Searchlight Based Feature Extraction for Brain Imaging

NCSP Research Seminar

Shahar Jamshy

Supervisors:
Prof. Nathan Intrator & Prof. Talma Hendler

In collaboration with the Functional Brain Center,
Tel Aviv Sourasky Medical Center

Brain Imaging Modalities

- MRI, fMRI, PET, SPECT, EEG, ECoG, MEG
- Many features
  - 3D images
  - High resolution (spatial or temporal)
- Few observations
  - Subject/patient availability
  - Limited experiment time
- Noise
  - Measuring equipment
  - Individual differences

[Beutel et al. 2000]

Computer Science and Brain Imaging

- Image reconstruction
- Noise reduction
- Registration
- Normalization
- Inference
  - Quantify and interpret pathologies and anomalies
  - Quantify and interpret brain states

  - From statistical inference to prediction

[Friston, 2006]

Machine Learning in Brain Imaging

- Difficult – many features, few observations
- Challenge – find only the relevant features
- My proposal
  - Utilize domain specific knowledge to select features
  - Combine local and global modeling
- Preliminary work
  - Feature extraction using spherical searchlights and local SVM modeling.
  - Holistic fMRI classification framework
  - Improvement over state-of-the-art methods

Can we predict?

- Is the subject seeing a sentence or a picture?
- Which of several categories of words or pictures is a subject seeing?
- Is the subject reading an ambiguous sentence?
- Will the subject answer correctly?
- What is the orientation of a stimulus visual grating?
- Is there a face/music/tools/… in a film clip being seen?
- What is the subject perceiving?
- Is the subject concealing information?

[Mitchell et al 2004, Haynes 2006, Norman 2006]

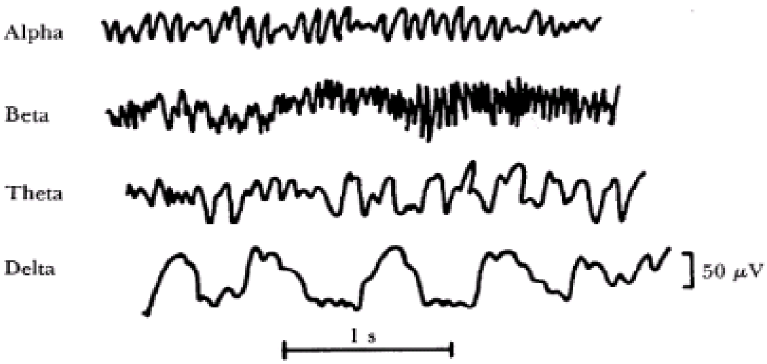
Questions asked

- **Where** in the brain is the information represented?
- **When** does the brain represent the information?
- **How** is the information represented in the brain?

- For a single subject
- For a group of subjects
- Between groups of subjects
  - Patients vs. Healthy
  - Trained vs. Untrained
  - Different prior condition

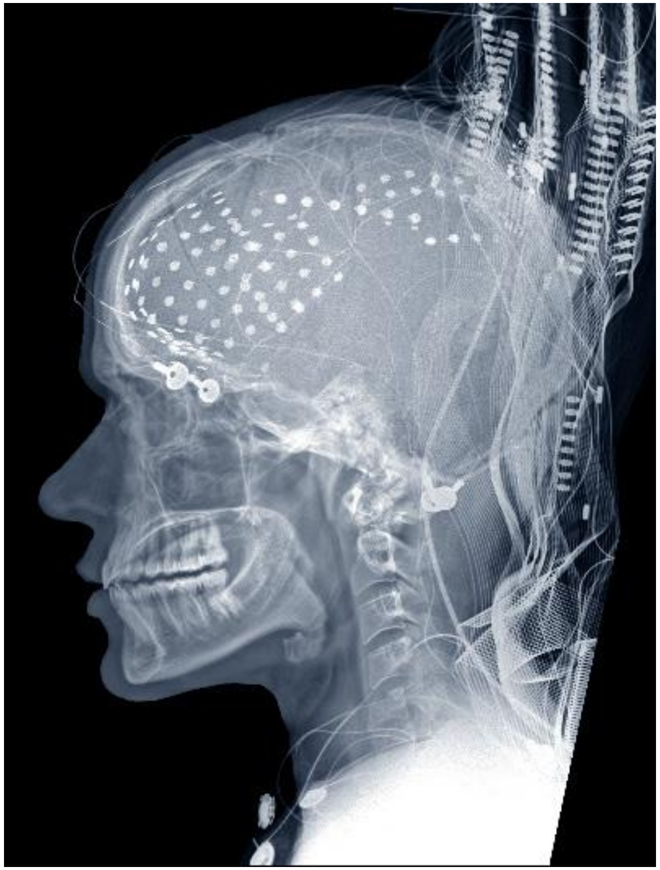
Electroencephalography (EEG)

- Hans Berger, 1924
- Electric activity measured on scalp
  - Amplitude ~100μV
  - Frequencies: 0.5-1kHz
- Different frequencies believed to represent underlying neural states
- Advantages
  - Cheap (~50K$ reseatch, ~500$ consumer)
  - High temporal acuity (1kHz)
- Disadvantages
  - Low spatial acuity (~cm)
  - Very noisy



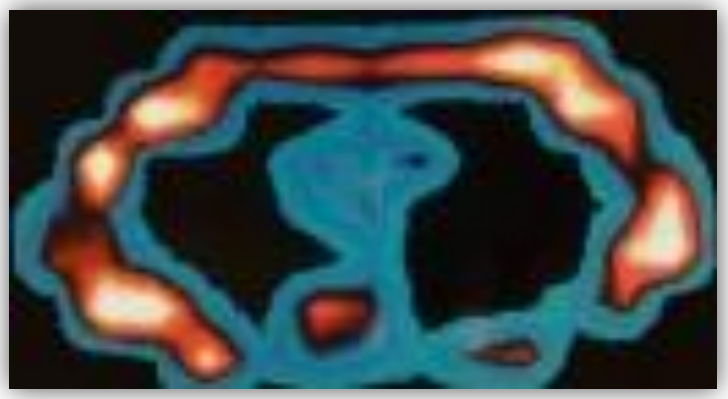
Electrocorticography (ECoG)

- Electric activity measured on the cortex
- Epileptic patients undergoing pre-surgical evaluation
- Advantages
  - High temporal acuity (1kHz)
  - Medium spatial acuity (~10mm)
  - Low noise
- Disadvantages
  - Invasive - low availability
  - Electrode location based on clinical considerations



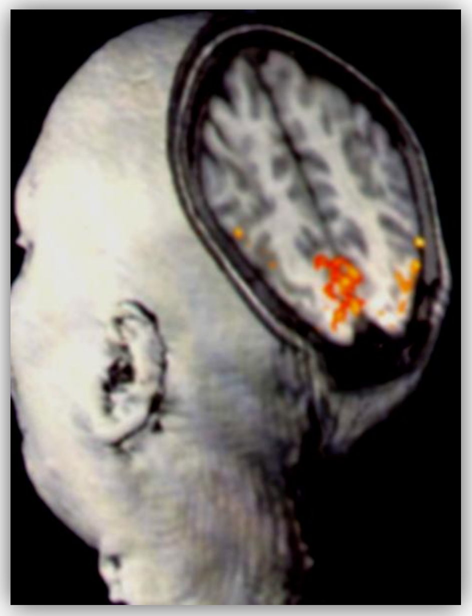
Magnetic Resonance Imaging (MRI)

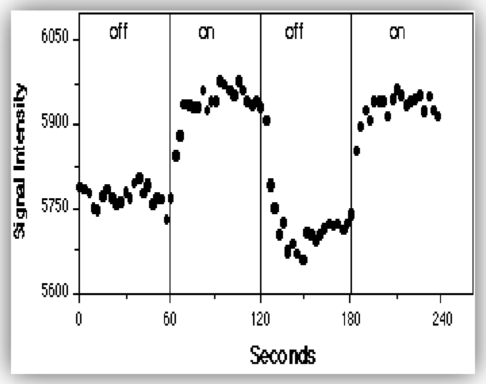
- Raymond Damadian, 1977
- Resonate hydrogen atoms in the body
- Reveals soft tissue structures
- Signal depends on organization of hydrogen
- In the brain – CSF, gray matter, white matter
- Very high spatial acuity (~mm)



