Neurotechnologies and Machine Learning

Maxim Sharaev
NeuroML Group, CDISE

Skoltech
Skolkovo Institute of Science and Technology

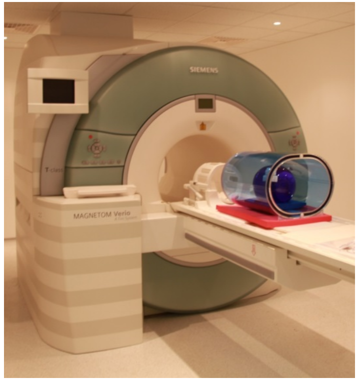
Agenda

- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- Educational tasks
- Neuroimaging in eSport
- Conclusions

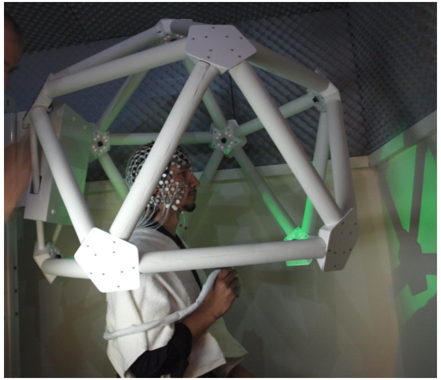
- **Neuroimaging data sources**
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- Educational tasks
- Neuroimaging in eSport
- Conclusions

