Classification of Autism Brain Imaging Data Using Machine Learning Algorithm

Colleen P. Chen[1, 2], Christopher L. Keown[1,2], Ralph-Axel Müller[2,3]

[1]Computational Science Research Center, SDSU
[2]Brain Development Imaging Lab, Dept. of Psychology, SDSU
[3]Joint Doctoral Program in Clinical Psychology, SDSU/UCSD

Motivation

- Can we diagnose autism using functional neuroimaging data?
  - Machine-learning (ML) application in classifying autism data from typically developing functional imaging data
- Experimental method: implement a ML algorithm to maximize classification accuracy while selecting for *features* that were most informative

Background: Autism Spectrum Disorders (ASD)

- Array of neurodevelopmental disorders characterized by sociocommunicative, cognitive, and sensory motor impairments.
- Prevalence 1 in 88: CDC statistics show increasing trend
- Current diagnostic methods rely on autism-specific behavior evaluations

Functional connectivity (FC) and networks in ASD

- FC is a type of fMRI analysis that looks at temporal correlation between brain activity from different regions
- Network refers to a collection of spatially distributed brain regions that participate in common functionality
- Convergent evidence implicates abnormal connectivity in autism

Machine learning: a data-driven approach

- Machine learning (ML) is programming computers to optimize a performance criterion using data or past experience
- Exploratory data analysis looks at the experimental results and attempts to make sense of them
- Machine learning finds pattern in complex data, but requires a priori knowledge to determine whether pattern is meaningful or relevant

A step toward diagnostic classification assay for autism

- Motivation: Identification of complex biomarkers
- Current classification efforts performance only moderately above chance
  - Anderson et al. (2011): 71 – 79% overall

Objective

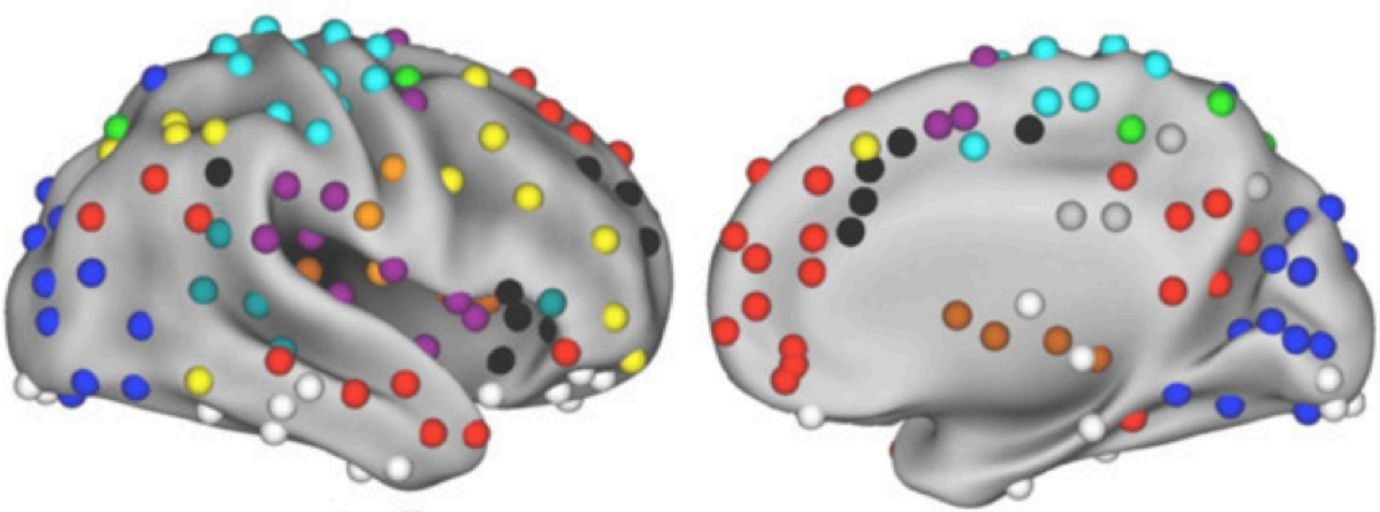
To find functional connections that are informative in classifying autism from typically developing imaging data, for identifying biomarkers

Data

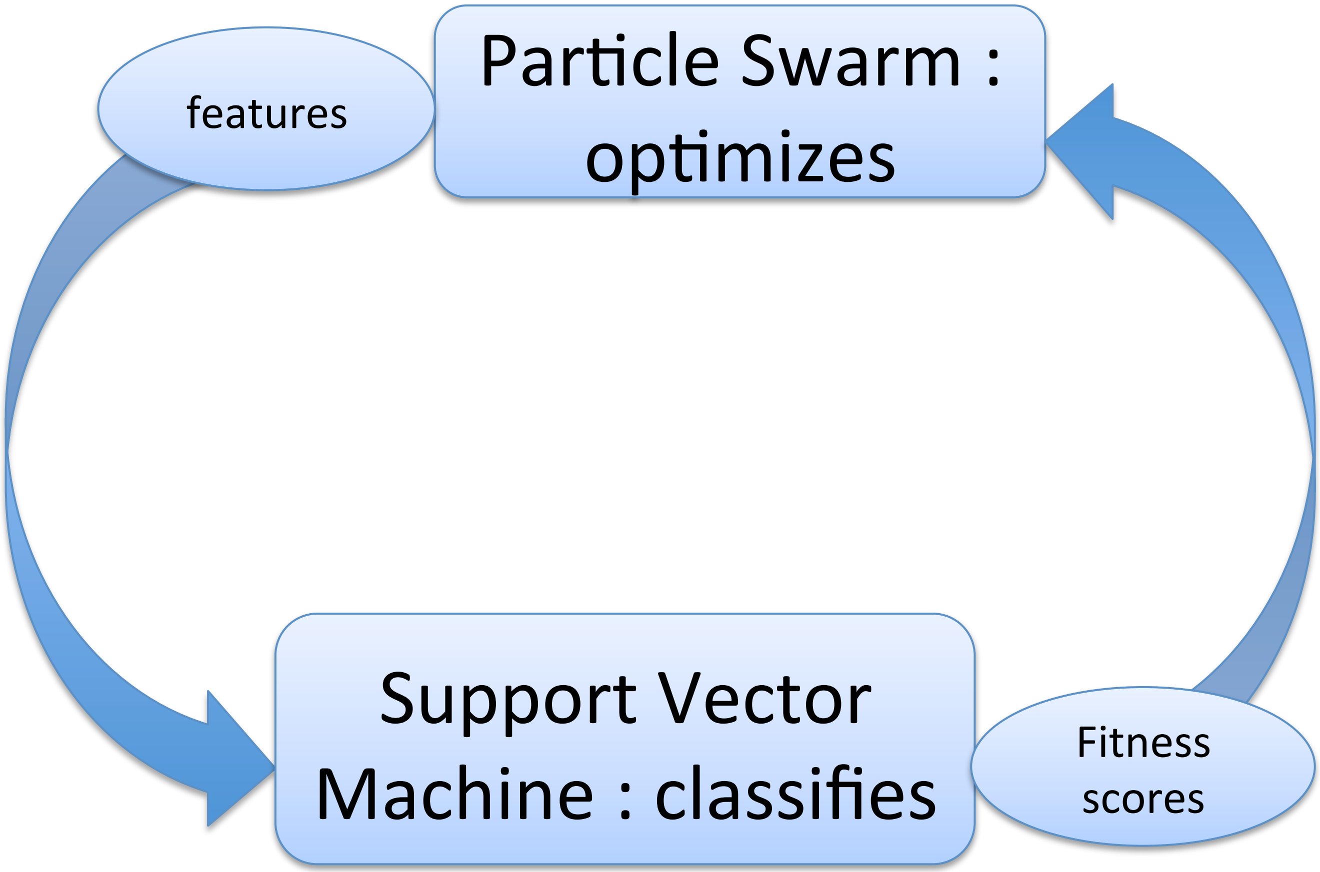
- 3 datasets collected at different sites:
  1. SDSU (24 TD, 24 ASD)
  2. UCLA (21 TD, 21 ASD)- [ABIDE data sharing consortium]
  3. YALE (20 TD, 20 ASD)
- Adolescents matched on age, IQ, gender, handedness, and motion during scan
- Resting state fMRI data
  - Intrinsic fluctuation of baseline brain activity

