深圳大学
SHENZHEN UNIVERSITY

Longitudinal Feature Selection and Feature Learning for Parkinson’s Disease Diagnosis and Prediction

Presenter: Zhongwei Huang

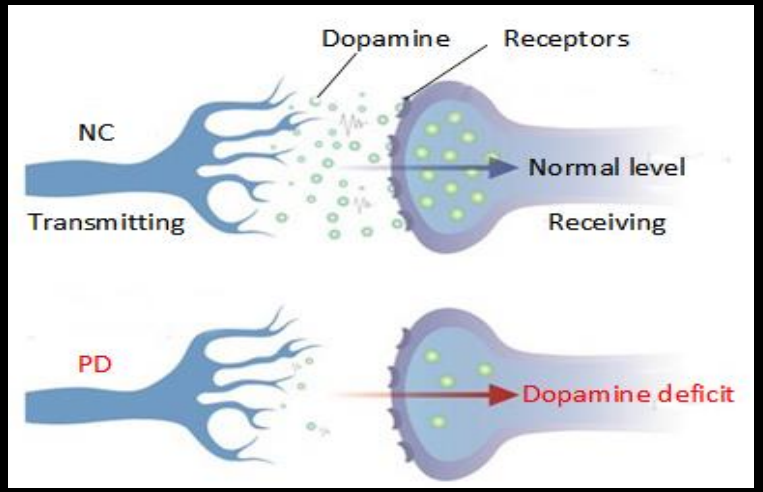
Shenzhen University

Background

- Parkinson's disease

Progressive neurodegenerative disease

Etiology



SWEDD

Normal control , NC
Parkinson's disease, PD
Scans without evidence of dopamine deficit, SWEDD

