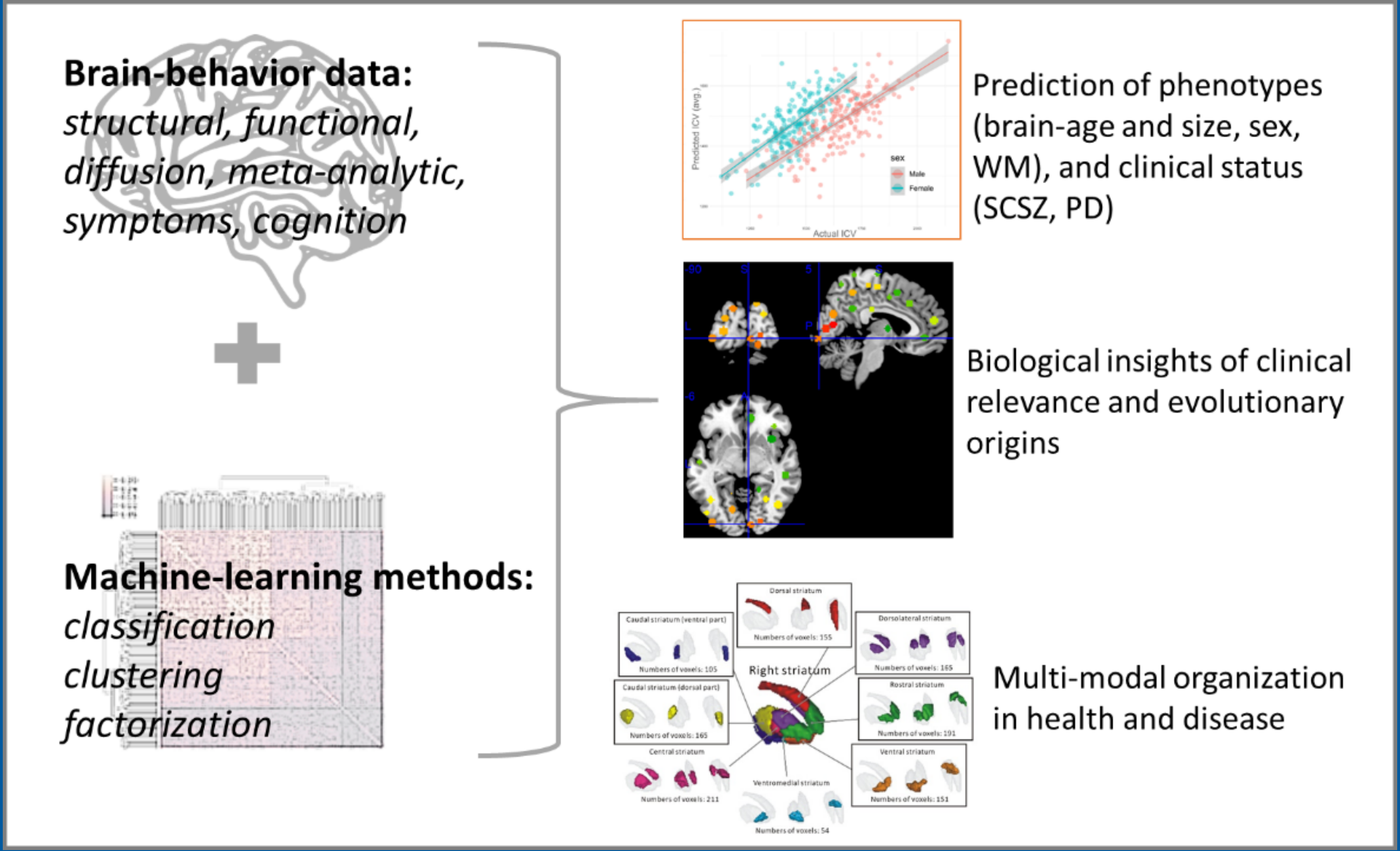
Machine learning applications and challenges

In the medical domain

Kaustubh R. Patil
k.patil@fz-juelich.de

Applied Machine Learning
Institute of Neuroscience and Medicine
INM-7: Brain and Behaviour
Research Centre Jülich

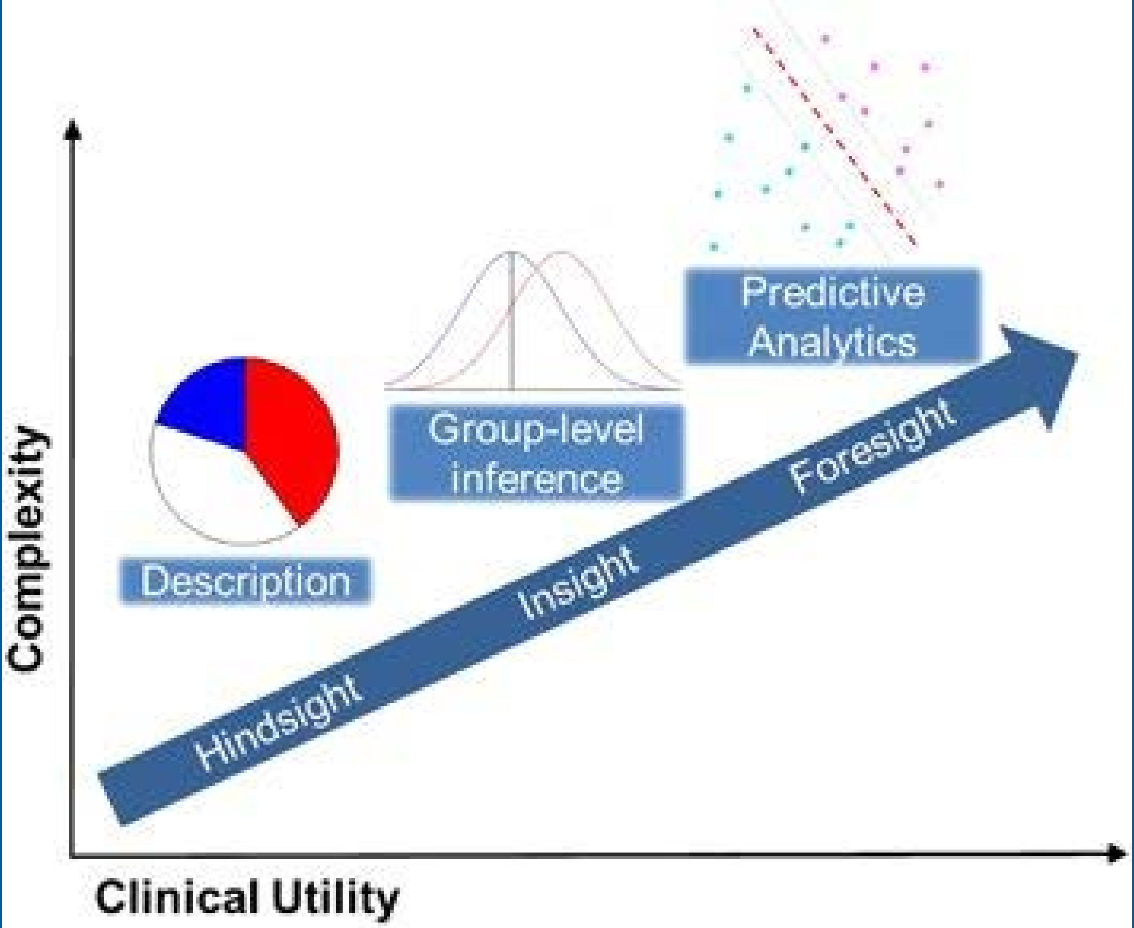
Institute of Systems Neuroscience
Heinrich Heine University Düsseldorf



“We become what we behold.
We shape our tools and then our tools shape us.”

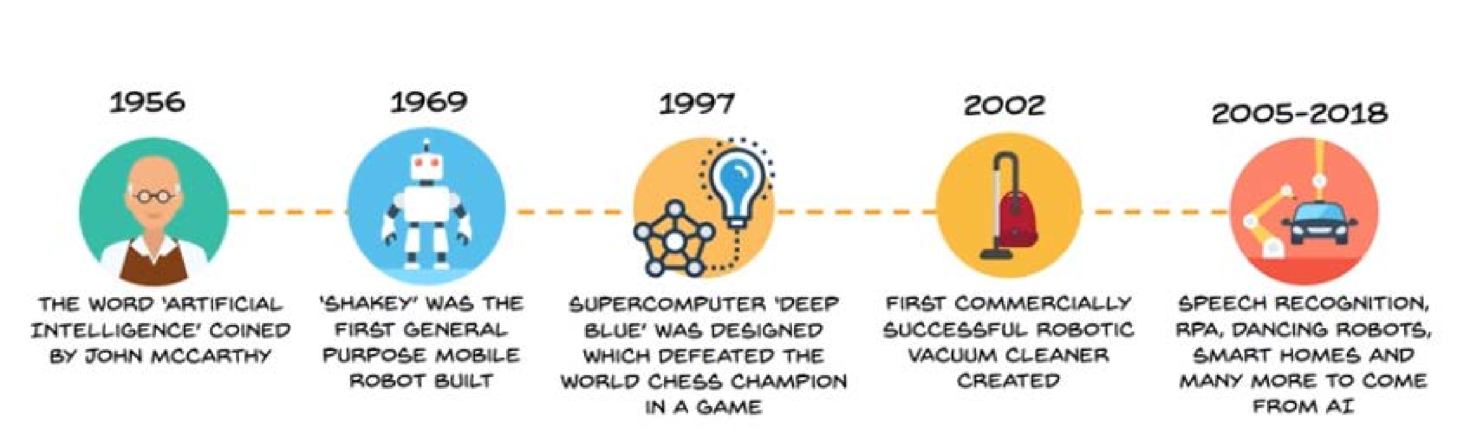
Marshall McLuhan

What is ML/AI?



Hahn et. el, 2017 Mol. Psych.

Since 60s: so why now?



https://towardsdatascience.com/

Lots of data

https://cs.stanford.edu/people/karpathy/cnnembed/

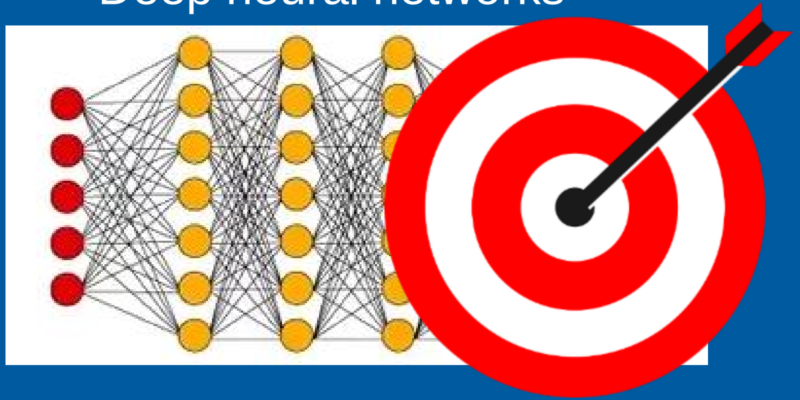
+

Fast computers

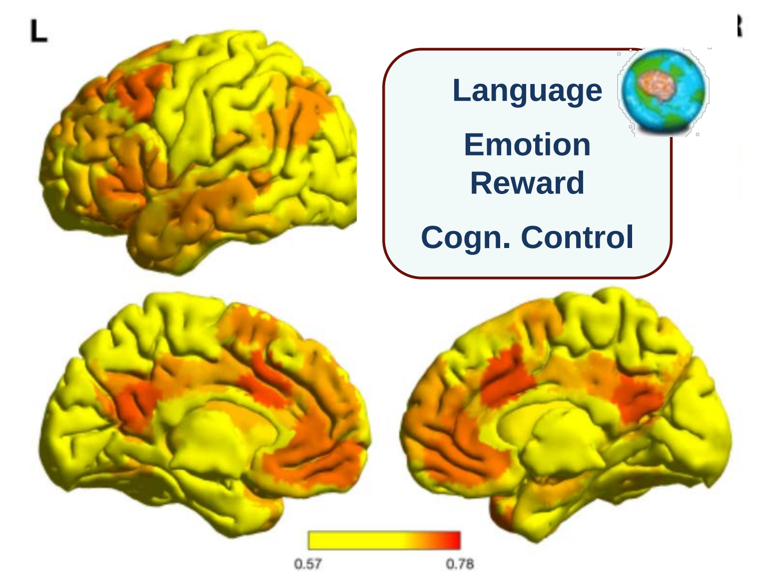
http://www.fz-juelich.de

+ **Good Learning Algorithms** =

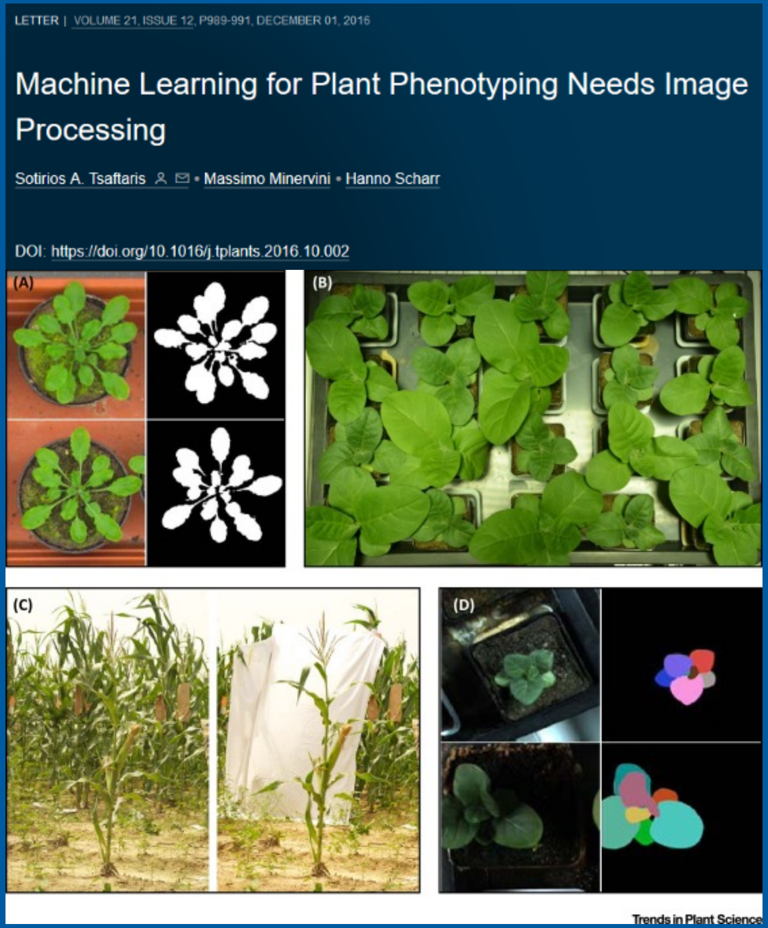
Deep neural networks



Utopia?



