Multi-task multiple kernel machines for personalized pain recognition from functional near-infrared spectroscopy brain signals

Daniel Lopez-Martinez[1,2] (dlmocdm@mit.edu)
Ke Peng[3,4] , Sarah C. Steele[3,4,5], Arielle J. Lee[3,4,5], David Borsook[3,4,5] , Rosalind Picard[2]

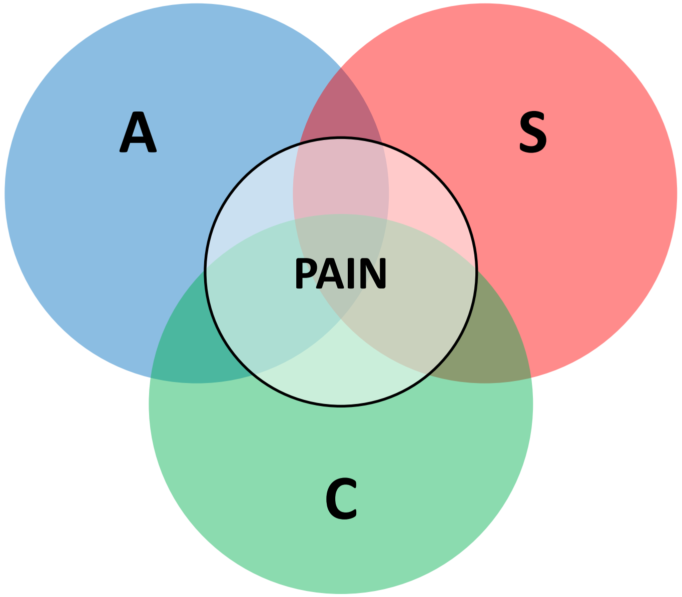
[1] Affective Computing group @ MIT Media Lab
[2] Harvard-MIT Health Sciences & Technology
[3] Center for Pain and the Brain, Harvard Medical School
[4] Department of Anesthesiology, Boston Children's Hospital
[5] Anthinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital

What is pain?

"Pain is a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive and social components."

A. C. de C Williams and K. D. Craig. Updating the definition of pain. Pain, 2016

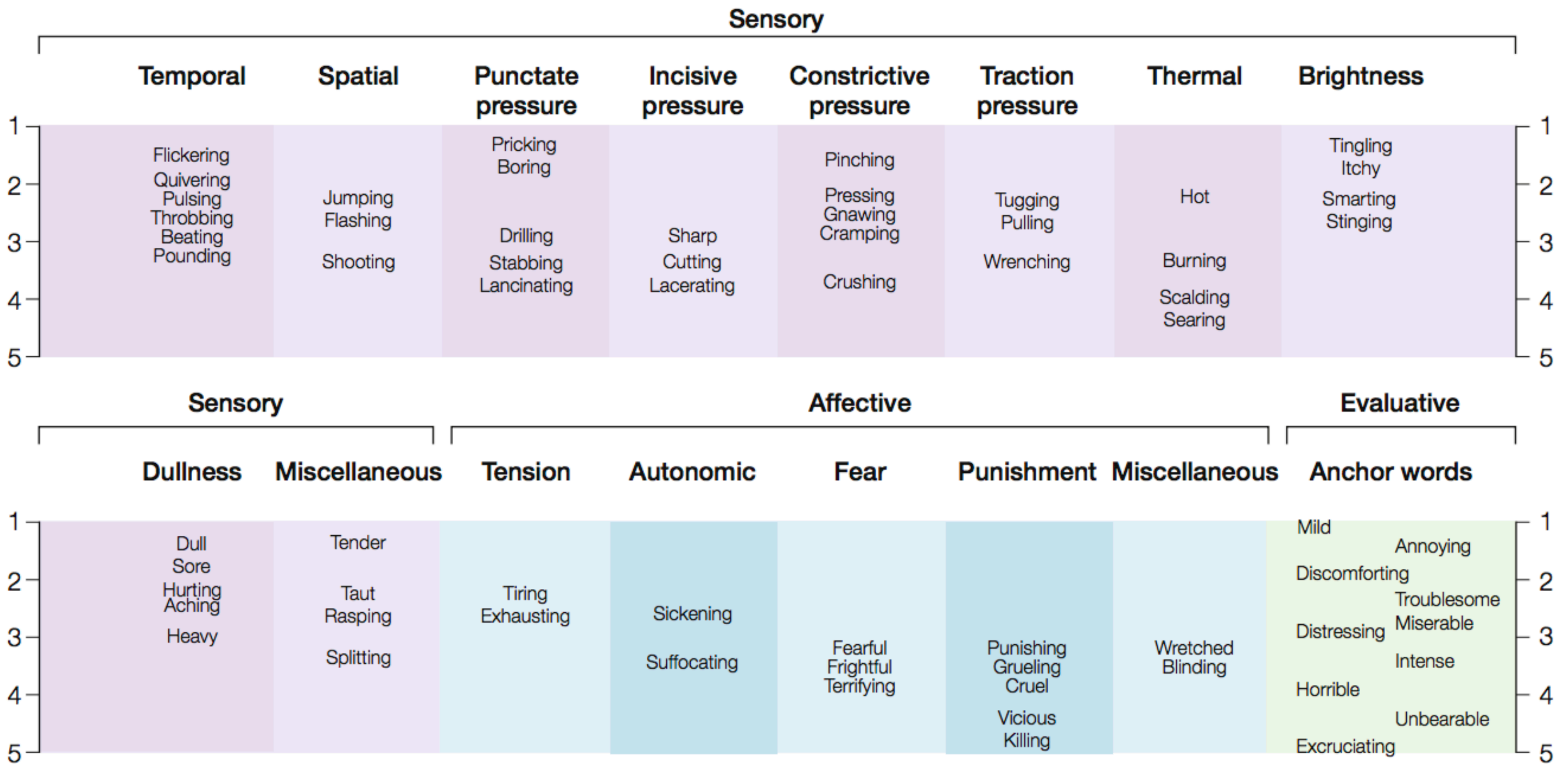
Affective: Negative emotion: anxiety, fear, unpleasant sensation.