Background

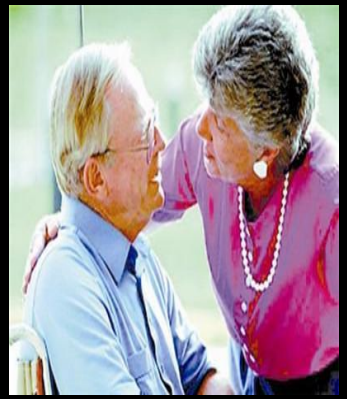
- Early stage

Depression

GDS :0~15

Sleep

ESS:0~24

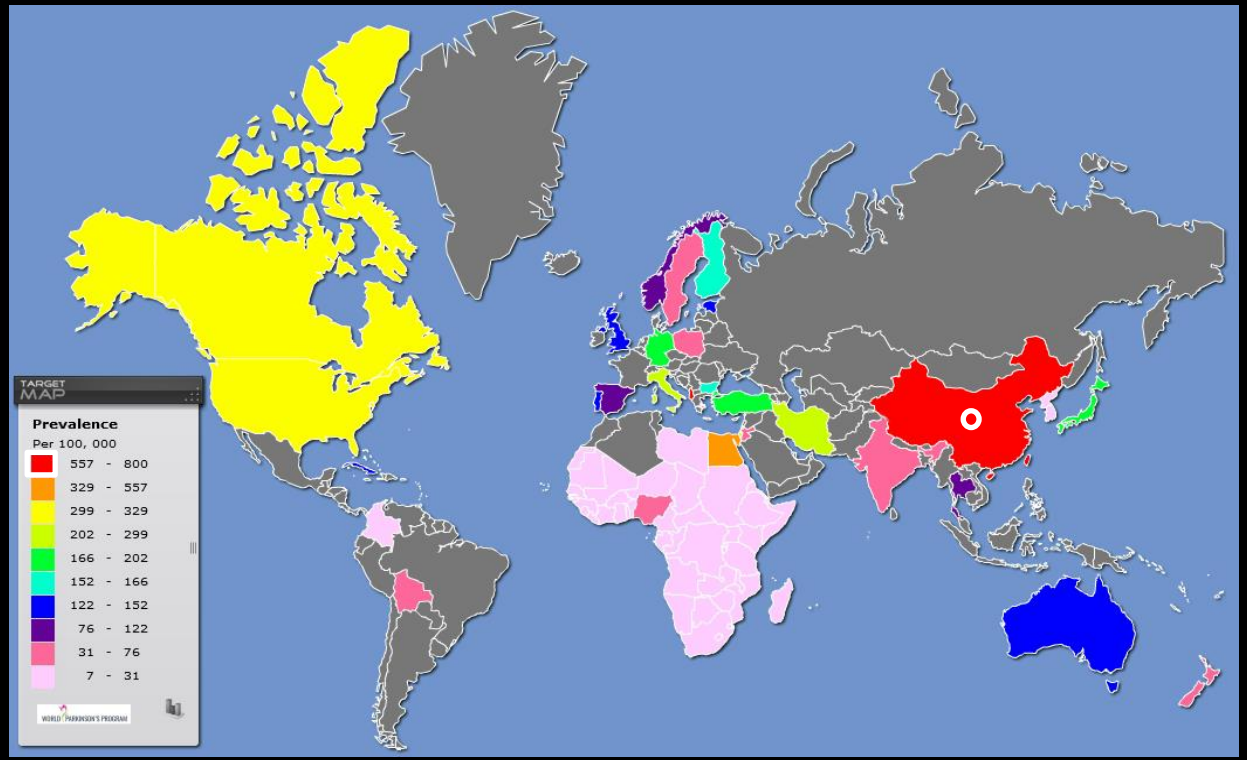


Cognition

MoCA:0~30

Geriatric Depression Scale, GDS
Epworth Sleepiness Scale, ESS
Montreal Cognitive Assessment, MoCA

Background



https://www.targetmap.com/viewer.aspx?reportId=53892

5.8 million people with PD around the world

2.6 million people with PD in China, ranking first in the world

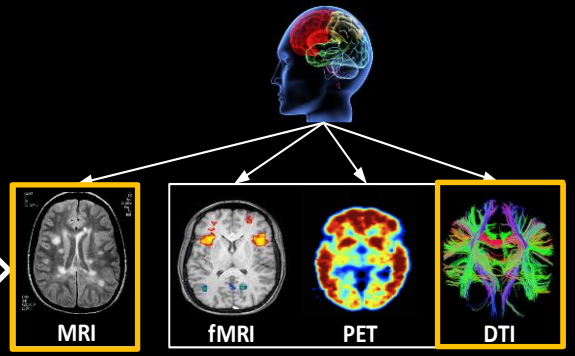
By 2030, the number of people with PD in China will reach 5 million

Background

- Early diagnosis

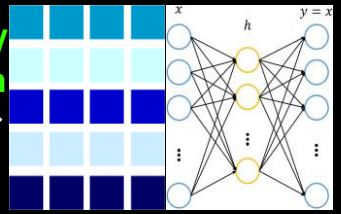
Clinical diagnosis

•Rely on doctors' clinical experience



Multi-modal data

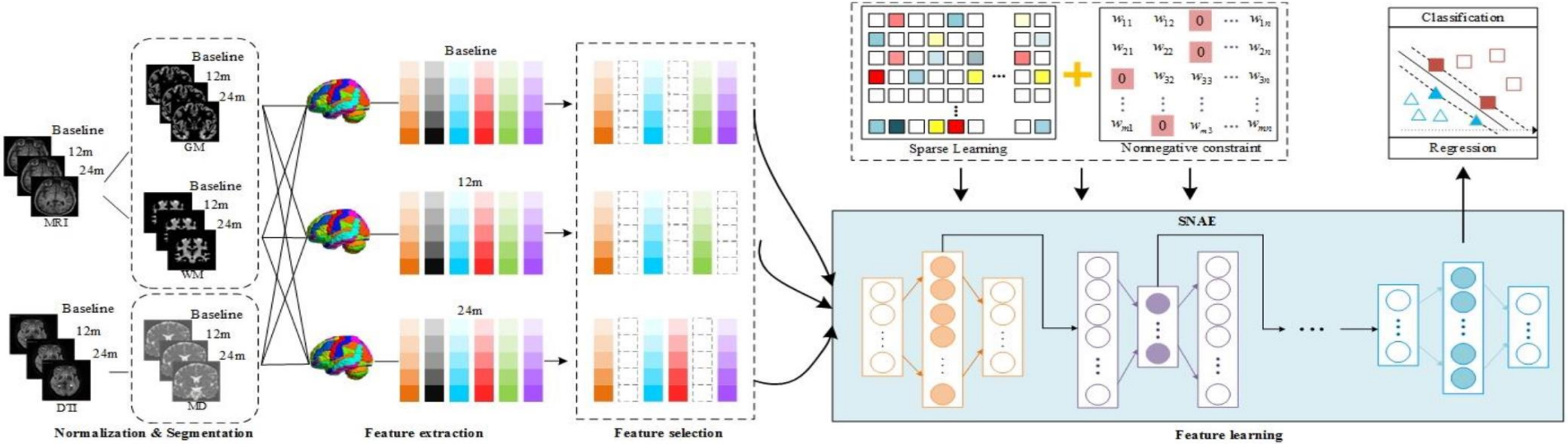
•Limited subjects
•High dimensionality
•Overfitting problem