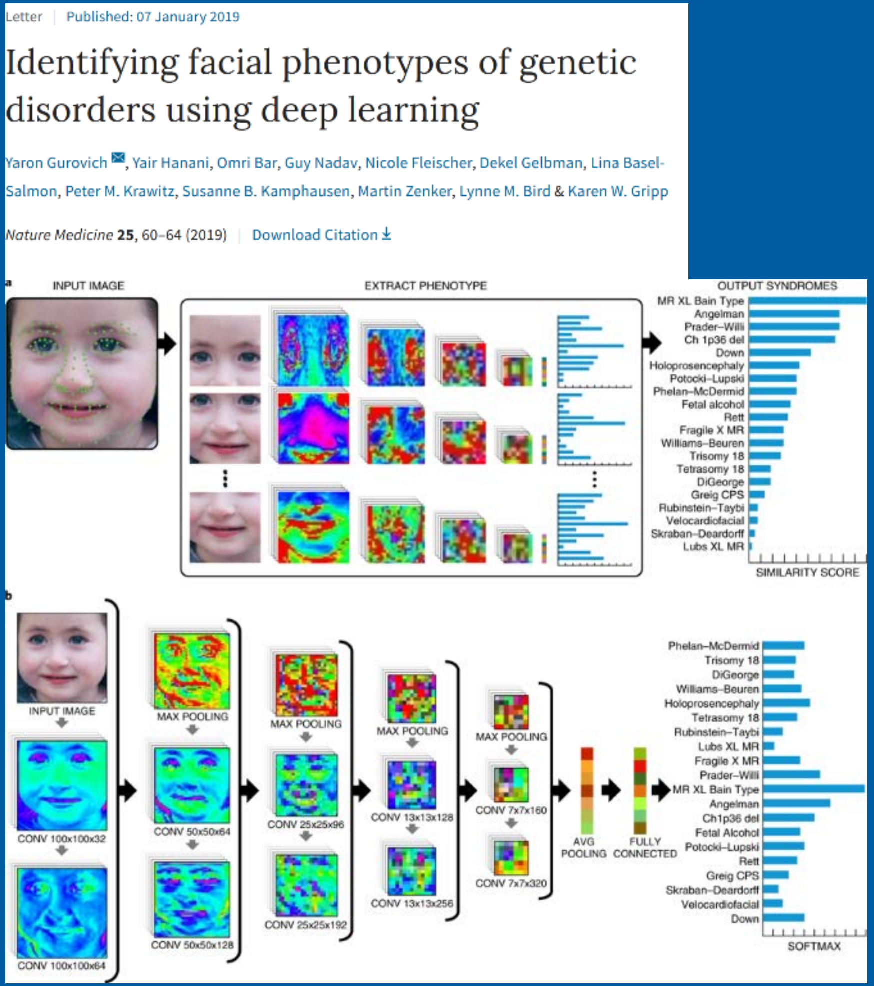
We can accurately predict sex of a new subject from region-wise FC profiles
(SVM, nested optimization, between-sample prediction, N=434/310)


LETTER | VOLUME 21, ISSUE 12, P989-991, DECEMBER 01, 2016

Machine Learning for Plant Phenotyping Needs Image Processing

Sotirios A. Tsaftaris • Massimo Minervini • Hanno Scharr

DOI: https://doi.org/10.1016/j.tplants.2016.10.002

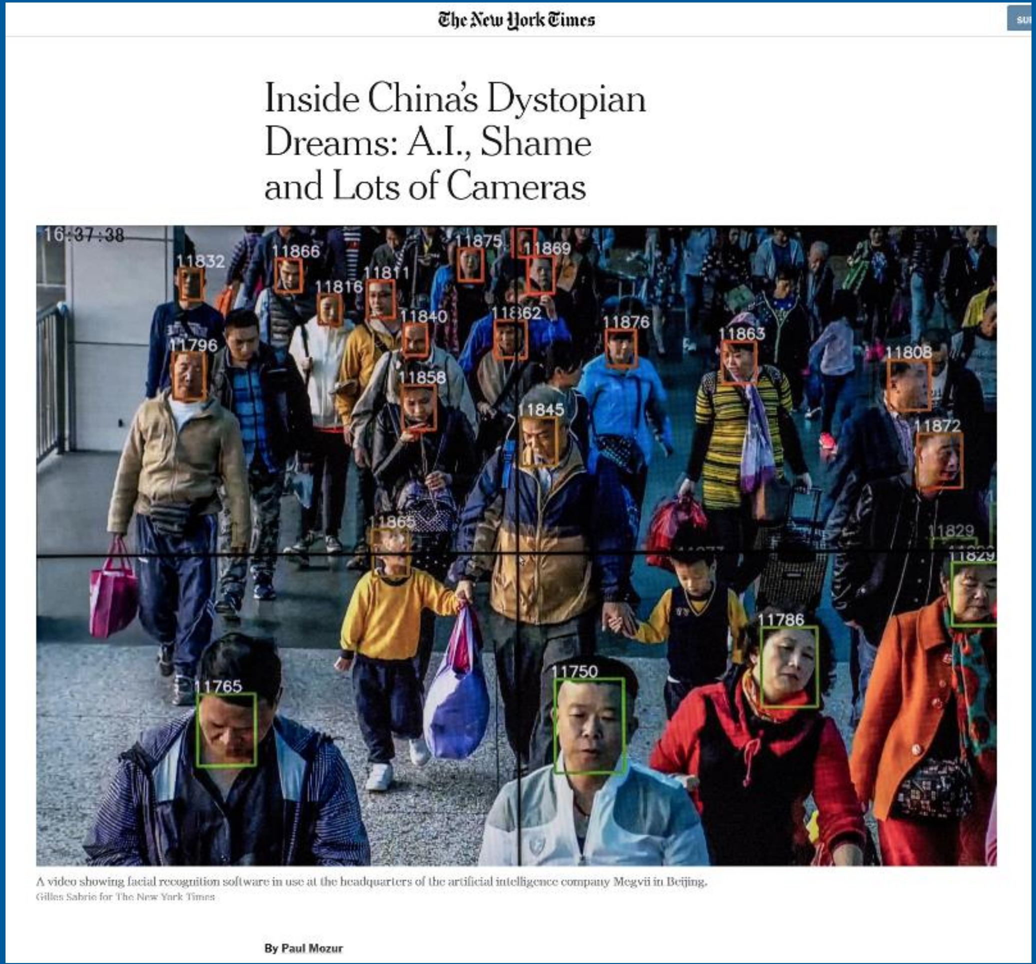

Letter | Published: 07 January 2019

Identifying facial phenotypes of genetic disorders using deep learning

Yaron Gurovich, Yair Hanani, Omri Bar, Guy Nadav, Nicole Fleischer, Dekel Gelbman, Lina Basel-Salmon, Peter M. Krawitz, Susanne B. Kamphausen, Martin Zenker, Lynne M. Bird & Karen W. Gripp

Nature Medicine **25**, 60–64 (2019) | Download Citation

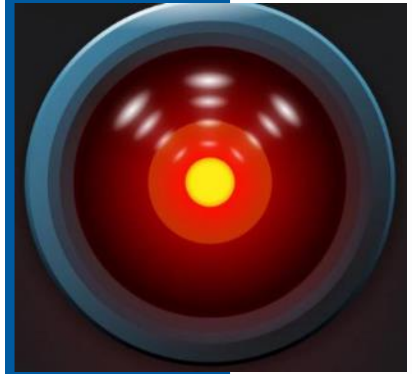
Dystopia?

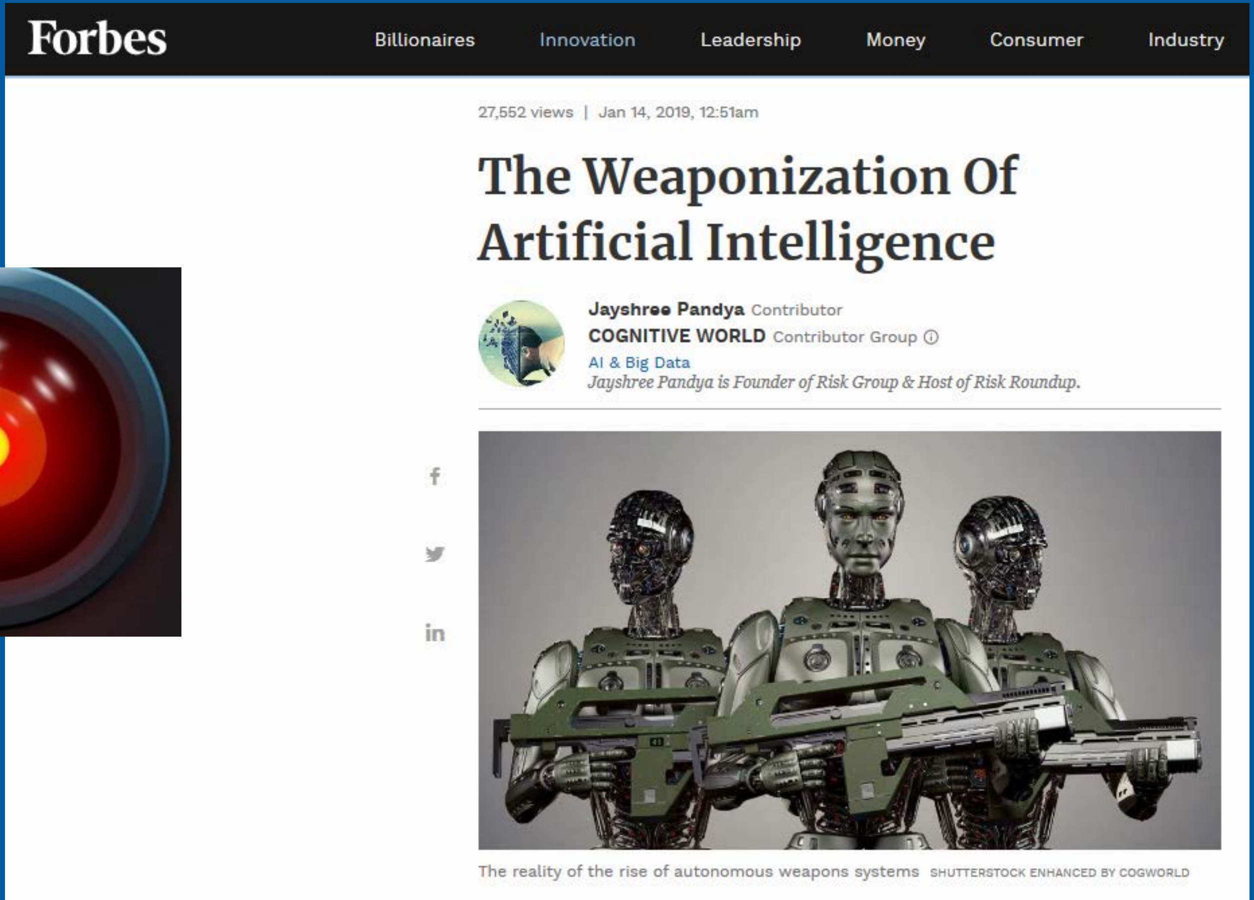

The New York Times

Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras

A video showing facial recognition software in use at the headquarters of the artificial intelligence company Megvii in Beijing.
Gilles Sabrie for The New York Times

By Paul Mozur




Forbes | Billionaires | Innovation | Leadership | Money | Consumer | Industry

27,552 views | Jan 14, 2019, 12:51am

The Weaponization Of Artificial Intelligence

Jayshree Pandya Contributor
COGNITIVE WORLD Contributor Group
AI & Big Data
Jayshree Pandya is Founder of Risk Group & Host of Risk Roundup.