Methods

- 264 regions of interest (ROIs) were defined by Power et al. (2011) using meta-analysis
  - Each ROI belongs to one of 14 functional networks
- Pearson's correlation matrix was computed to quantify connectivity between ROIs



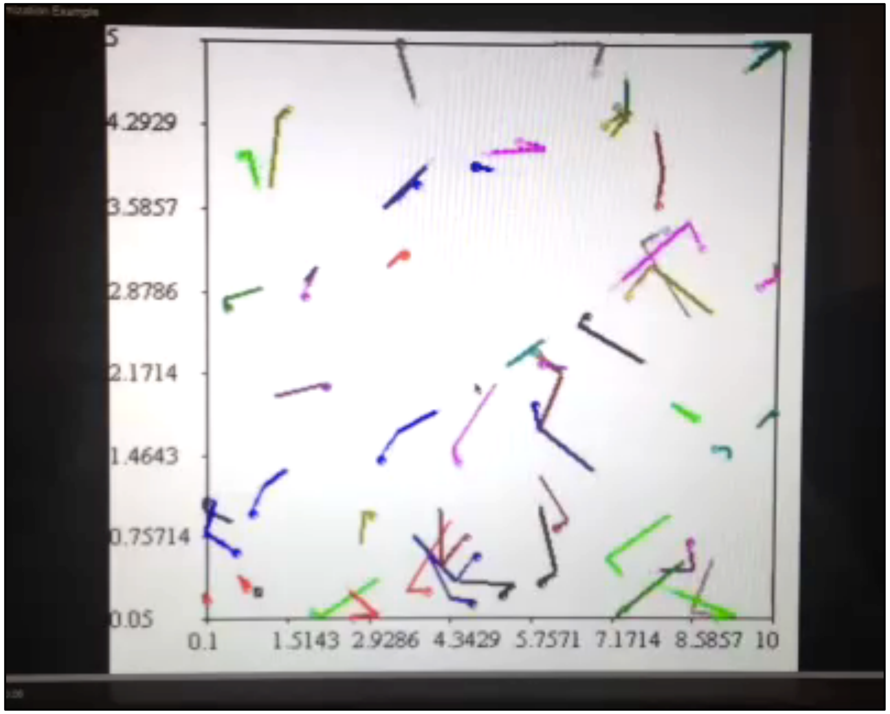
The feature selection process



"Features" refer to connections

Particle Swarm Optimization (PSO): population-based stochastic search algorithm

PSO appears like a disorganized population of flying particles that tends to cluster together, while each particle seems to be moving in a random direction



Results : PSO-SVM classification performance on SDSU dataset

N = 48 (24 ASD, 24 TD)

runs = 100

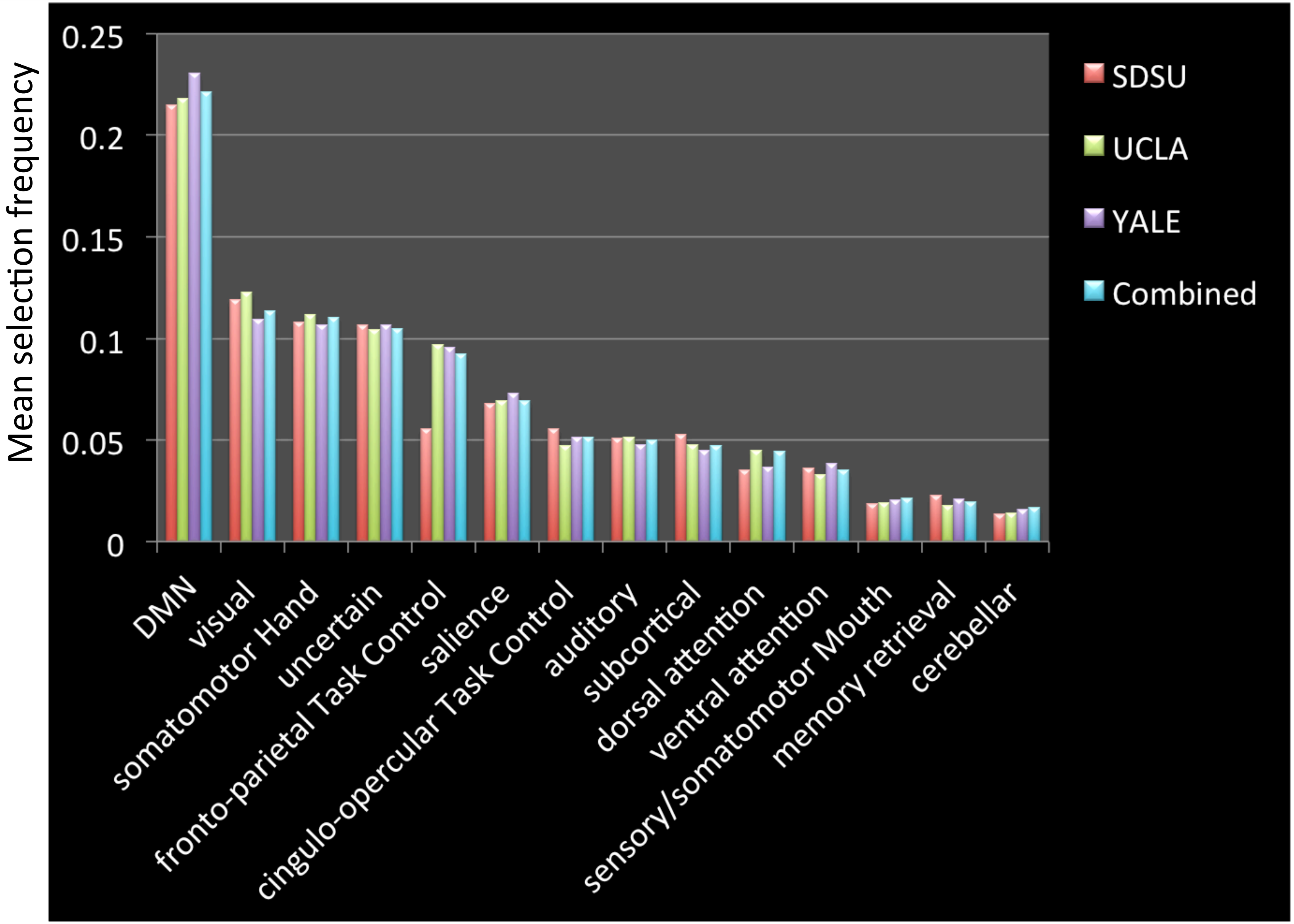
PSO-SVM	Best performing run
Sensitivity	96% (23/24)
Specificity	92% (22/24)
Overall	94% (45/48)