Sensory: Perception of pain characteristics: intensity, quality, location.

Cognitive: Interpretation of pain.

Self-report: the gold standard of pain measurement



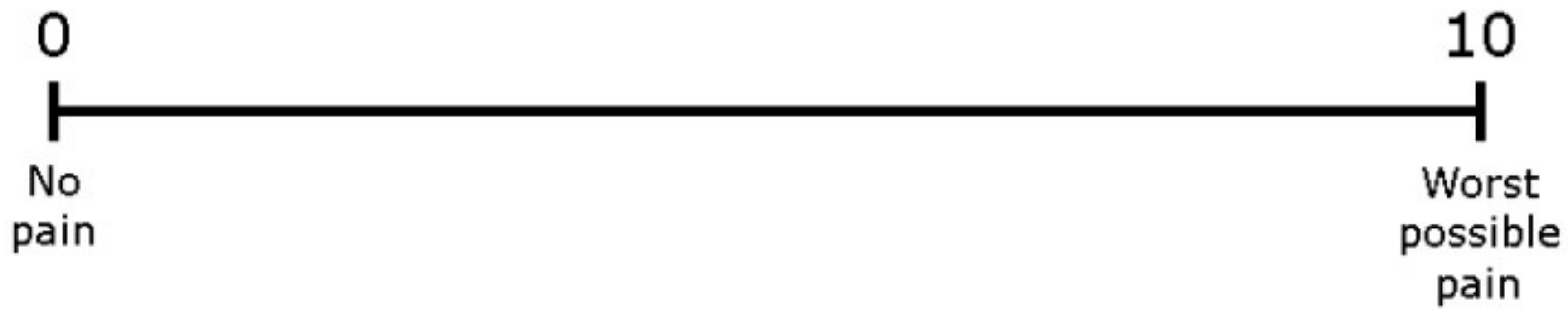
Pain descriptors based on intensity ratings by patients

Wall & Melzack's Textbook of Pain, Sixth Edition

Self-report: the gold standard of pain measurement

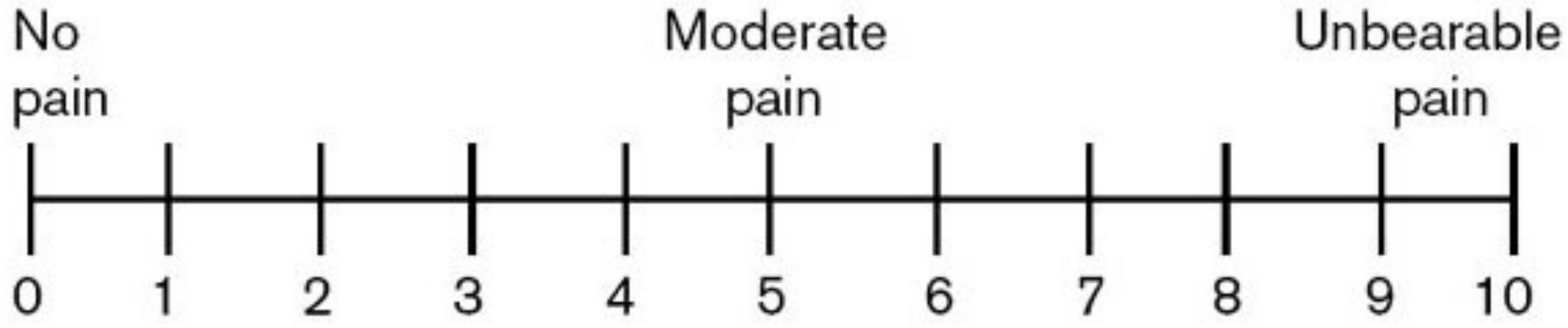
VAS

Visual Analog Scale



NRS

Numerical Rating Scale

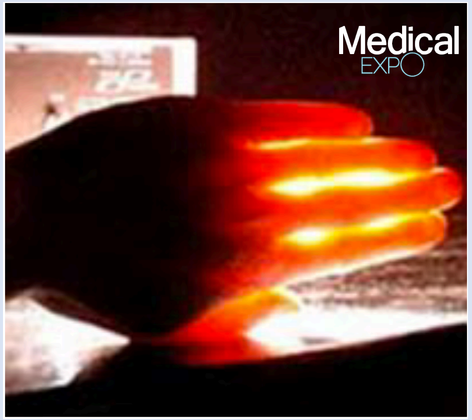


Overview of Methods Informing on Brain Function: Functional Near Infrared Spectroscopy (fNIRS)

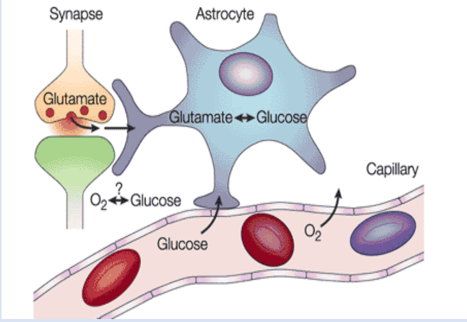
fNIRS device

fNIRS is a portable, optical imaging technology that measures hemodynamic response in the brain. Subjects wear a band/hat of sensors that project + detect light enabling real-time, changes in oxy- and deoxy-hemoglobin during rest, stimulation or task performance.

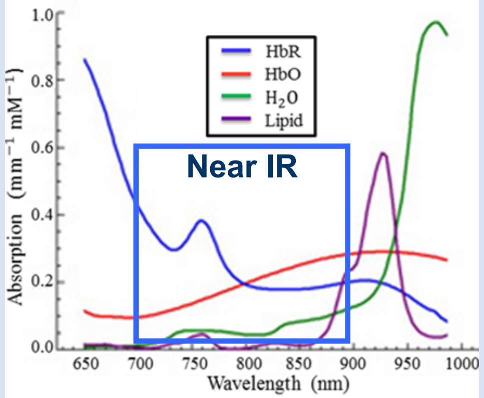
1. Red and near infrared light penetrate biological tissues



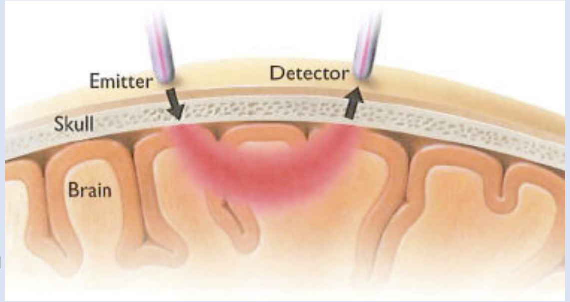
2. Oxy- to deoxy-Hemoglobin are correlated with brain activity through O_2 consumption by neurons.