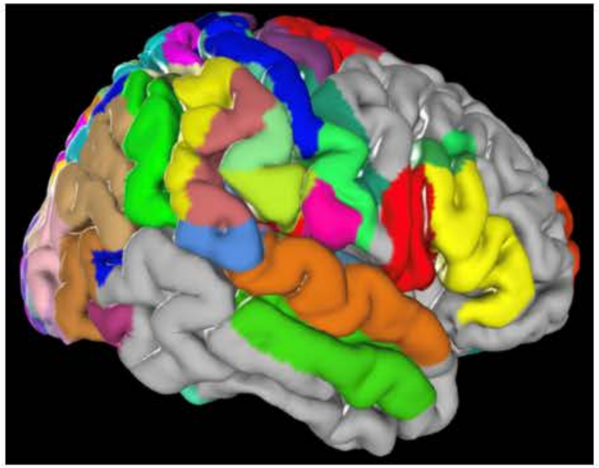
The reality of the rise of autonomous weapons systems SHUTTERSTOCK ENHANCED BY COGWORLD

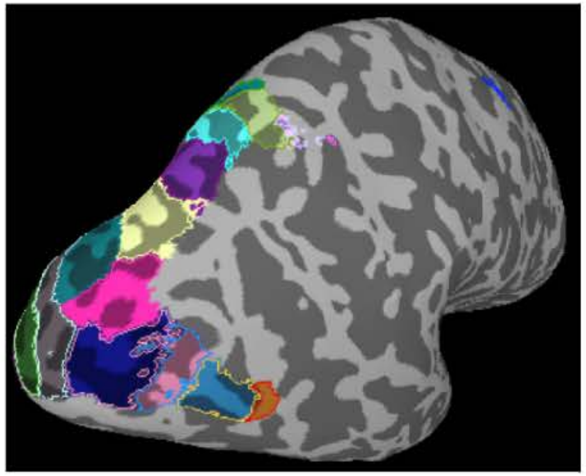
https://www.forbes.com/sites/cognitiveworld/2019/01/14/the-weaponization-of-artificial-intelligence/#20d4a6873686

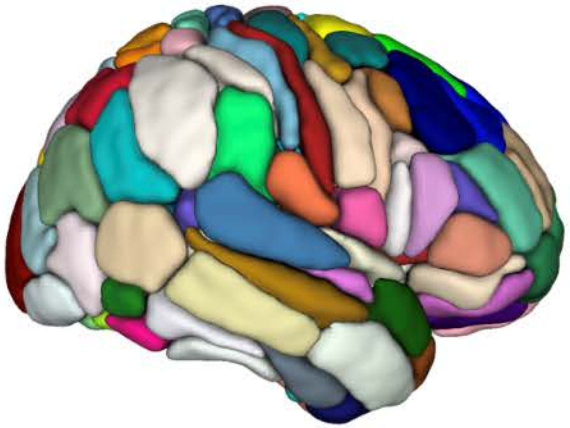
https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

Knowledge on brain organization: functional neuroanatomy

Machine-learning to predict behavioral or clinical phenotypes from MRI-data







Outline

- Machine learning
- Sex classification: replication
- Schizophrenia sybtypes: data separation

Machine learning: why to use it?

Machine learning is not magic; it can't get something from nothing. (Domingos, 2012)

Patterns are often more informative than individual variables

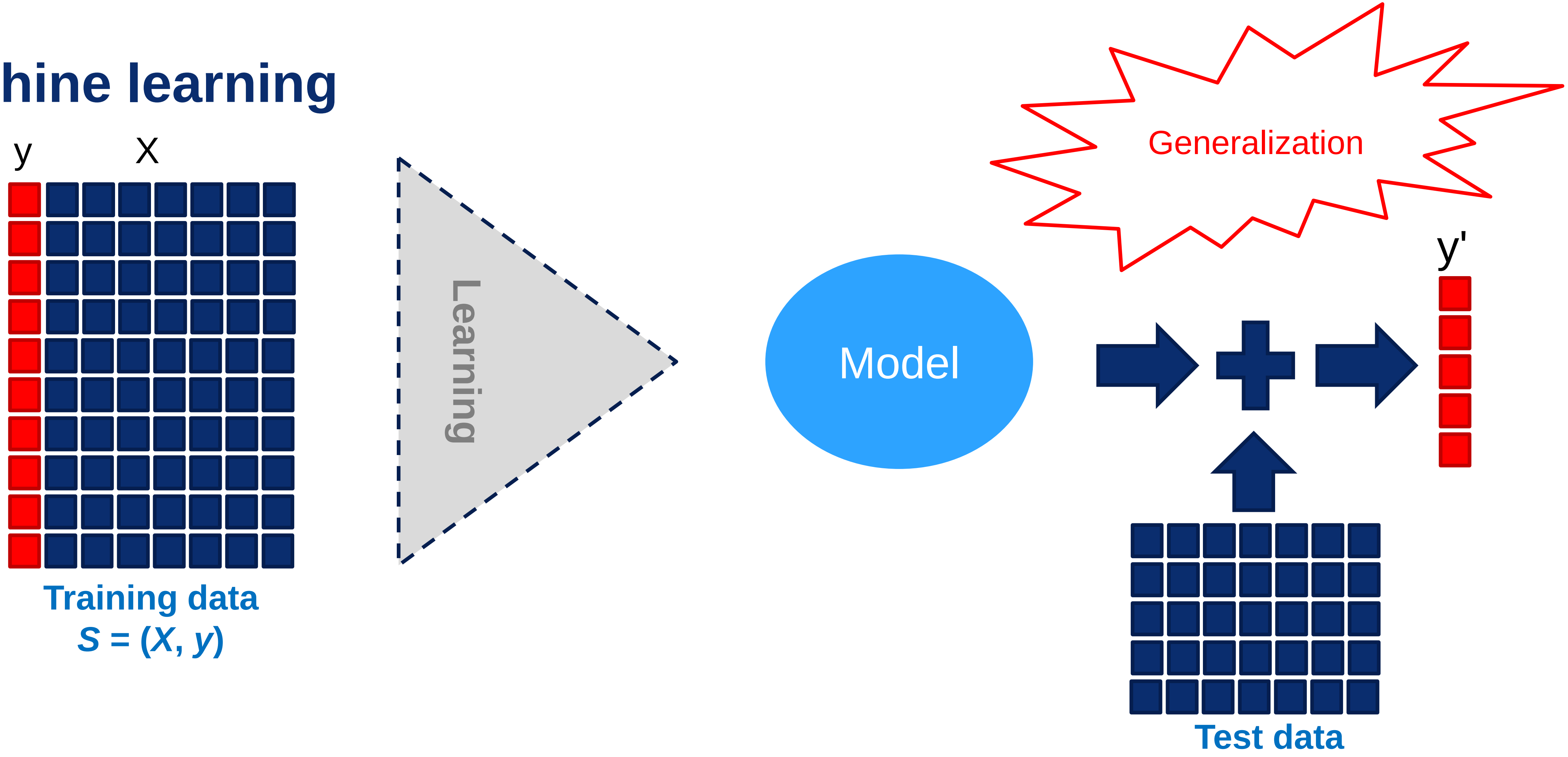
- e.g. any logical function
- Univariate methods cannot identify those
  - Multiple testing correction issue (lower power)

Generalizable solutions that work on unseen data

- Fewer false positives (not guaranteed)
- Predictive analytics
- Practical applications: e.g. clinical status/score prediction

Domingos, 2012 , A Few Useful Things to Know about Machine Learning, CACM

Machine learning



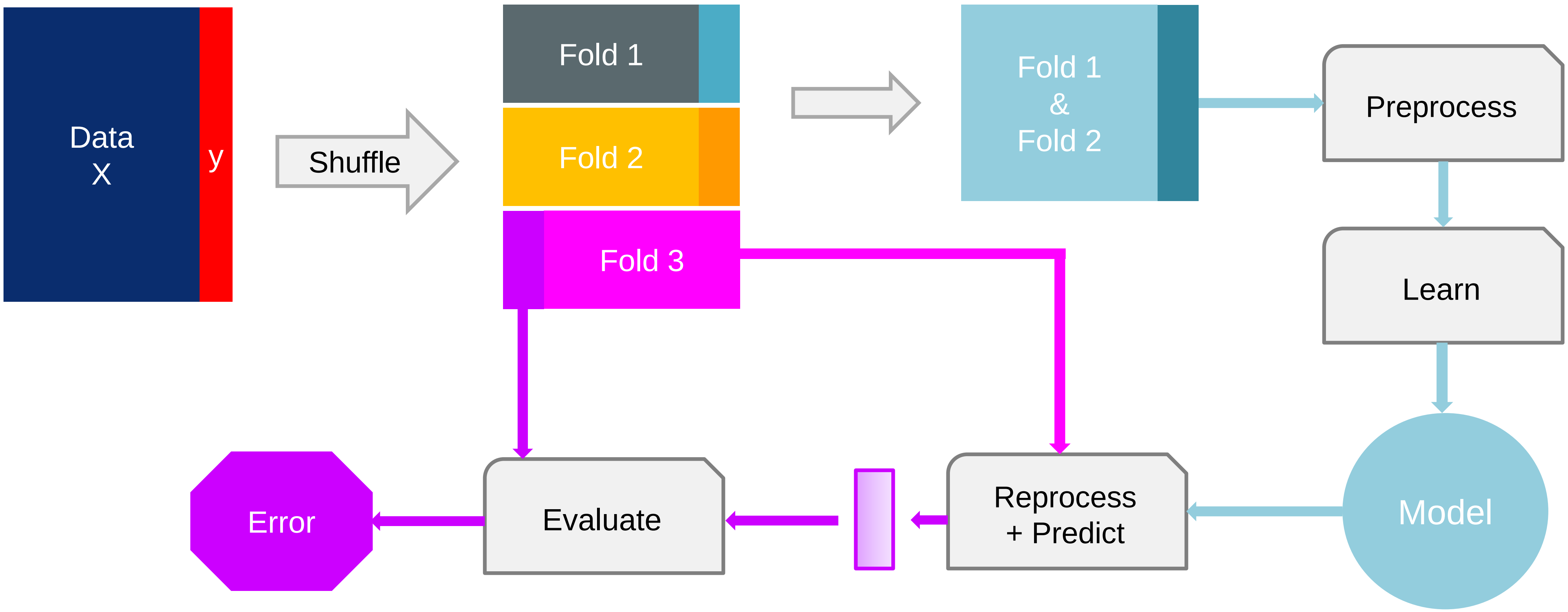
“The fundamental goal of machine learning is to generalize beyond the examples in the training set. This is because, no matter how much data we have, it is very unlikely that we will see those exact examples again at test time.” (Domingos, 2012, A Few Useful Things to Know about Machine Learning)

Challenge: Generalization (avoid over-fitting)

- **The fundamental goal of machine learning is to generalize beyond the examples in the training set. This is because, no matter how much data we have, it is very unlikely that we will see those exact examples again at test time.** (Domingos, 2012)
- But we only one dataset!
- Fit on the complete data
  - Model describes the „training“ data well
  - Fails to generalize on „unseen“ data

K-fold cross-validation

Estimate performance on “unseen” data

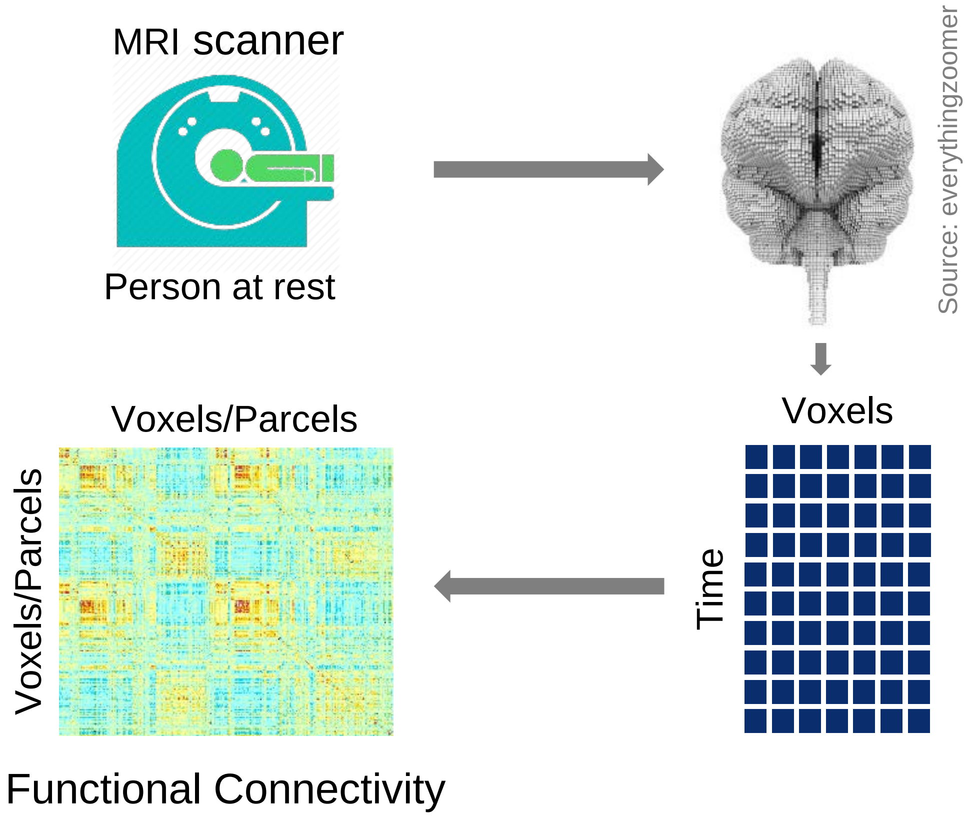


This becomes part of the model.
Mean-centering, Z-score, PCA etc. applied only to the training data.
The parameters should be retained and applied to the test data.

