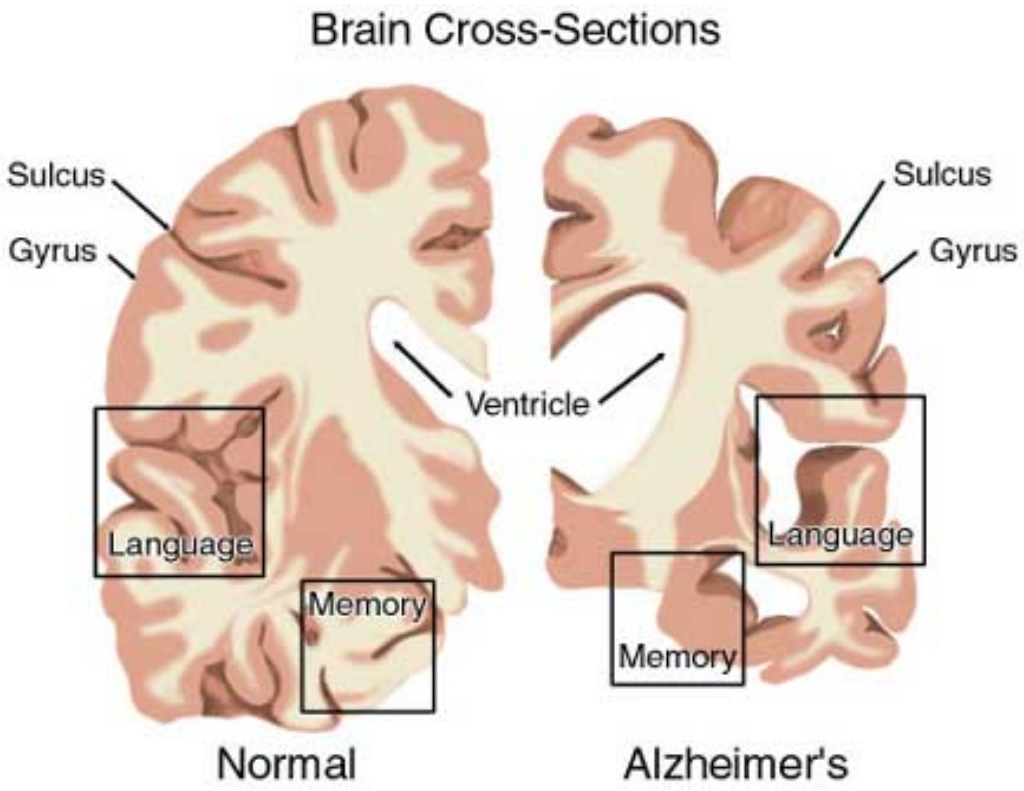
Celebrities with AD





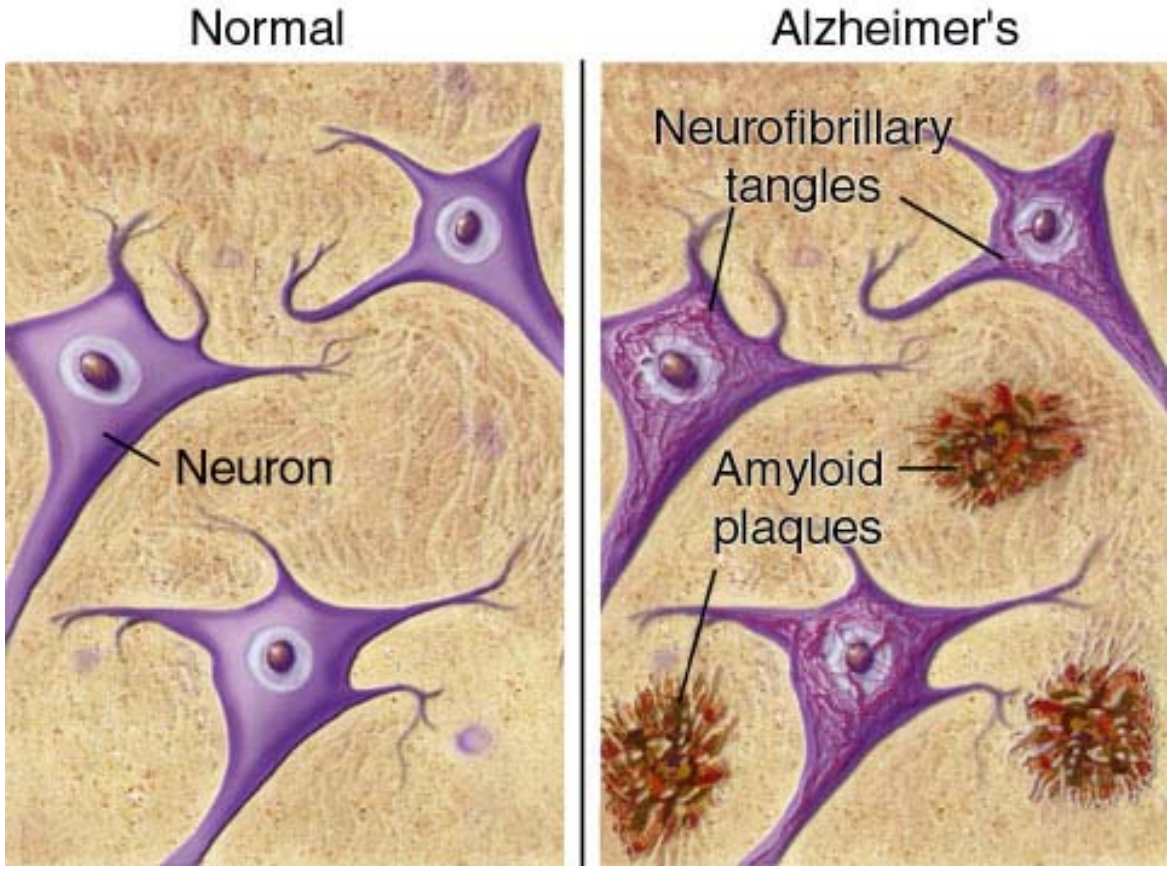


Normal vs. AD Brain



- 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

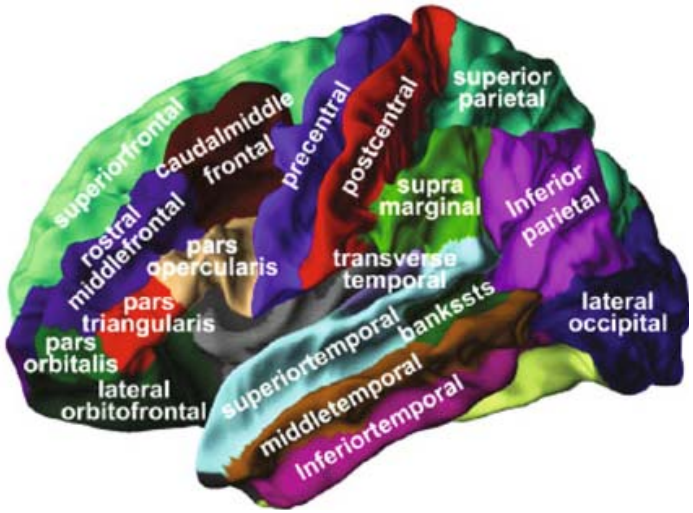
Normal vs. AD Brain

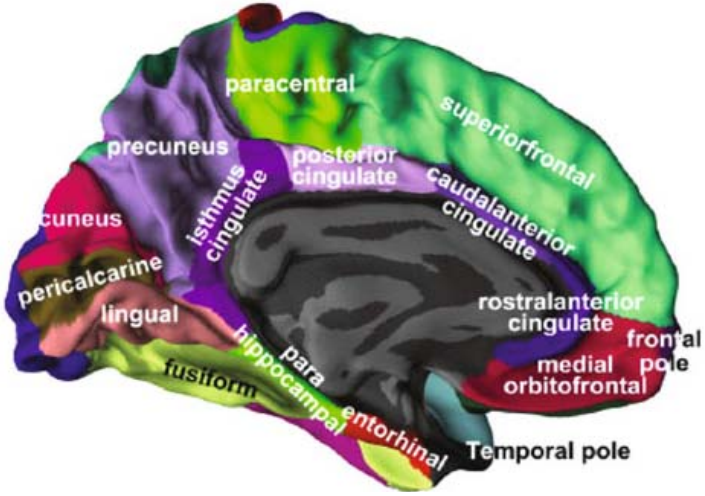


- Forms abnormal clumps called amyloid plaques and tangled bundles of fibers called neurofibrillary tangles in the brain