Data Sources

MRI/
fMRI



EEG

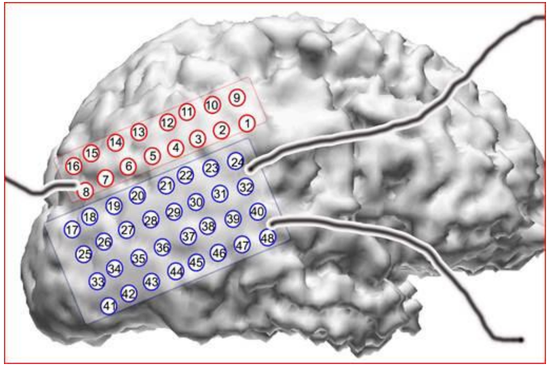


fNIRS

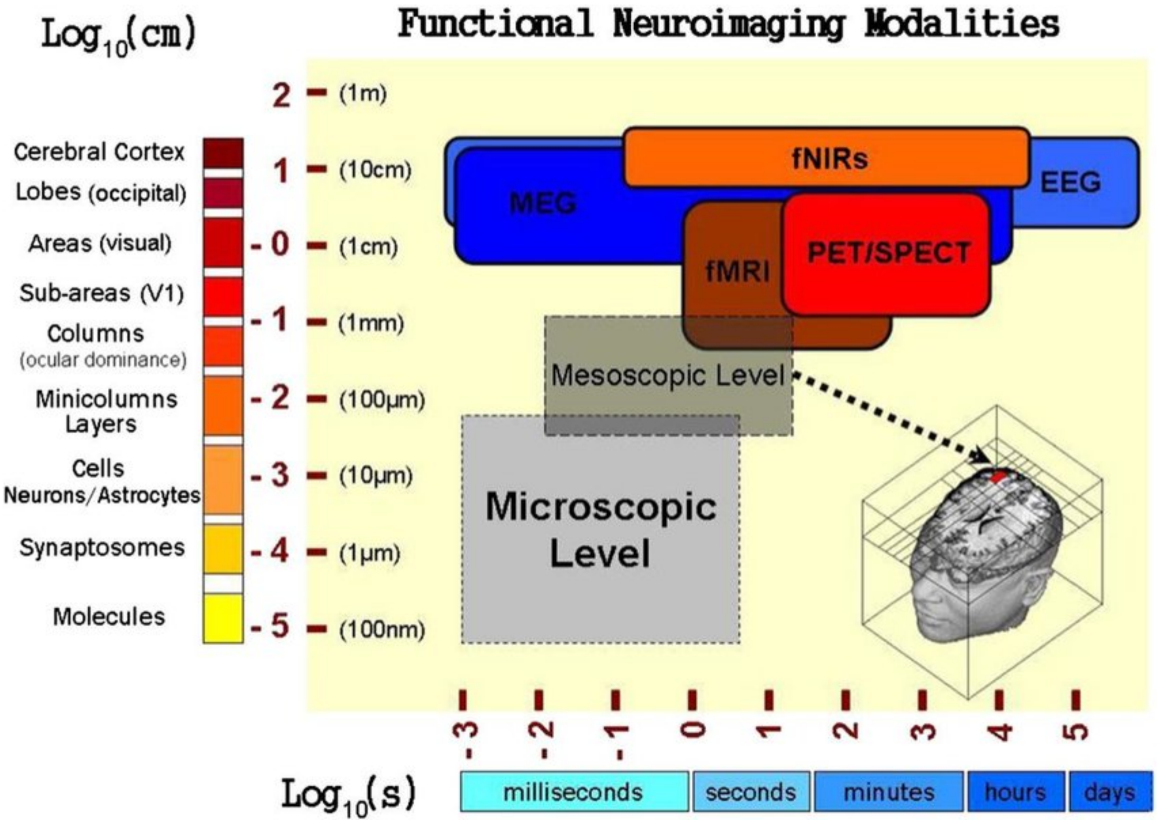
MEG

Eye-
tracking

iEEG



Data Sources: scales



- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- Educational tasks
- Neuroimaging in eSport
- Conclusions

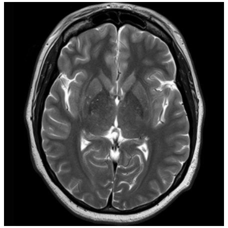
Neuroimaging data peculiarities: MRI

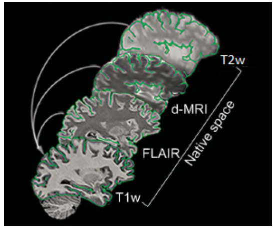
MRI: protocol of medical examination:

- structural MRI (sMRI)

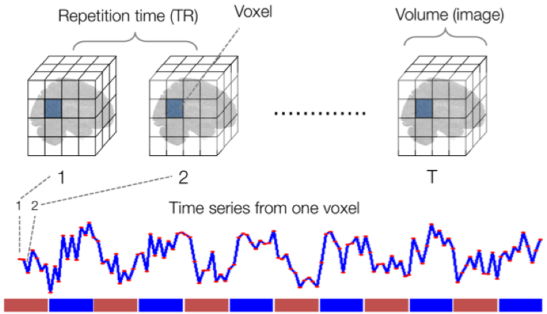
in various modalities (T1, T2, DTI, FLAIR)

3D data
$8\text{x}10^6$ voxels
~100 Mb





- functional MRI (fMRI)

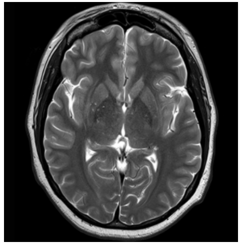


4D (3D time-dependent data)
1.5×10^6-dimensional MRI-measurements
received per a few seconds
>100 Mb

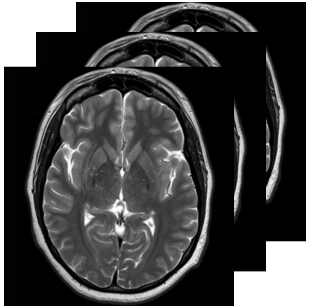
Skoltech
Skolkovo Institute of Science and Technology