Predict biological traits / clinical status using neuroimaging data

- **Aim:** Generalization models
- **Aim:** Interpretable results
- **Data:** Resting-state data
  - Easy to acquire
  - Intrinsic properties of brain function
- **Issue:** High dimensions
  - Leads to over-fitting

Resting state data



Issue 1: High dimensions

- Over-fitting
  - Curse of dimensionality
- Results might not be interpretable

- **… our intuitions, which come from a three-dimensional world, often do not apply in high-dimensional ones.** (Domingos, 2012)

Our approach

a priori feature reduction: parcel-wise or pre-defined networks

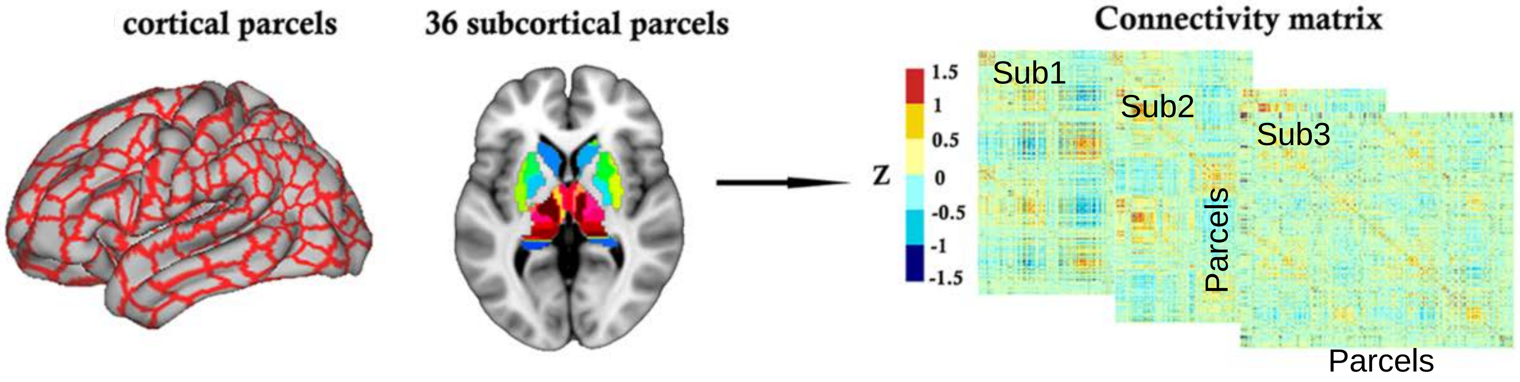
Whole-brain analysis

- Too many features
  - 200 nodes: ~20,000 features
  - #features **>>** #subjects
- Machine learning
  - Need for feature selection
  - Accuracy can suffer
- Interpretation is difficult

Parcel/Network-based analysis

- Reasonable number of features
  - 200 nodes: ~200 features
  - #features **>** #subjects
- Machine learning
  - a priori feature selection
  - Better predictions
- Interpretable results

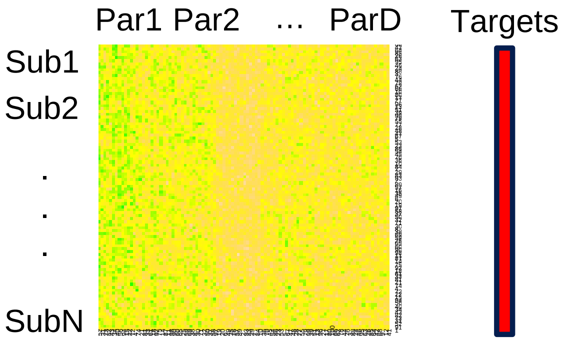
Parcel-based classification



Extract parcel-wise connectivity

Parcel1 Sub1 

Parcel1

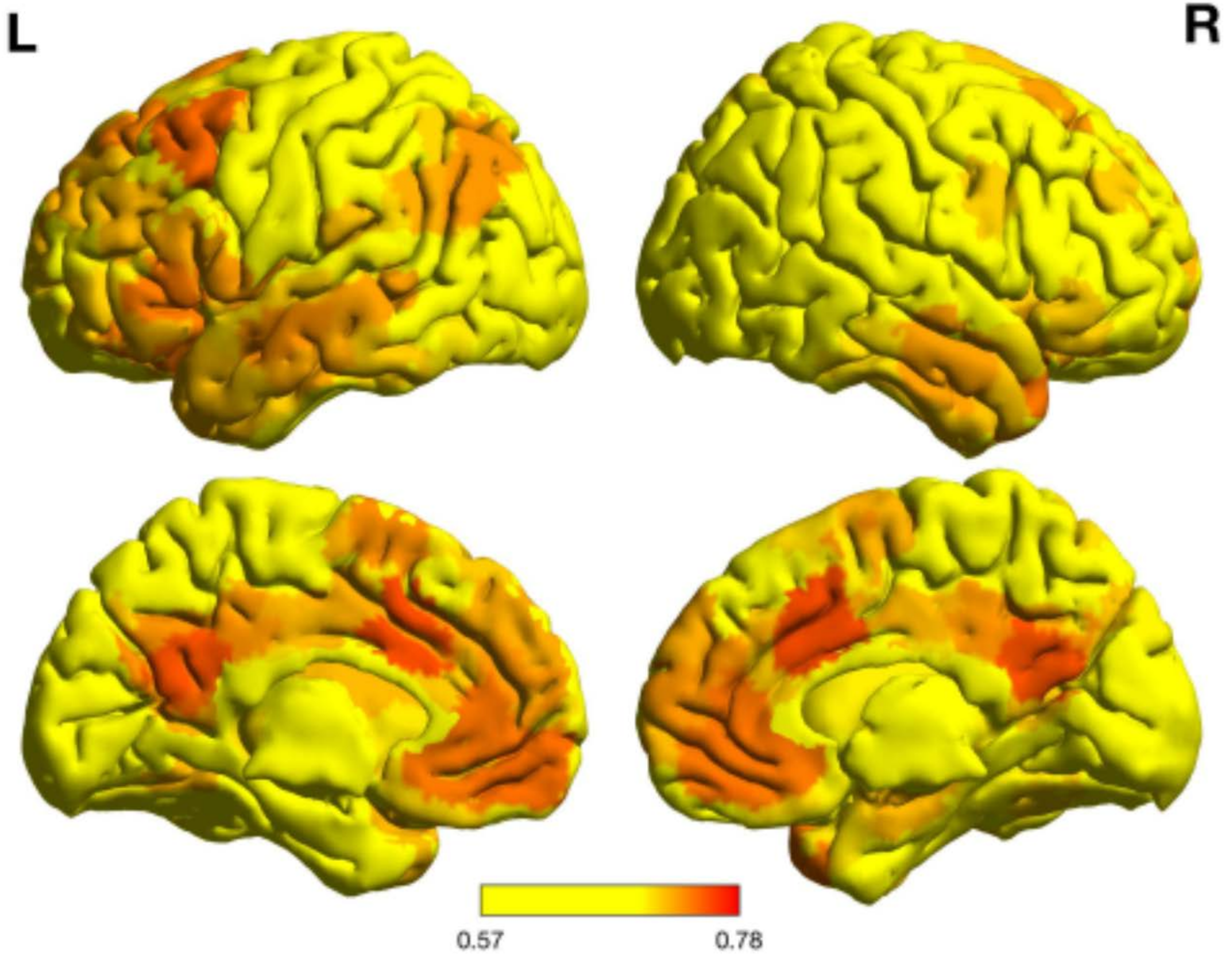


Cross-validation analysis

Accuracy of Parcel 1

Repeat for other parcels

Mapping fingerprint – phenotype relationships



BrainMap database

Language

Emotion
Reward

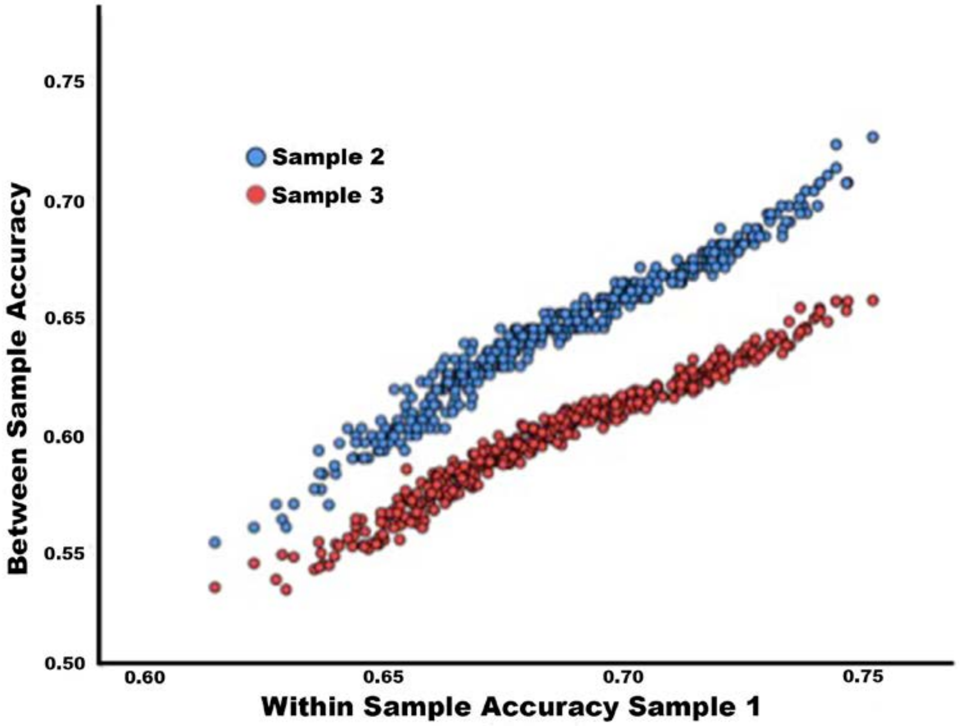
Cogn. Control

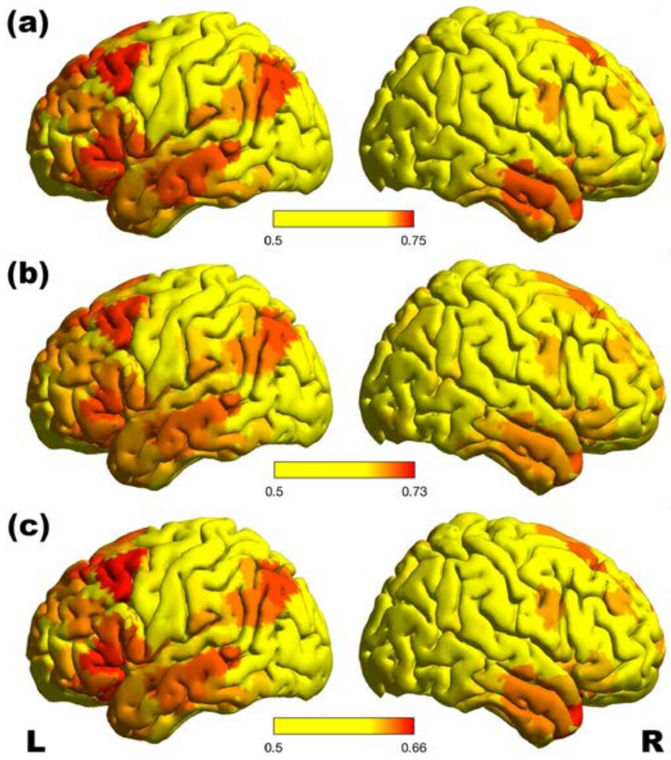
Accurate prediction sex of a new subject from region-wise FC profiles
(SVM, nested optimization, between-sample prediction, N = 434 / 310)

Cross-sample prediction: generalization check

Susanne Weis

Weis, Patil et. al,
Sex Classification by Resting State Brain Connectivity,
Cerebral Cortex 2019





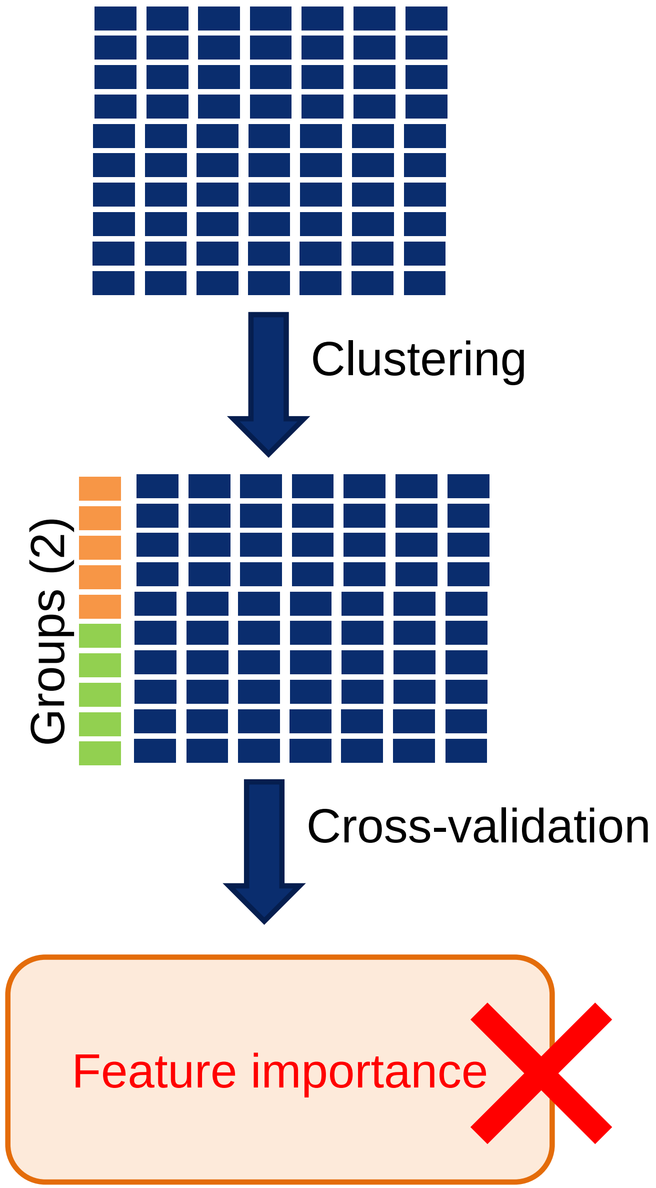
Similar performance on other datasets. Solution is generalizable.