Feature selection and learning

Methods

Architecture of Feature Selection and Feature Learning

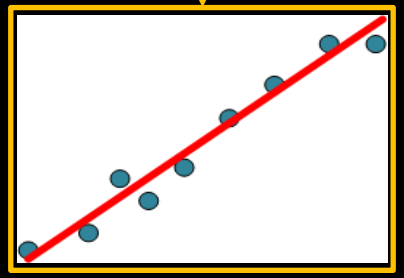


Magnetic resonance imaging, MRI
Gray matter, GM
White matter , WM
Mean diffusion, MD
Diffusion tensor imaging, DTI

Methods

- **Joint Learning from Multiple Modalities and Relations (JLMMR)**

$$\min_{\mathbf{W}} \underline{\|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2} + \lambda_1 \sum_{i,j}^{s} \left\|(\mathbf{y}_i - \mathbf{y}_j) - (\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j)\right\|_2^2 + \lambda_2 \sum_{i,j}^{n} \left\|(\mathbf{y}^i - \mathbf{y}^j) - (\mathbf{x}^i\mathbf{W} - \mathbf{x}^j\mathbf{W})\right\|_2^2 + \lambda_3 \sum_{i,j}^{d} \mathrm{h}_{ij} \left\|\mathbf{w}^i - \mathbf{w}^j\right\|_2^2 + \lambda_4 \|\mathbf{W}\|_{2,1}$$

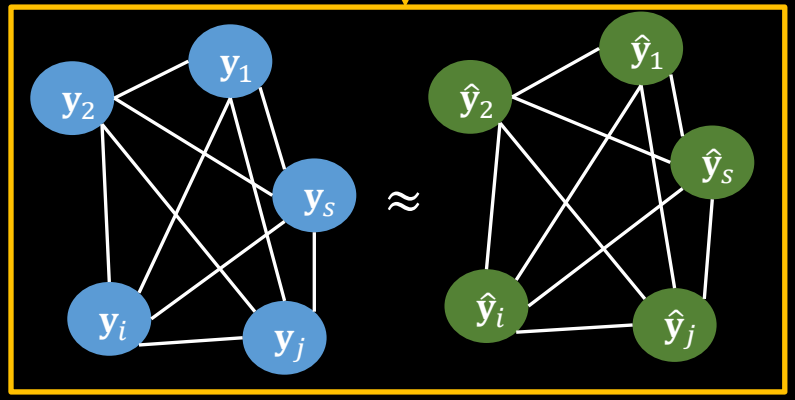


Least square regression

Methods

- **Joint Learning from Multiple Modalities and Relations (JLMMR)**

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2 + \lambda_1 \sum_{i,j}^{s} \left\|(\mathbf{y}_i-\mathbf{y}_j)-(\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j)\right\|_2^2 + \lambda_2 \sum_{i,j}^{n} \left\|(\mathbf{y}^i-\mathbf{y}^j)-(\mathbf{x}^i\mathbf{W}-\mathbf{x}^j\mathbf{W})\right\|_2^2 + \lambda_3 \sum_{i,j}^{d} h_{ij} \left\|\mathbf{w}^i-\mathbf{w}^j\right\|_2^2 + \lambda_4 \|\mathbf{W}\|_{2,1}$$