AD 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

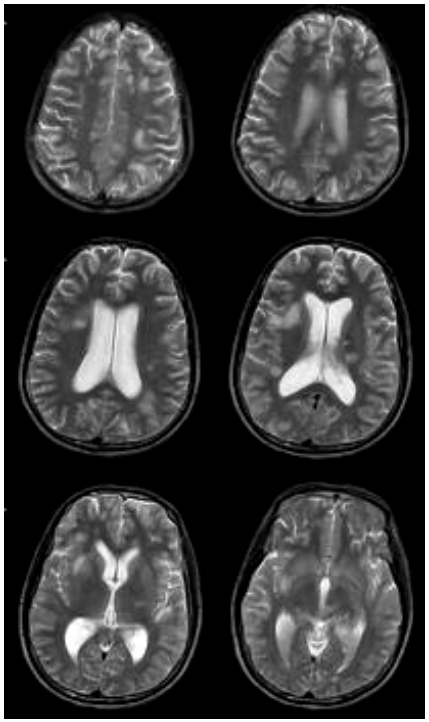




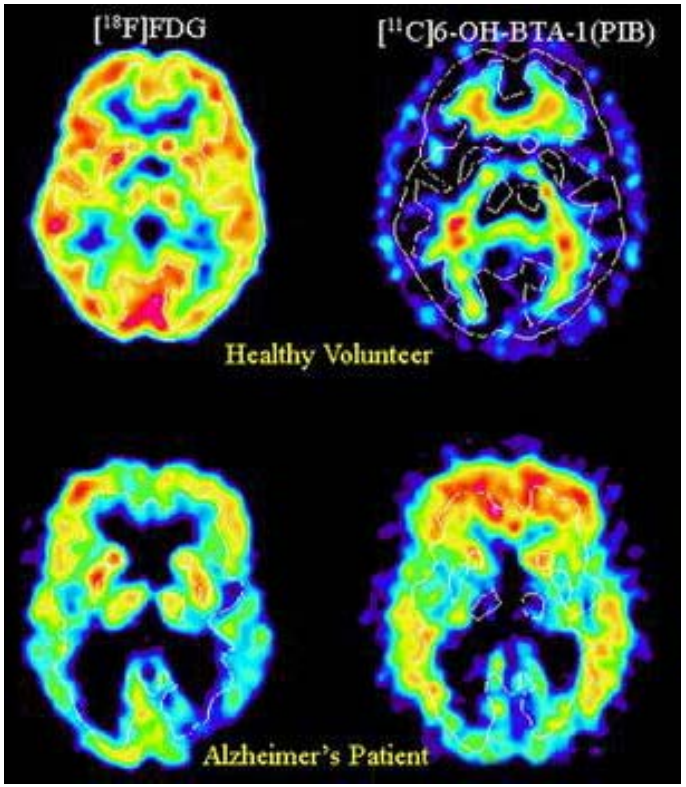
AD Biomarkers

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

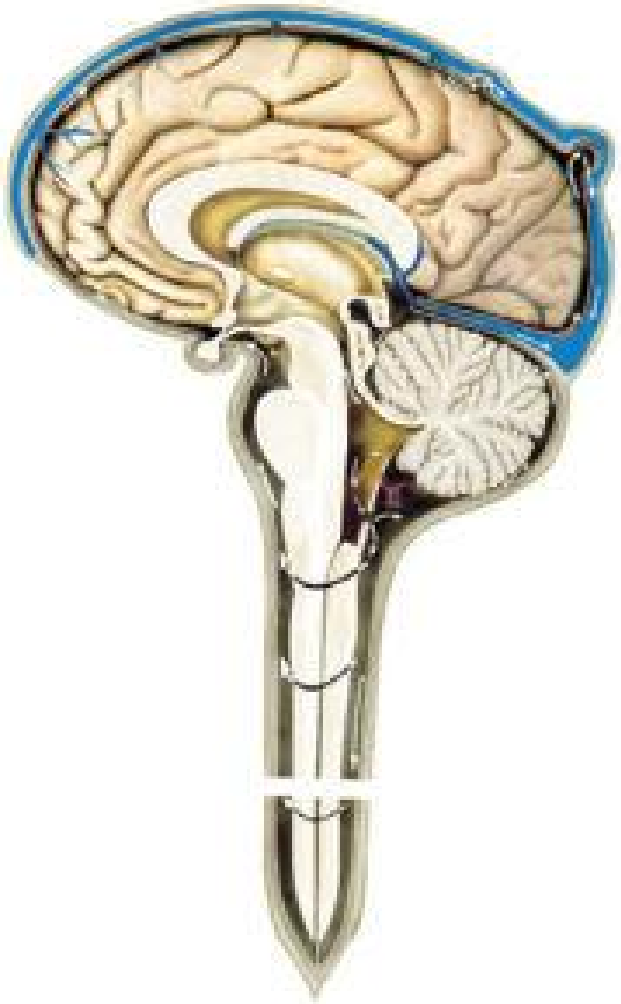
MRI



PET


[18F]FDG
[11C]6-OH-BTA-1(PIB)
Healthy Volunteer
Alzheimer's Patient

CSF

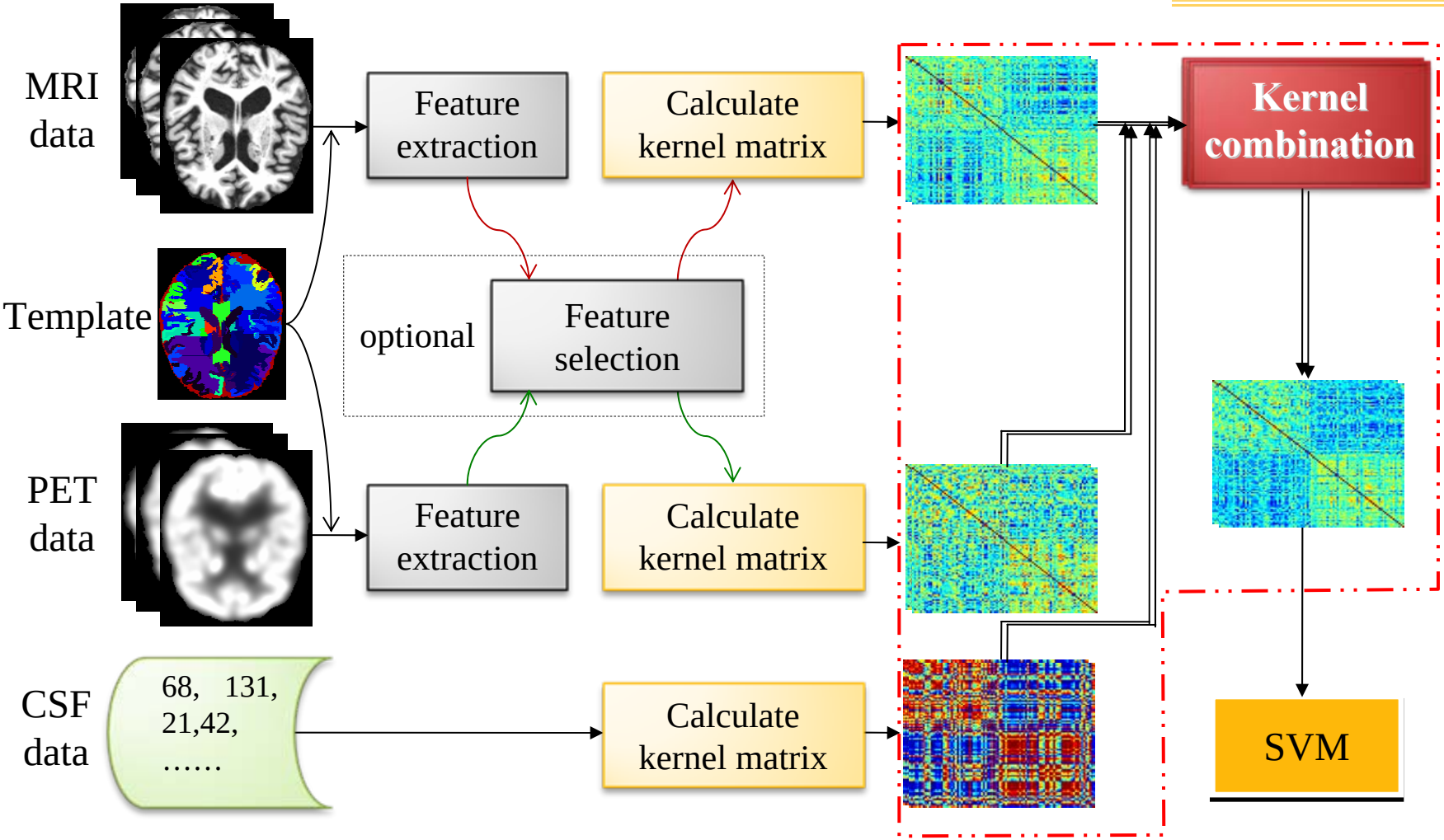


Outline

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)

Multi-kernel SVM

- Objective function

$$\min_{\mathbf{w}^{(m)},b,\xi} \quad \frac{1}{2}\sum_{m=1}^{M}\beta_m \left\|\mathbf{w}^{(m)}\right\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$s.t. \quad y_i\left(\sum_{m=1}^{M}\beta_m\left(\left(\mathbf{w}^{(m)}\right)^T\phi^{(m)}(\mathbf{x}_i^{(m)})+b\right)\right)\geq 1-\xi_i$$

$$\xi_i \geq 0, i=1,\ldots,n.$$

$\mathbf{x}_i^{(m)}$ data in the *m*-th modality

$\mathbf{w}^{(m)}$ normal vector of hyperplane of *m*-th modality

$\phi^{(m)}$ kernel-induced mapping function of *m*-th modality

β_m combining weight on the *m*-th modality

Materials

- 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

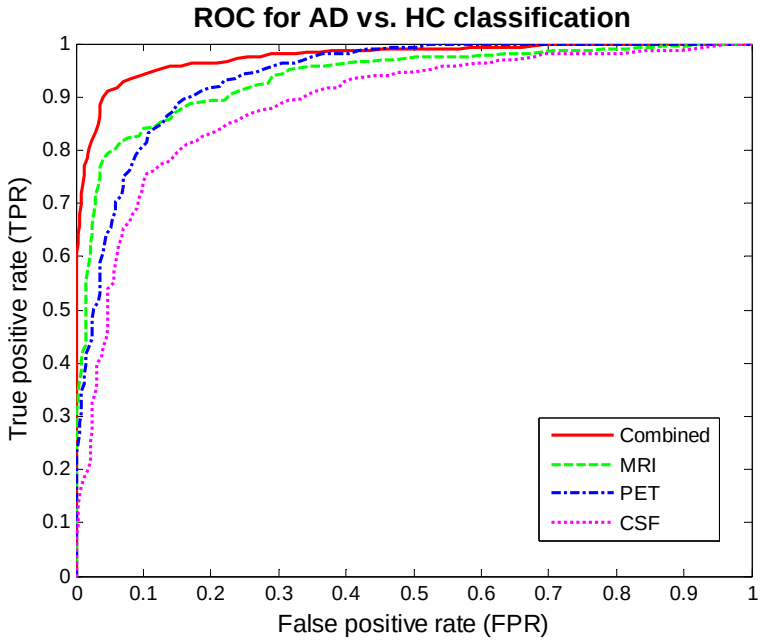
Results

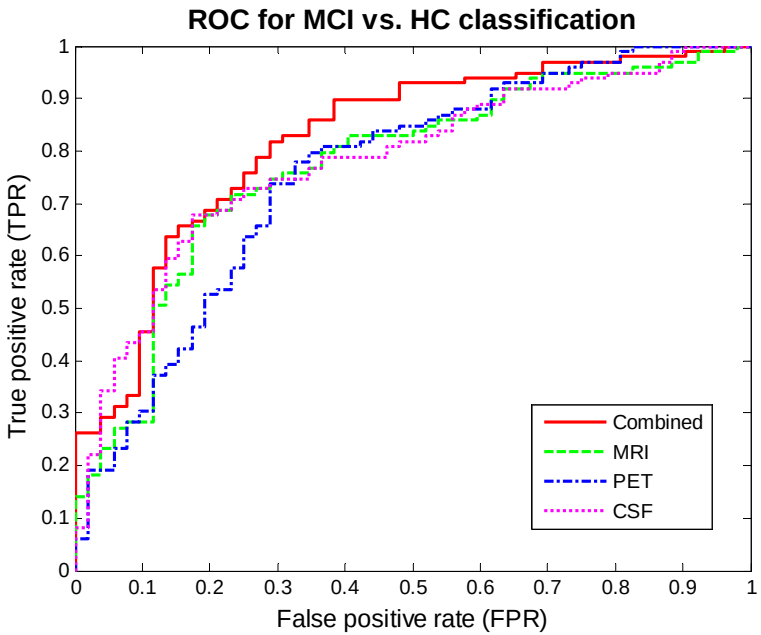
Comparison of performance of single-modal and multimodal classification methods

	AD vs. HC			MCI vs. HC		
Methods	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)
CSF	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)
PET	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)