3. Oxy- to deoxy-Hb are the main absorbers of light within near infrared spectra.



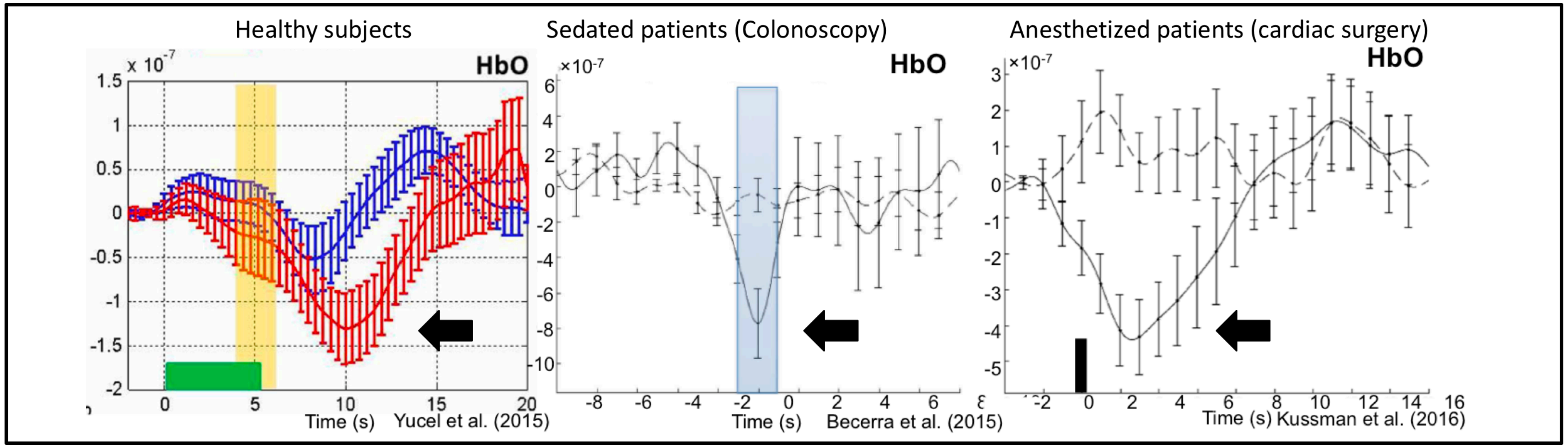
4. When a light source is applied in the brain, cortical regions of activation or deactivation can be observed based on extracted Oxy/Deoxy Hb concentration changes.

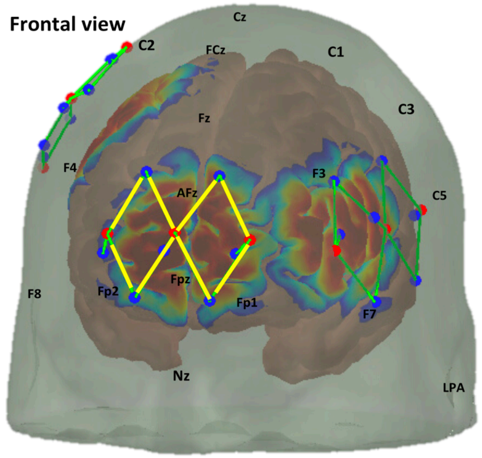


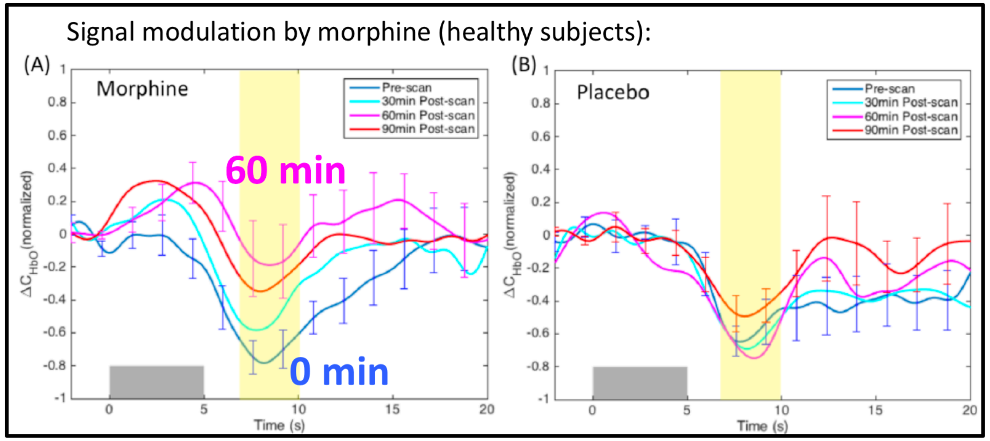
5. Broad + safe application of fNIRS towards mapping brain functions.