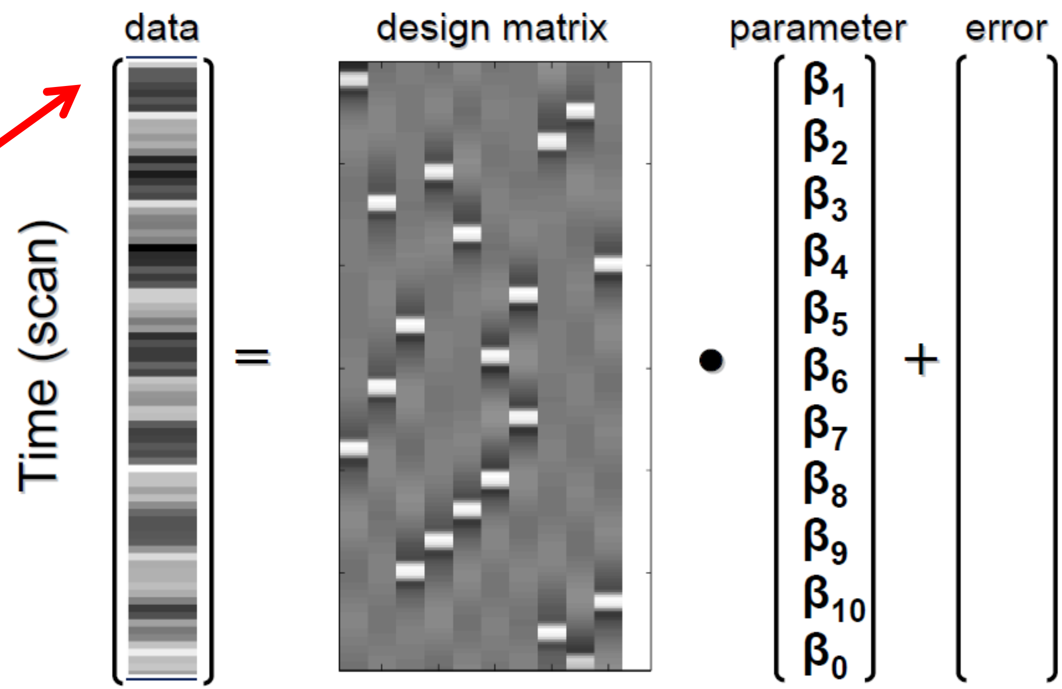
Functional MRI (fMRI)

- John Belliveau, 1991
- Small changes in MR signal caused by blood oxygenation level (BOLD)
- Advantages
  - High spatial acuity (~3mm)
- Disadvantages
  - Low temporal resolution (~2 sec.)
  - Expensive (~3M$)
  - Very Noisy (3% of MR signal strength)





General Linear Model (GLM)

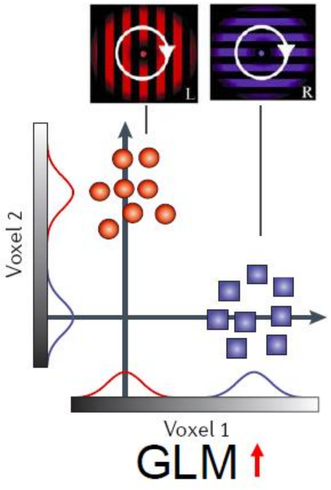


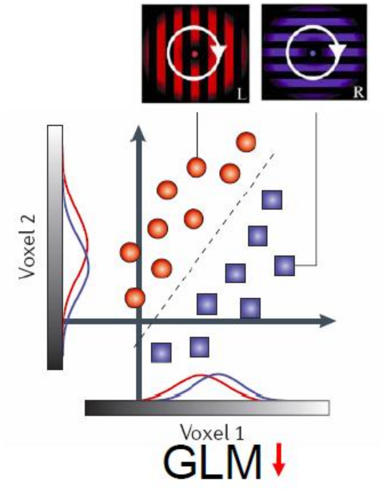
$$y = X \bullet \beta + \varepsilon$$

GLM: separate model fitting for each voxel mass-univariate analysis!

Limitations of GLM

- Multiple comparisons (60k comparisons)
  - FDR or other FWER eliminate most results
- Individual differences
- Cognitive/Sensori-motor states can be expressed in the brain as ***distributed patterns of brain activity***





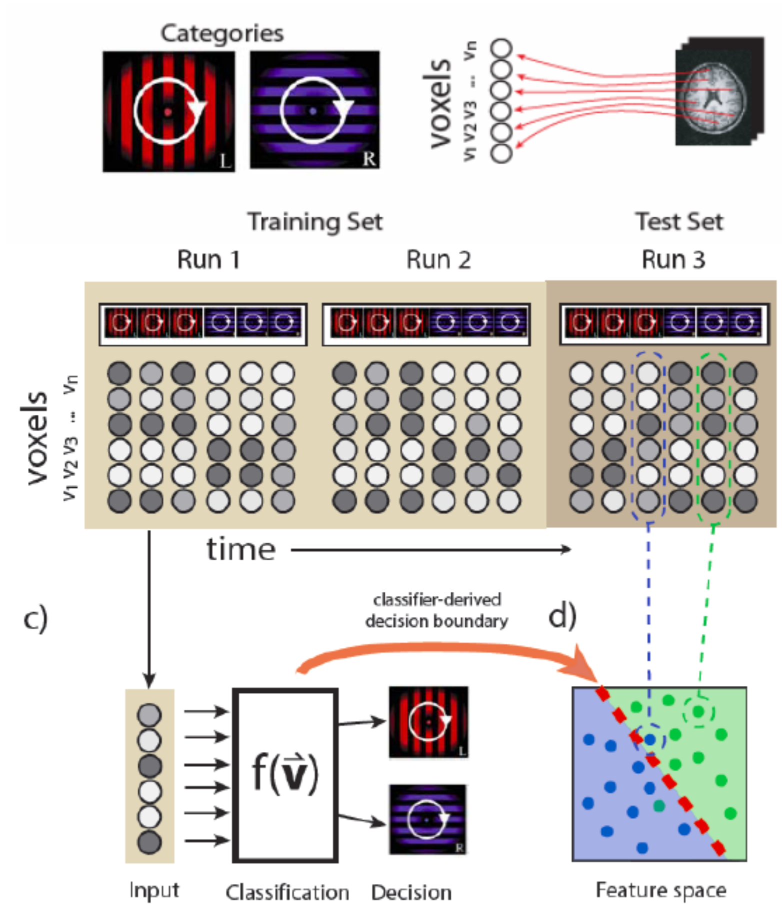
Multivariate pattern analysis (MVPA)

- Take advantage of the information contained in activity patterns from multiple features
- Increase in sensitivity: weak information in single voxels is accumulated
- Multiple regions/voxels may only carry info when jointly analyzed
- Can prevent information loss due to spatial smoothing
- Can preserve temporal resolution and prevent averaging

[Haxby et al. 2001, Mitchell et al. 2004, Mur et al 2009]

Processing Stream

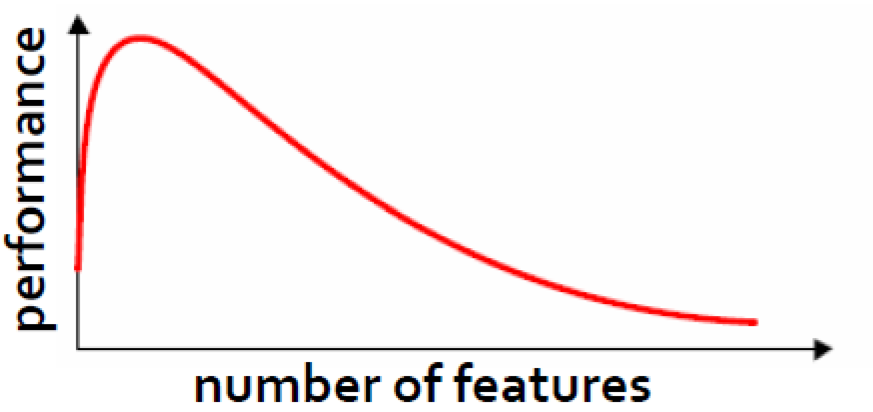
- Acquire data
- Preprocess data
- Select features
- Generate patterns
- Label patterns
- Train the classifier
- Validate/Cross-validate
- Statistical inference



[Heynes & Rees 2005]

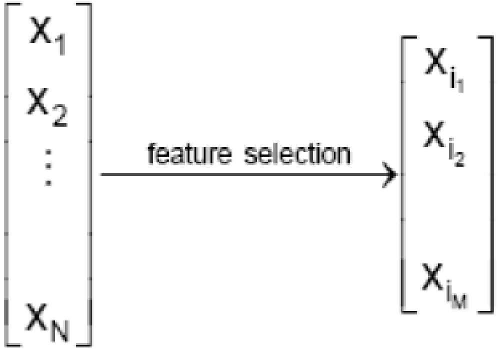
Feature Reduction – The Problem

- Data is typically sparse, high-dimensional and noisy
- Number of observations is low
- Classification is sensitive to information content in all voxels
- Many uninformative voxels = poor classification (due to over fitting)