Algorithm applied to independent replication samples and a combined dataset

	SDSU rs-fMRI	UCLA rs-fMRI	YALE rs-fMRI	combined
N	48	42	40	130
median	83%	82%	80%	73%
range	69 - 94%	68 - 93%	62 - 93%	64 - 80%
Number of runs	100	100	100	100

- Consistent classification rates reproduced in datasets collected at other institutions
- Combining the three samples produces lower classification rate

functional networks that included the most informative connections



What can we learn from machine learning?

- Decreased classification rates from the combined set (N= 130) suggest latent variability in data collected from different sites
  - variations likely due to scanning instrumentation
- Despite this inherent variability, which we have no control over, ML still classifies well above chance
  - ML may be suitable for future studies requiring pooled-data from different sites

Conclusion

- High classification accuracy using PSO-SVM by identifying connections that are most informative
  - Improved accuracy over conventional SVM [Anderson et al. 2011]
  - Reproducible results
- Variability across datasets reduces classification performance when combined
  - Hardware difference
  - Cohort variability

Future research

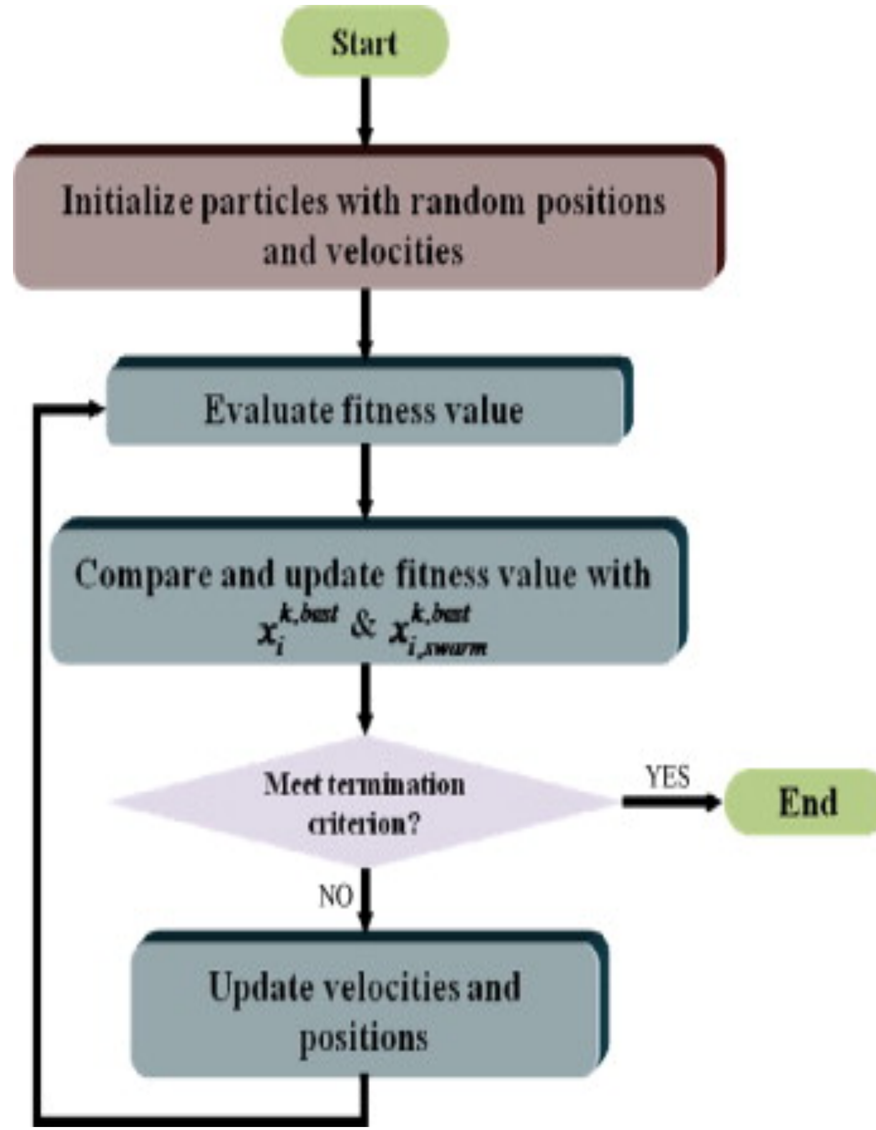
- Compare with other feature selection methods
- Make better sense of the connections selected and what it means in the biological aspect
- To better understand the differences across the datasets in order to improve the pooled-data classification

Acknowledgments

- Brain Development Imaging Lab
- Christopher Keown, Mark Mulvey, Omar Maximo
- Funding:
  - NIH R01-MH081023 (R-AM)

Swarm intelligence: population-based stochastic search algorithm

- Bio-inspired computer algorithm mimics *social behavior* of swarming particles
- Requires *cooperation* and *competition* among population to move potential solutions toward better search areas
- Position encodes features selected