Sample 1 and 2 = Human Connectome Project

Sample 3 = FZJ 1000 brains study

Issue 2: Double-dipping (limited data)

- Over-fitting
  - Data-leakage
- Misleading results (false positives)



Our approach

Two-step solution

Step 1: Symptomatology

- Subtypes
  - Clinical symptom scales
  - Factorization
  - Clustering analysis

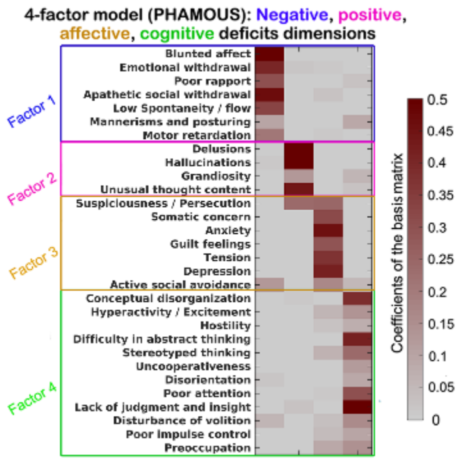
Step 2: Neuroimaging

- Resting-state data
- Subtypes from Step 1
- Cross-validation analysis
- Interpretable results

Mapping fingerprint – pathology relationships

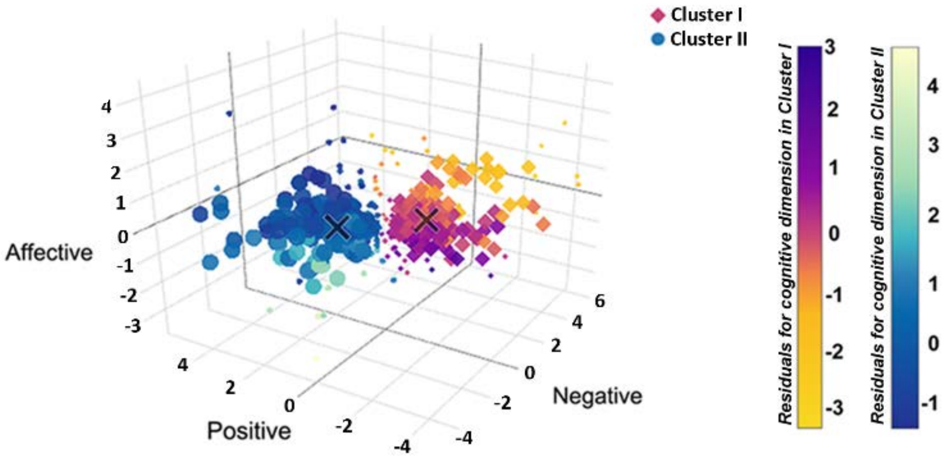
Step 1: Symptomatology → Groups

Robust low-rank description of SCZ psychopathology from >2000 patients



Orthogonal Non-Negative Matrix Factorization

Two core phenotypical subtypes



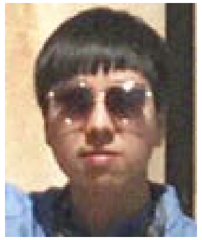
K-means
Gaussian Mixture Modelling

Schizophrenia subtypes and brain basis

Step 2: Groups → Brain regions

Chen, Patil et. al,
Neurobiological divergence of the positive and negative schizophrenia subtypes identified upon a new factor-structure of psychopathology using non-negative factorization: An international machine-learning study,
Biological Psychiatry 2019 (accepted)

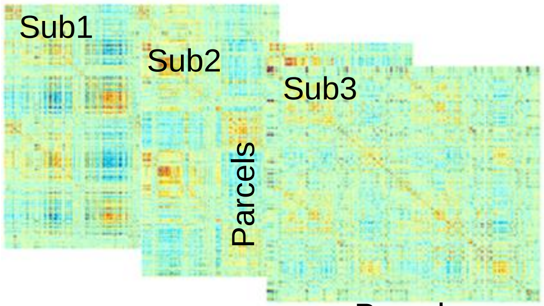
Ji Chen

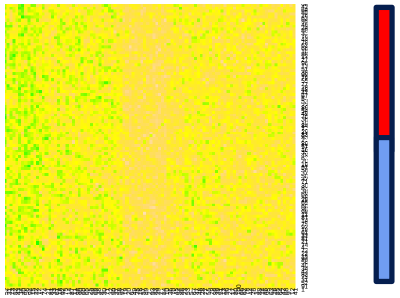


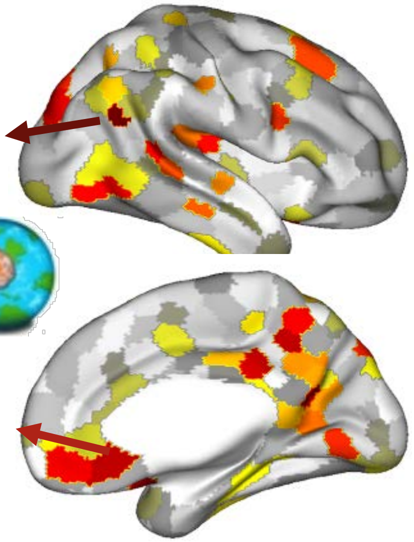
rs-fMRI data N=84

Parcel-wise classification

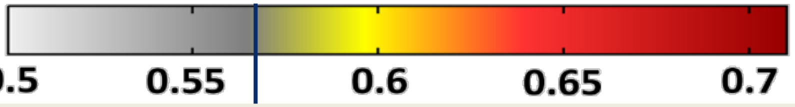
Connectivity matrix







Region-based classification accuracy



Whole brain connectome classification accuracy: 0.57

Regional FC profiles support subtype discrimination

Other issues

- **Feature selection / construction**
  - “Over-optimization is root of all non-generalization”
  - Solutions are often over-fitted
- **Data privacy**
  - Fingerprint analysis
    - 95% identification with high quality scan
  - Deep networks can retain too much information
- **… developing successful machine learning applications requires a substantial amount of “black art” that is difficult to find in textbooks.** (Domingos, 2012)

Thank you!

Funding

HBP

HORIZON 2020

DFG

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

NATIONAL INSTITUTES OF HEALTH

Helmholtz Portfolio Theme
SMHB
Supercomputing and Modeling for the Human Brain

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF
SYSTEMS NEUROSCIENCE

Brain & Behaviour

JÜLICH
Forschungszentrum

Rank Selection in Non-negative Matrix Factorization

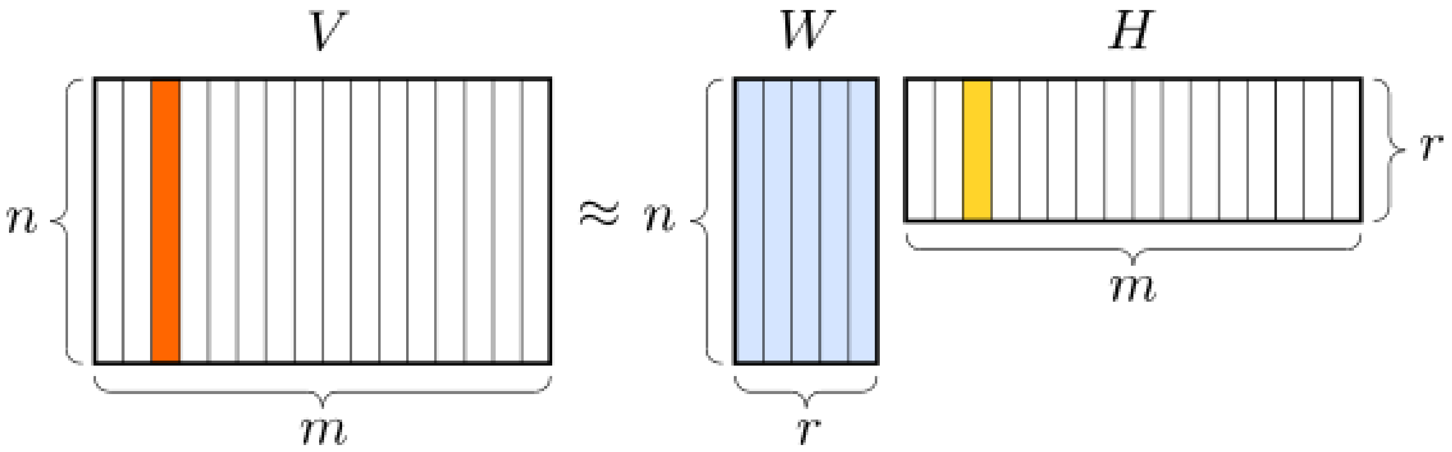
Laura Muzzarelli

Muzzarelli, Weis, Eickhoff & Patil,
Rank Selection in Non-negative Matrix Factorization: Systematic Comparison and a New MAD Metric
IJCNN 2019

- Non-Negative Matrix Factorization
  - Powerful dimensionality reduction
  - Rank Selection Problem
- Rank Selection Methods
  - Stability vs. imputation
  - Our proposal: MADimput
- Data properties impact rank selection
  - Sparsity
  - Intrinsic dimensionality

Non-negative matrix Factorization (NMF)

Basic properties of NMF



n = features

m = observations

r = reduced components

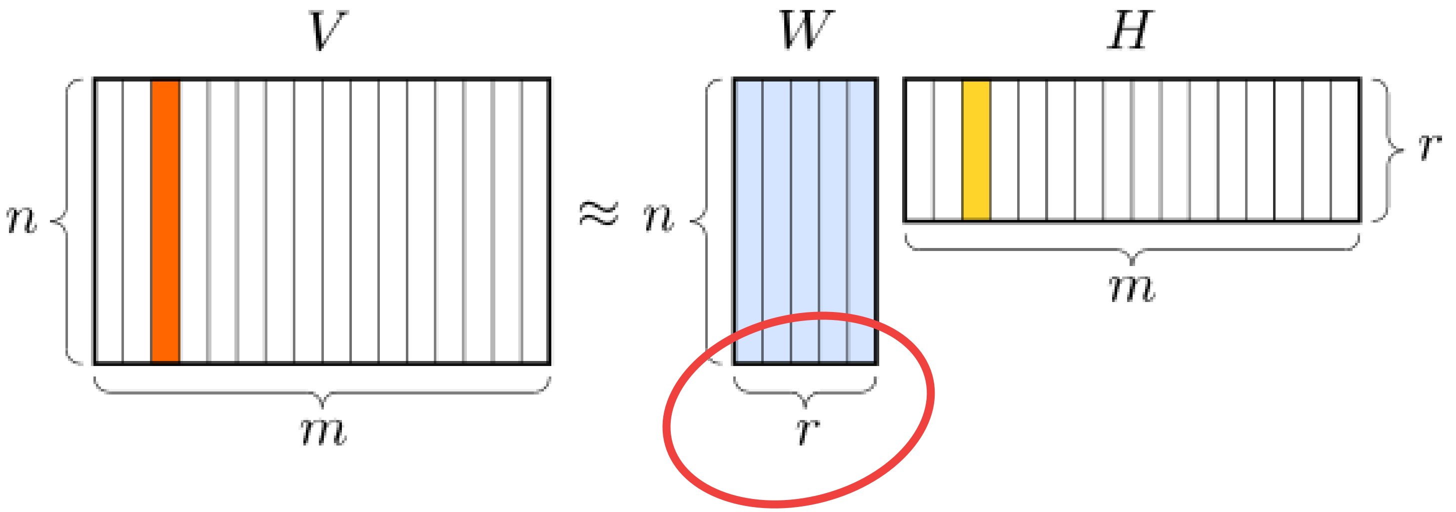
$$||V - WH||_2 \quad subject\ to\ W \geq 0,\ H \geq 0$$

part-based representation

interpretability

Rank selection problem

Need to find dimensionality of reduced representation



n = features

m = observations

r = reduced components

Crucial to select "best" rank

mostly when no prior knowledge

what is "best"
& which metric to use?