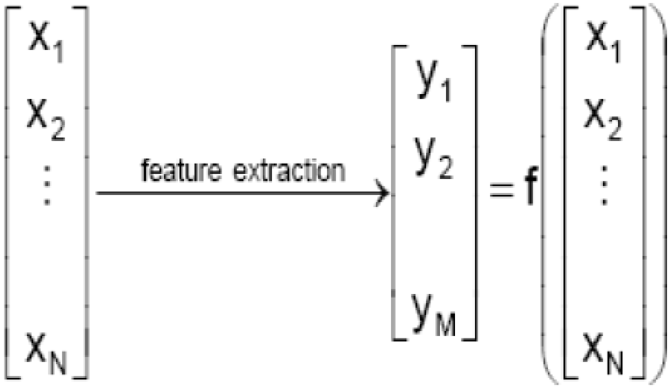


Feature Reduction – Solutions

- **Solution 1: Feature selection**
  - Select subset from all available features
  - Original features remain unchanged



- **Solution 2: Feature extraction**
  - Create new features as a function of existing features
  - Linear functions (PCA, ICA,…)
  - Nonlinear functions (hidden units in a neural network)
  - Domain specific methods



[Guyon & Elisseefi, 2003]

Feature Selection in Brain Imaging

- **‘External‘ Solutions**
  - Anatomical regions of interest
  - Independent functional localizer
- **‘Internal‘ univariate solutions**
  - Activation vs. baseline (t-Test)
  - Mean difference between conditions (ANOVA)
  - Single voxel classification accuracy
- **‘Internal’ multivariate solutions**
  - Searchlight classification

[Norman et al. 2006, Pereira et al. 2009]

Feature Selection in Brain Imaging

method	number of voxels					
	100	200	400	800	1000	all
accuracy	0.81	0.81	0.75	0.73	0.74	0.65
searchlight	0.81	0.82	0.82	0.77	0.79	0.65
activity	0.79	0.80	0.77	0.73	0.74	0.65
ANOVA	0.77	0.75	0.75	0.73	0.71	0.65

Pereira et al. (2009)

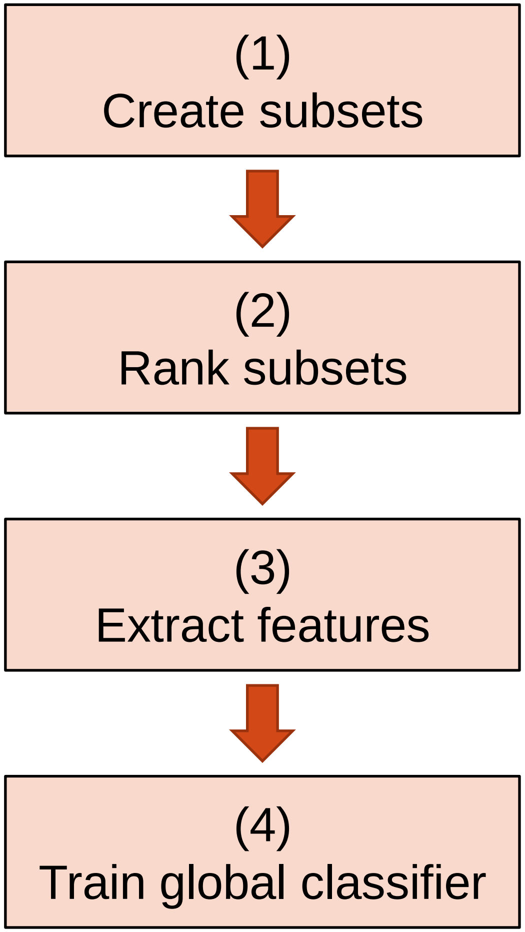
Contribution

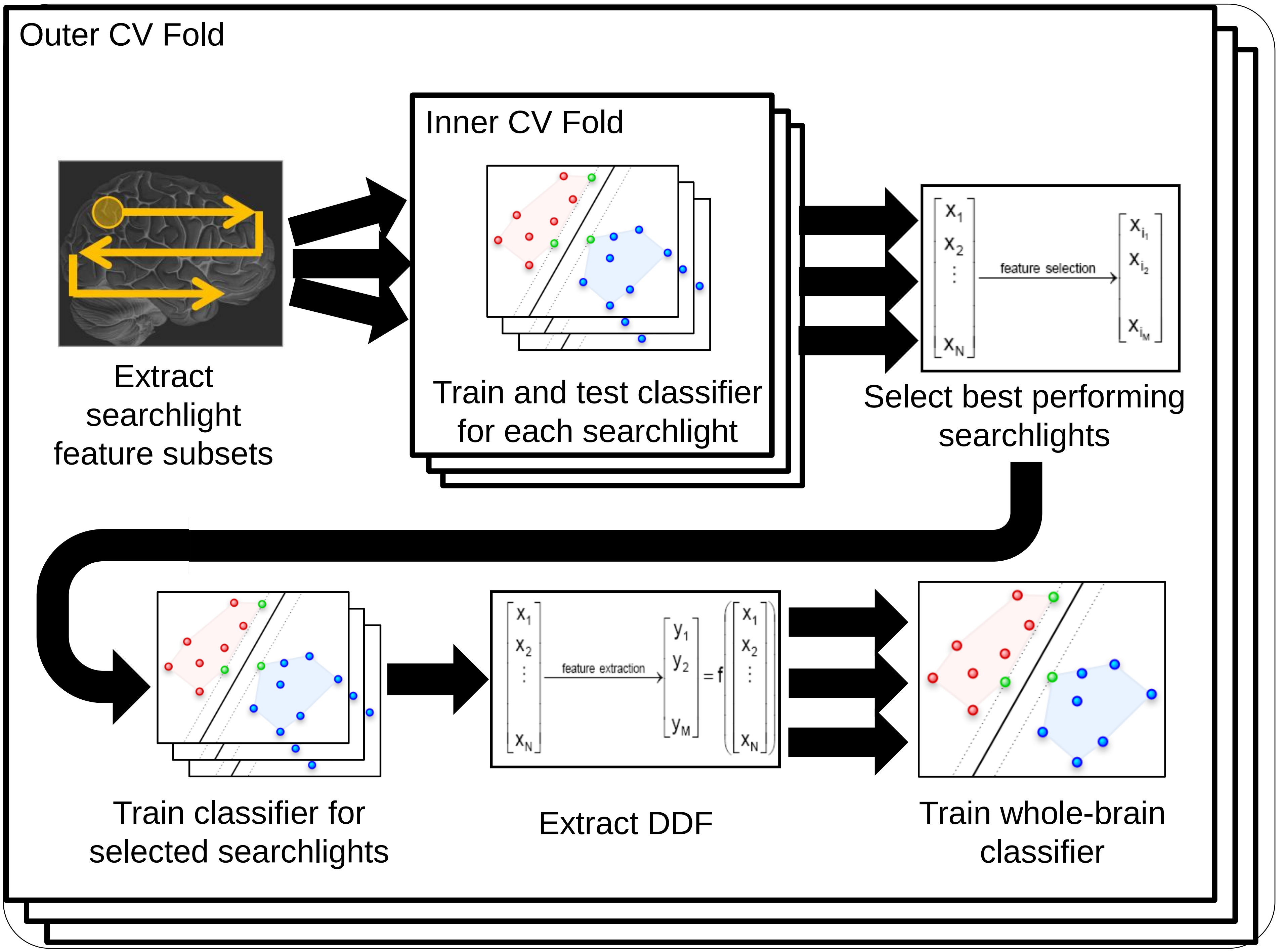
- Formalize the MVPA framework
- Select features based on local modeling when number of observations is low
- Introduce a feature extraction method based on local modeling
- Improved whole-brain classification results on challenging fMRI data

Suggested Framework

Given the set of features $F = \{v_i\}$ we consider the supervised learning task $\{\bar{x}_k, y_k\}$ where $\bar{x}_k = \{x_{k,v} \mid v \in F\}$

1. Create $S = \{S_j \subseteq F\}$, a set of subsets and the observation vectors $\bar{x}_{k,S_j} = \{x_{k,v} \mid v \in S_j\}$
2. Find a function $r : \|S\| \rightarrow \Re$ based on the classification performance of $\{\bar{x}_{k,S}, y_k\}$ and select the best subsets $\hat{S} = \{S_j \mid r(j) > \hat{r}\}$
3. For each S_j create a feature extraction operator $f_j : S_j \rightarrow \Re^{\hat{n}}$ where $\hat{n} << \|S_j\|$
4. Create a new learning problem $\{\hat{x}_k, y_k\}$ where $\hat{x}_k = \{f_j(\bar{x}_{k,S_j}) \mid S_j \in \hat{S}\}$