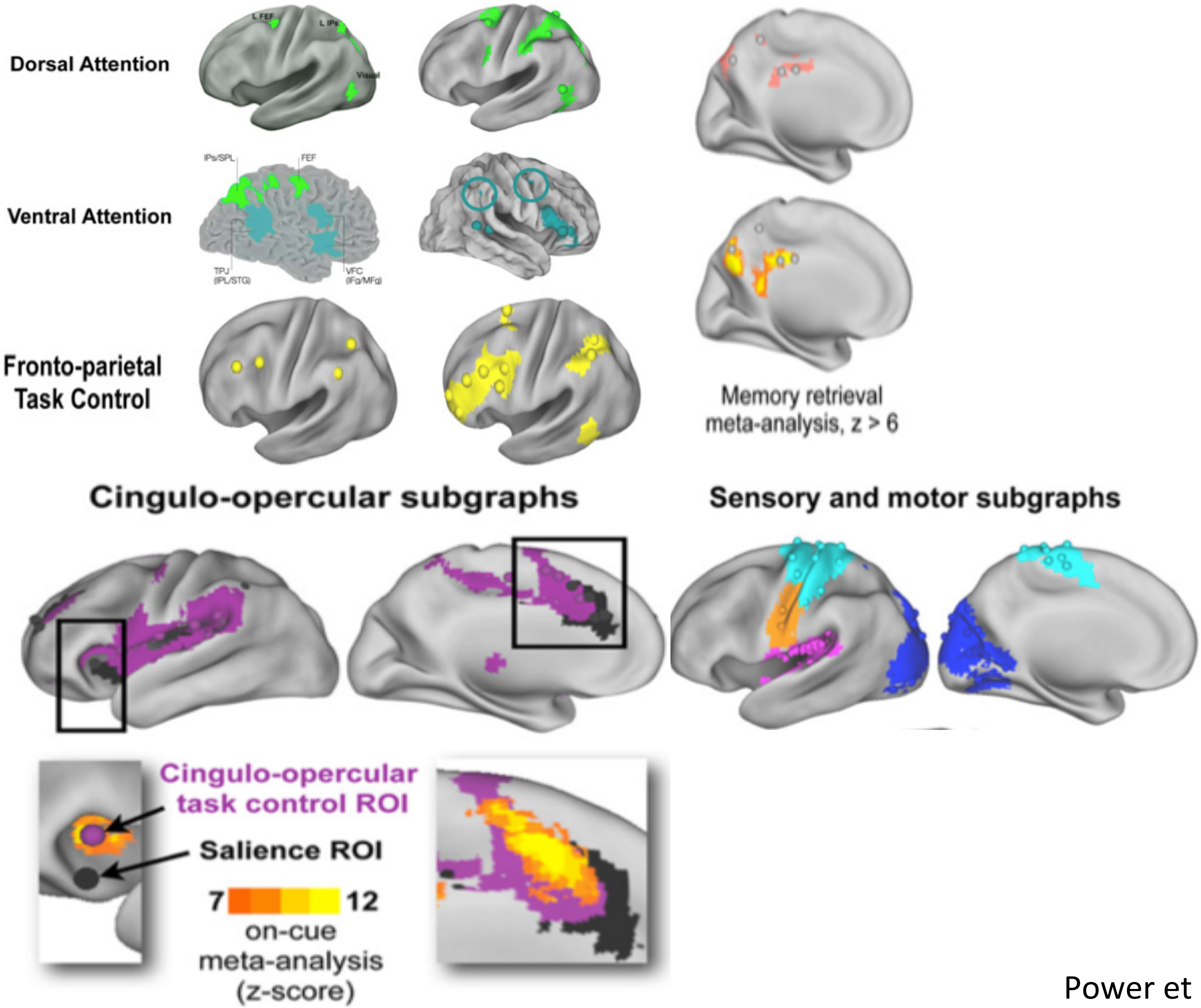
Support Vector Machine: an approach for classifying

- Leave-one-out cross validation:
  - Train model on N-1 subjects
  - Classify the 1 subject left out
  - Repeat for N times
- Fitness function = correct classification / N

- Machine learning models map features to a common categorical variable
- Feature selection algorithms operate within ML algorithms to identify subsets of features which are both non-redundant and mutually distinct of each other, yet predictive of the outcome variable of interest

Outline

- In the neuroimaging field (fMRI):
  - Classification prevalent in literature
- Given a binary classification problem: ASD or TD ?
  - using fMRI data collected from various institutions:
    - SDSU, UCLA, YALE
- Using a machine learning algorithm that seeks to **optimize** the classification accuracy of autistic children
  - Significance of early diagnosis for at-risk, or likely individuals to be autistic (siblings → genes?)
  - Help researchers understand biological differences in ASD and TD brains
- Convergent evidence point to dysfunctional network connectivity
  - many findings (inconsistent)
- Introduction to machine learning algorithms:
  - Feature selection and classification
- Classification results: Discussion and Conclusion



Power et al. (2011)