Pain and Analgesic Response Measured with fNIRS

Deactivation over the anterior prefrontal cortex in healthy volunteers and surgical patients

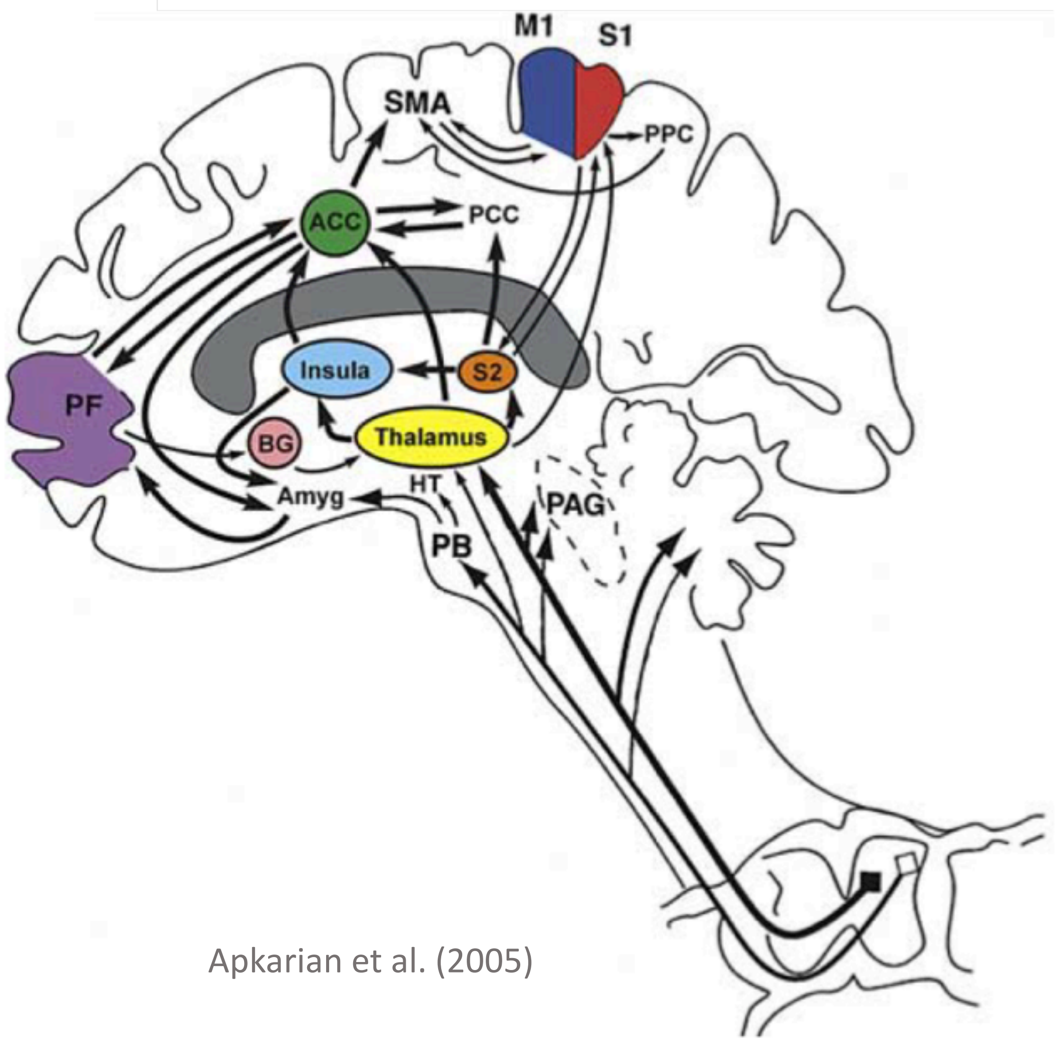






Study arms consists of N=10-20 healthy subjects or patients.

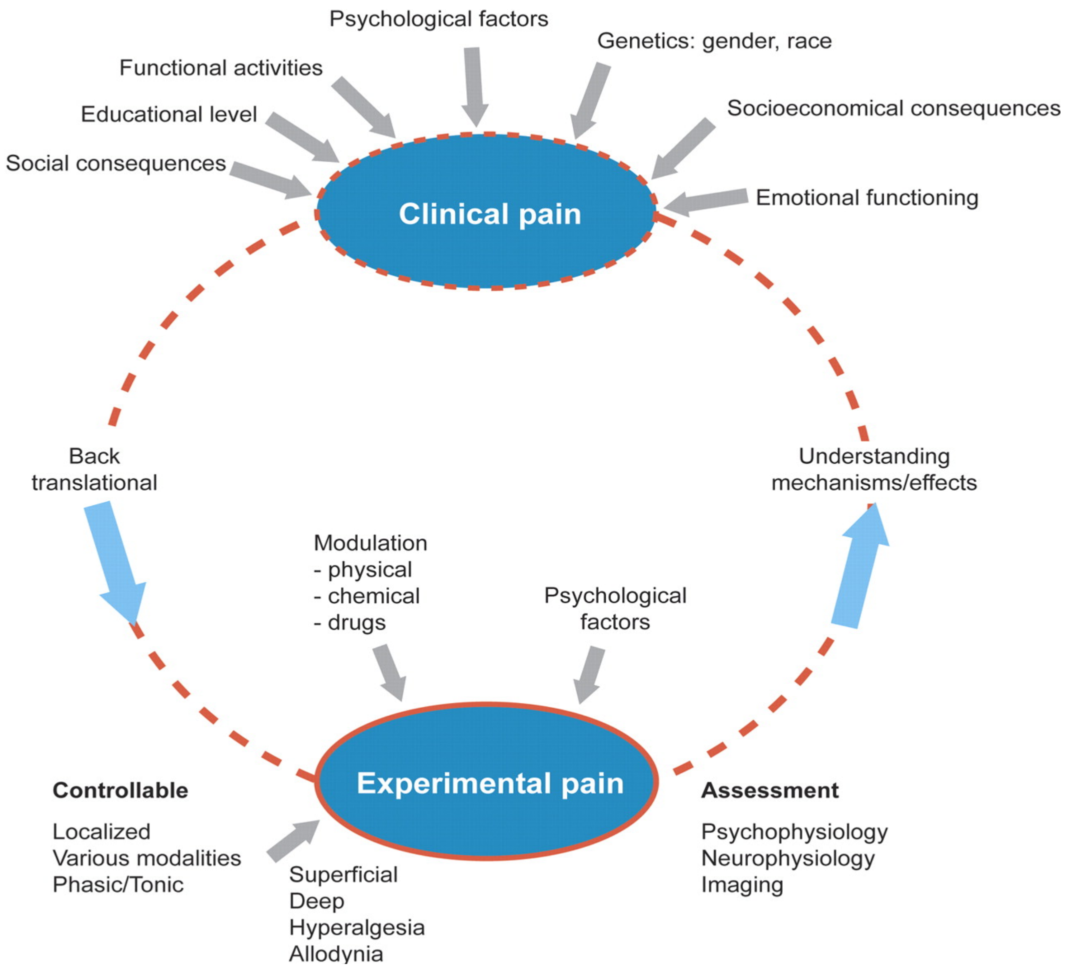
Pain pathways and the medial prefrontal cortex



The medial prefrontal cortex has been suggested to be a supramodal region that integrates the processing of sensory, emotional and attentional information of pain.