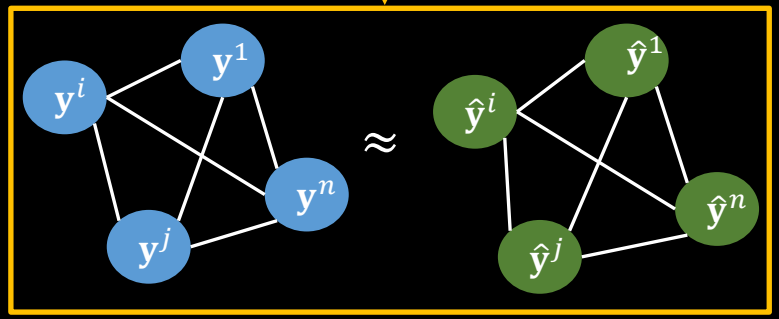


Column-wise relation in response variables

Methods

- **Joint Learning from Multiple Modalities and Relations (JLMMR)**

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2 + \lambda_1 \sum_{i,j}^{s} \left\|(\mathbf{y}_i-\mathbf{y}_j)-(\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j)\right\|_2^2 + \lambda_2 \sum_{i,j}^{n} \left\|(\mathbf{y}^i-\mathbf{y}^j)-(\mathbf{x}^i\mathbf{W}-\mathbf{x}^j\mathbf{W})\right\|_2^2 + \lambda_3 \sum_{i,j}^{d} \mathrm{h}_{ij} \left\|\mathbf{w}^i-\mathbf{w}^j\right\|_2^2 + \lambda_4 \|\mathbf{W}\|_{2,1}$$

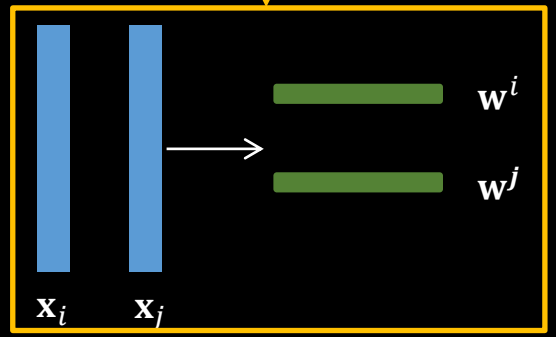


Row-wise relation in response variables

Methods

- **Joint Learning from Multiple Modalities and Relations (JLMMR)**

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2 + \lambda_1 \sum_{i,j}^{s} \left\|(\mathbf{y}_i-\mathbf{y}_j)-(\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j)\right\|_2^2 + \lambda_2 \sum_{i,j}^{n} \left\|(\mathbf{y}^i-\mathbf{y}^j)-(\mathbf{x}^i\mathbf{W}-\mathbf{x}^j\mathbf{W})\right\|_2^2 + \lambda_3 \sum_{i,j}^{d} \mathrm{h}_{ij} \left\|\mathbf{w}^i-\mathbf{w}^j\right\|_2^2 + \lambda_4 \|\mathbf{W}\|_{2,1}$$

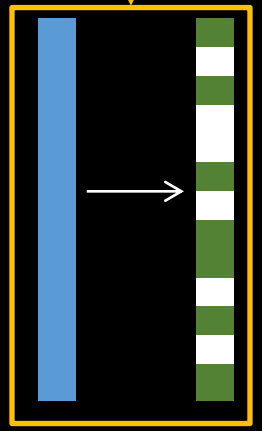


Feature-feature relation

Methods

- **Joint Learning from Multiple Modalities and Relations (JLMMR)**

$$\min_{\mathbf{W}} \|\mathbf{Y}-\mathbf{X}\mathbf{W}\|_F^2 + \lambda_1 \sum_{i,j}^{s} \left\|(\mathbf{y}_i-\mathbf{y}_j)-(\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j)\right\|_2^2 + \lambda_2 \sum_{i,j}^{n} \left\|(\mathbf{y}^i-\mathbf{y}^j)-(\mathbf{x}^i\mathbf{W}-\mathbf{x}^j\mathbf{W})\right\|_2^2 + \lambda_3 \sum_{i,j}^{d} \mathrm{h}_{ij} \left\|\mathbf{w}^i-\mathbf{w}^j\right\|_2^2 + \lambda_4 \|\mathbf{W}\|_{2,1}$$