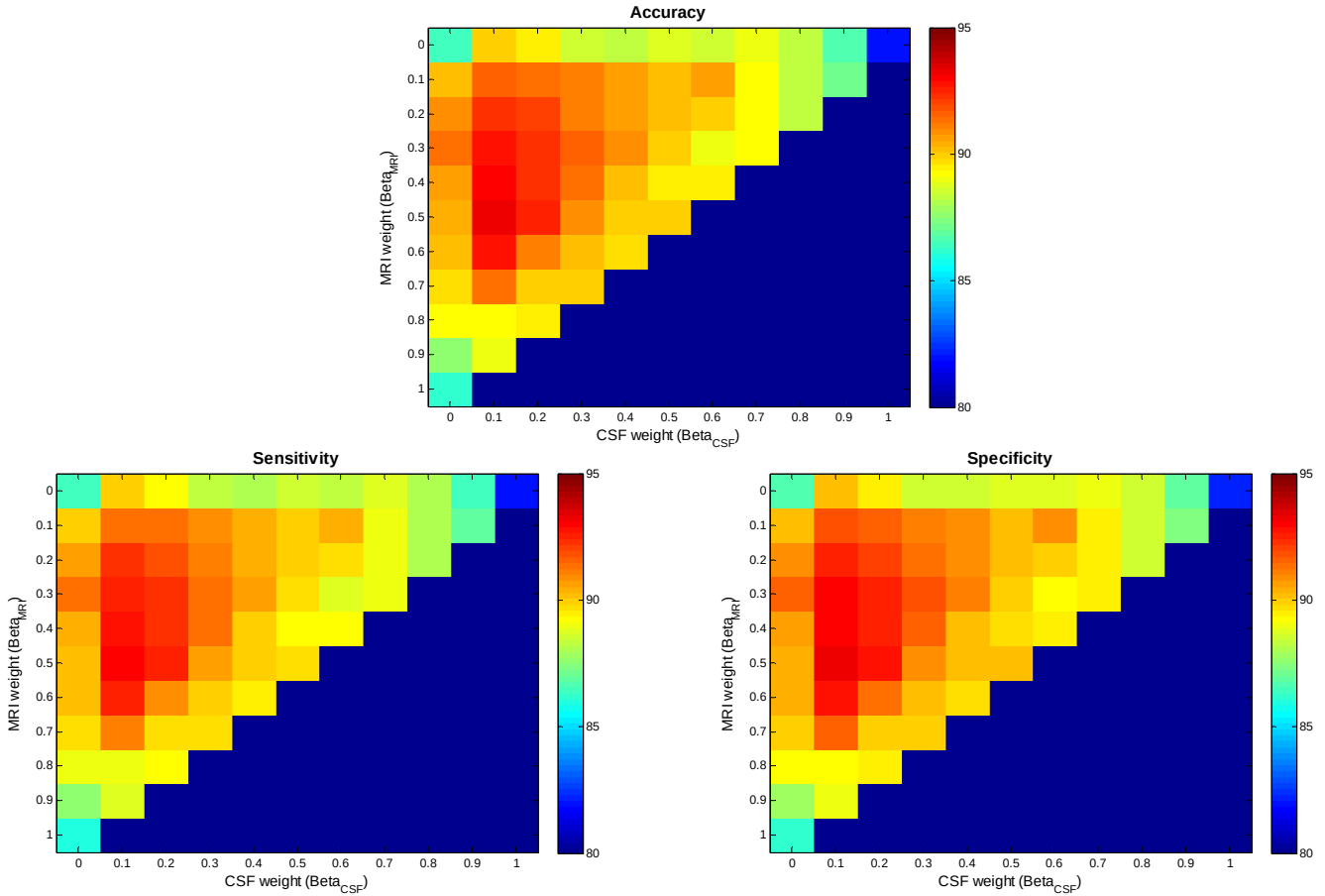




ROC curves of different methods for AD and MCI classification

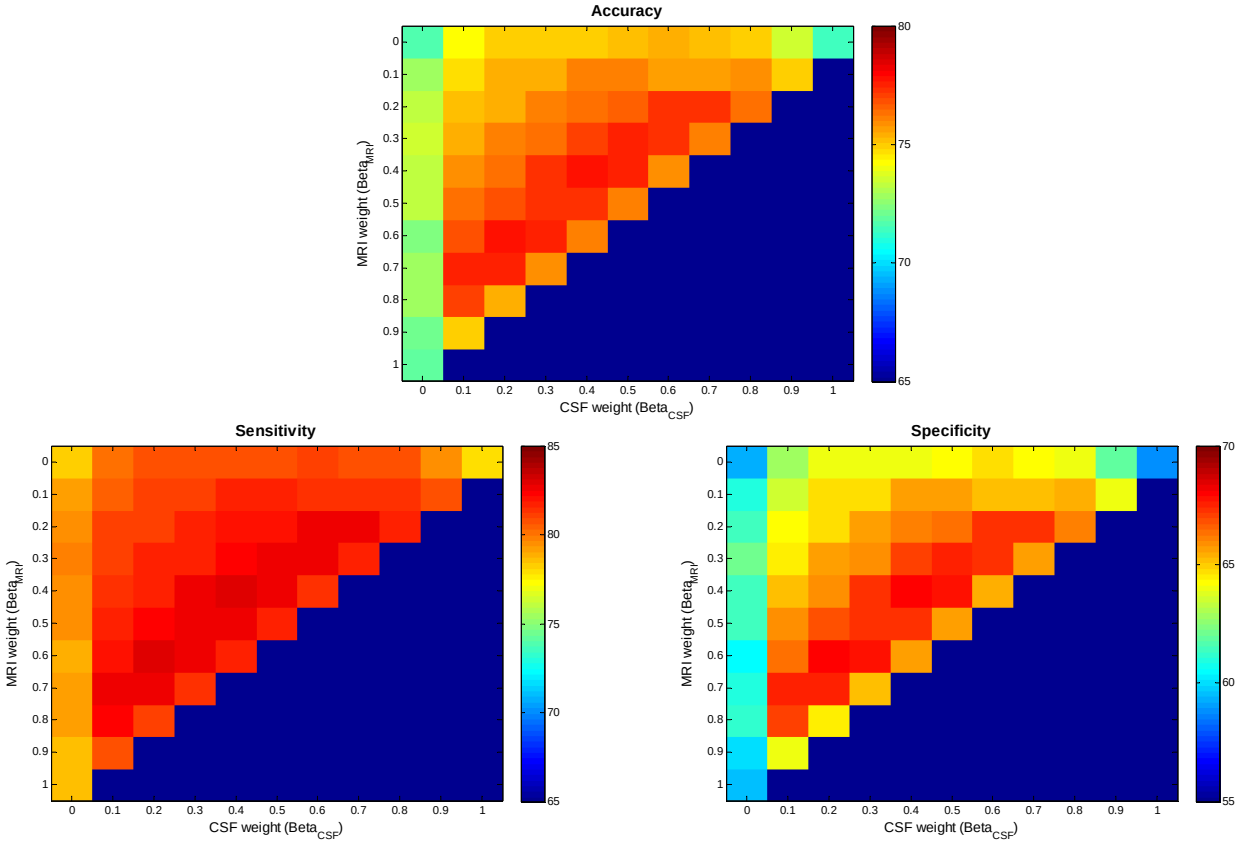
(D. Zhang, et al. Neuroimage, 2011)

Results (cont'd)



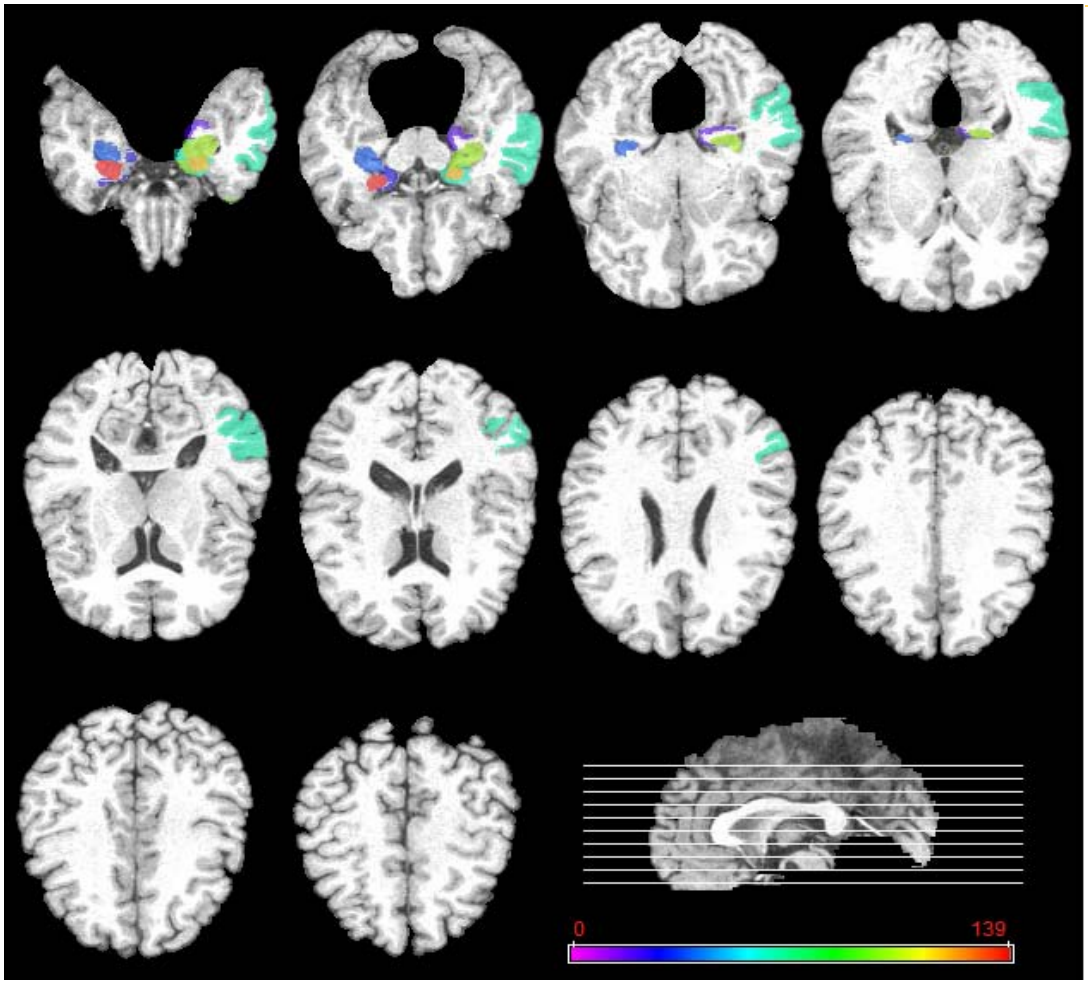
AD Classification results with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint: $\beta_{PET} + \beta_{CSF} + \beta_{MRI} = 1$

Results (cont'd)