fMRI data preprocessing

Resting-state imaging time series

↓

Image realignment

↓

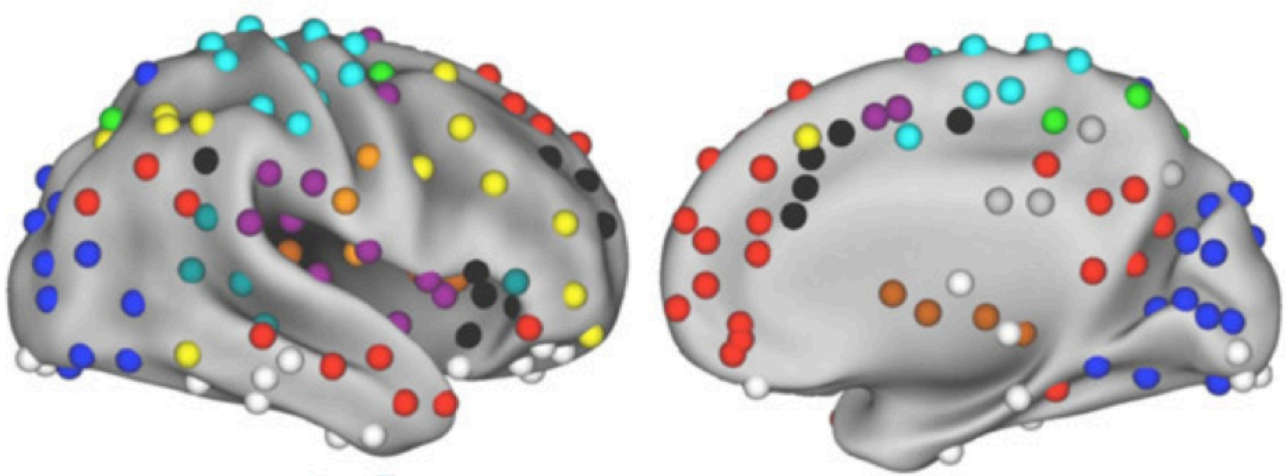
Spatial Normalization

↓

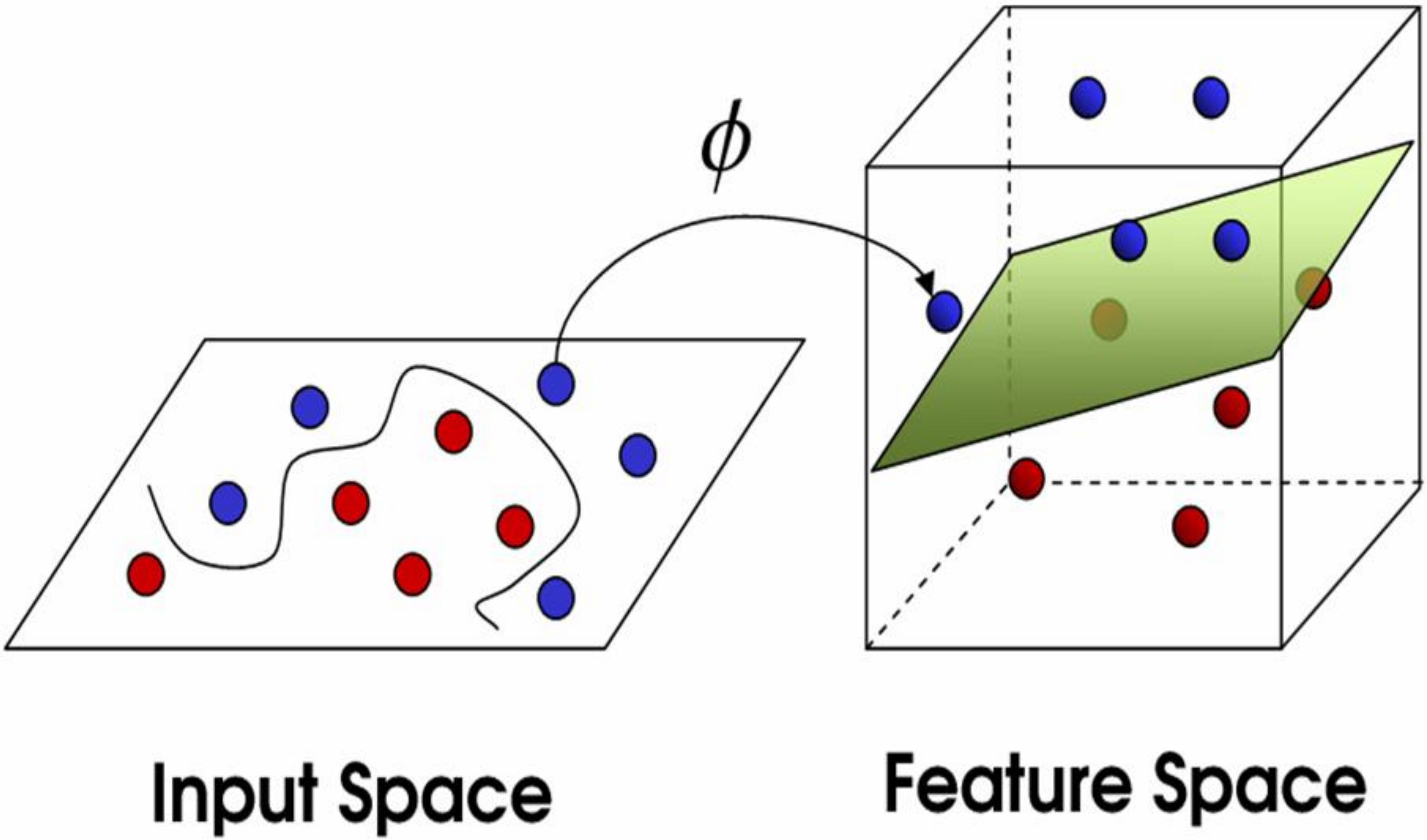
Motion filtering

- Analysis of Functional NeuroImage (AFNI) software (Cox, 1996) for preprocessing and quality control

To implement a machine learning algorithm
for **optimizing** the performance of the supp
vector machine (SVM) as classifier whereby