Sparse term

Methods

- **Sparse Nonnegative Autoencoder (SNAE)**

  - **Loss function**

$$J(\boldsymbol{W},\boldsymbol{b}) = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{d}(y_{ij}-x_{ij})^2 + \frac{\lambda}{2}\Omega_{weights} + \beta\Omega_{sparsity}$$

n: sample size

d : dimension

  - **Sparse learning**

$$\Omega_{sparsity} = \sum_{i=1}^{k}\rho\, log\frac{\rho_i'}{\rho} + (1-\rho)\, log\frac{1-\rho_i'}{1-\rho}$$

k:amount of hidden nodes

ρ_i':average activation

ρ: desired activation

  - **Nonnegative constraint**

$$\Omega_{weights} = \sum_{l=1}^{2}\sum_{i=1}^{t_l}\sum_{j=1}^{t_{l+1}} f(w_{ij}^{(l)}),\quad f(w_{ij}^{(l)}) = \begin{cases} w_{ij}^2, & w_{ij} < 0 \\ 0, & w_{ij} \geq 0 \end{cases},$$

t_l , t_{l+1}: amount of adjacent layer nodes

Experiments

- **Data distribution**

Time	Information	NC	PD	SWEDD
0m	Number	62	142	34
	Depression scores	5.1±1.2	5.3±1.5	5.6±1.3
	Sleep scores	6.5±3.9	5.9±3.4	8.7±4.2
	Cognition scores	28.2±1.1	27.5±2.1	26.9±2.9
12m	Number	54	123	9
	Depression scores	4.9±1.4	5.3±1.3	5.6±1.0
	Sleep scores	6.1±3.7	6.6±4.3	7.2±3.7
	Cognition scores	27.5±1.9	26.8±2.9	26.4±2.9
24m	Number	7	98	22
	Depression scores	4.9±0.4	5.7±1.6	5.4±1.5
	Sleep scores	7.1±3.6	7.9±4.4	7.2±3.6
	Cognition scores	28.3±1.0	26.6±2.9	26.0±2.7

Experiments

- **Classification performance**

TABLE 1. Classification performance. Boldface denotes the best performance.(mean± standard deviation)

Time	Feature selection	Feature learning	NC vs. PD				NC vs. SWEDD			
			ACC	SEN	SPEC	F-score	ACC	SEN	SPEC	F-score
Baseline	MMSL	DBN	75.98±10	40.48±20	91.57±2	48.17±20	78.11±11	98.33±1	40.00±20	85.49±10
	JESR		76.40±12	44.76±22	92.86±3	51.42±18	80.33±12	98.57±1	54.17±21	87.98±11
	JLMMR		74.85±13	42.60±23	90.48±1	47.38±21	77.35±18	93.88±4	38.51±15	84.57±9
	MMSL	SSAE	78.98±15	39.29±29	96.43±2	49.74±20	83.78±10	95.00±3	63.33±19	88.49±8
	JESR		78.45±10	51.67±16	90.05±3	**57.02±15**	83.67±12	93.57±2	**67.50±15**	88.23±10
	JLMMR		79.02±12	42.14±24	95.00±1	52.41±13	84.44±13	**98.33±1**	59.17±13	89.09±9
	MMSL	Ours	79.00±9	39.05±23	**96.43±2**	49.16±21	83.78±10	95.00±2	63.33±14	88.49±10
	JESR		78.50±18	42.38±19	94.38±3	51.82±20	**84.78±11**	96.9.0±3	64.17±15	**89.27±8**
	JLMMR		**79.95±19**	**45.71±28**	95.00±2	55.55±20	82.56±14	95.00±2	59.17±10	87.85±6
12m	MMSL	DBN	73.54±15	24.55±29	95.06±2	34.02±23	90.77±8	100±0	40.00±13	94.93±3
	JESR		72.76±10	22.36±28	**99.20±0**	35.29±25	91.57±7	100±0	50.00±15	96.44±3
	JLMMR		73.11±20	25.68±23	96.54±1	36.15±22	89.16±10	100±0	40.00±12	95.24±4
	MMSL	SSAE	78.54±16	36.91±20	96.80±2	49.28±20	93.85±6	100±0	60.00±13	96.52
	JESR		75.10±12	34.91±20	92.73±2	43.24±26	93.85±5	100±0	60.00±15	93.59±2
	JLMMR		79.04±13	**51.45±19**	91.13±3	58.23±20	95.38±4	100±0	70.00±20	97.39±2
	MMSL	Ours	**80.71±10**	46.36±18	95.79±2	**58.42±21**	95.38±4	100±0	70.00±19	97.39±2
	JESR		76.26±16	42.36±16	91.13±3	51.74±19	93.85±5	100±0	60.00±22	96.52±2
	JLMMR		78.13±17	37.45±18	96.00±2	47.29±18	**95.38±4**	**100±0**	**70.00±15**	**97.46±1**
24m	MMSL	DBN	93.41±10	30.00±20	99.00±1	10.00±29	68.95±15	80.00±10	64.00±13	50.00±10
	JESR		92.14±12	40.00±30	100±0	73.33±20	69.48±18	100±0	73.00±15	74.00±15
	JLMMR		91.67±13	60.00±25	100±0	50.00±23	70.18±16	100±0	75.00±18	**74.00±10**
	MMSL	SSAE	97.23±18	60.00±29	100±0	**66.67±18**	75.62±12	100±0	68.00±19	70.67±15
	JESR		96.18±12	60.00±19	98.89±1	60.00±12	75.81±10	100±0	69.00±20	70.00±10
	JLMMR		97.09±11	50.00±22	100±0	53.33±19	78.00±15	80.00±0	76.00±21	56.67±15
	MMSL	Ours	97.23±19	**70.00±21**	99.00±1	62.00±15	80.81±10	80.00±10	75.00±15	65.18±16
	JESR		96.32±13	60.00±20	99.00±1	63.33±21	**80.95±11**	**100±0**	74.00±10	66.67±15
	JLMMR		**97.99±16**	60.00±17	**100±0**	60.00±18	80.76±12	80.00±10	**79.00±11**	60.76±19