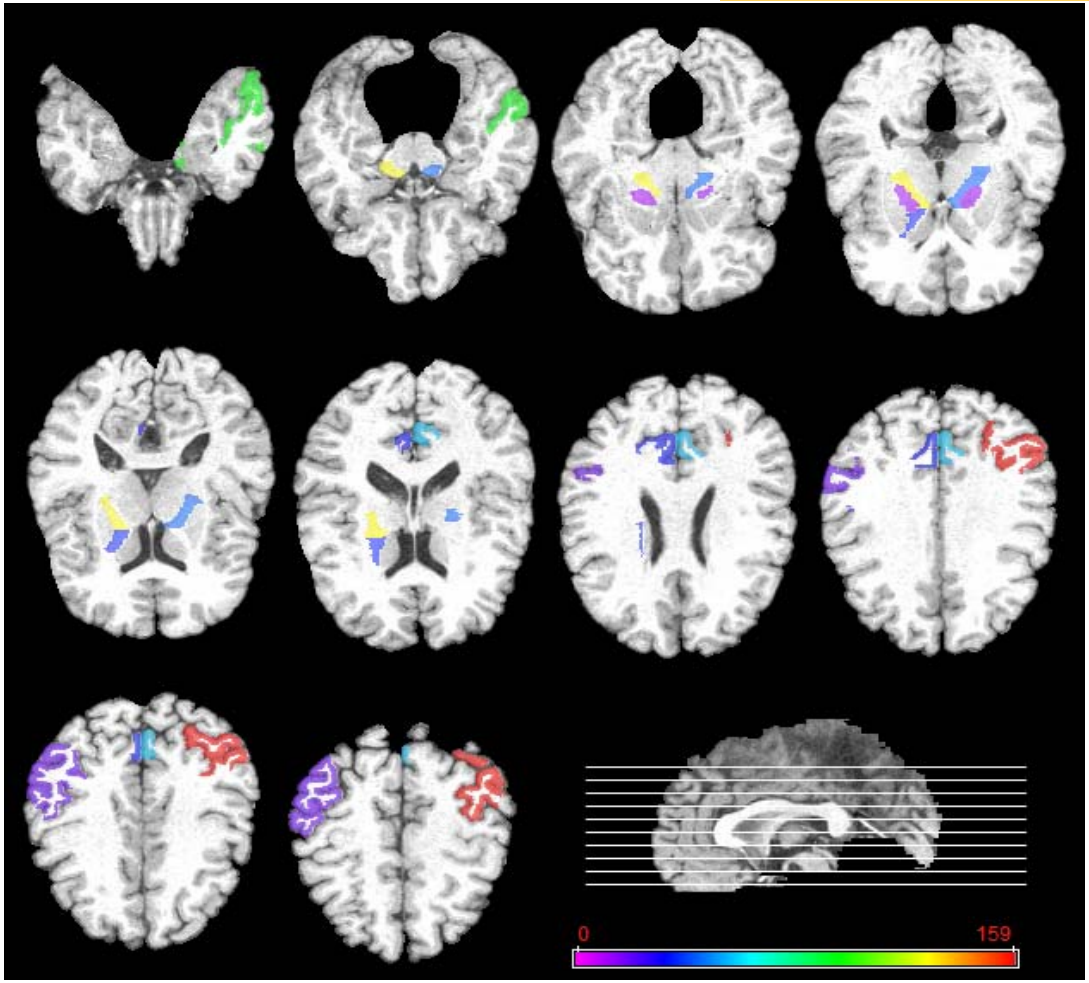
MCI Classification with respect to different combining weights of MRI, PET and CSF. Only the squares in the upper triangular part have valid values, due to the constraint: $\beta_{PET} + \beta_{CSF} + \beta_{MRI} = 1$

Results (cont'd)



Top 11 brain regions selected for MCI classification detected from MRI

Results (cont'd)



Top 11 brain regions selected for MCI classification detected from PET

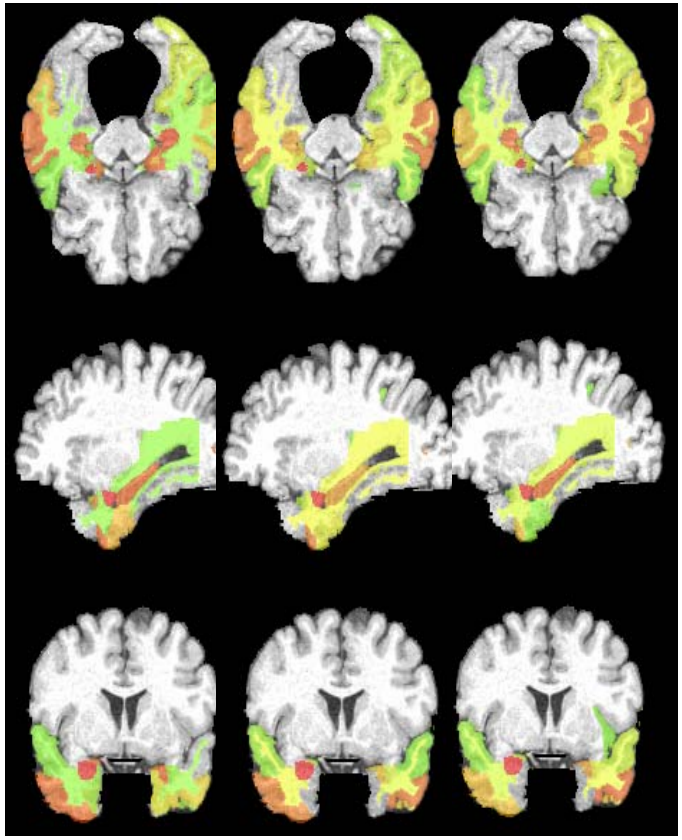
Results (cont'd)

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

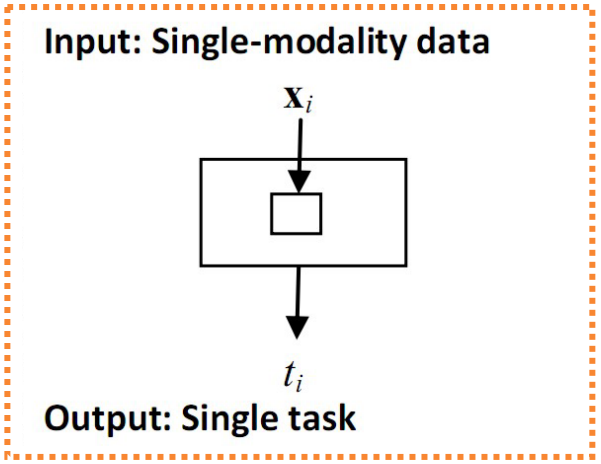
- Motivation
  - Besides classification, there also exist regression tasks which estimate continuous clinical scores to evaluate the stage of AD pathology and predict future progression
  - Both regression and classification tasks are essentially related due to the same underlying pathology
- **Question: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



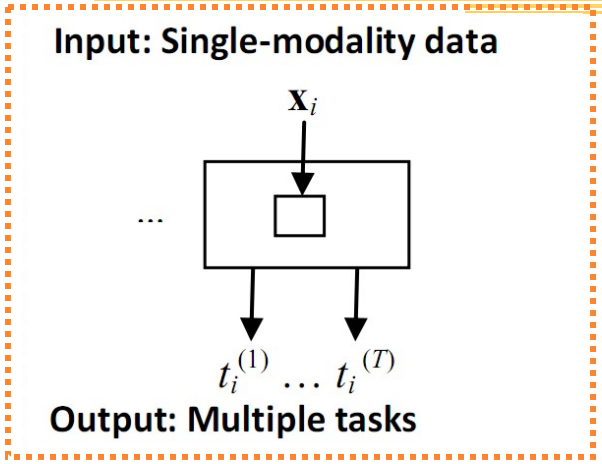
AD/MCI/HC MMSE ADAS-Cog

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

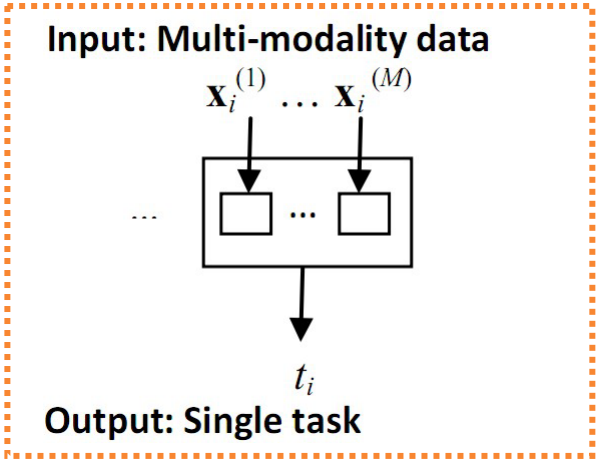
Four Learning Problems



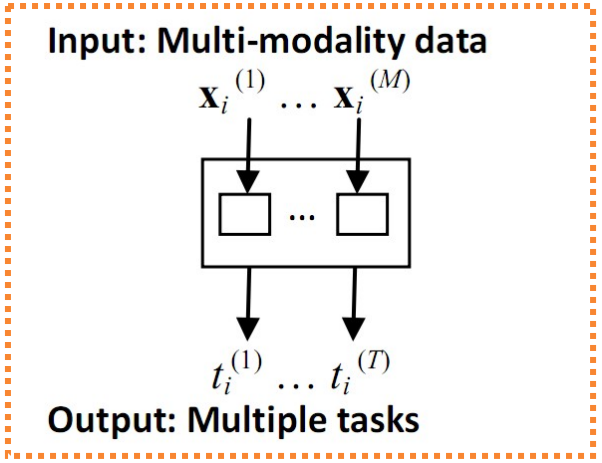
(a) Single-model single task (SMST)