Apkarian et al. (2005)

Experimental pain elicitation methods and human models of pain



Experimental pain models are important for study of mechanisms which could not be studied in patients but could be standardized and modified for clinical use.

Pain elicitation methods:

- Heat
- Cold
- Electrical
- Chemical
- Mechanical

Pharmacol Rev 64:722–7?9, 2012

fNIRS dataset

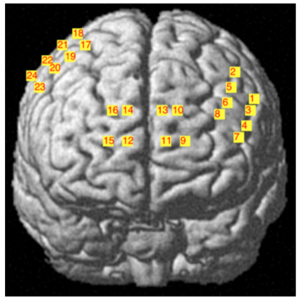
- Collected at Boston Children's Hospital by D Borsook et al.
- 20 healthy participants (right handed, males, ages 19 to 38, mean age 28.7)
- fNIRS: 24 channels, 690- and 830-nm.
- Two types of stimulus:

Electrical stimulus

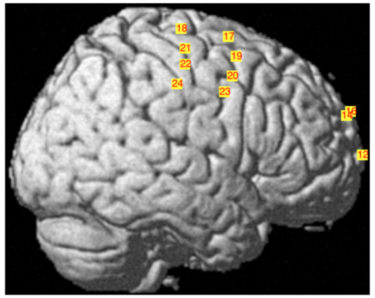
- 38 fNIRS sessions from 20 participants
- 5sec 5Hz electrical stimulus to left thumb
- Two stimulation levels:
  - 3/10 (innocuous)
  - 7/10 (noxious)

Heat stimulus

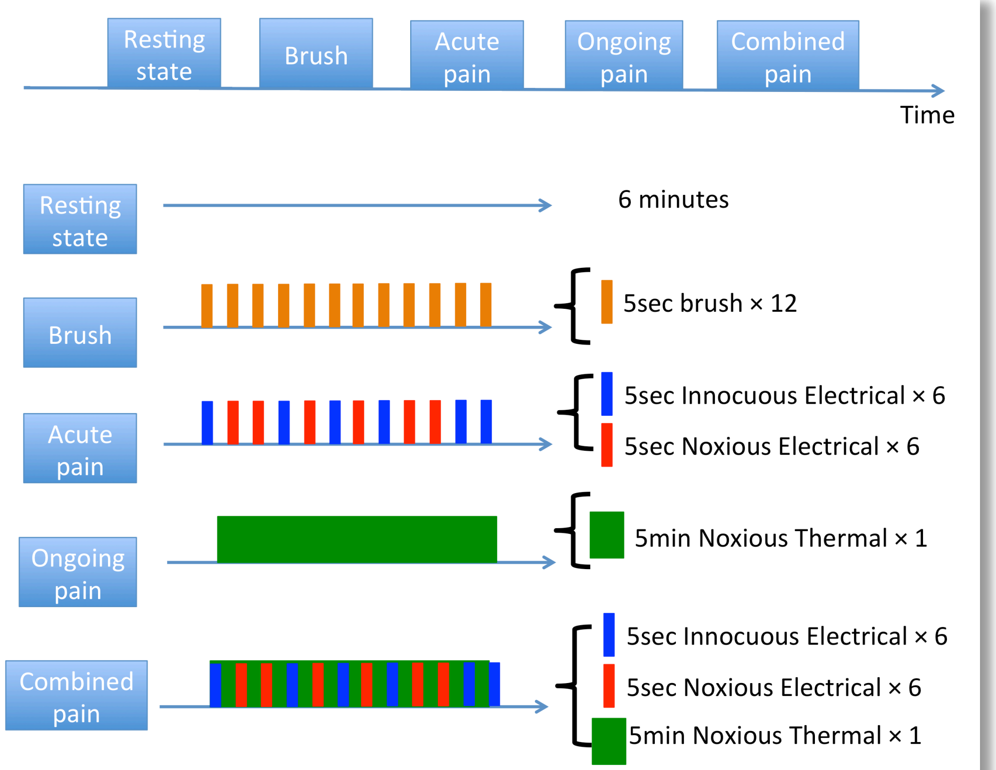
- 10 fNIRS sessions from 10 participants
- 5-min thermal stimulus to left hand
- One stimulation level:
  - 5/10 (moderately painful)



Frontal view

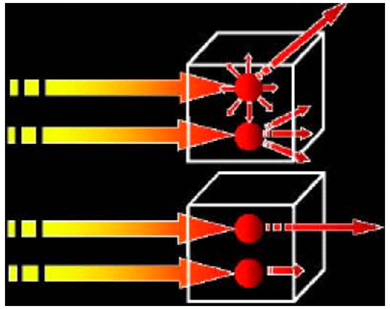


Lateral view



Data preprocessing

Photons that enter tissue undergo different trajectories.



Scattering (cell membranes), assumed to be constant.

Absorption (variations mainly due to Hb concentration changes).