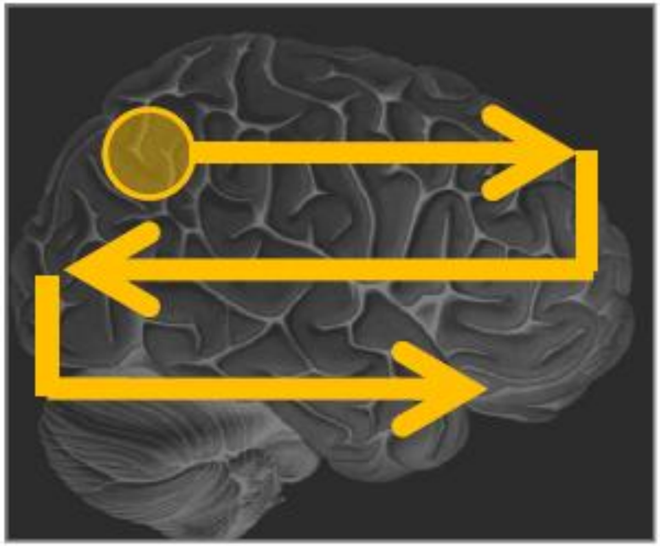




Outer CV Fold
Inner CV Fold
Extract searchlight feature subsets
Train and test classifier for each searchlight
feature selection
Select best performing searchlights
Train classifier for selected searchlights
feature extraction
Extract DDF
Train whole-brain classifier

Create Subsets – Searchlight Analysis

- Classification is performed on a voxel and its (spherical) neighborhood
- Classification accuracy is assigned to centre voxel
- Searchlight is moved across entire dataset to obtain accuracy estimates for each voxel
- Is used for feature selection or to generate a brain map of p-values
- Given a distance on the features $d : F \times F \rightarrow \Re^{+}$ we define $S_j^{SL} = \{v \in F \mid d(v, v_j) < \hat{d}\}$

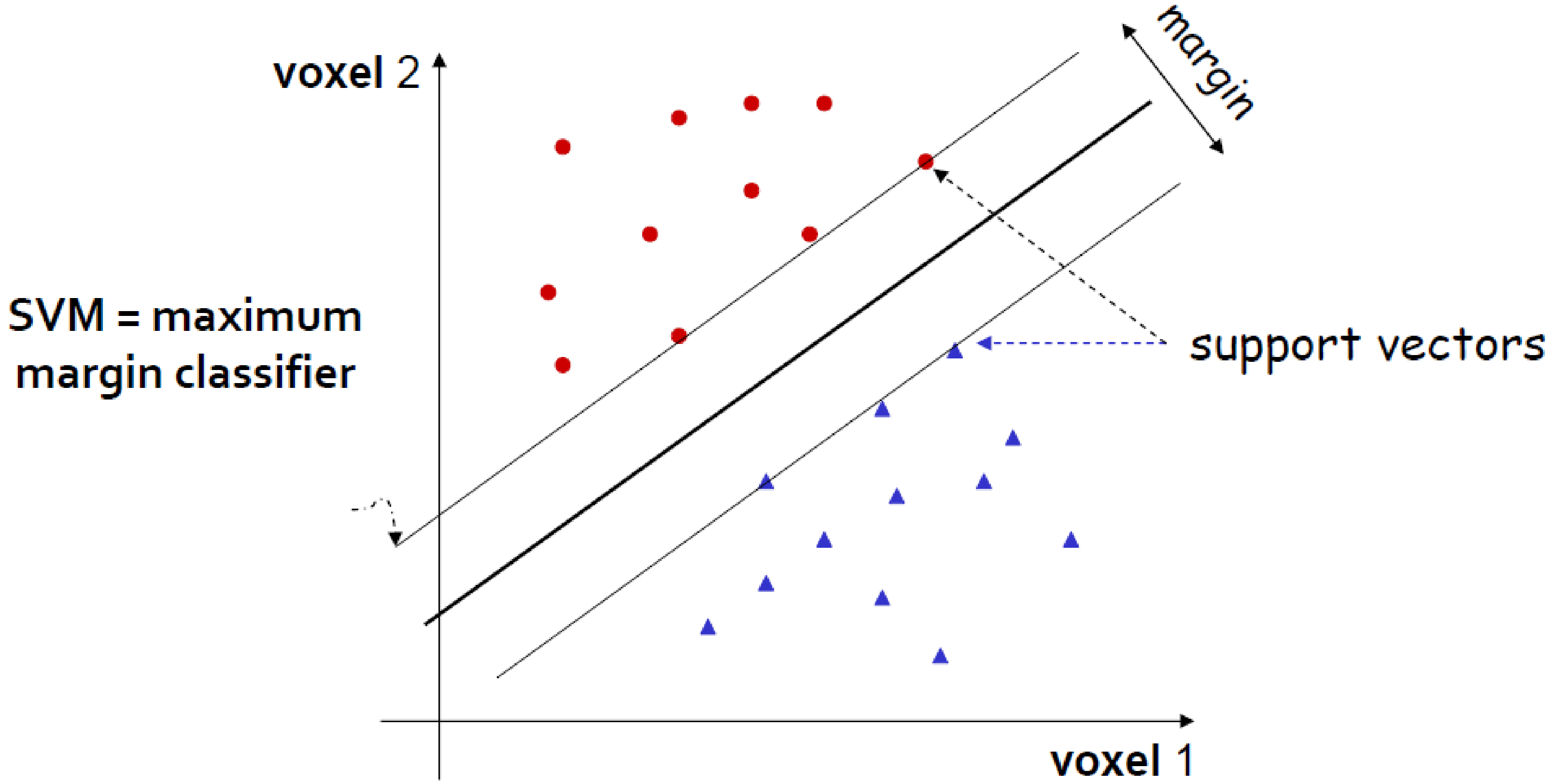


[Kriegeskorte et al. 2006]

Subset Ranking

- Use local modeling with an inner cross-validation tier
- Use a balanced leave-k-out scheme
- In each round:
  1. Randomly divide dataset into groups of k, each group contains an equal number of observations from each label
  2. For each group: train on all other groups, test with remaining group
- Use overall performance of several rounds to rank each subset

Support Vector Machine (SVM)



If classes have overlapping distributions, SVM's account for misclassification errors by introducing additional slack variables

[Vapnik, 1998]

Extract Features – Discriminating Distance Features (DDF)

- Given $\{\bar{x}_{k,S_j}, y_k\}$ many classifiers find a discriminating hyper-plane:

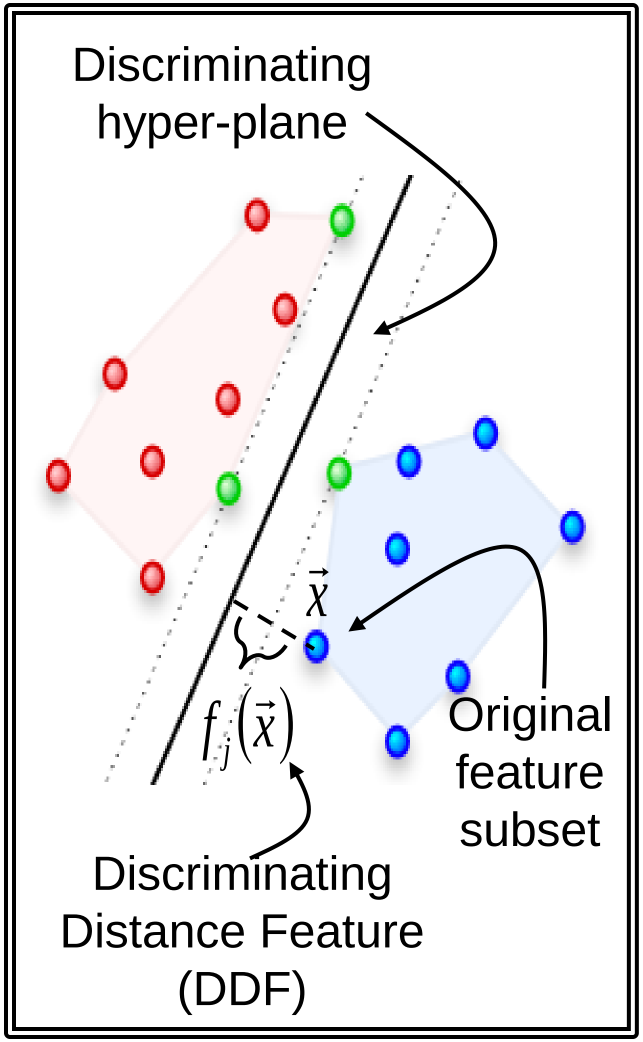
$$\bar{w}_j \cdot \phi(\bar{x}) - b_j = 0$$