Experiments

- **Classification performance**

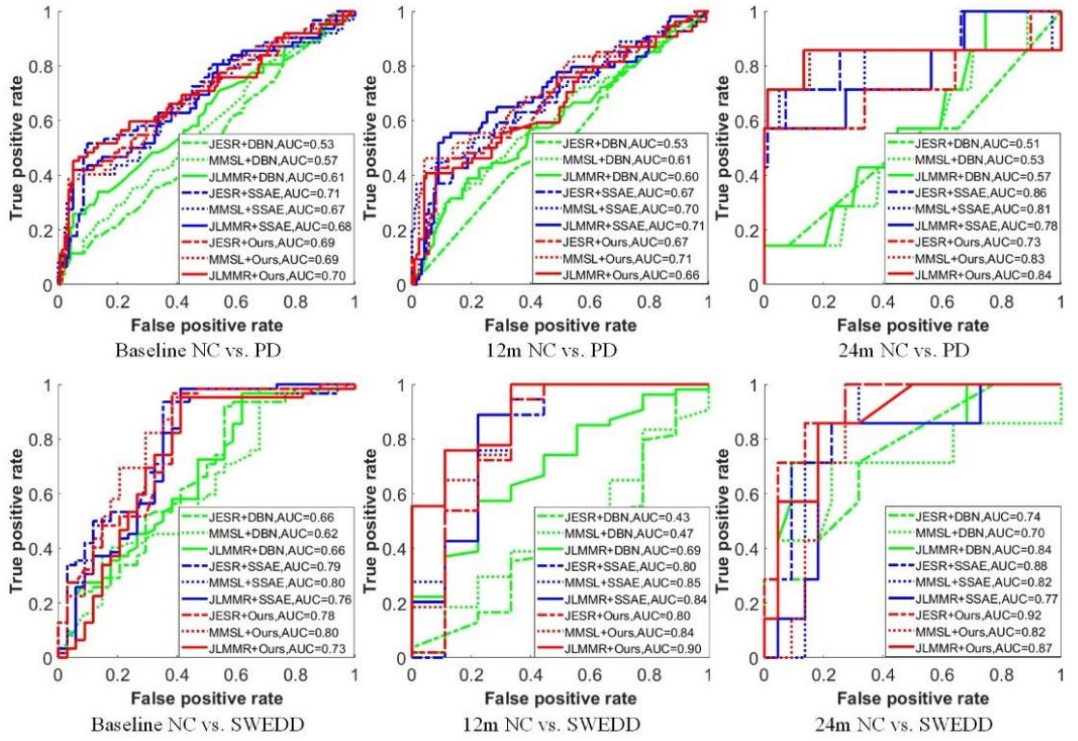


Fig. 3. ROC curves for all methods.