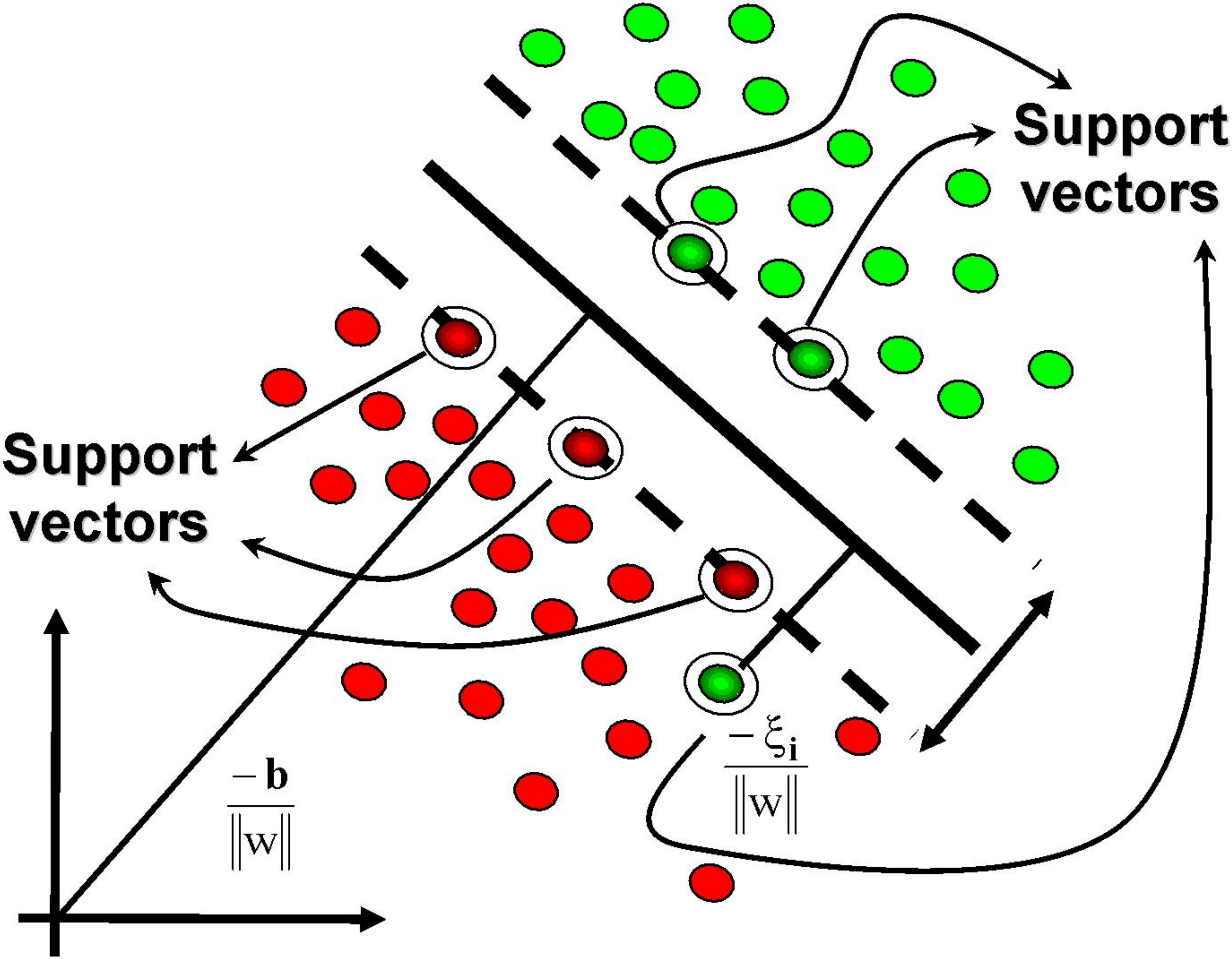
ions betwe
ative of AS



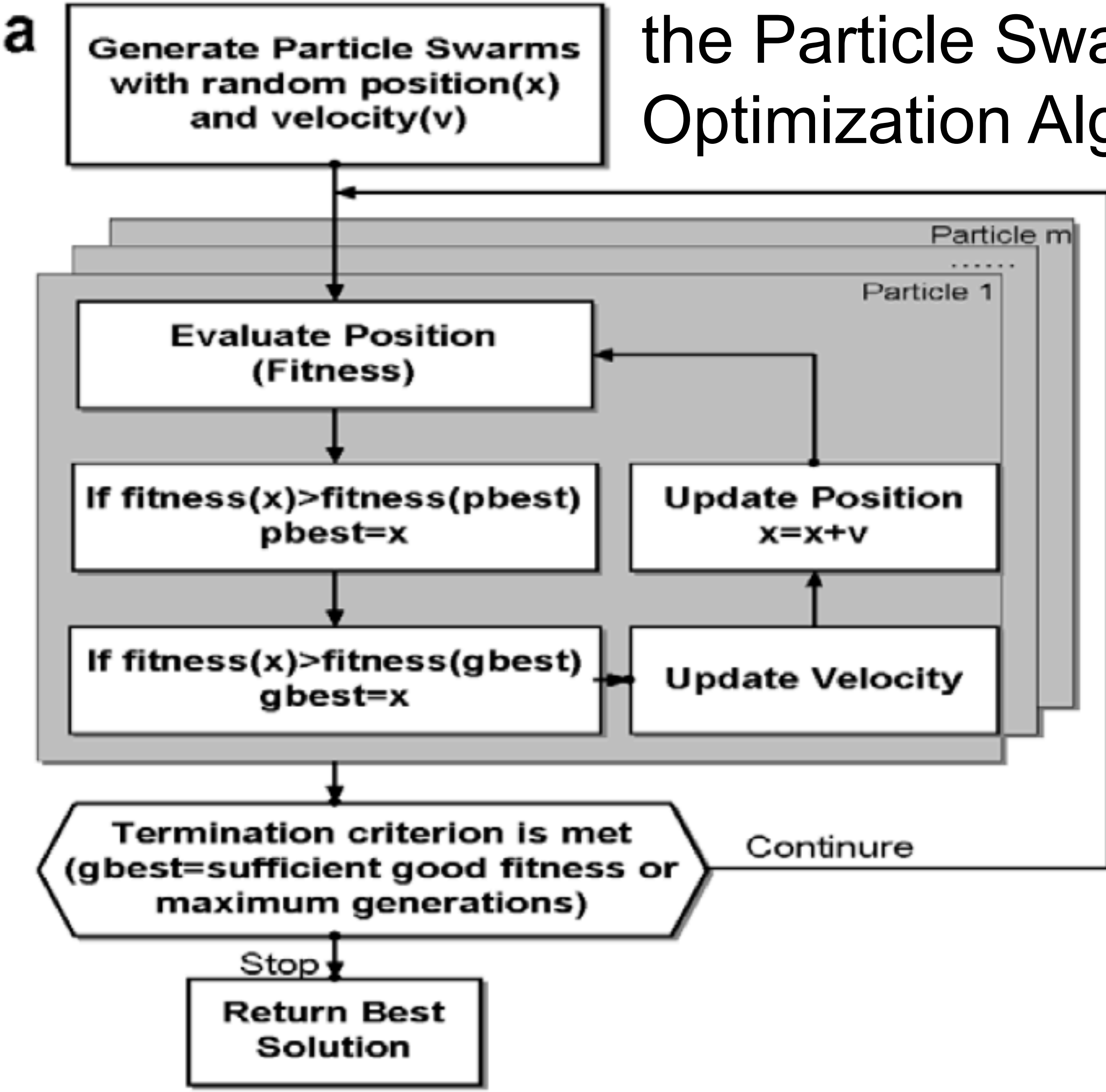
ϕ
Input Space
Feature Space

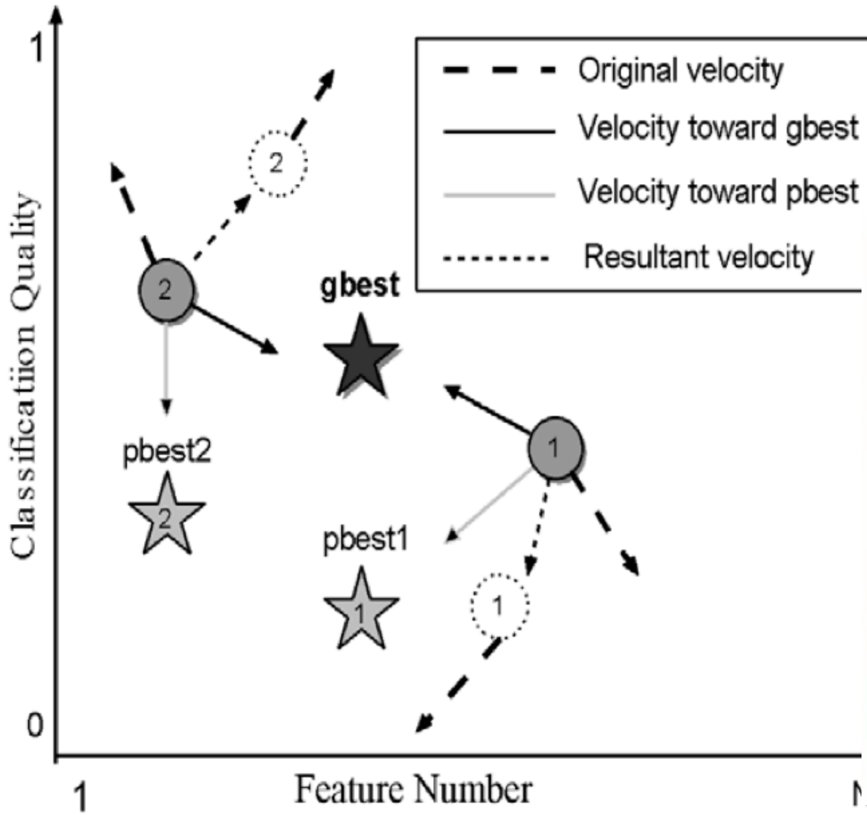
Neuroimaging studies of ASD increasingly focused on connectivity analysis