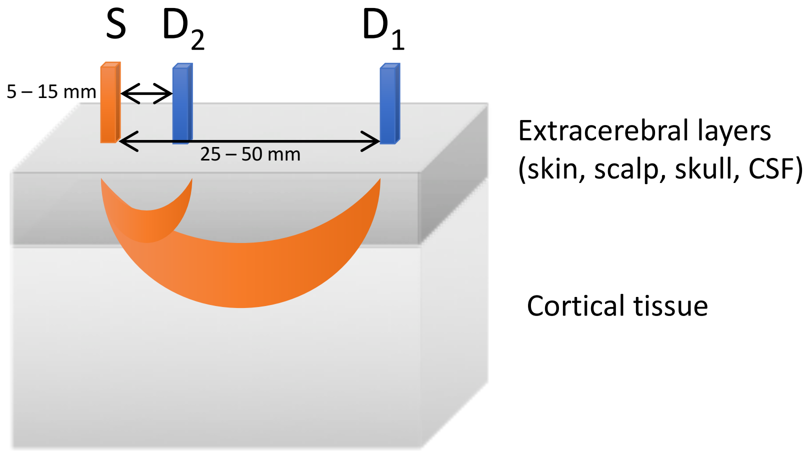
Modified Beer-Lambert Law (MBLL): reconstruction of Hb concentration changes with optical data

$$\Delta OD = d \cdot PPF_{\lambda} \cdot [\varepsilon_{HbO,\lambda}\, \Delta C_{HbO} + \varepsilon_{HbR,\lambda}\, \Delta C_{HbR}]$$

OD: optical density
d: source detector distance
PPF: partial pathlength factor

ε: extinction coefficient
C: Hb concentration
λ: light wavelength

Short distance regression technique



The hemoglobin concentration changes derived from the short distance detector (D_2) may be used to filter out the physiological noise (e.g. heart beat, respiration) from the long distance channel data (D_1) for a more accurate estimation of cortical oxygenation activity.

NIRS data pre-processing pipelines

Raw NIRS data

↓

Optical density change calculation

↓

Motion Artifact identification

↓

Temporal bandpass filtering (0.01-0.50Hz)

↓

Hemoglobin Concentration change calculation

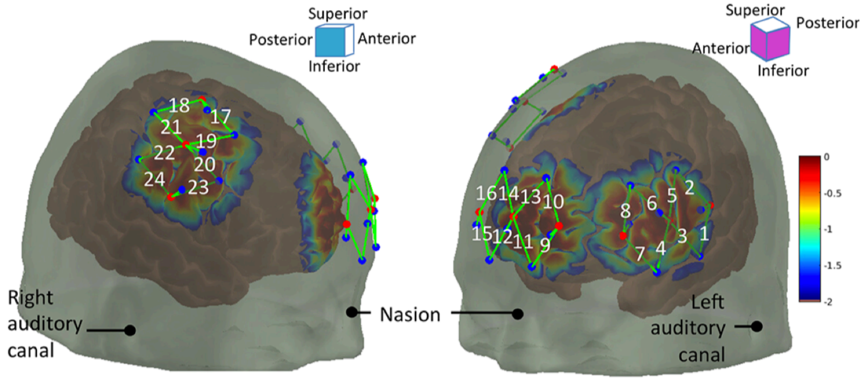
↓

Drift removal — Short distance data regression — Hemodynamic response function estimation

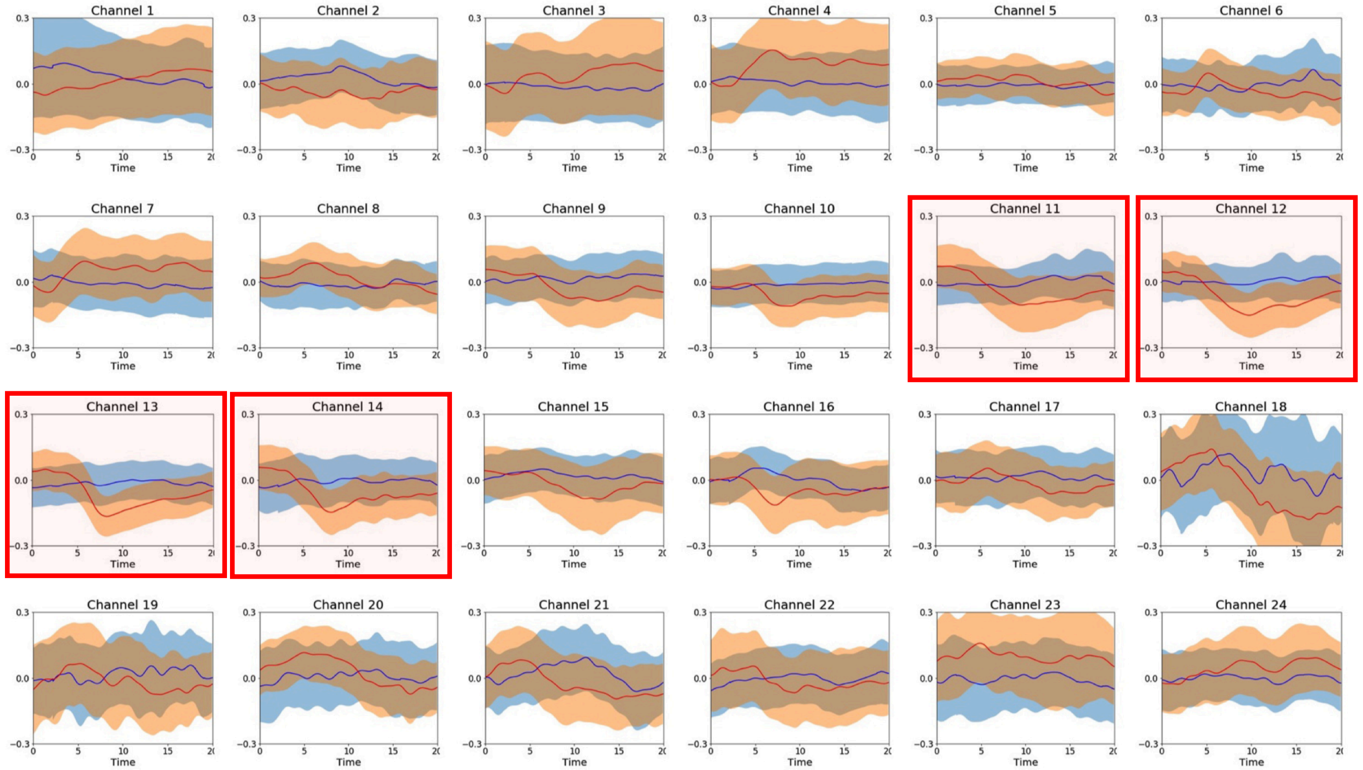
↓

Hemodynamic response function to painful and nonpainful events

fNIRS with electrical stimulus (no pain vs electrical pain VAS7)



Arrangement of the fNIRS optodes and the corresponding sensitivity profile in mm −1 , showing the location of the detectors (blue dots), the sources (red dots) and the 24 channels (green lines).