Rank selection approaches

Stability – based

Perform multiple NMF runs at each rank

Check stability of factor matrix

does NOT ensure accuracy

degeneracy

Imputation – based

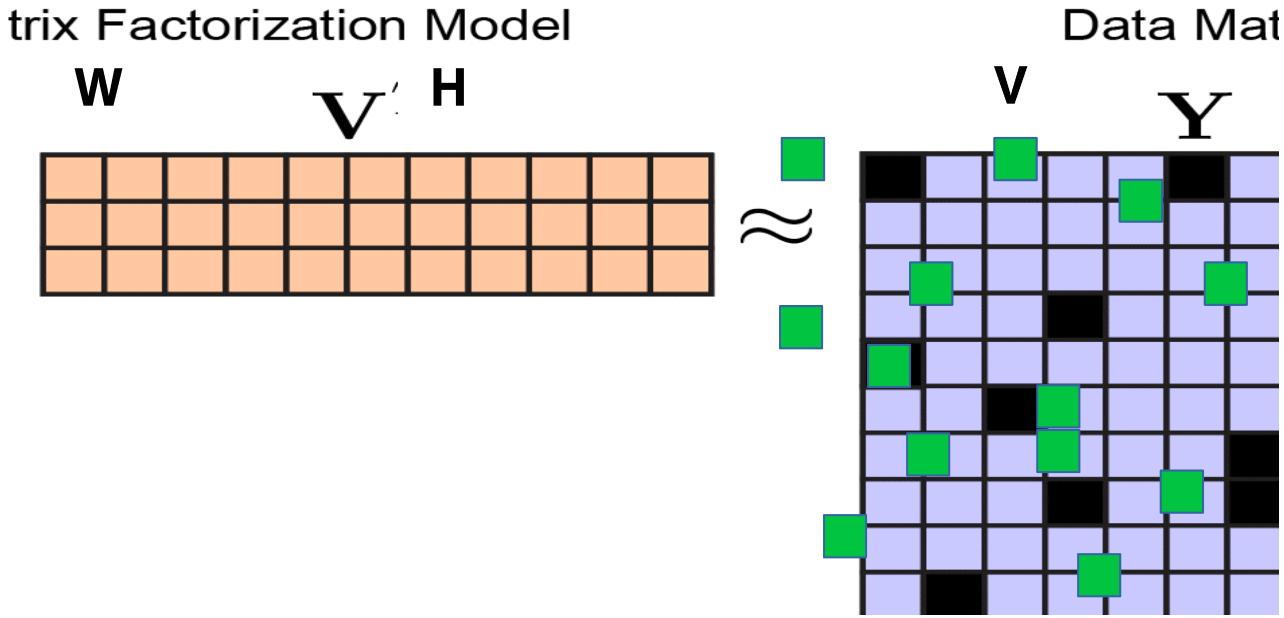
Exclude random data points in multiple CV runs

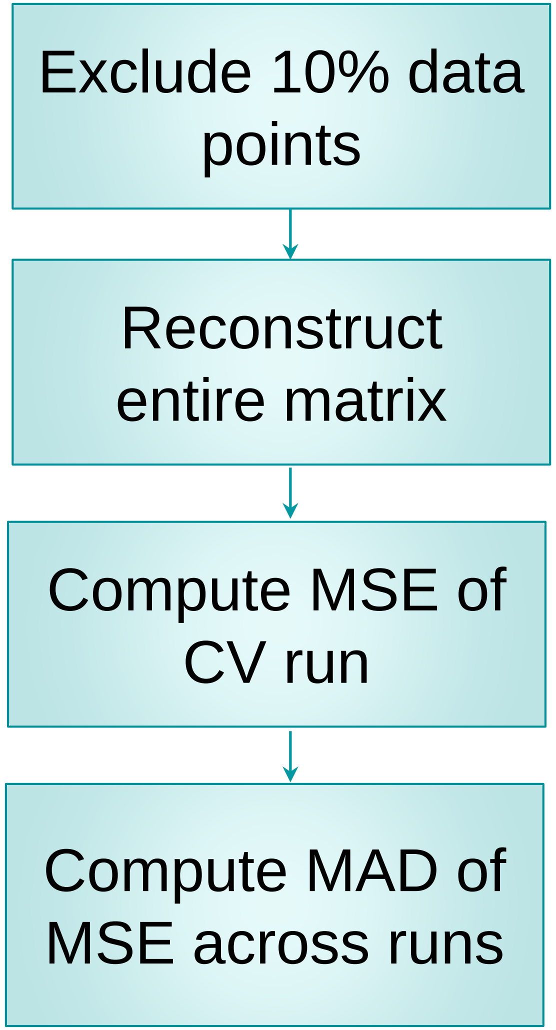
Compute reconstruction error in imputed points

Calculates Accuracy

Our proposal: MADimput in ImputationCV

Good reconstruction, and homogeneous





Systematic evaluation

Performance comparison of 6 metrics

Stability

- Consensus
  - **coph -** cophenetic coefficient
  - **disp -** dispersion
- Stability in split-half CV
  - **aRI -** adjusted Rand Index
  - **inner -** inner product

Imputation

- **MSE -** mean of MSE in CV runs
- **MAD -** MAD of MSE in CV runs

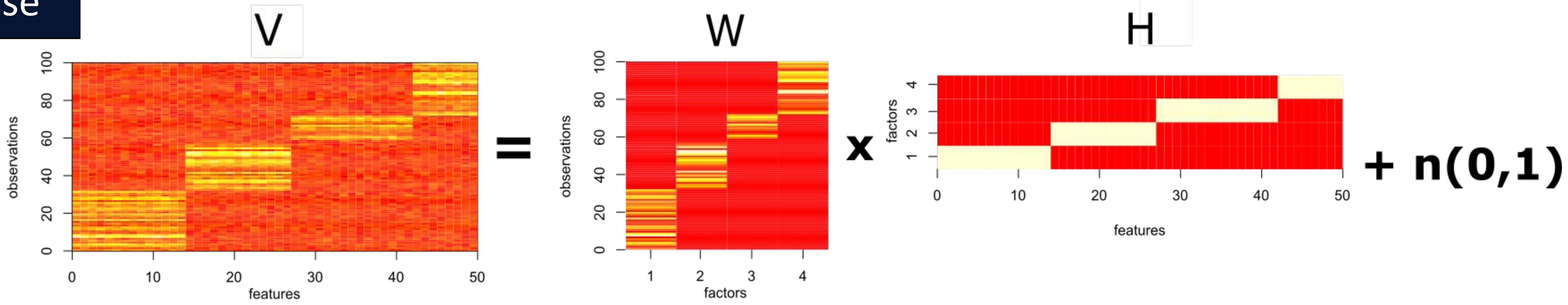
Permutation of underlying distribution

- **perm** – error slope comparison with permuted data

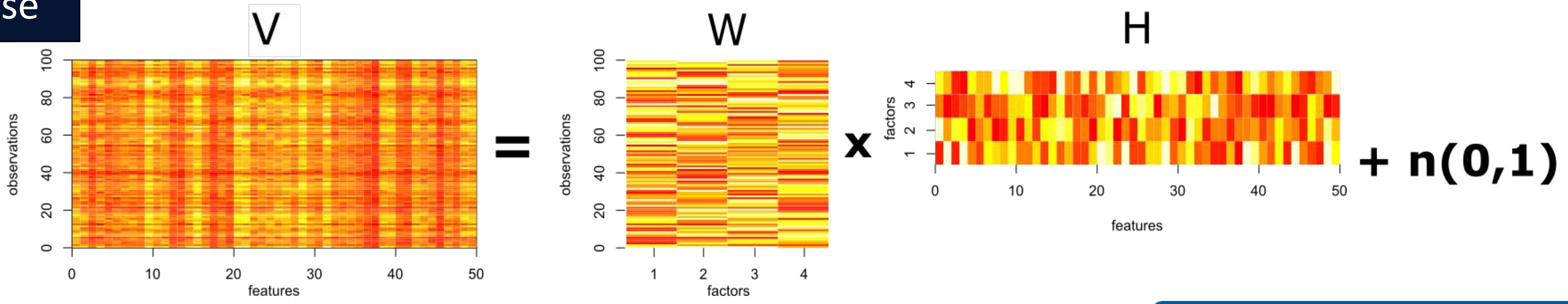
Effect of data properties

Simulated data : Manipulation **sparsity** + **latent dimensionality**

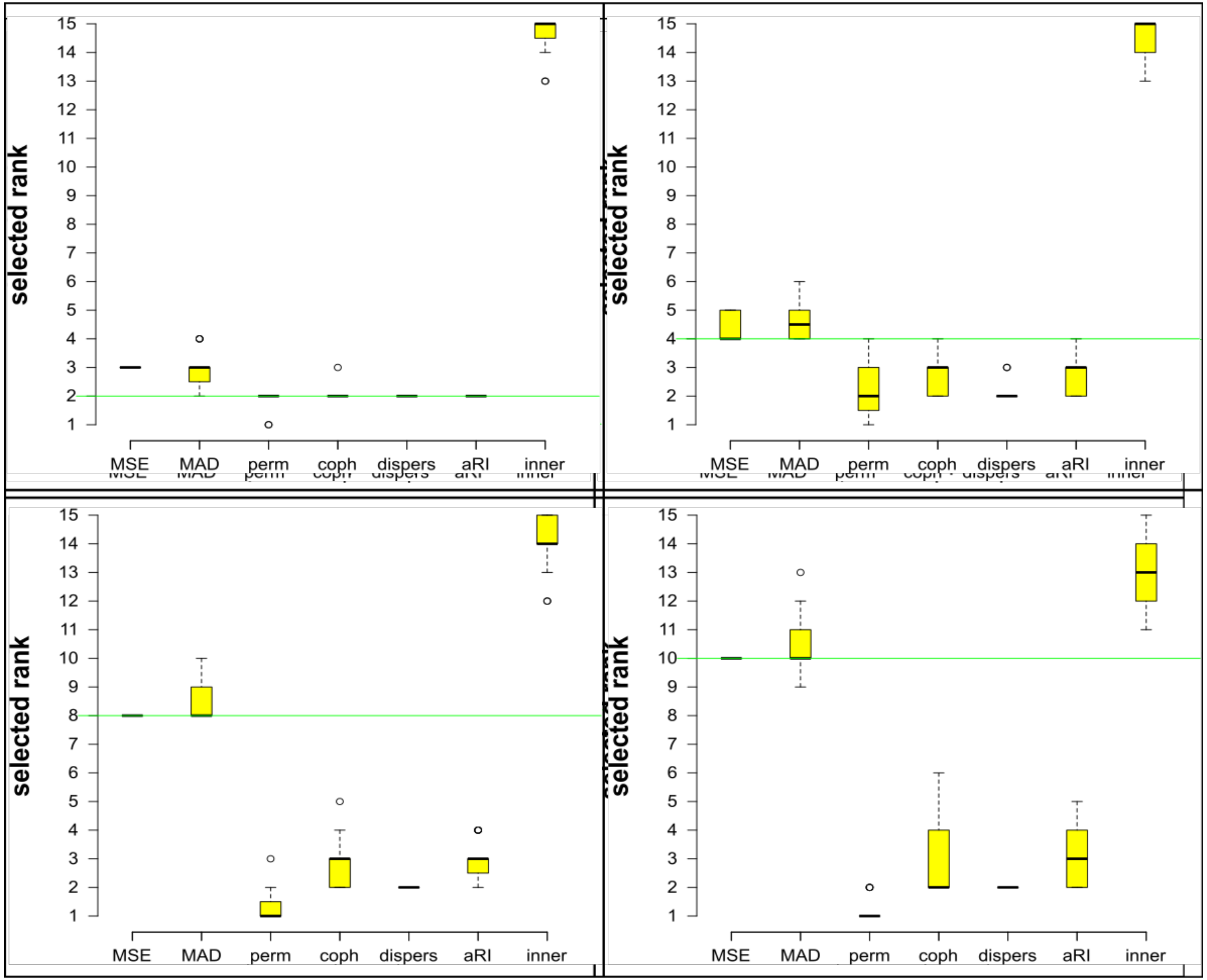
sparse



dense



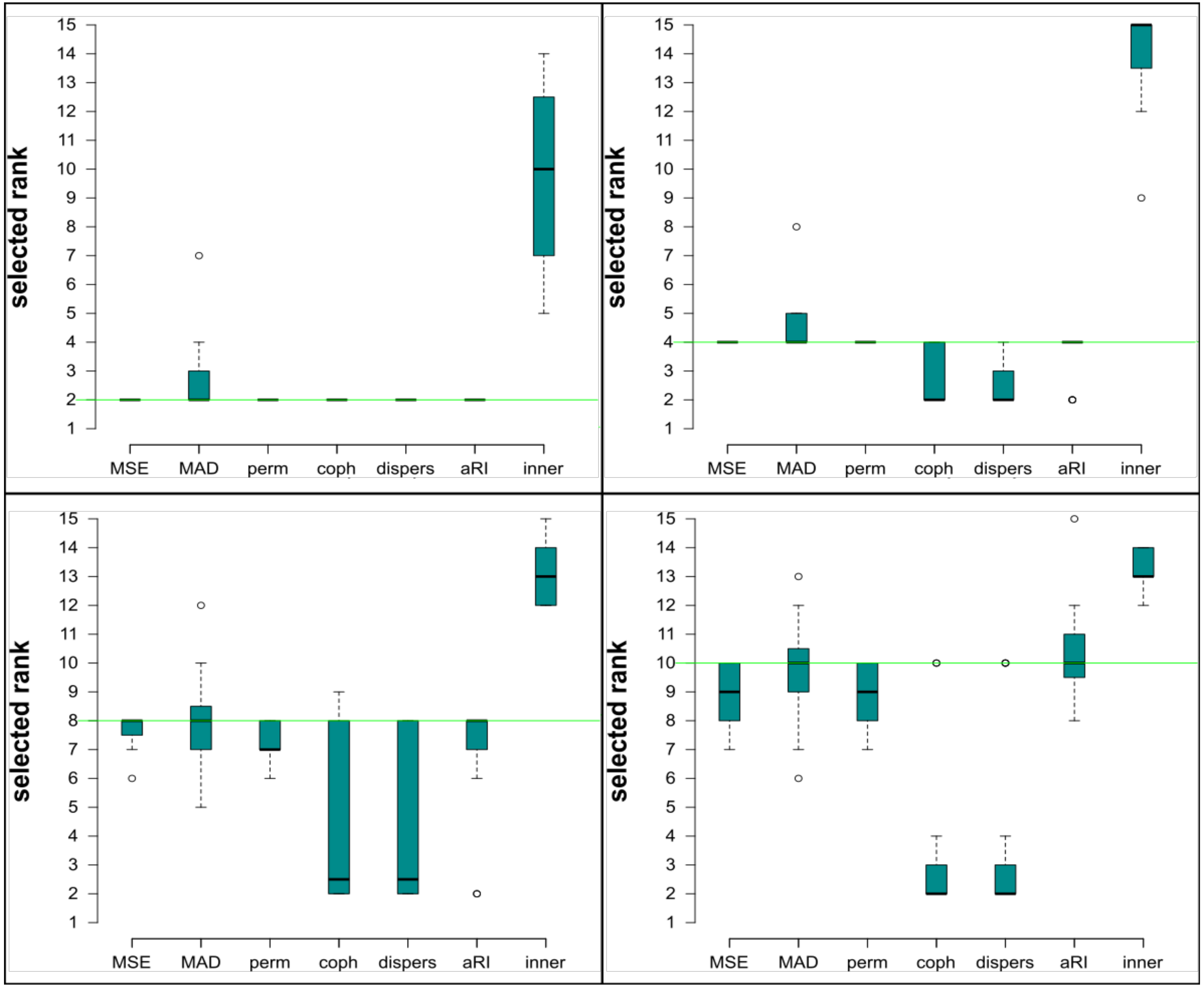
Results: Simulated data dense



Only imputation methods are accurate

(but struggle at low true rank)