- Convergent evidence implicate abnormal connectivity in autism
    - Diffusion tensor imaging (DTI) studies show anatomical white-matter compromise (Courchesne et al., 2003)
- However, inconsistent findings from fMRI studies of network dysfunction, due to varying methodological parameters (Müller et al., 2011)


Support vectors
Support vectors
$\frac{-b}{\|w\|}$
$\frac{-\xi_i}{\|w\|}$

Graphical illustration of the Particle Swarm Optimization Algorithm



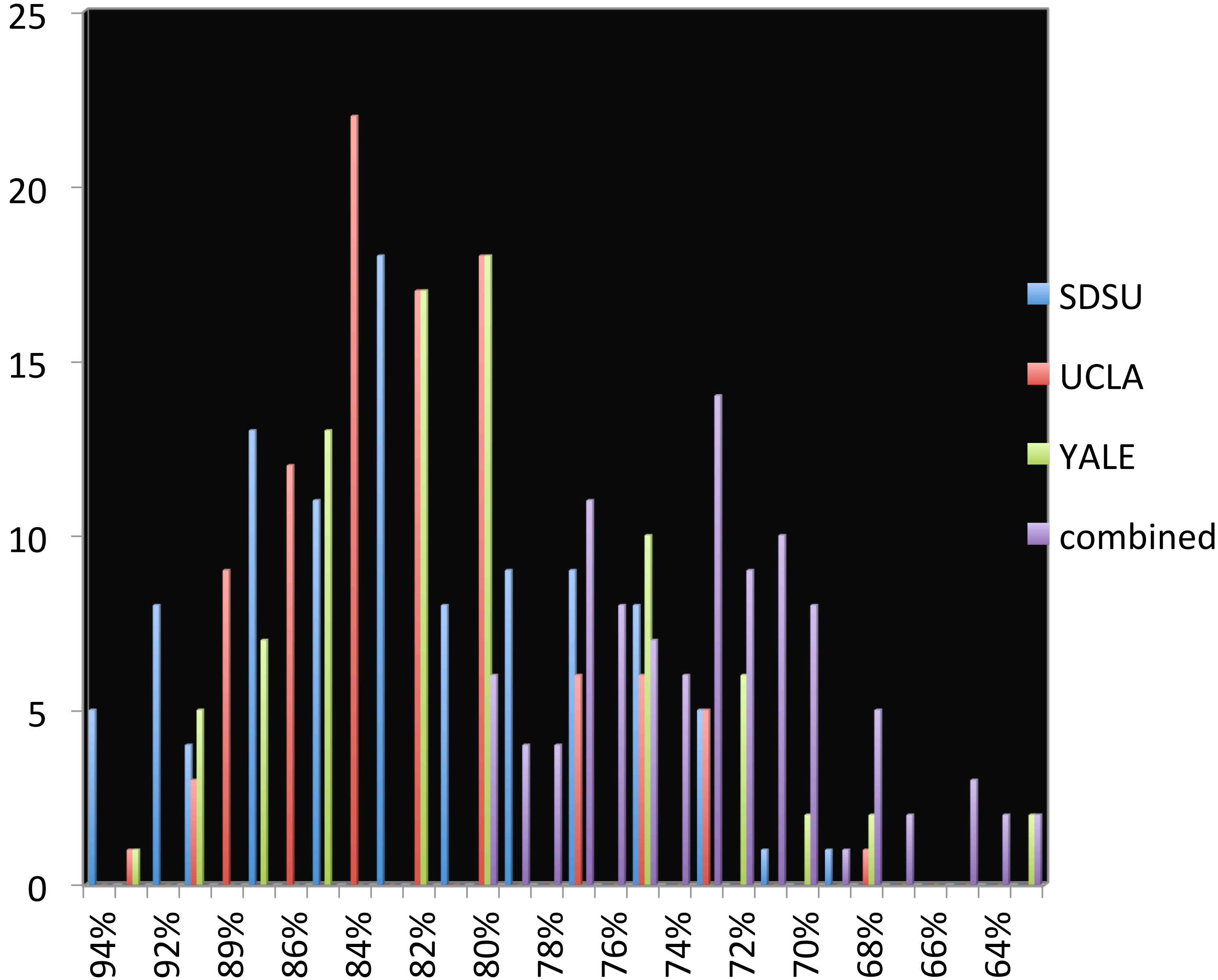


Identified Prevalence of Autism Spectrum Disorders
ADDM Network 2000-2008
Combining Data from All Sites

Surveillance Year	Birth Year	Number of ADDM Sites Reporting	Prevalence per 1,000 Children (Range)	This is about 1 in X children...
2000	1992	6	6.7 (4.5-9.9)	1 in 150
2002	1994	14	6.6 (3.3-10.6)	1 in 150
2004	1996	8	8.0 (4.6-9.8)	1 in 125
2006	1998	11	9.0 (4.2-12.1)	1 in 110
2008	2000	14	11.3 (4.8-21.2)	1 in 88

$$\boxed{v_{i,j}^{n+1} = w \times v_{i,j}^{n} + c_1 \times r_1(p_{i,j}^{n} - x_{i,j}^{n}) + c_2 \times r_2(p_{g,j}^{n} - x_{i,j}^{n})}$$