Shadow thickness indicates 0.5xSDEV. Blue represents no pain. Red represents electrical pain VAS 7.

Features

B-spline coefficients

Coefficients $\mathbf{c} = [c_1, ..., c_{10}]$ such that:

$$s(t) \approx \sum_{k=1}^{K} c_k \phi_k(t) = \boldsymbol{c}^\top \boldsymbol{\phi}$$

where $\boldsymbol{\phi} = [\phi_1, ..., \phi_{10}]$ is the cubic B-spline basis system.

This is achieved by minimizing a penalized squared error cost function (PENSSE)

$$\mathtt{PENSSE} = \underbrace{\sum_{i=1}^{W} [w_i - s(t_i)]^2}_{\mathtt{SSE}} + \lambda \underbrace{\int \left[\frac{d^2 s(t)}{dt^2}\right]^2}_{\mathtt{PEN}_2}$$

Statistical features

These include:

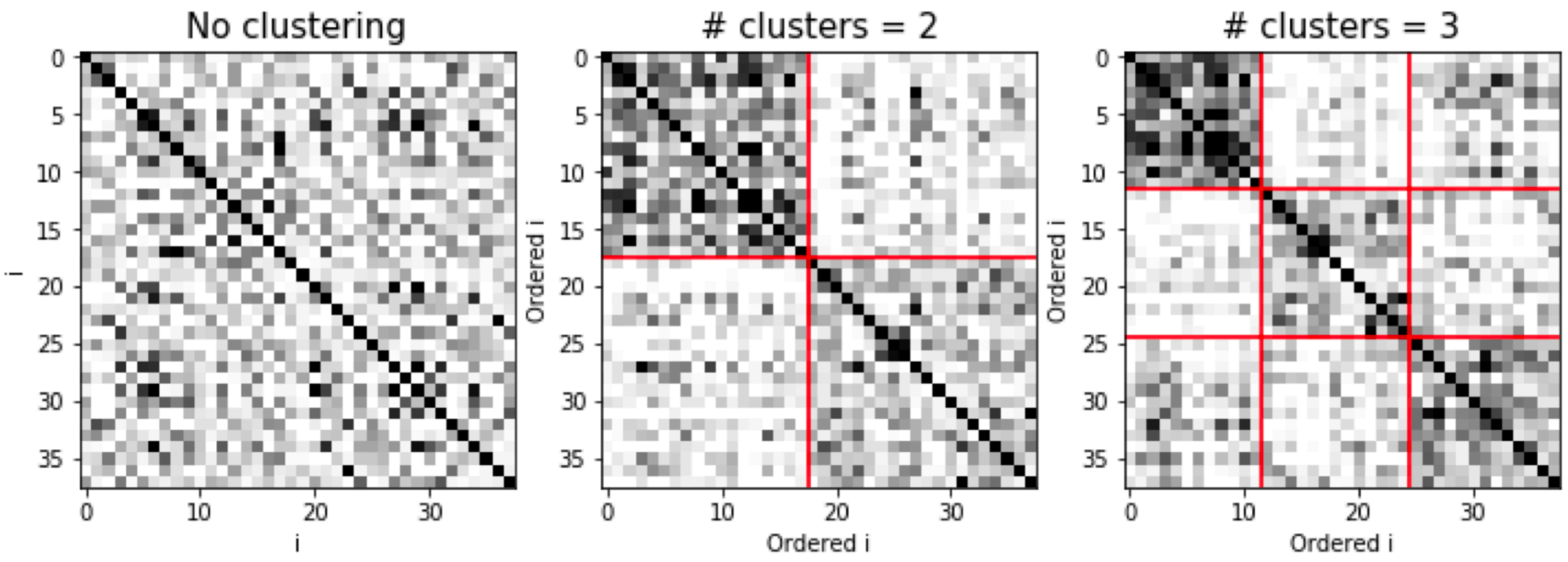
(1) mean
(2) standard deviation
(3) maximum
(4) minimum
(5) range
(6) the slope of the linear regression of w in its time series
(7) the location of maximum in the window w
(8) the location of minimum in the window w
(9) kurtosis
(10) skewness
(11) area under the curve.

Results: single-task learning

CLASSIFICATION PERFORMANCE OF SINGLE-TASK MODELS OVER 10-FOLD CROSS-VALIDATION.

Features & model	Accuracy	Sen.	Sp.	F_1
Spline coefficients				
Logistic regression (L1)	0.68	0.65	0.69	0.57
Logistic regression (L2)	0.71	0.67	0.73	0.60
Linear Lasso	0.69	0.61	0.74	0.57
SVM (linear kernel)	0.70	0.65	0.73	0.59
SVM (rbf kernel)	0.72	0.81	0.67	0.65
Statistics				
Logistic regression (L1)	0.63	0.64	0.62	0.54
Logistic regression (L2)	0.63	0.64	0.63	0.53
Linear Lasso	0.65	0.64	0.66	0.54
SVM (linear kernel)	0.64	0.64	0.64	0.53
SVM (rbf kernel)	0.65	0.76	0.60	0.58
All features				
Logistic regression	0.70	0.67	0.71	0.59
Logistic regression (L2)	0.71	0.72	0.70	0.6
Linear Lasso	0.71	0.70	0.71	0.61
SVM (linear kernel)	0.71	0.72	0.70	0.62
SVM (rbf kernel)	0.71	0.84	0.65	0.66

Inter-subject variability calls for subject clutering



Machine learning model: spectral clustering + MT-MKL

Spectral clustering

Normalized spectral clustering:

1. Generate descriptor vector **p**
2. construct a fully connected similarity graph and the corresponding weighted adjacency matrix **W.**
    1. Use radial basis function.
3. Compute normalized graph Laplacian, **L= I-D^{-1}W**
4. Calculate first **T** eigenvectors $\mathbf{U} \in \mathbb{R}^{S \times T}$
5. Cluster vectors using k-means algorithm