Results: Simulated data sparse



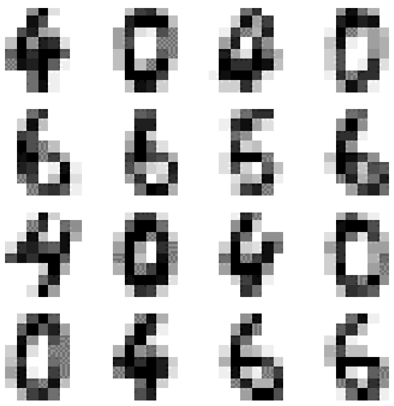
Imputation, permutation and aRI methods have good performance,

but all struggle at higher rank

Real data: sources

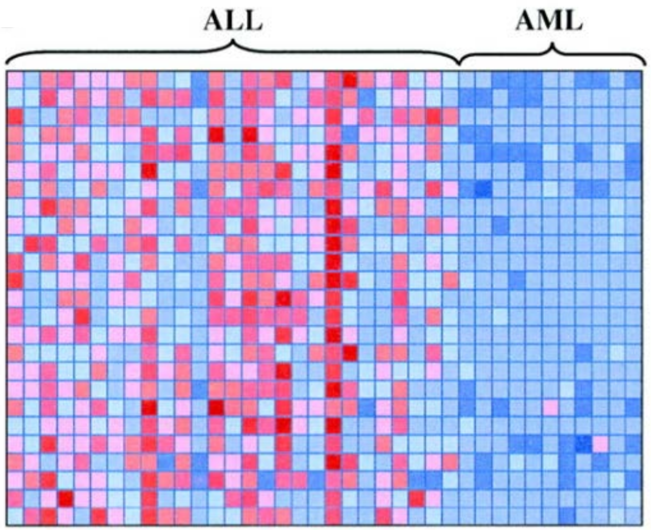
MED5 - medical abstracts

1159 terms x 124 abstracts
5 underlying human-labelled topics



dig0246 – handwritten digits recognition

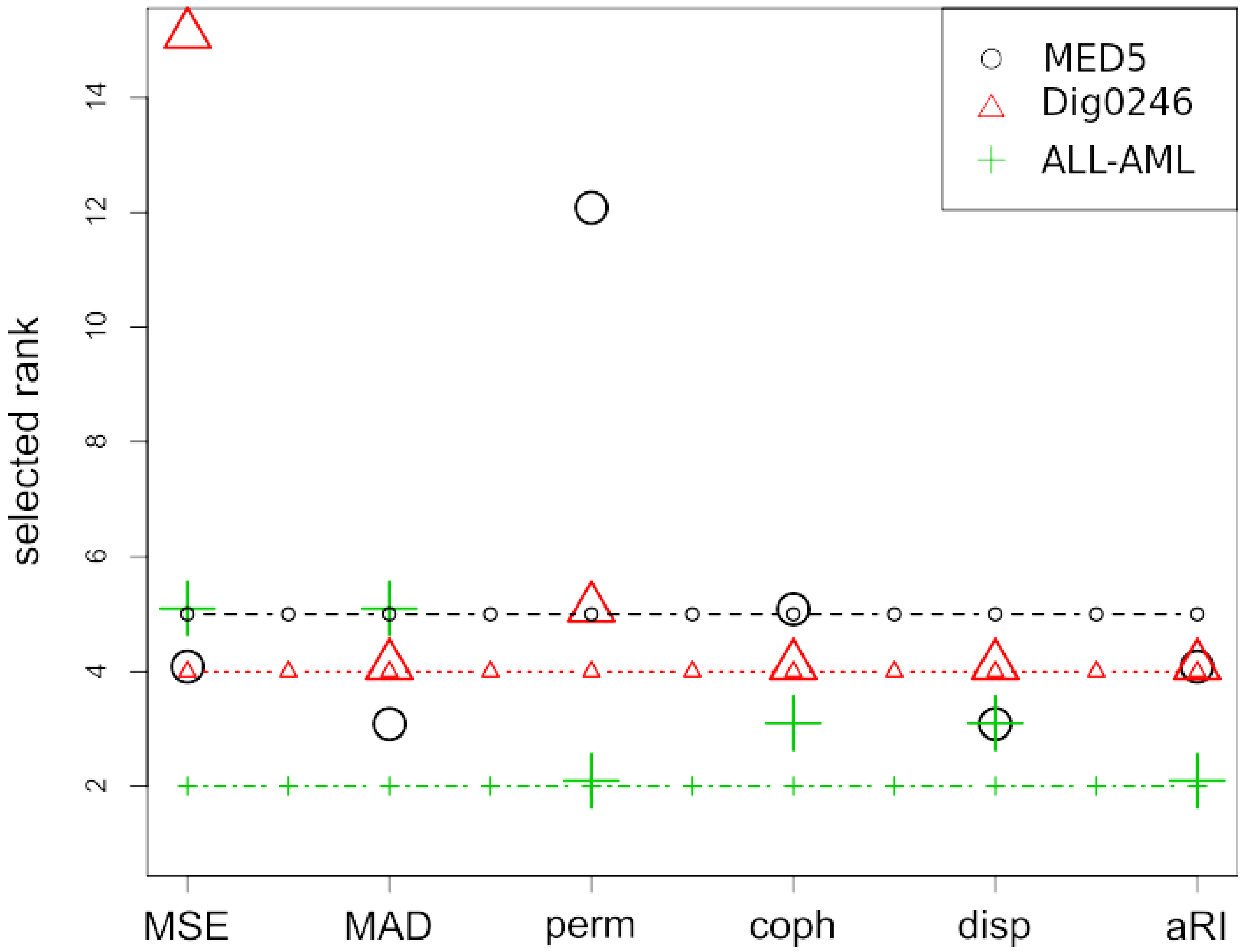
64 attributes and 1520 samples representing digits {0, 2, 4, 6}
→ true rank: 4



ALL-AML – cancer gene expression

5000 genes x 38 samples
2 (more?) myeloma types

Results: Real data



MADimput, aRI and consensus methods close to expected true rank

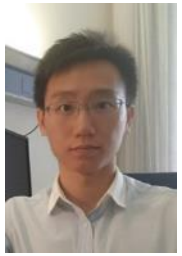
but consensus methods failed badly in simulated !
→ false hit for low rank ?

NMF rank selection: discussion

- No method is perfect, and most are just bad
    - No methods works best in all data type & dimensionality scenarios
- ImputationCV–based methods are better
    - Imputation CV-based MSE and MAD overall more reliable
    - MAD captures more complex properties (as expected) ?
- Data properties do impact rank selection
    - Both sparsity and latent dimensionality
    - Tip of the iceberg ?

Multi-modal parcellation of the human striatum

Xiaojin Liu

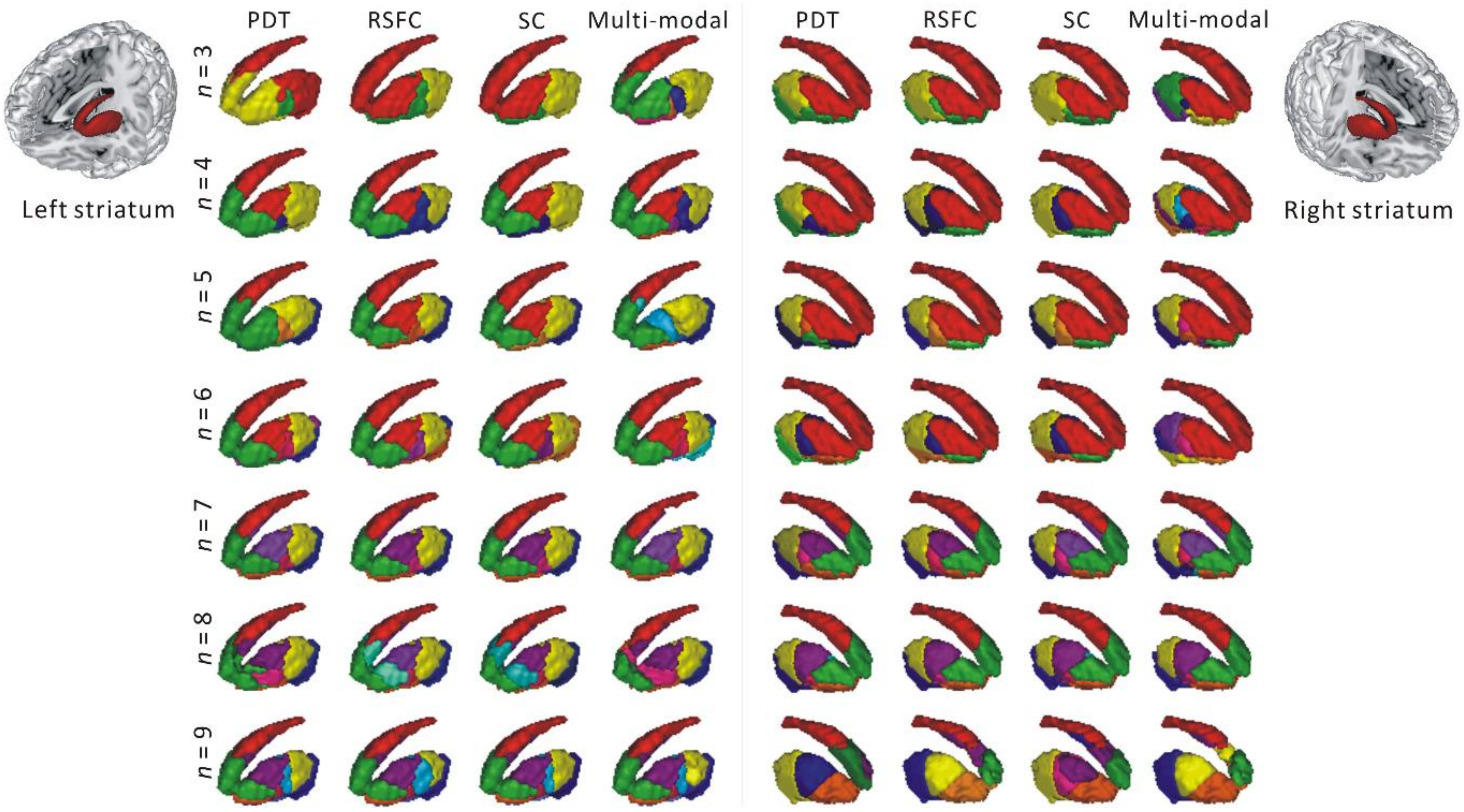


- Most parcellation studies are based on single modality
- Fundamental organization convergent across modalities not known
- Three modalities
  - Resting-state (RS)
  - Probabilistic Diffusion Tractography (PDT)
  - Structural Covariance (SC)
- Context-dependent-clustering (CDC, Gabasova et. al, 2017)
  - Can cluster across contexts (i.e. modalities)
  - Model selection based on several criteria
- Behavioral decoding
- Clinical relevance: VBM analysis Parkinson’s and Schizophrenia