(b) Multi-task learning



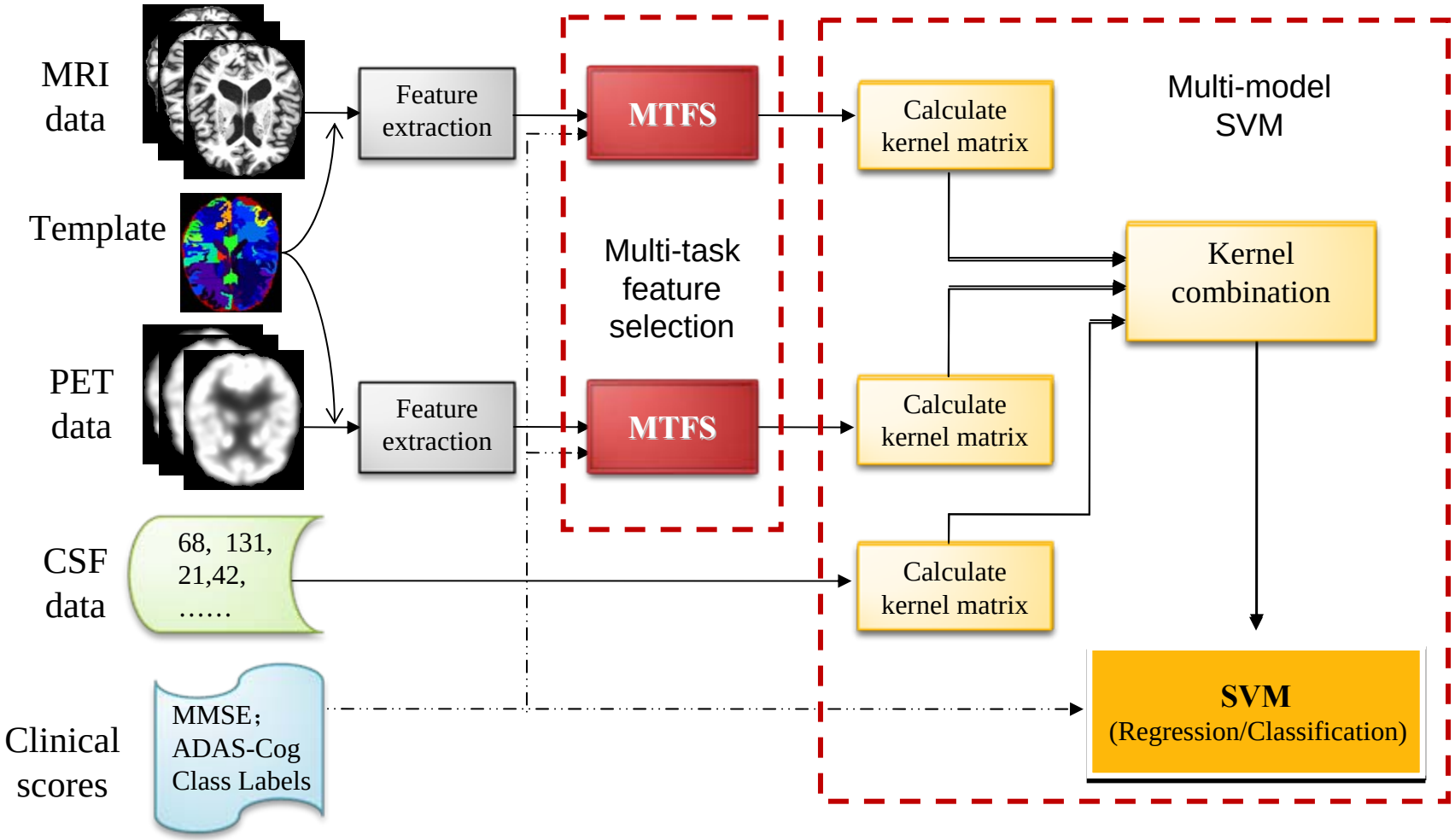
(c) Multi-model learning



(d) Multi-model Multi-task (M3T)

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

Flowchart



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

Multi-Task Feature Selection

- Objective function

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

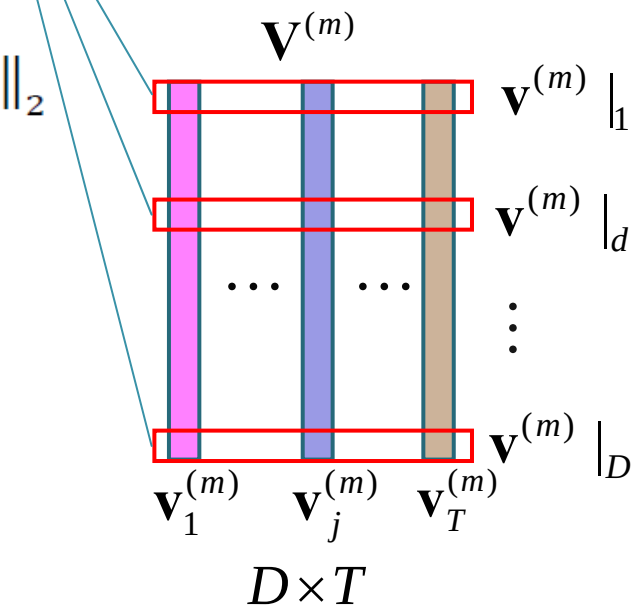
$$= \frac{1}{2}\sum_{j=1}^{T}\left\|\mathbf{y}^{(j)} - \mathbf{X}^{(m)}\mathbf{v}_j^{(m)}\right\|_2^2 + \lambda \sum_{d=1}^{D^{(m)}} \left\|\mathbf{V}^{(m)}|_d\right\|_2$$

$$\mathbf{X}^{(m)} = [\mathbf{x}_1^{(m)}, \ldots, \mathbf{x}_i^{(m)}, \ldots, \mathbf{x}_N^{(m)}]^T$$

$$\mathbf{y}^{(j)} = [t_1^{(j)}, \ldots, t_i^{(j)}, \ldots, t_N^{(j)}]^T$$

$$\mathbf{V}^{(m)} = [\mathbf{v}_1^{(m)}, \ldots, \mathbf{v}_j^{(m)}, \ldots, \mathbf{v}_T^{(m)}]$$

Group sparsity



Weight matrix

Materials

- ADNI Subjects
  - 186subjects (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

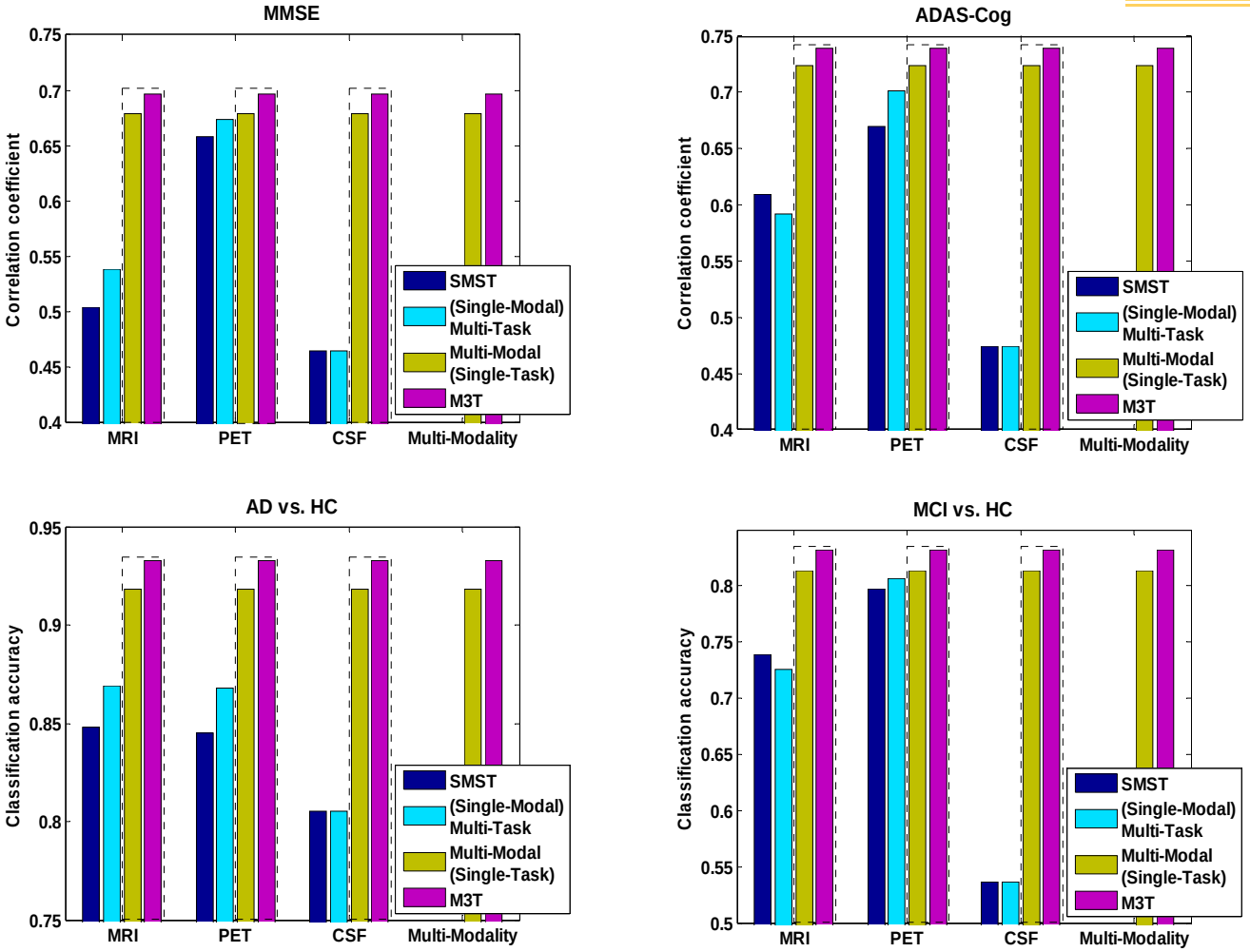
- Experiment 1
  - Estimating clinical stages
  - MMSE, ADAS-Cog, and class label (AD/MCI/HC)
- Experiment 2
  - Predicting 2-year MMSE and ADAS-Cog changes and MCI conversion

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)



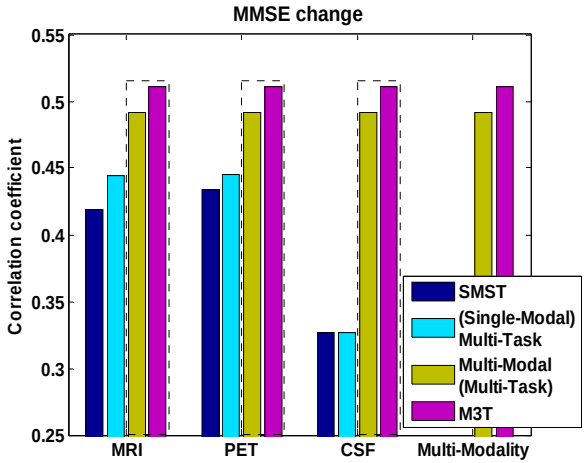
Comparison of performances of four different methods on Experiment 1

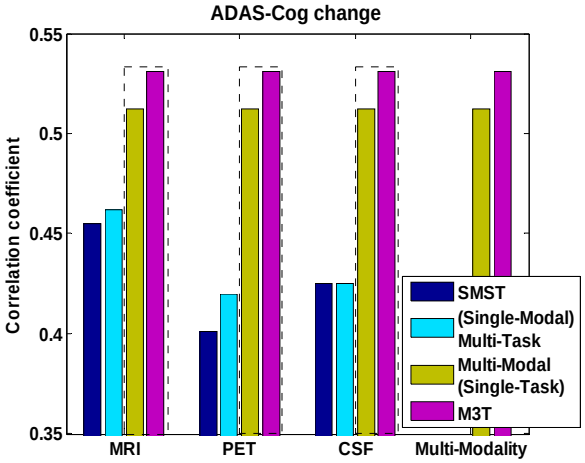
Results (cont'd)