MT-MKL

- MT-MKL combines both concepts:
    - multiple kernel learning to account for different fNIRS channels
    - multi-task learning to personalize the models according to different clusters or tasks
- Each channel is represented by a kernel k_m:

$$k_\eta(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\eta}) = \sum_{m=1}^{M} \eta_m k_m(\boldsymbol{x}_i^{(m)}, \boldsymbol{x}_j^{(m)})$$

- Max-min optimization problem:

$$\underset{\{\boldsymbol{\eta}^{(r)} \in \mathcal{E}\}_{r=1}^{T}}{\text{minimize}} \underbrace{\left\{ \underset{\{\boldsymbol{\alpha}^{(r)} \in \mathcal{A}^{(r)}\}_{r=1}^{T}}{\text{maximize}} \ \Omega(\{\boldsymbol{\eta}^{(r)}\}_{r=1}^{T}) + \sum_{r=1}^{T} J^{(r)}(\boldsymbol{\alpha}^{(r)}, \boldsymbol{\eta}^{(r)}) \right\}}_{\mathcal{O}_\eta}$$

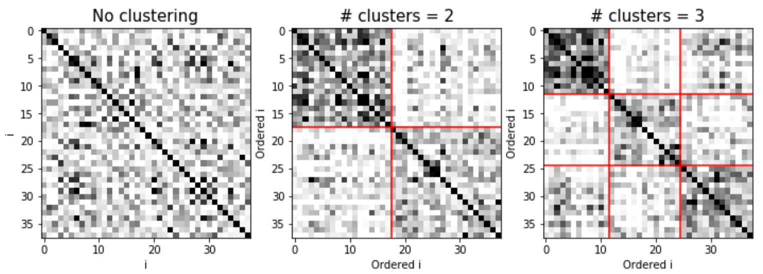
- Two regularizers:

$$\Omega_1(\{\boldsymbol{\eta}^{(r)}\}_{r=1}^{T}) = -\nu \sum_{r=1}^{T} \sum_{s=1}^{T} \boldsymbol{\eta}^{(r)\top} \boldsymbol{\eta}^{(s)} \qquad \Omega_2(\{\boldsymbol{\eta}^{(r)}\}_{r=1}^{T}) = -\nu \sum_{r=1}^{T} \sum_{s=1}^{T} ||\boldsymbol{\eta}^{(r)} - \boldsymbol{\eta}^{(s)}||_2$$

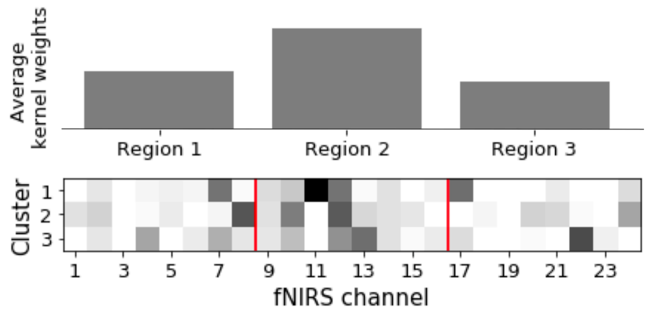
Personalization in the fNIRS heat dataset using mutitask multiple kernel machines

Data and features
- Classification between 7/10 and baseline.
- Features extracted from 20sec HbO windows.
- Two types of features:
  - B-splines
  - Statistical

Multitask multiple kernel learning (MT-MKL)
MT-MKL combines:
- multiple kernel learning to account for different fNIRS channels
- multi-task learning to personalize the models according to different clusters or tasks



Subject profiling using spectral clustering



The model is able to learn the importance of the channels

MT-MKL model			B-spline coefficients				Statistical features				All features			
T	Kernel	Reg	ACC	Sen.	Sp.	F1	ACC	Sen.	Sp.	F1	ACC	Sen.	Sp.	F1
1	Linear	L1	0.68	0.74	0.66	0.60	0.65	0.68	0.63	0.56	0.72	0.77	0.71	0.64
	Linear	L2	0.70	0.71	0.70	0.61	0.66	0.69	0.64	0.56	0.72	0.75	0.70	0.63
	RBF	L1	0.74	0.75	0.74	0.65	0.70	0.77	0.67	0.62	0.74	0.76	0.74	0.66
	RBF	L2	0.73	0.79	0.71	0.66	0.69	0.72	0.68	0.61	0.74	0.76	0.74	0.65
2	Linear	L1	0.74	0.77	0.72	0.66	0.65	0.70	0.64	0.56	0.71	0.71	0.71	0.62
	Linear	L2	0.74	0.75	0.73	0.65	0.64	0.65	0.63	0.54	0.71	0.73	0.70	0.62
	RBF	L1	0.77	0.79	0.75	0.69	0.70	0.76	0.67	0.62	0.72	0.79	0.69	0.65
	RBF	L2	0.76	0.76	0.75	0.68	0.70	0.77	0.67	0.63	0.73	0.73	0.73	0.64
3	Linear	L1	0.75	0.76	0.74	0.66	0.68	0.70	0.67	0.59	0.75	0.75	0.75	0.66
	Linear	L2	0.75	0.79	0.74	0.68	0.69	0.67	0.70	0.59	0.76	0.76	0.75	0.67
	RBF	L1	**0.80**	0.83	0.78	0.73	0.72	0.78	0.68	0.65	0.76	0.81	0.74	0.69
	RBF	L2	0.79	0.85	0.76	0.72	**0.73**	0.75	0.72	0.64	**0.77**	0.81	0.75	0.70

Performance of the MT-MLK models for different number of clusters.

# clusters	Cluster	Acc.	Sen.	Sp.	F1
NC (T=1)	1	0.74	0.75	0.74	0.65
T=2	1	0.81	0.87	0.77	0.74
	2	0.67	0.77	0.61	0.60
T=3	1	0.87	0.94	0.83	0.82
	2	0.73	0.82	0.69	0.66
	3	0.80	0.88	0.77	0.74

Model performance for each cluster.

Thanks!

dlmocdm@mit.edu