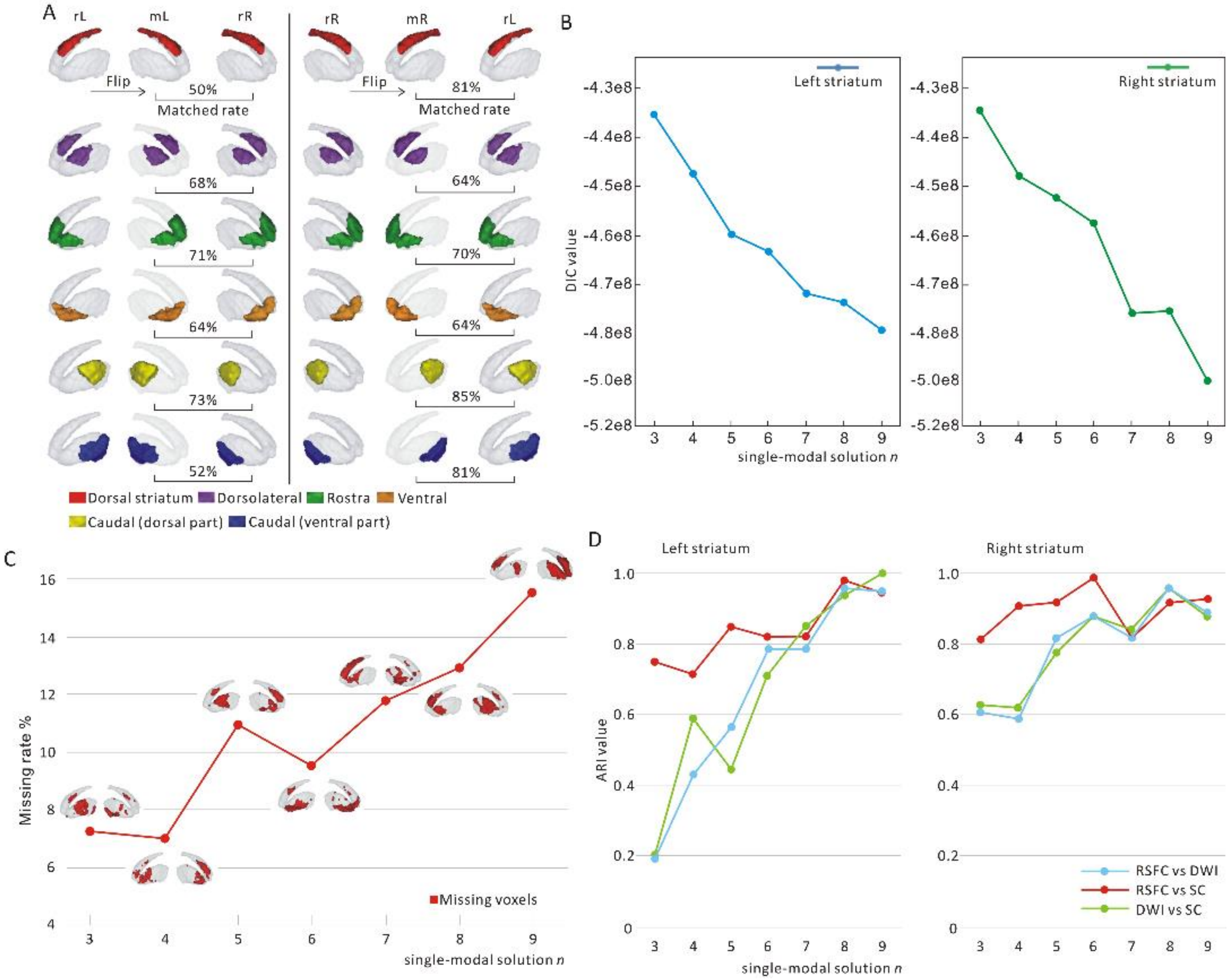
N = 324 (164 Female) Human Connectome Project

CONNECTOME
COORDINATION FACILITY

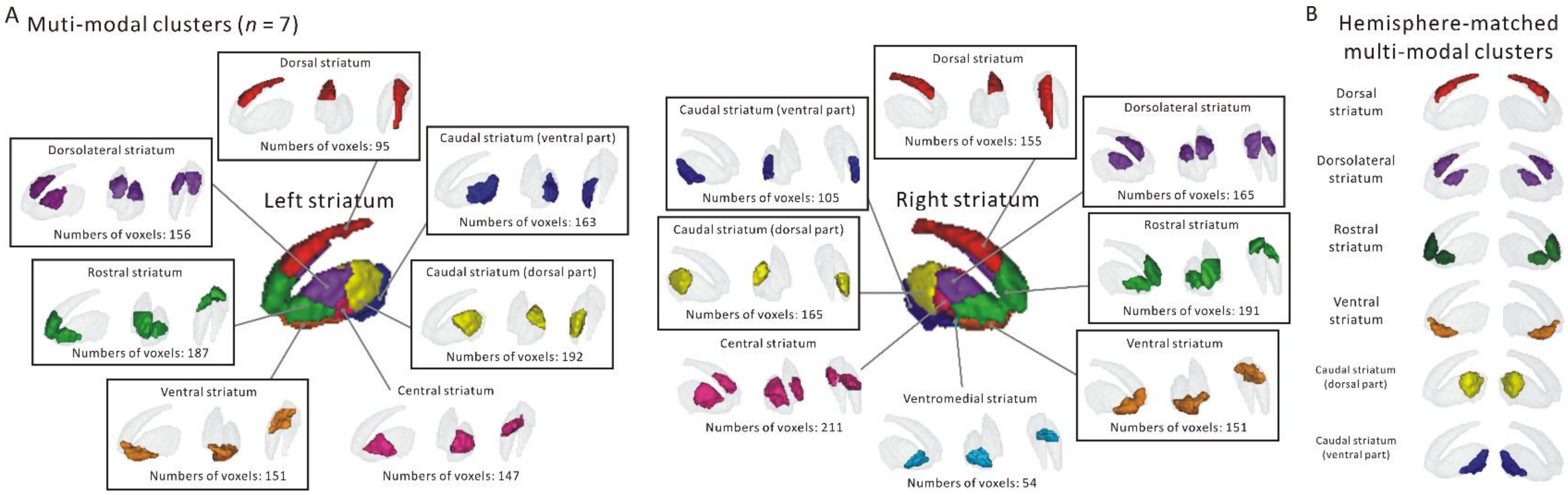
Multi-modal striatum parcellation



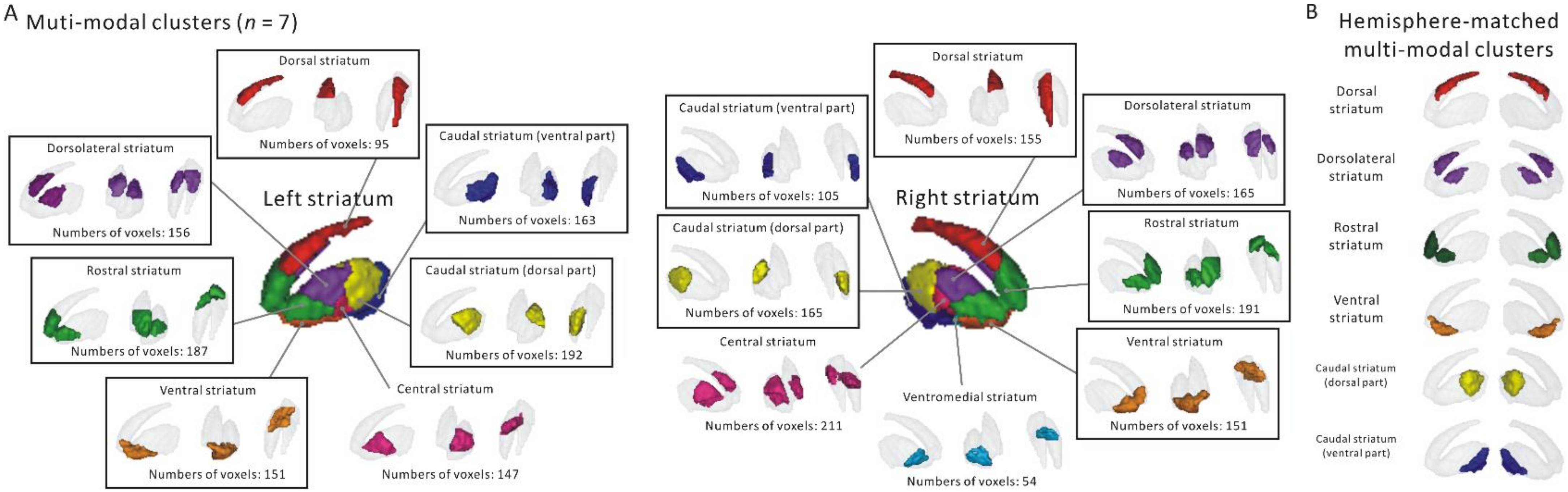
Multi-modal striatum parcellation: model selection



Multi-modal striatum parcellation: selected solution

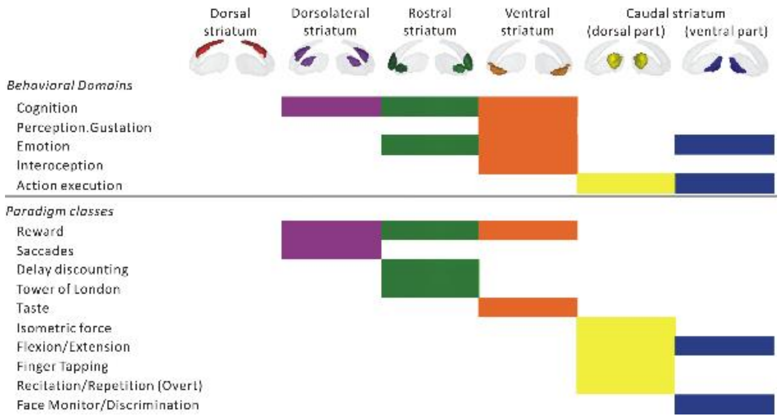


Multi-modal striatum parcellation: selected solution n=7

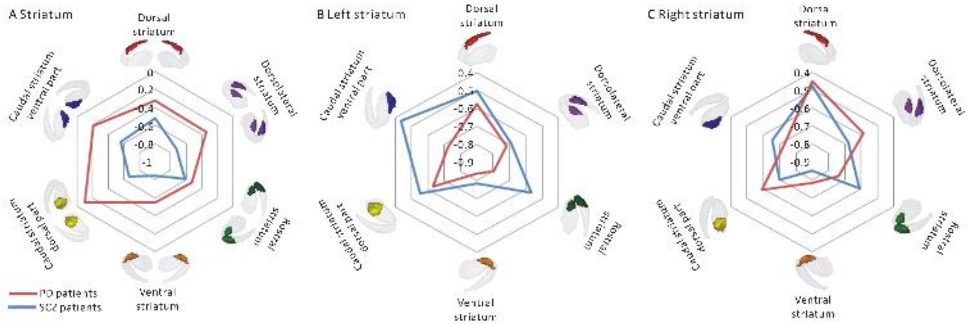


Behavioral decoding

brainmap



PD vs. SCZ
Averaged
Z-scores of
GM volume



Ji Chen: SCZ study

Sample details

Characteristics	PHAMOUS sample (N=1545)	International dataset from 9 centers (N=490)	International dataset with imaging (N=147)	Statistics	p-value
Demographic					
Age (years)[a]	44.15 (11.42)	33.82 (10.28)	34.89 (11.67)	183.51	<.001
Gender (male/female)	1108/437	333/157	102/45	2.45	.292
Illness during (years)[b]	18.22 (10.54)	9.13 (8.98)	11.37 (10.36)	134.71	<.001
PANSS					
Positive[c]	12.48 (4.91)	14.24 (5.76)	15.36 (5.50)	37	<.001
Negative	14.60 (6.20)	14.67 (7.21)	15.07 (6.06)	0.375	.687
General[d]	26.70 (8.16)	29.10 (11.34)	30.93 (10.97)	23.67	<.001
Illness severity (Total PANSS)[e]	53.78 (16.35)	58.01 (21.87)	61.36 (19.57)	19.48	<.001
P3 item (hallucinations)[f]	2.30 (1.47)	2.66 (1.83)	3.22 (1.91)	28.18	<.001
Medication[g]					
Atypical antipsychotics	NA	167 (34.1%)	110 (74.8%)		
Typical antipsychotics	NA	26 (5.3%)	8 (5.4%)		
Both A & T	NA	16 (3.3%)	9 (6.1%)		
None or unknown	NA	281 (57.3%)	20 (25.9%)		
Current antipsychotic medication[h]	NA	19.64 (14.15)	19.30 (12.57)		

Samples: sex prediction

HCP functional imaging parameters:

Siemens 3T Skyra scanner, 1200 volumes, voxel size= 2 x 2 x 2 mm^3, FoV= 208x180 mm^2, 72 slices, TR = 720 ms; TE= 33.1 ms, FA=52°)

- Sample 1: 434 subjects (217 males, age range: 22-37, mean age: 28.6 years),
- Sample 2: 310 subjects (155 males, age range: 22-36, mean age: 28.5 years).

1000 brains functional imaging parameters:

Siemens 3T TRIO scanner, 297 volumes, voxel size= 3.1 x 3.1 x 3.1 mm^3, FoV= 200x200 mm^2, 36 slices, TR = 2200 ms; TE= 30 ms, FA=90°)

- Sample: 1115 subjects (508 males, age range: 18-88, mean age: 63.5 years)

SCZ study

Sample details

- Slide Text

Cohort imaging: Large, multi-modal datasets

Initial sample: 131 subjects
Enhanced sample: 900+ subjects
45 publications

10,000 subjects currently
30,000 subjects final

900 related subjects
17 publications

1,100 subjects
Longitudinal design
7 publications

>1,000 subjects currently
15,000 subjects final

1,000+: Schizophrenia, Parkinson
Major Depression, Stroke
50+ publications

10,000 subjects currently
100,000 subjects final