Experiments

- **Regression performance**

Table 2. Regression performance of all methods. Boldface denotes the best performance.

Time	Feature selection	Feature learning	NC vs. PD						NC vs. SWEDD					
			Depression		Sleep		MoCA		Depression		Sleep		MoCA	
			CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC	RMSE
Baseline	MMSL	DBN	0.51	2.41	0.46	4.35	0.47	2.70	0.64	3.04	0.68	5.53	0.71	4.06
	JESR		0.50	2.30	0.51	4.27	0.50	2.68	0.67	3.02	0.64	5.24	0.65	3.56
	JLMMR		0.50	2.41	0.49	4.38	0.46	2.91	0.61	3.18	0.59	5.38	0.63	3.92
	MMSL	SSAE	0.53	1.97	0.56	5.17	0.55	2.60	0.77	2.42	0.71	4.61	0.83	2.45
	JESR		0.51	2.20	0.49	4.36	0.47	2.92	0.68	1.72	0.69	4.81	0.70	3.28
	JLMMR		0.56	1.46	0.55	4.06	0.58	2.36	0.73	1.69	0.71	**3.99**	0.82	2.36
	MMSL	Ours	0.54	1.95	0.57	4.59	0.58	**2.16**	0.75	2.79	0.70	4.67	0.84	1.85
	JESR		0.54	2.12	0.57	5.17	0.52	2.68	0.69	1.71	**0.75**	4.45	0.71	2.82
	JLMMR		**0.57**	**1.38**	**0.57**	**3.56**	**0.58**	2.34	**0.80**	**1.19**	0.69	6.11	**0.84**	**1.79**
12m	MMSL	DBN	0.33	2.25	0.34	4.89	0.44	3.50	0.67	2.10	0.57	4.13	0.66	2.85
	JESR		0.43	2.05	0.40	4.93	0.49	3.44	0.68	2.03	0.62	4.17	0.65	3.84
	JLMMR		0.39	3.29	0.39	5.29	0.51	2.89	0.65	2.39	0.60	4.93	0.60	4.59
	MMSL	SSAE	0.48	1.72	0.44	4.05	0.63	3.28	0.84	1.28	0.73	3.96	0.67	**2.07**
	JESR		0.40	1.91	0.38	**4.01**	0.60	4.10	0.77	1.31	0.68	**3.56**	0.69	4.22
	JLMMR		0.53	**1.42**	0.50	5.91	0.64	2.77	0.79	1.28	0.76	3.73	0.75	3.93
	MMSL	Ours	0.49	1.73	0.59	4.06	0.66	3.72	0.85	**1.28**	0.71	3.68	0.72	4.01
	JESR		0.43	2.67	0.45	4.92	0.58	**2.68**	0.79	1.31	0.64	4.71	0.71	4.07
	JLMMR		**0.53**	1.72	**0.65**	4.86	**0.68**	4.09	**0.86**	1.31	**0.81**	4.41	**0.75**	2.44
24m	MMSL	DBN	0.45	3.62	0.45	6.58	0.47	3.60	0.80	4.03	0.74	5.53	0.81	4.30
	JESR		0.44	3.49	0.51	6.21	0.50	3.54	0.82	3.97	0.77	6.23	0.89	3.92
	JLMMR		0.49	3.91	0.48	6.72	0.52	3.35	0.85	3.27	0.78	5.67	0.89	3.20
	MMSL	SSAE	0.64	**1.70**	0.53	**5.20**	0.67	3.31	0.89	**1.63**	0.89	**3.11**	0.98	**2.52**
	JESR		0.63	2.37	0.55	5.62	0.59	3.58	0.99	2.18	0.87	4.96	0.98	3.30
	JLMMR		0.62	2.03	0.52	5.40	**0.65**	2.88	0.98	2.39	0.91	4.73	0.91	2.67
	MMSL	Ours	0.65	2.61	0.55	6.28	0.60	2.91	0.96	3.29	0.93	4.28	0.98	3.67
	JESR		0.67	2.34	0.56	6.10	0.63	**2.84**	0.97	4.60	0.91	5.17	0.99	4.29
	JLMMR		**0.68**	2.68	**0.58**	6.61	0.62	2.96	**0.99**	2.38	**0.97**	3.35	**0.99**	2.74