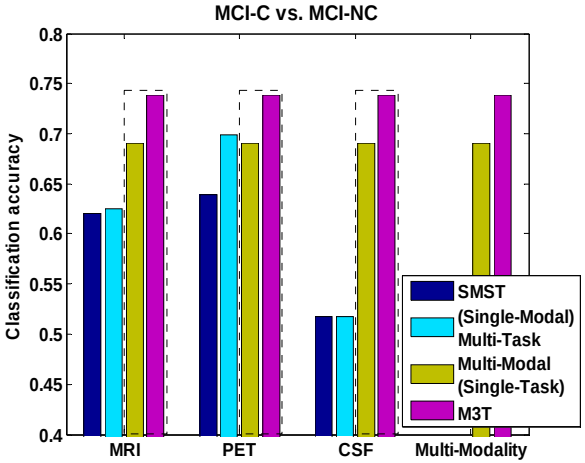
- Comparison of performances of different methods on Experiment 2

Methods	Correlation coefficient		Classification accuracy
	MMSE change	ADAS-Cog change	MCI-C vs. MCI-NC
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

Results (cont'd)







Comparison of performances of four different methods on Experiment 2

More Works

- Using subjects with other related diseases
  - Semi-Supervised multimodal classification of AD (D. Zhang, D. Shen. ISBI'11)
  - Semi-Supervised multimodal relevance vector regression (B. Cheng, D. Zhang, et al. Neuroinformatics, 2013)
  - Domain transfer SVM for MCI conversion prediction (B. Cheng, D. Zhang, et al. MICCAI'12)
- Using longitudinal data
  - Longitudinal feature extraction/selection (D. Zhang, D. Shen. PLoS ONE, 2012)
  - Temporally-guided group Lasso (D. Zhang, et al. MICCAI'12)
- Using structure-based regularization
  - Manifold regularized group Lasso (B. Jie, D. Zhang, et al. MICCAI'13)
  - Tree-guided Lasso (M. Liu, D. Zhang, et al. MICCAI'12)

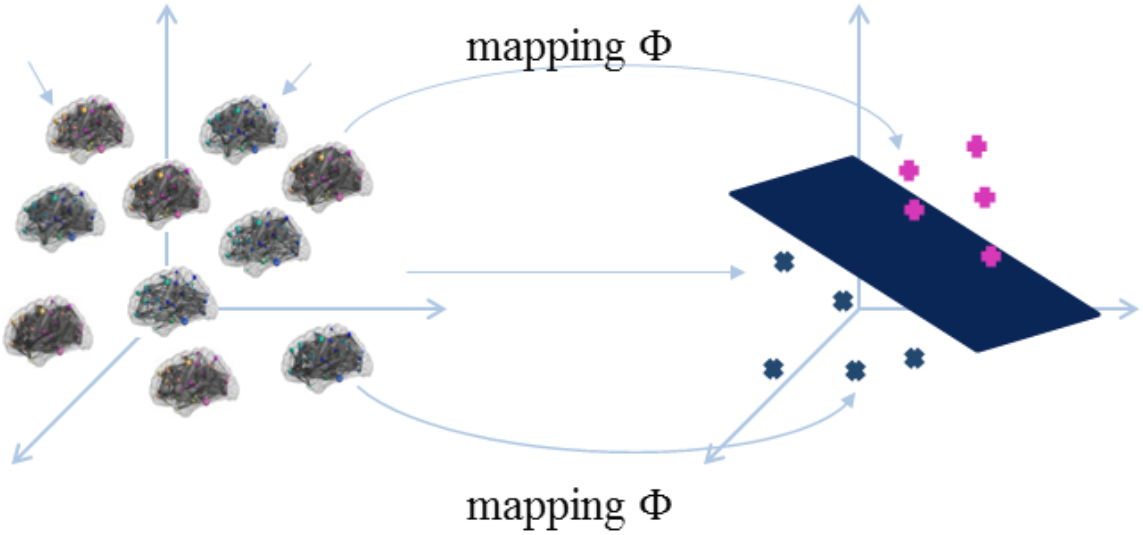
Outline

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

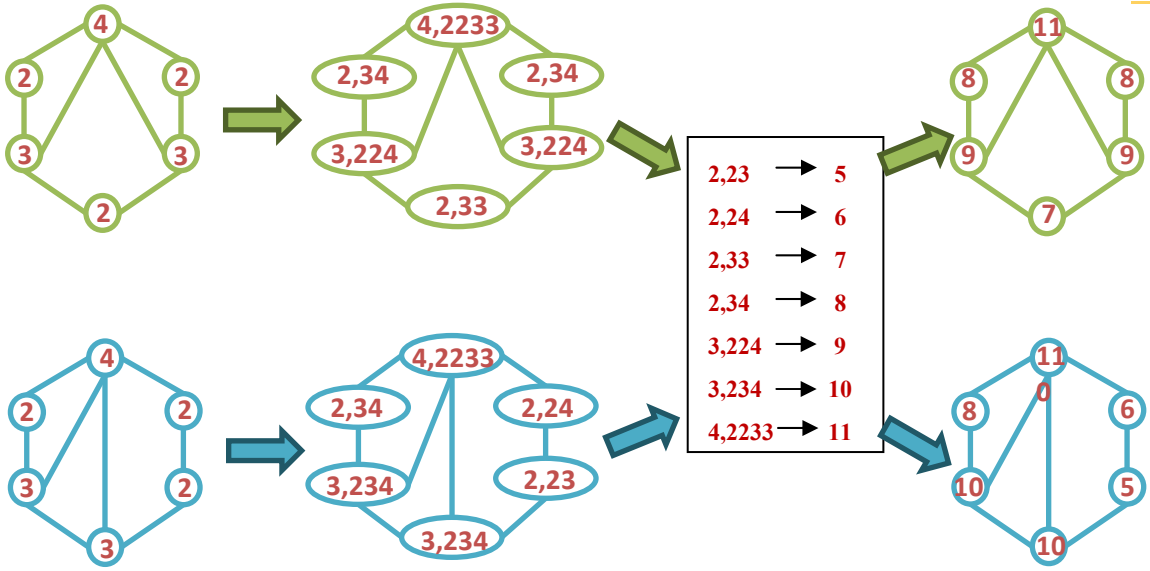


Subtree-based Graph Kernel

- 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



(a) (b) (c) (d)

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

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

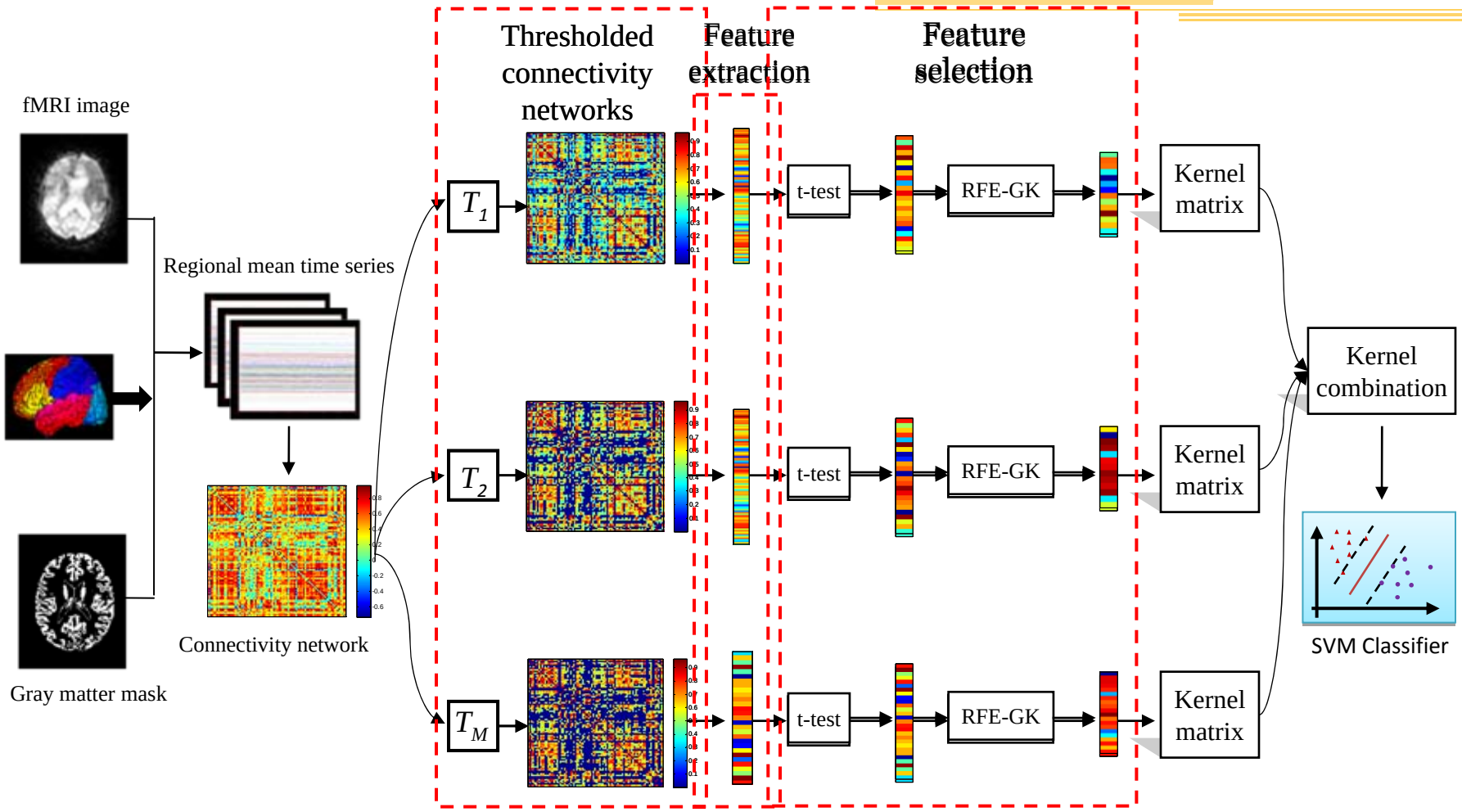
$$\emptyset(H) = (\ 3,\ 2,\ 1,\ 1,\ 1,\ 0,\ 1,\ 0,\ 2,\ 1)$$

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

(e)