- For a new observation $\bar{x}_N$ the model will classify:

$$y_N = sign(\bar{w}_j \cdot \phi(\bar{x}_N) - b_j)$$

- DDF operator is defined by:

$$DDF_j(\bar{x}_N) = \bar{w}_j \cdot \phi(\bar{x}_N) - b_j$$



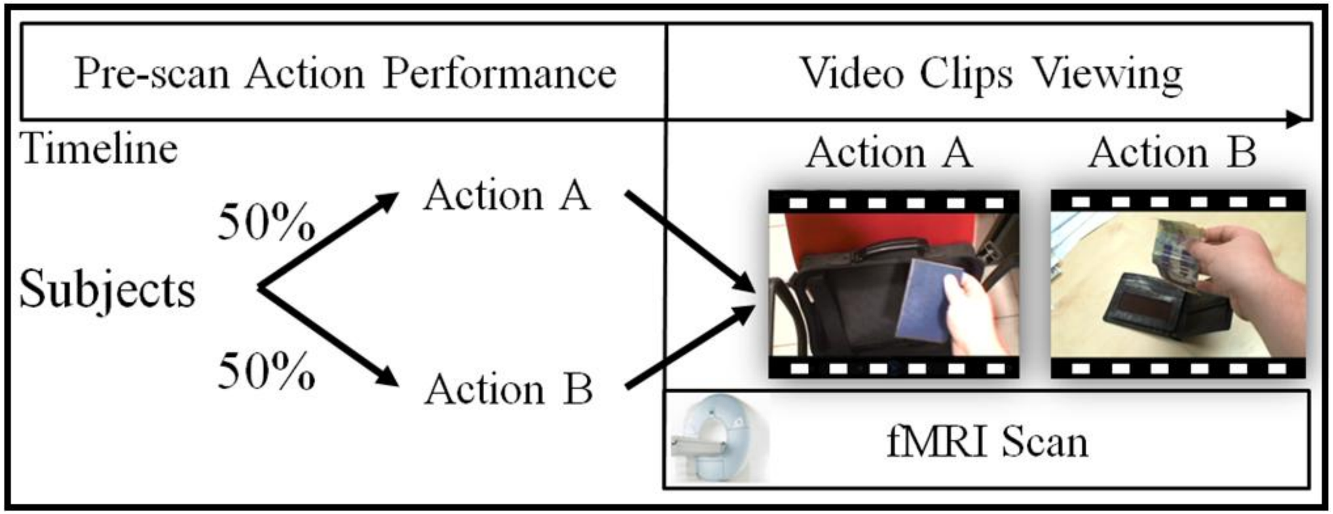
Related methods

Method	Realization	Pros/Cons
Single voxel accuracy feature selection	Searchlight with radius of 0	Low features/ Not multivariate
Searchlight center feature selection	Extraction operator selects central voxel	Low features/ No use of all searchlight data
Searchlight full feature selection	Extraction operator selects all voxels	Use of all searchlight data/ High features

Experimental setup

- In collaboration with Omri Perez
- Aim – to identify the neural trace of recently performed actions
- A neural trace of our prior recent experience is necessary for the sense of reality and continuity in our life.
- Ongoing debate whether neural marking of prior experience is mediated by action perception or cognitive-memory systems.
- Means – fMRI study that shows subjects clips of recently performed actions

Experimental Setup – cont.



- 36 movies
  - 18 for each action type, 6 for each of the 3 sub actions
  - 9 seconds each with short fixation screen periods of 6-9 seconds
  - Due to inattention/adaptation only the first two clips of each type are included in analysis.
- 26 subjects participated, 1 removed due to excessive movement

Preprocessing

- Preprocessing (Brain Voyager QX)
  - Time and motion correction
  - Detrending
  - Talairach normalization
  - Gaussian smoothing

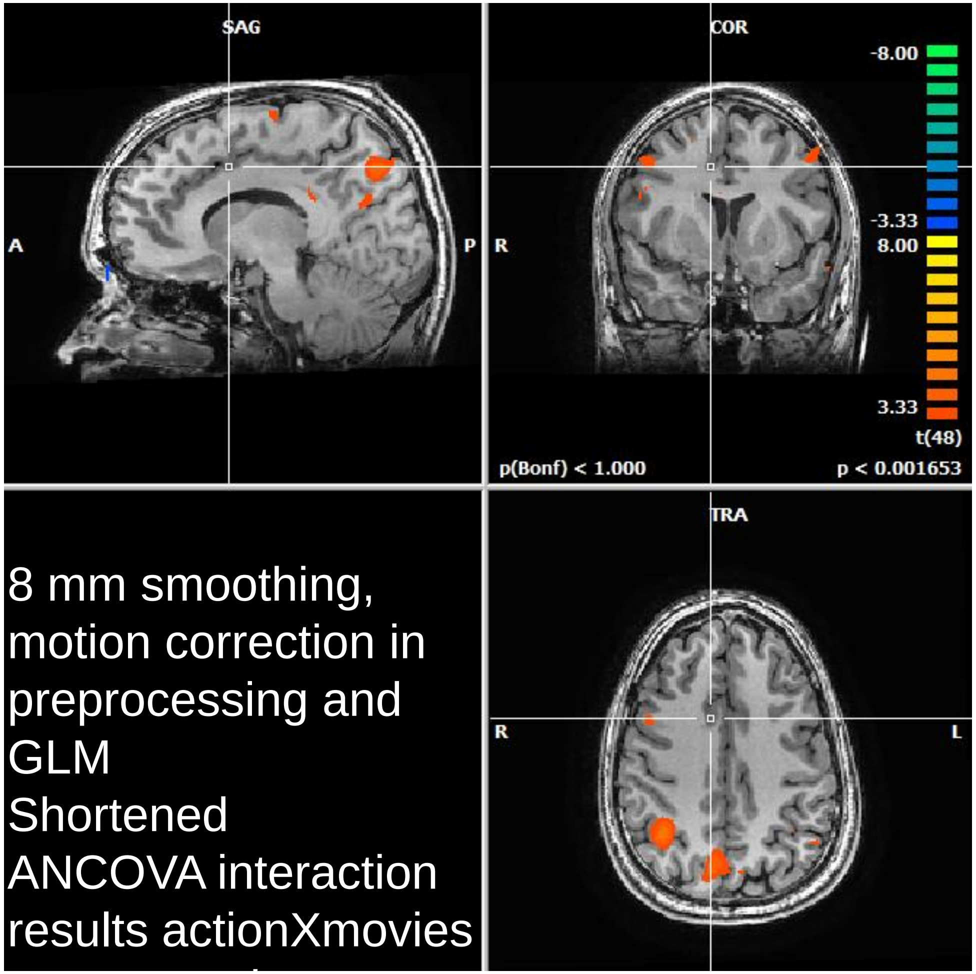
- Calculate difference in response between viewing a clip of task A and a clip of task B
  - Correlate the data with the appropriate regressors
  - Take into account the hemodynamic response