Experiments

- **Brain Connectivity Map**

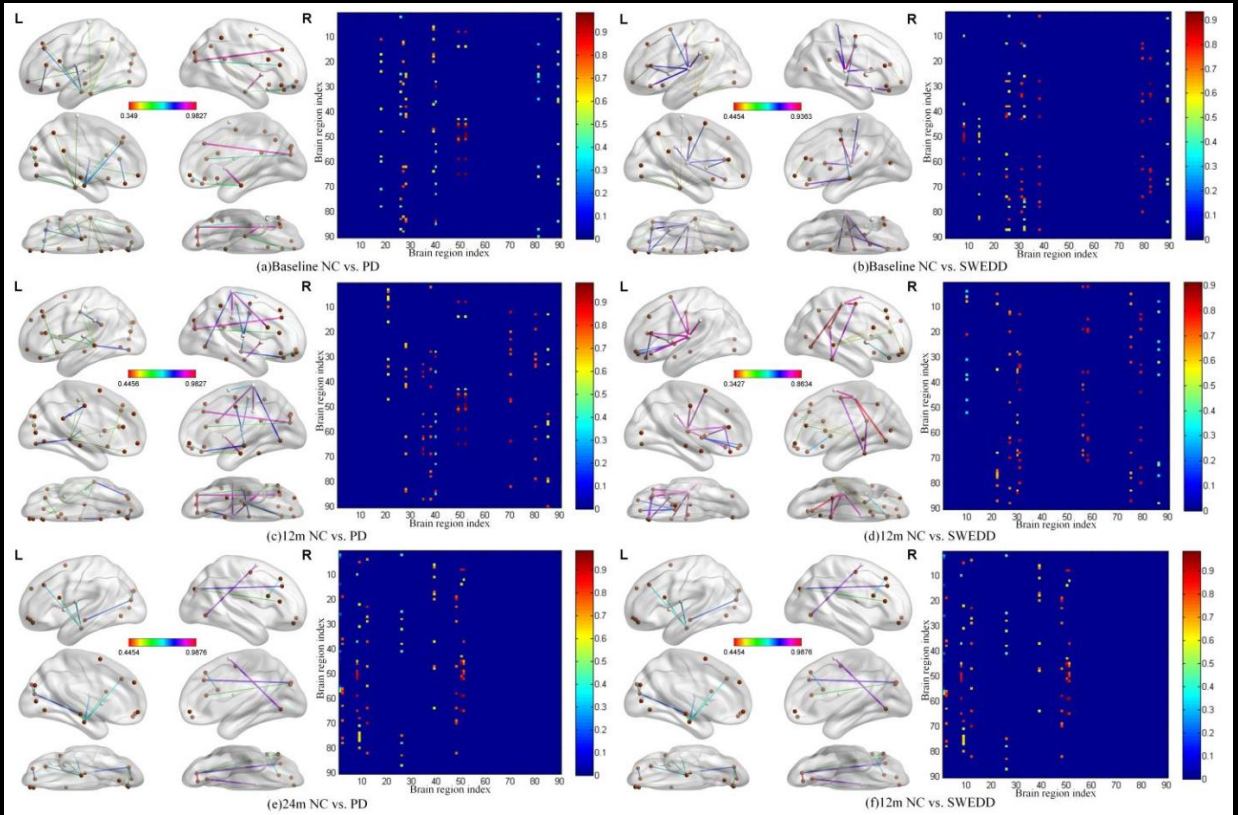


Fig. 4. Brain connectivity map.

Conclusions

- **Joint Feature Selection and Feature Learning**
  - Add relation regularization to feature selection
  - Add sparse learning and non-negative constraint to feature learning
  - Obtain discriminative feature representation
  - Have good performance for PD diagnosis

Thanks very much!