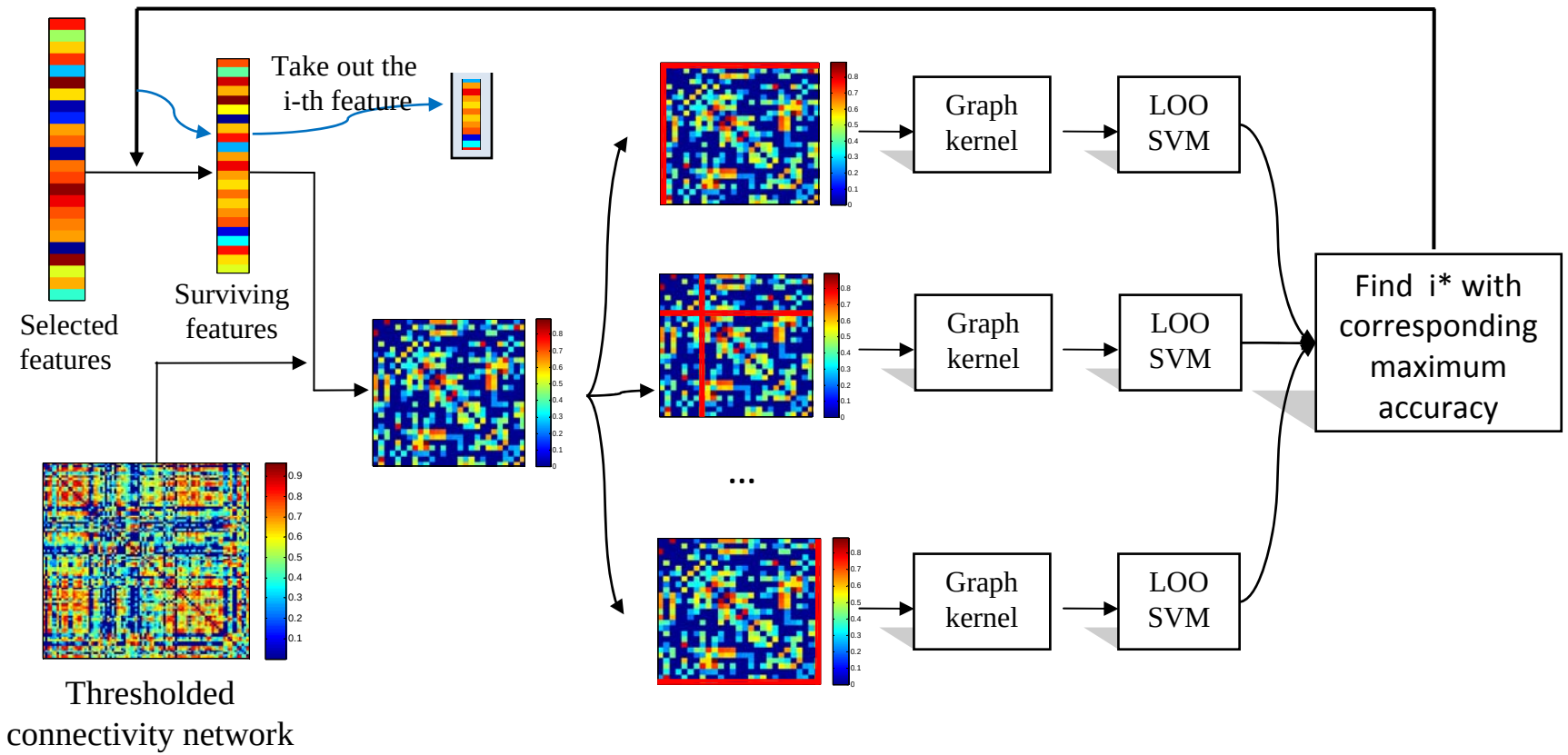
Illustration of the construction process of the Weisfeiler-Lehman subtree kernel with for two graphs *G* and *H*. Here, (a) the initial labeled graph *G* and *H*, (b) augmented labels on graph *G* and *H* , (c) label compression, (d) relabeled graph *G* and *H*, (e) computation of the kernel on Graph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

Flowchart



(B. Jie, D. Zhang, et al., Human Brain Mapping, 2013)

RFE-GK Flowchart



(B. Jie, D. Zhang, et al., Human Brain Mapping, 2013)

Experiment & Results

- Dataset: 12 MCI patients and 25 healthy controls

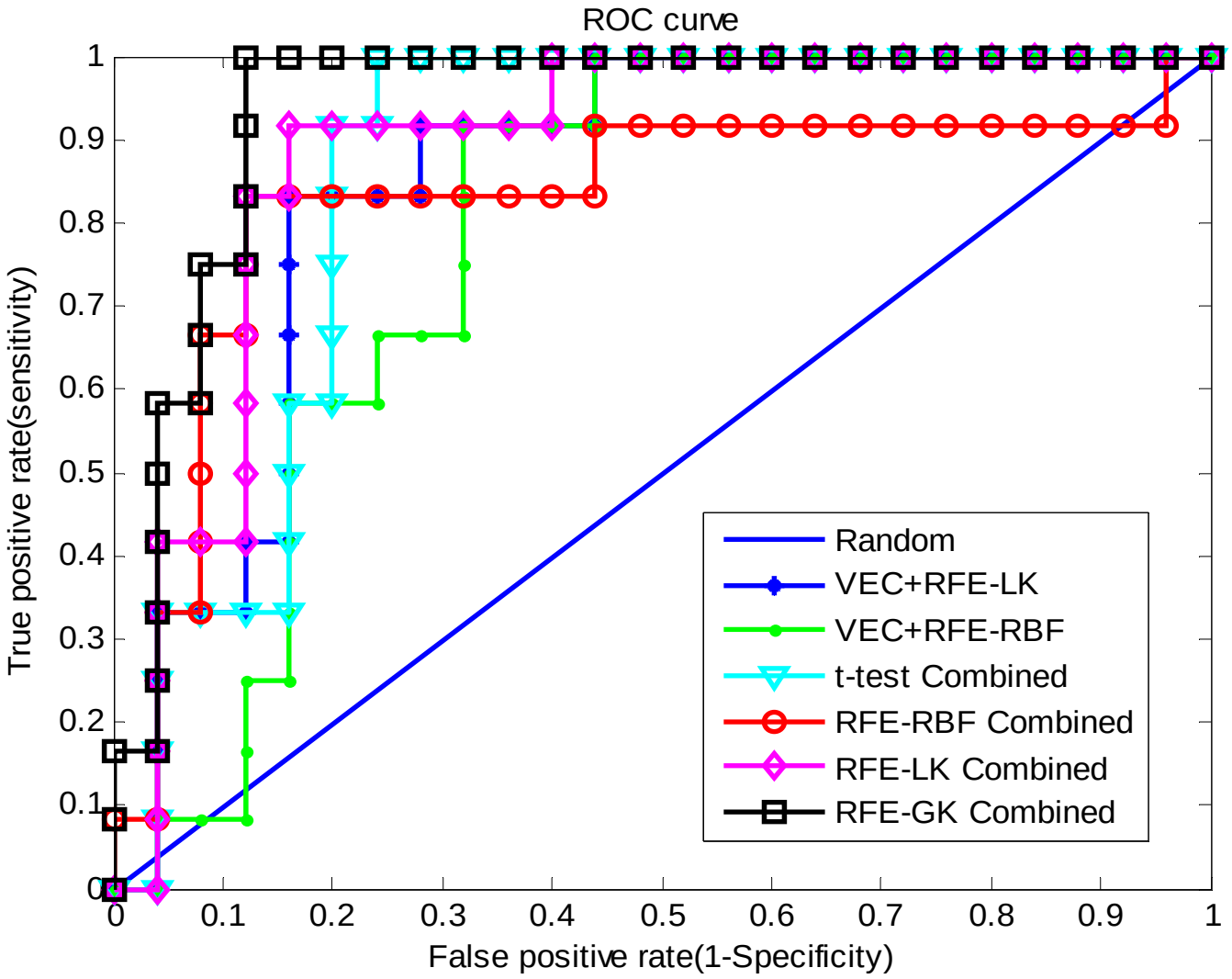
Group	MCI	Normal
No. of subjects (male/female)	6/6	9/16
Age (mean ± SD)	75.0 ± 8.0	72.9 ± 7.9
Years of edu (mean ± SD)	18.0 ± 4.1	15.8 ± 2.4
MMSE (mean ± SD)	28.5 ± 1.5	29.3 ± 1.1

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

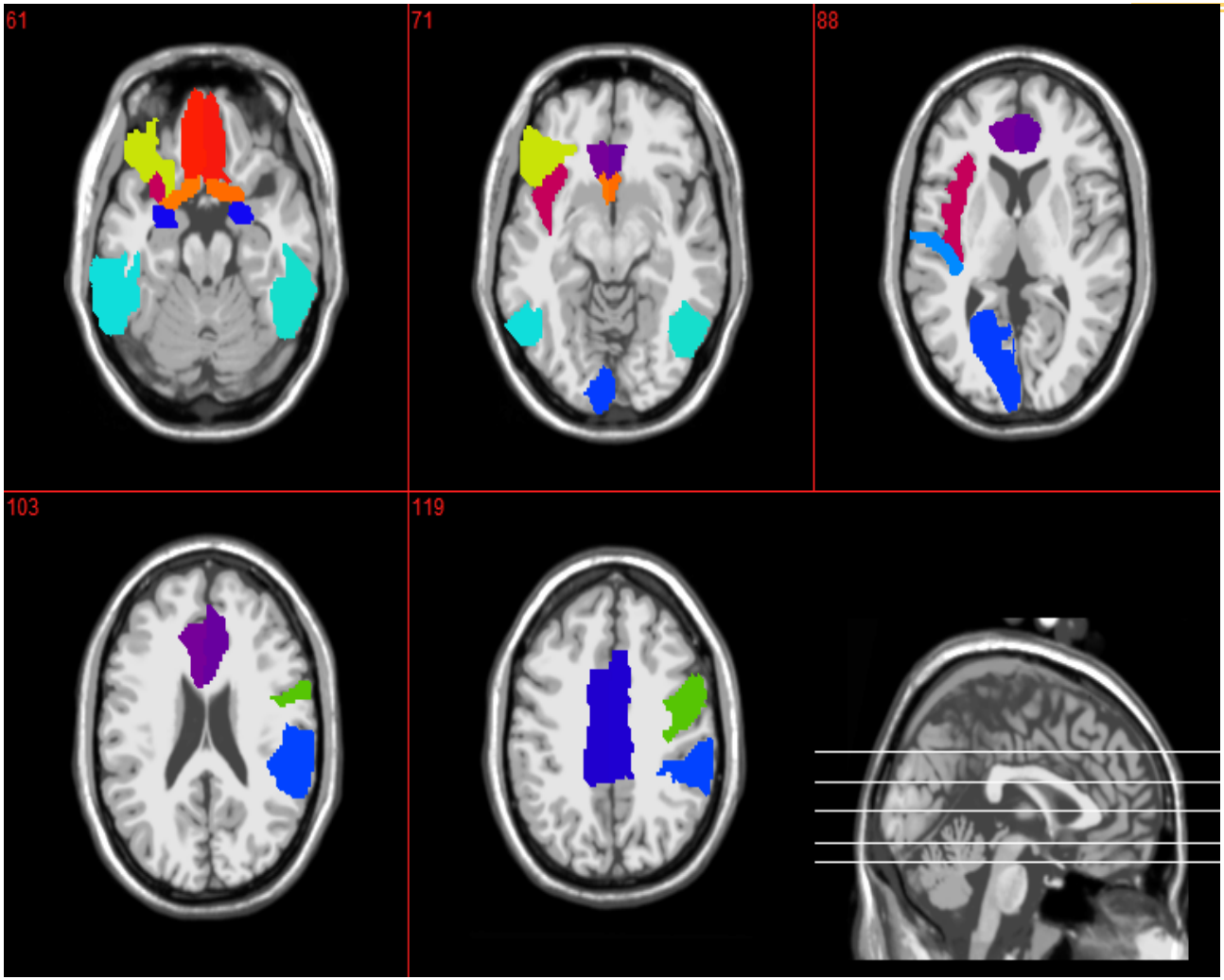
Classification results

	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


ROC curve
True positive rate(sensitivity)
False positive rate(1-Specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Random
VEC+RFE-LK
VEC+RFE-RBF
t-test Combined
RFE-RBF Combined
RFE-LK Combined
RFE-GK Combined