Experimental Setup - Summary

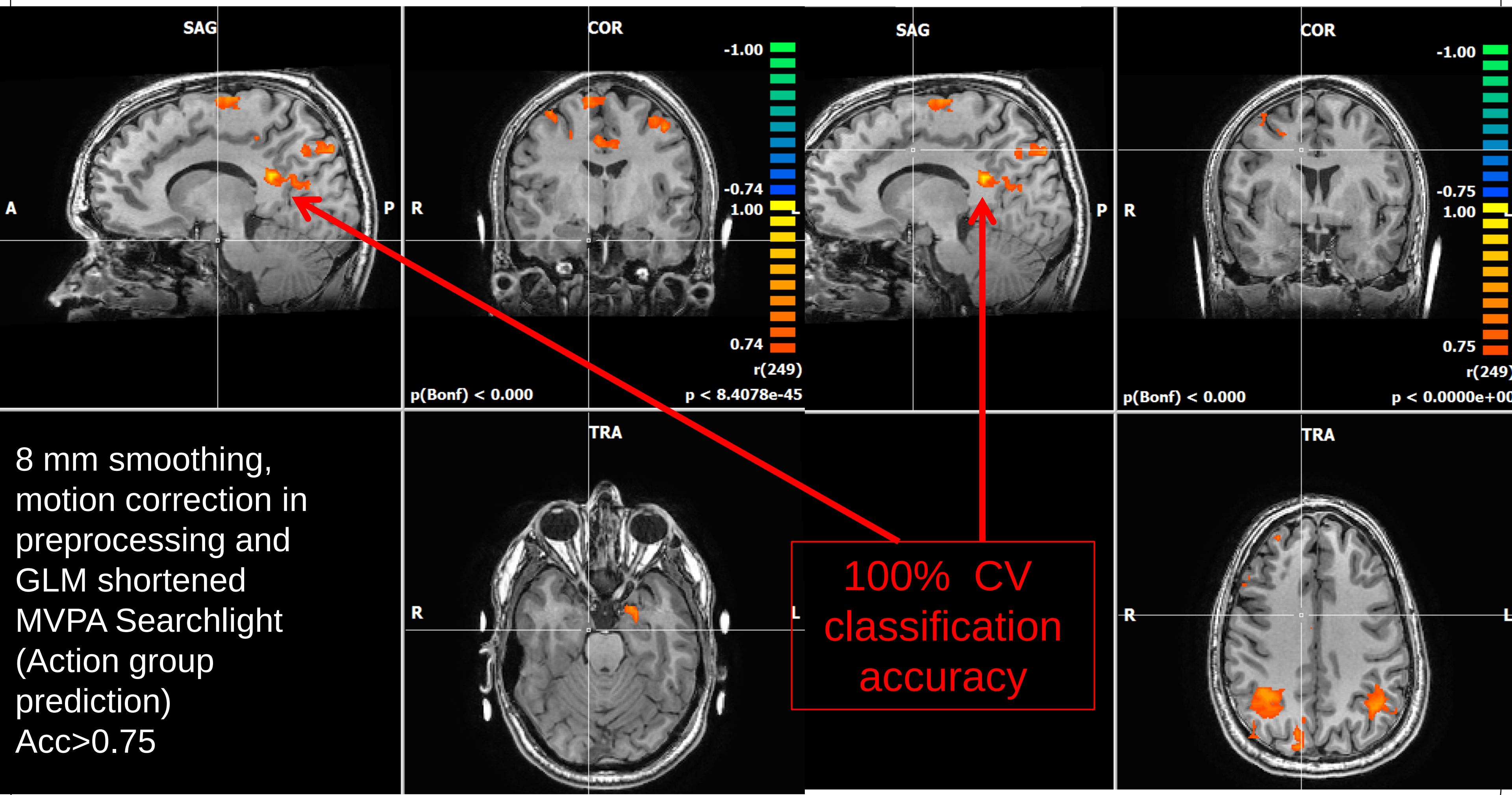
- For each subject
  - One brain map
  - Average difference in response between tasks
- Task
  - Classify subjects by performed action
- Dataset
  - 60k features
  - 25 observations

GLM Results

- No results when applying a multiple comparisons correction

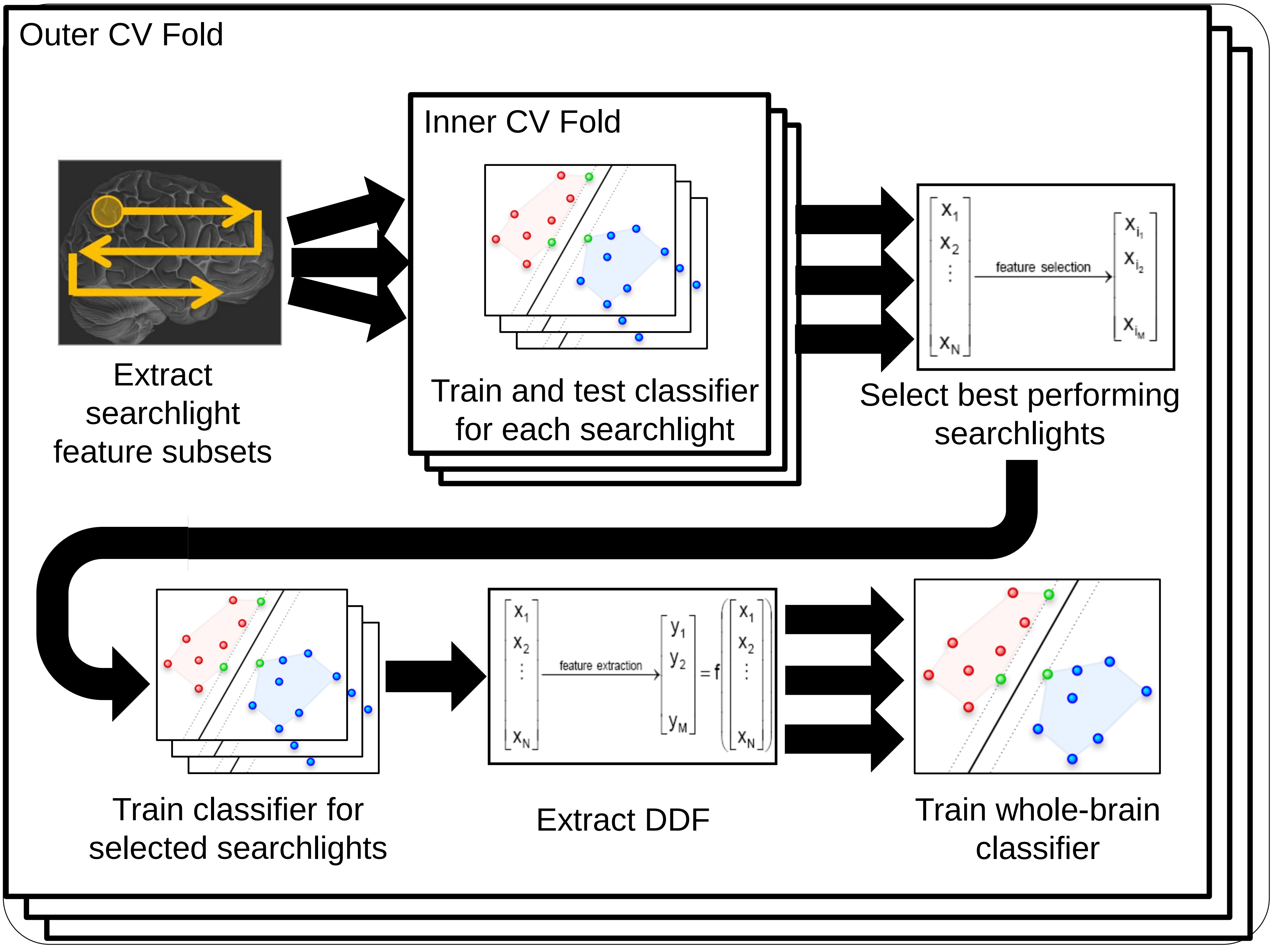


Searchlight Map Results



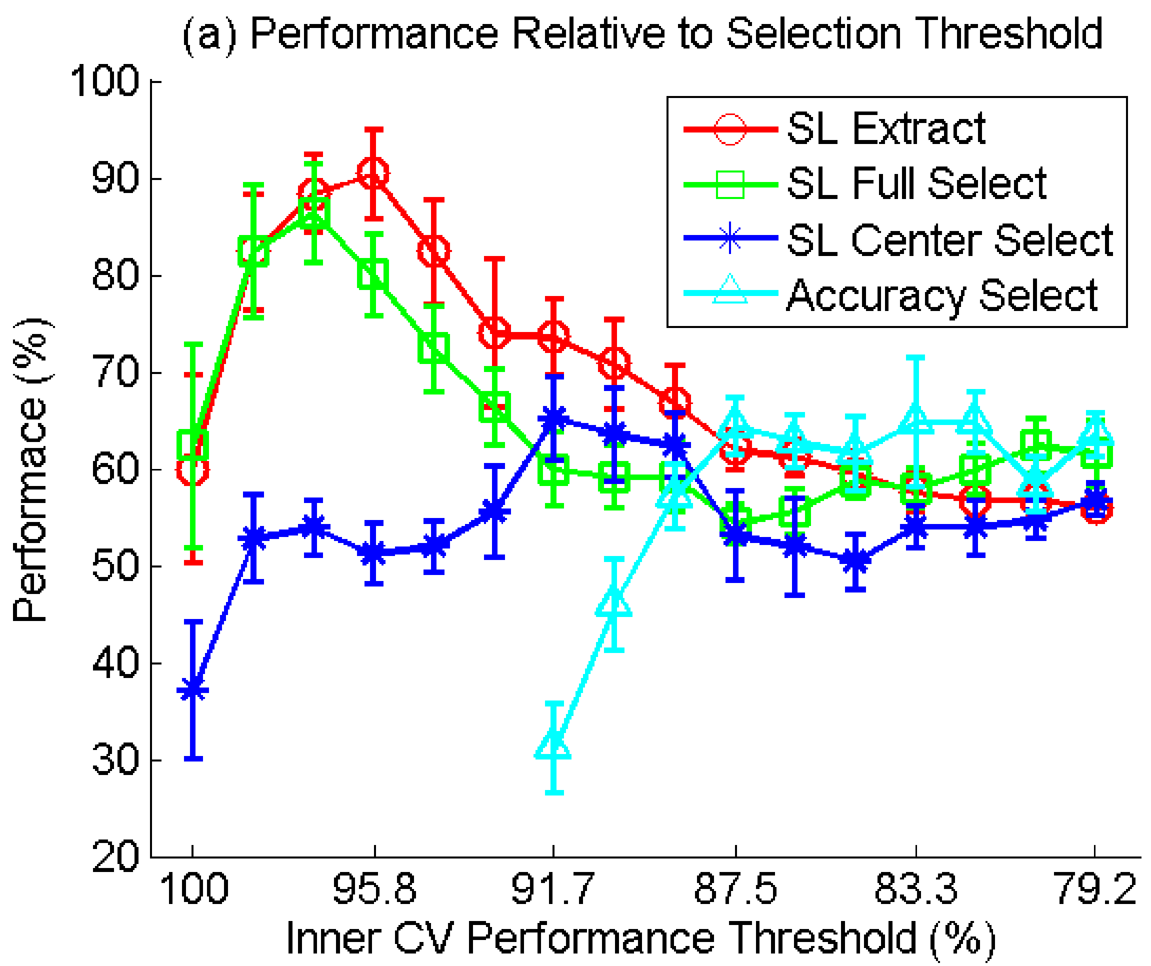
8 mm smoothing, motion correction in preprocessing and GLM shortened MVPA Searchlight (Action group prediction) Acc>0.75