$$\boxed{x_{i,j}^{n+1} = x_{i,j}^{n} + v_{i,j}^{n+1}, j = 1, 2, \cdots, d.}$$

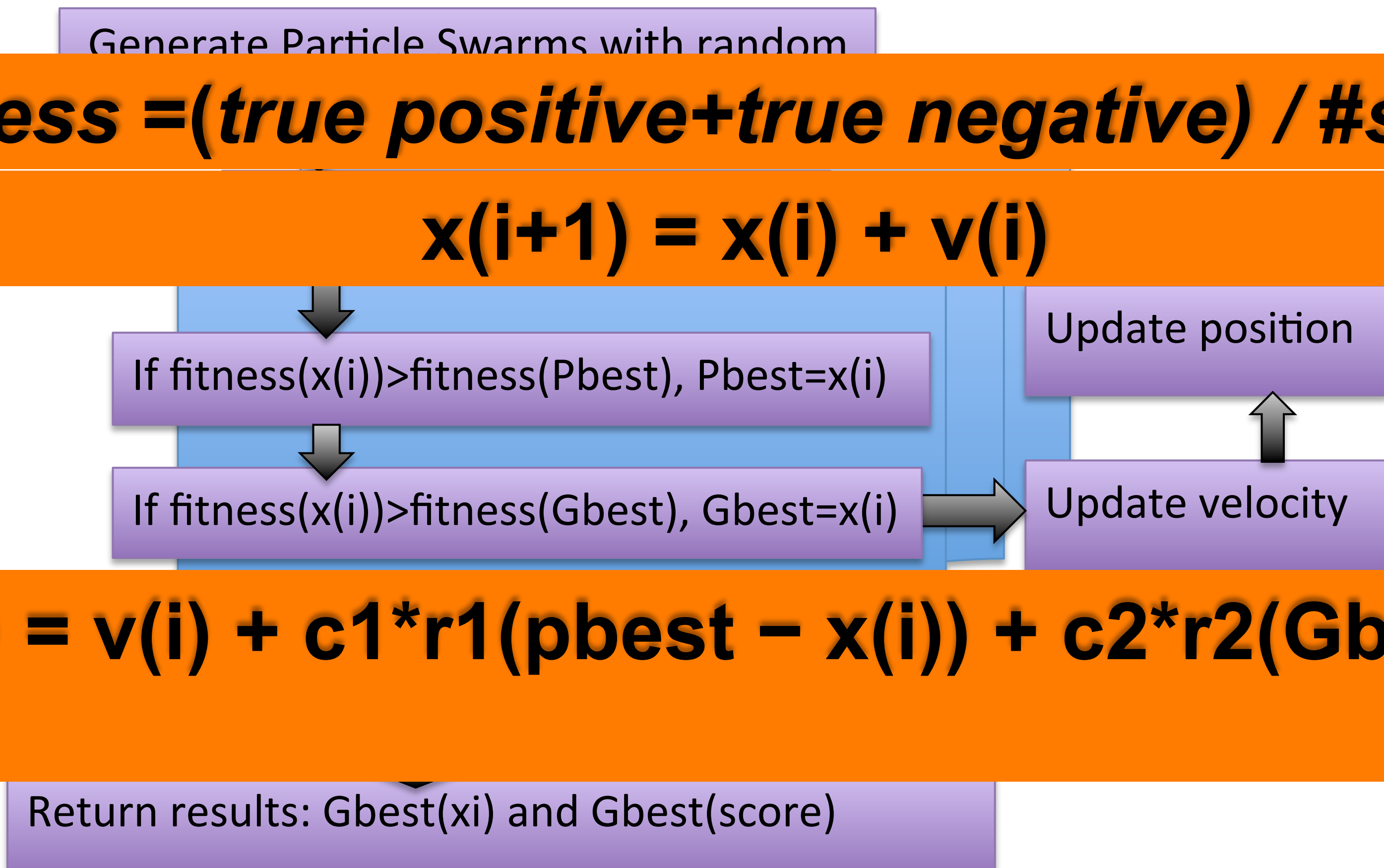


25
20
15
10
5
0
94%
92%
89%
86%
84%
82%
80%
78%
76%
74%
72%
70%
68%
66%
64%
SDSU
UCLA
YALE
combined

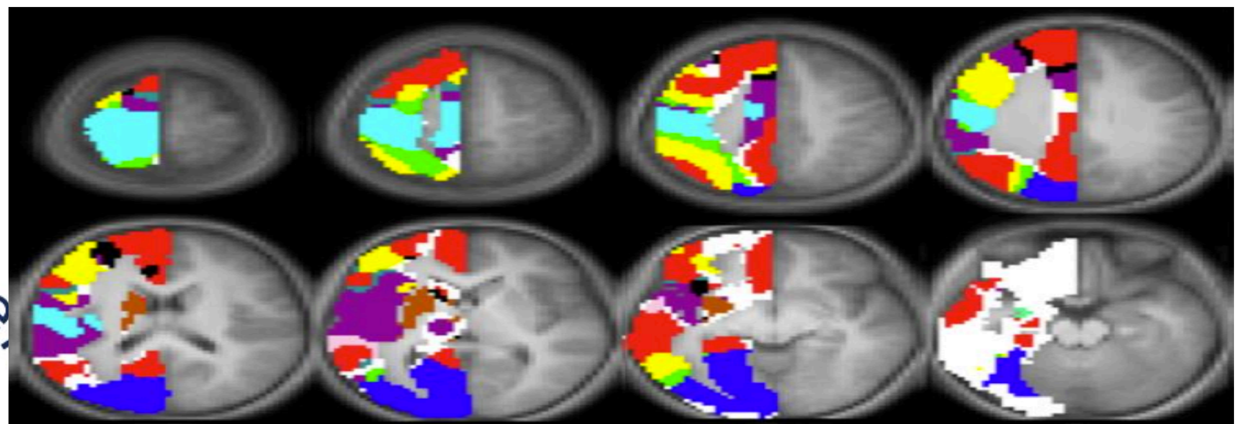
Evolutionary Computation

$$\%fitness = (true\ positive + true\ negative) / \#subjects$$

$$x(i+1) = x(i) + v(i)$$



$$v(i+1) = v(i) + c1*r1(pbest - x(i)) + c2*r2(Gbest - x(i))$$




Autism Aware
month
April is

SDSU (N=48)
5:0.94
8:0.92
4:0.90
13:0.88
11:0.85
18:0.83
8:0.81
9:0.79
9:0.77
8:0.75
5:0.73
1:0.71
1:0.69
total number of runs: 100
avg = 0.8296
std = 0.0608843911927276
median = 0.83
minimum = 0.69
maximum = 0.94

UCLA (N=42)
1:0.93
3:0.91
9:0.89
12:0.86
22:0.84
17:0.82
18:0.80
6:0.77
6:0.75
5:0.73
1:0.68
total number of runs: 100
avg = 0.8226
std = 0.0470894270661063
median = 0.82
minimum = 0.68
maximum = 0.93

YALE (N=40)
1:0.93
5:0.90
7:0.88
13:0.85
17:0.82
18:0.80
17:0.78
10:0.75
6:0.72
2:0.70
2:0.68
2:0.62
total number of runs: 100
avg = 0.8006
std = 0.0581502755406588
median = 0.8
minimum = 0.62
maximum = 0.93

combined (SDSU+UCLA+YALE)
N = 132

6 : 0.80
4 : 0.79
4 : 0.78
11 : 0.77
8 : 0.76
7 : 0.75
6 : 0.74
14 : 0.73
9 : 0.72
10 : 0.71
8 : 0.70
1 : 0.69
5 : 0.68
2 : 0.67
3 : 0.65
2 : 0.64
total number of runs: 100
avg = 0.7338
std = 0.0388412984418192
median = 0.73
minimum = 0.64
maximum = 0.80