Top Selected ROIs



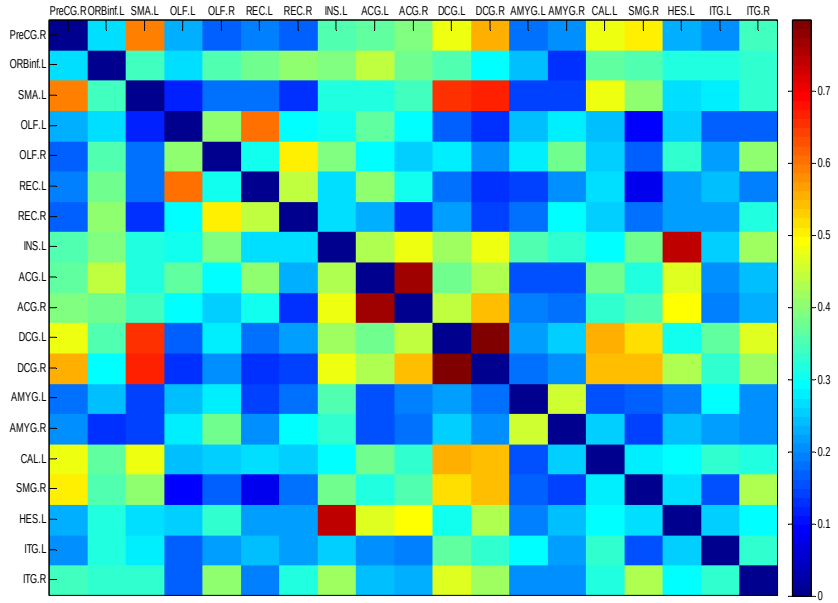
(B. Jie, D. Zhang, et al., Human Brain Mapping, 2013)

Selected ROIs

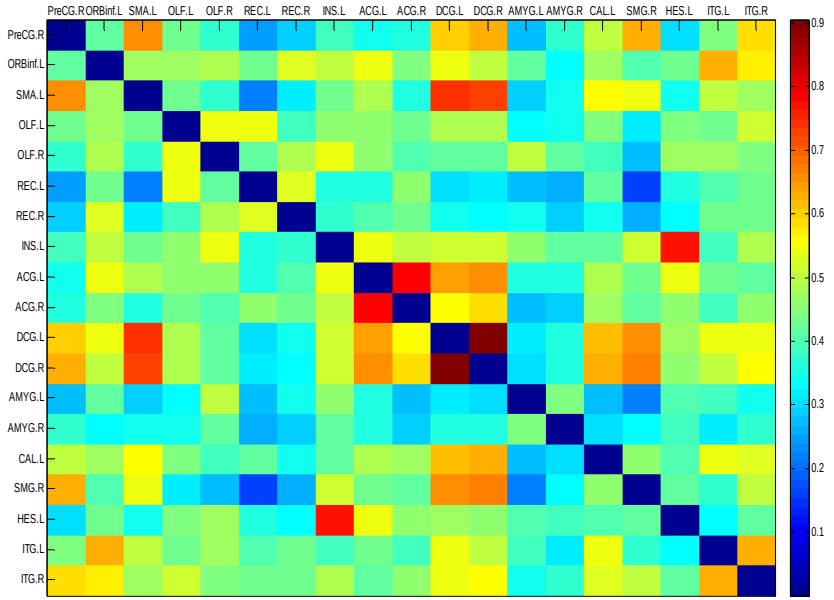
T_1	T_2	T_3
L olfactory cortex	L inferior temporal gyrus	L anterior cingulate gyrus
R inferior temporal gyrus	L olfactory cortex	L olfactory cortex
L inferior temporal gyrus	R inferior temporal gyrus	R middle cingulate
L anterior cingulate gyrus	R gyrus rectus	L amygdala
R supramarginal gyrus	R amygdala	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

T_4	T_5	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**
		L amygdala
		R amygdala
		L calcarine sulcus
		R supramarginal gyrus
		L heschl gyrus
		L inferior temporal gyrus
		R inferior temporal gyrus

Connectivity Sub-networks



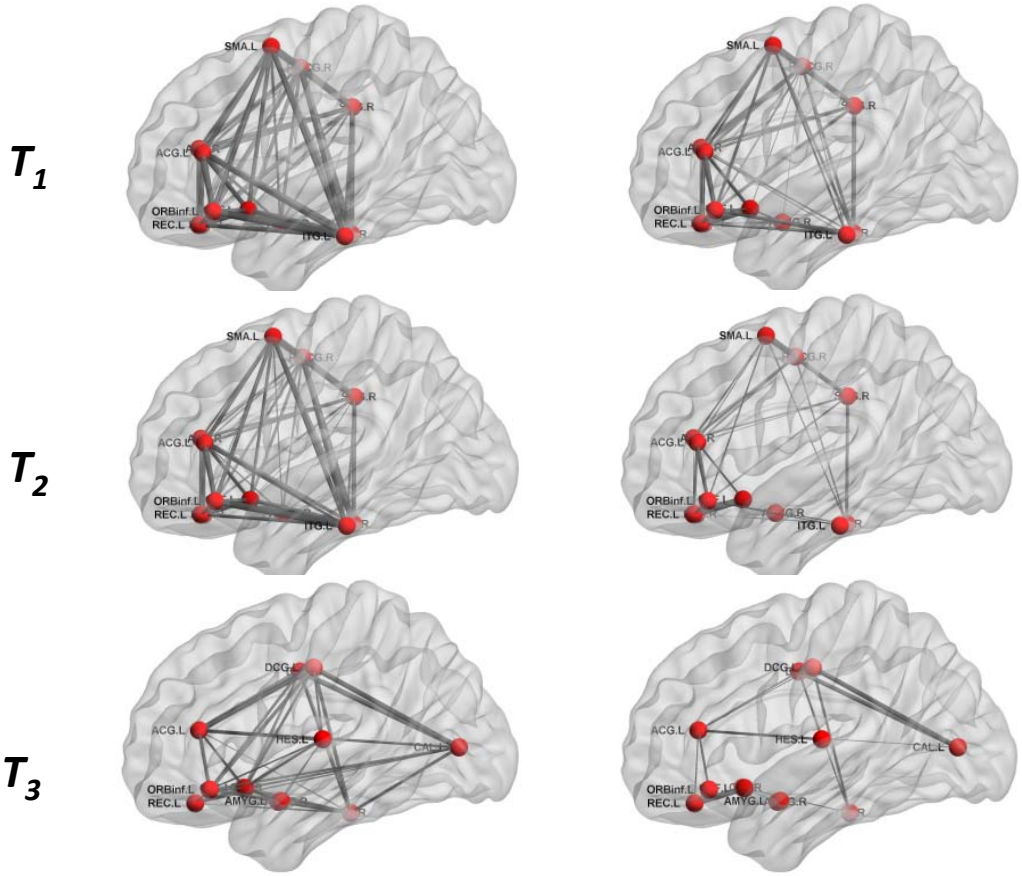
(a) MCI group



(a) NC group

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

Brain Sub-networks

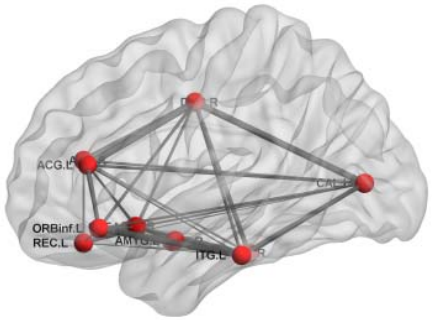


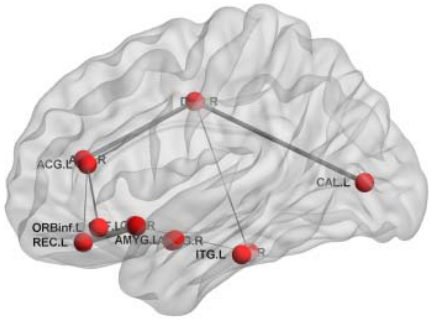
(a) NC (b) MCI

Thresholded average connectivity sub-network based on top selected ROIs

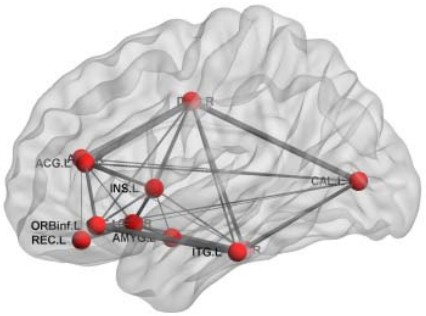
Brain Sub-networks

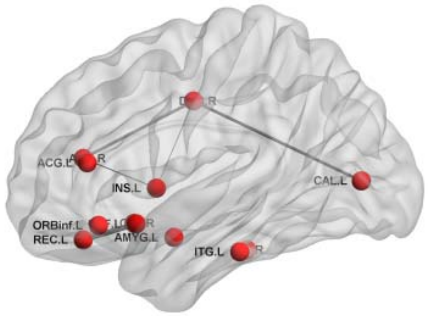
T_4





T_5





(a) NC (b) MCI

Thresholded average connectivity sub-network based on top selected ROIs

More Works

- Using both structural and functional connectivity networks
  - See (C.Y Wee, et al., Neuroimage, 2012)
- Integration of vector kernel and graph kernel for network-based classification
  - See (B. Jie, D. Zhang, et al., IEEE Trans. BME, 2013)
- Discriminative and frequent sub-network selection for network-based classification
  - Submitted for ISBI'14

Outline

1. Backgrounds on Alzheimer's Disease
2. Multi-modality based Classification
3. Brain-network based Classification
4. Summary

Summary

- Multi-modality and brain network are two hot topics in recent neuroimaging-based analysis
- Machine learning 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
  - Bo Cheng
  - Biao Jie
  - Mingxia Liu
  - Xiaoke Hao
  - Chen Zu
  - Fei Fei
  - Lipeng Wang
  - …

Thanks for your attention!

For more details, contact:

dqzhang@nuaa.edu.cn

http://parnec.nuaa.edu.cn/zhangdq