100% CV classification accuracy

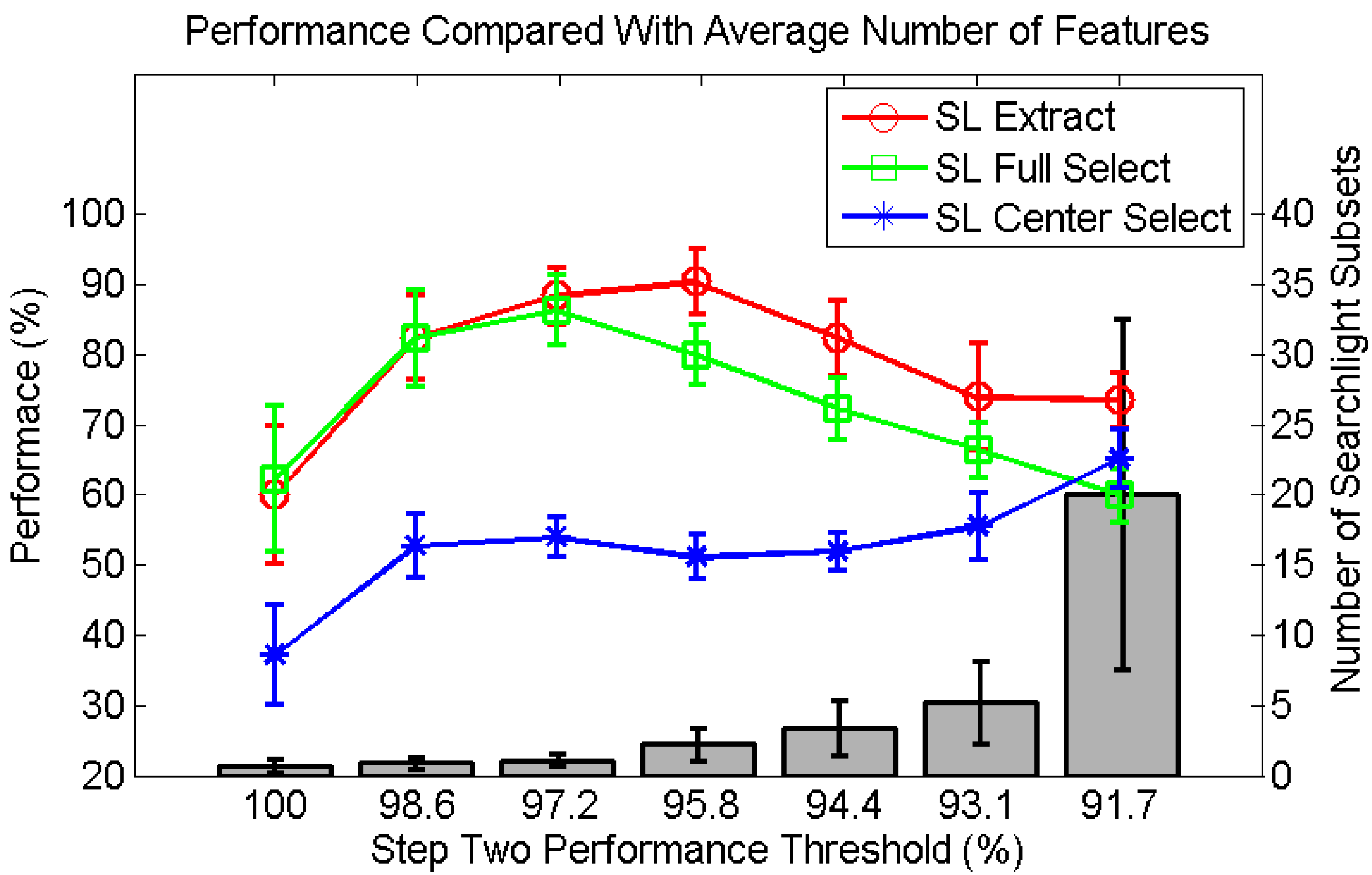


Outer CV Fold
Inner CV Fold
Extract searchlight feature subsets
Train and test classifier for each searchlight
feature selection
Select best performing searchlights
feature extraction
Train classifier for selected searchlights
Extract DDF
Train whole-brain classifier

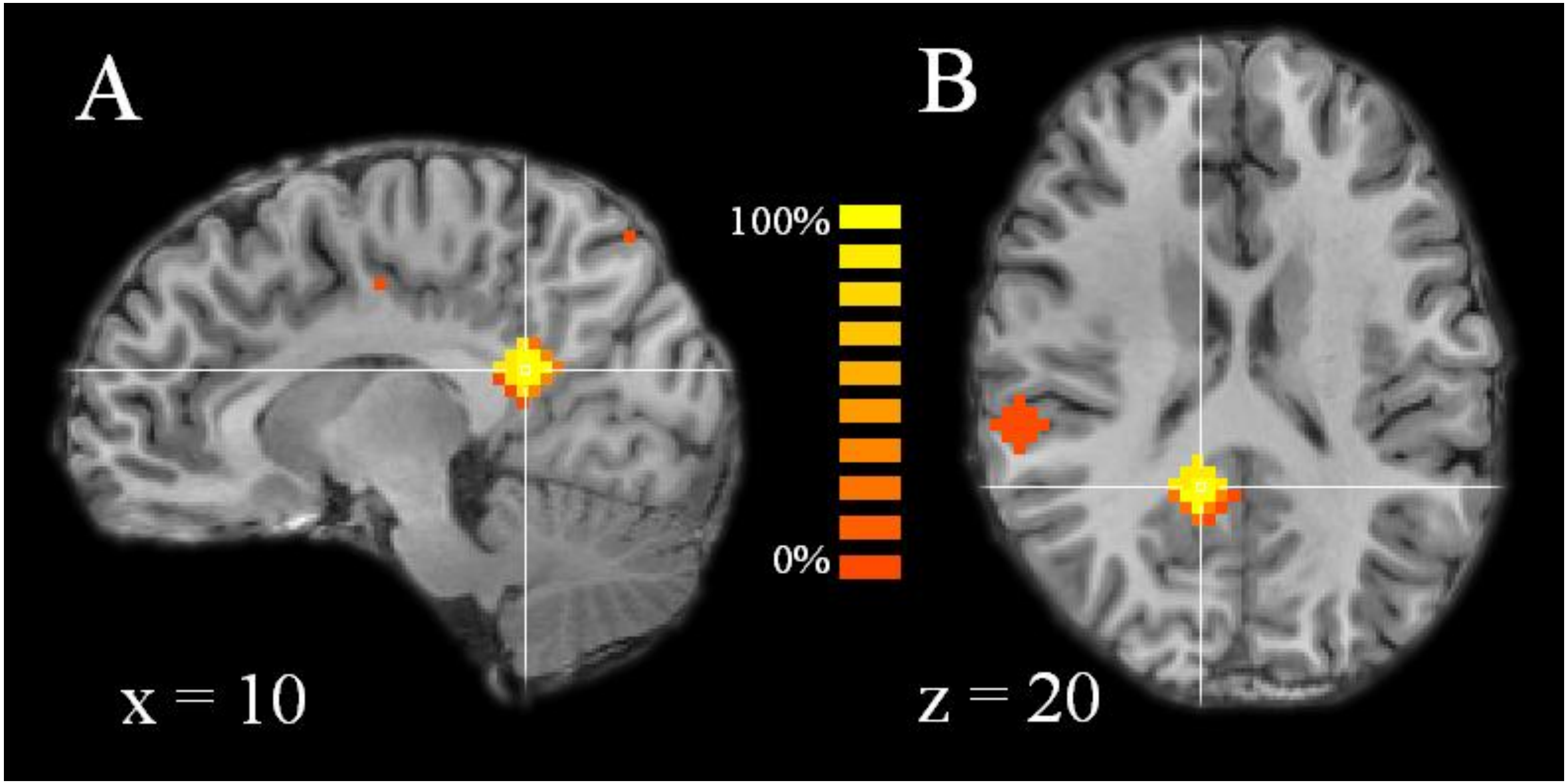
Results – Whole-brain



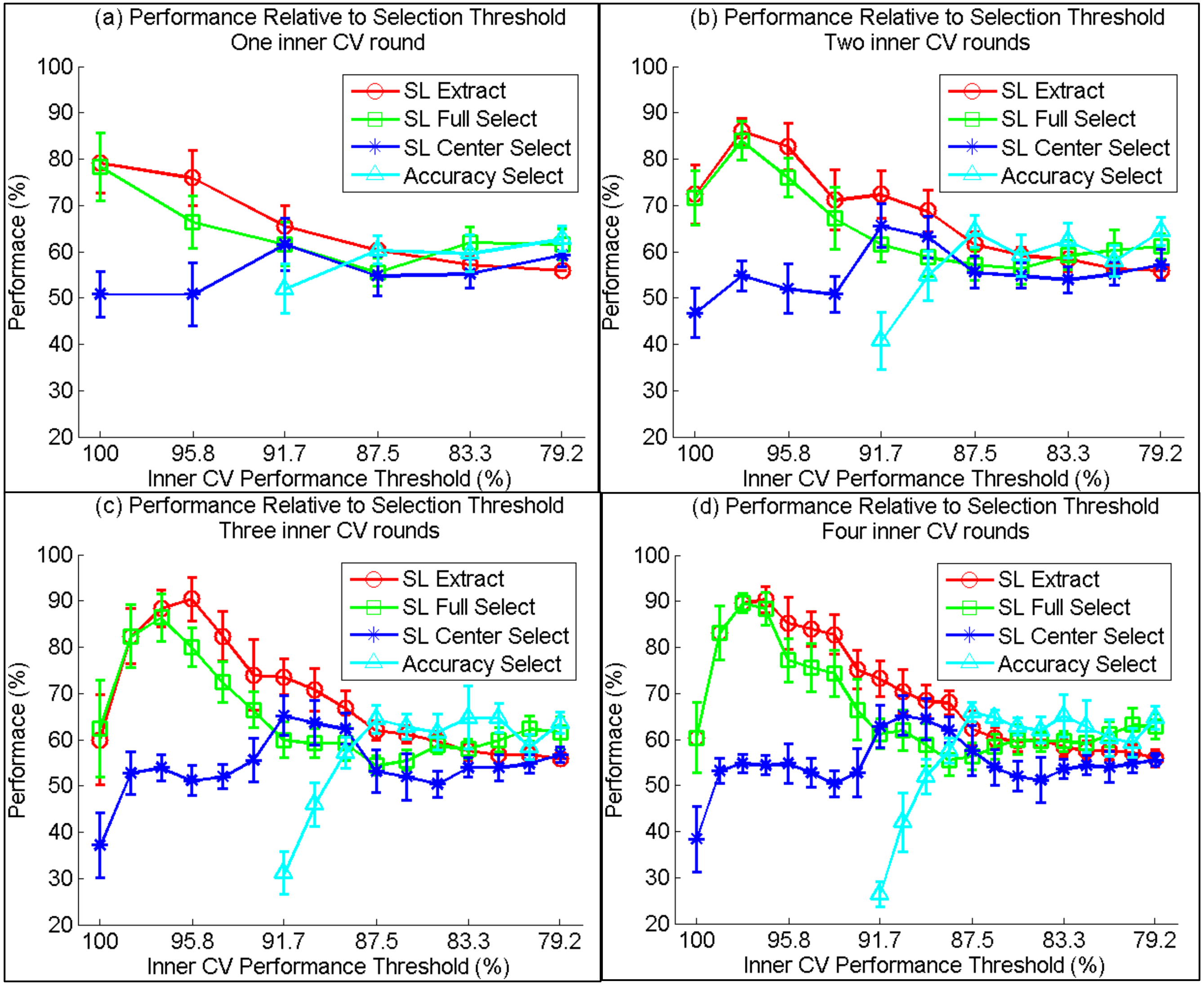
Results – Features Number



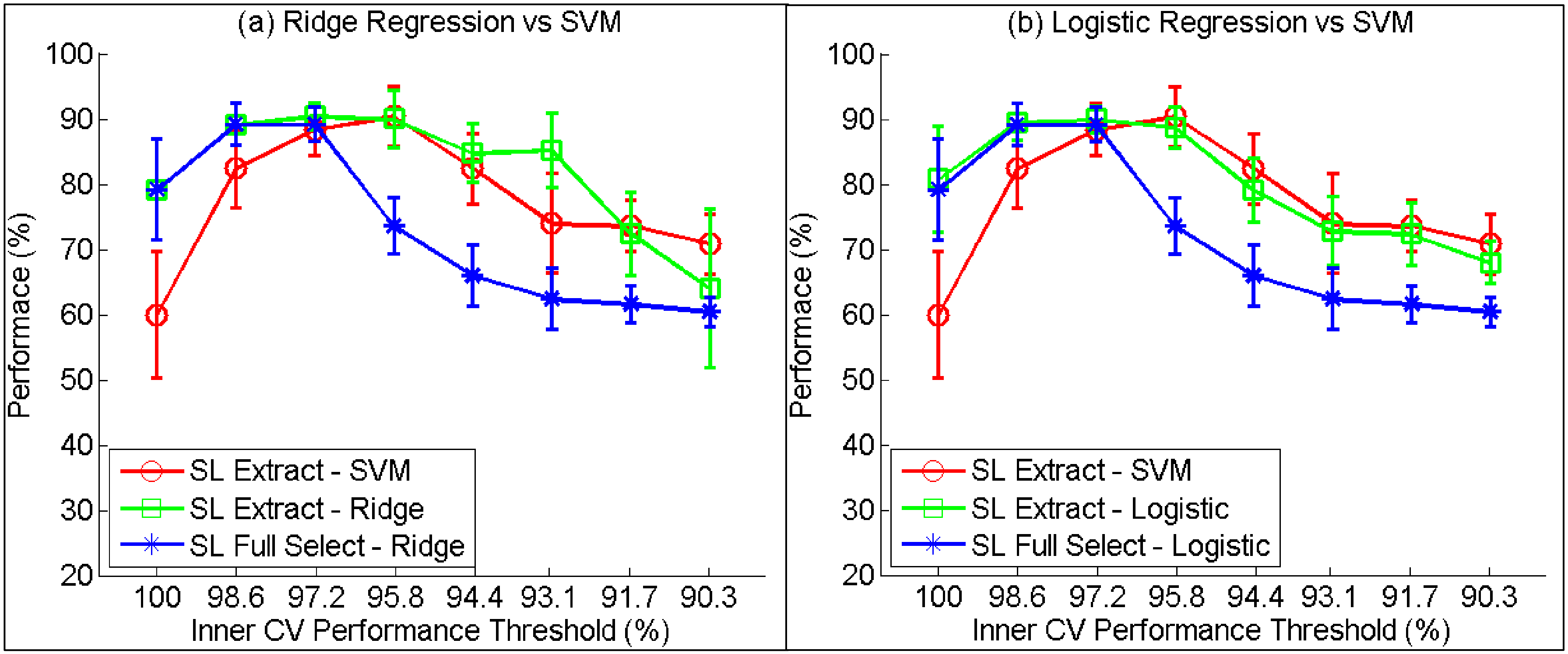
Results – Searchlight Location



Results – Inner Cross-validation



Results – Whole-brain Classifier



Summary of Results

- High whole-brain classification accuracy on chalenging fMRI classification task

Method	Results
Searchlight center selection	Low performance due to inappropriate features
Searchlight full selection	Performance drops as number of features grow
Searchlight-DDF	Best results with a lower number of whole-brain features.

- Importance of choosing enough inner-CV folds
- Improvement independent of whole-brain classifier

Thank You

Questions?