Neuroimaging data peculiarities: MRI

MRI



fMRI

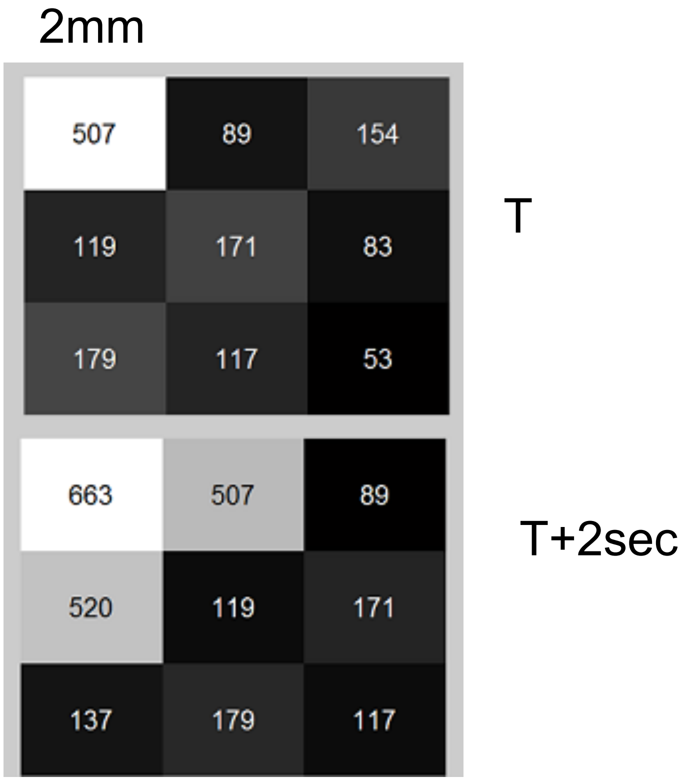


MRI: structural and functional

- High dimensional 3D/4D
- Good spatial, bad temporal resolution
- Noisy
- Different contrasts
- fMRI – correlational nature
- various Magnetic field strength characteristics (1.5T and 3T as usual, 7T maximum in clinics),
- various spatial characteristics (voxel size/image resolution), etc.

Noise and signal in fMRI data

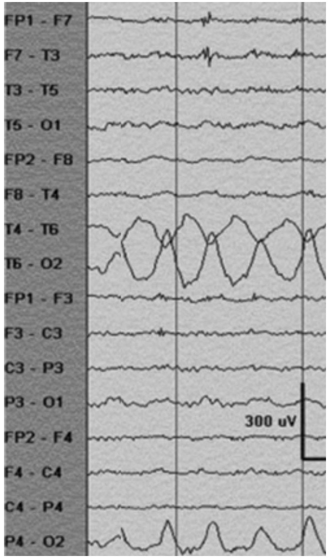
- ❑ Up to 80-90% of noise in raw data
- ❑ Most common sources
  - **Non-physiological**
    1. Thermal noise
    2. Magnetic field inhomogenity
    3. Imperfection in RF pulse timings
    4. Many others
  - **Physiological**
    1. Cardiac noise
    2. Respiratory noise
  - **Motion**

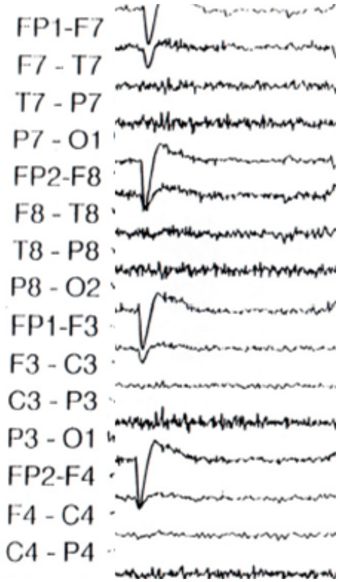


Neuroimaging data peculiarities

Electroencephalography (EEG) data:

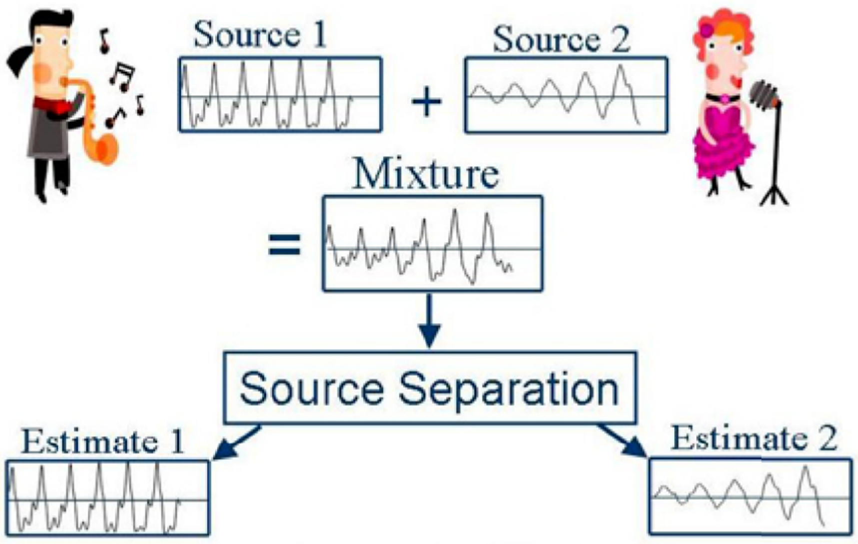
- various spatial characteristics (number of channels – 8, 16, … , 128)
- Good temporal, bad spatial resolution
- **Noisy:**
    - Eye movements
    - Muscular action potentials
    - Sweating
    - ECG
    - Poor electrode contact
    - Movement
    - Power supply (50Hz)





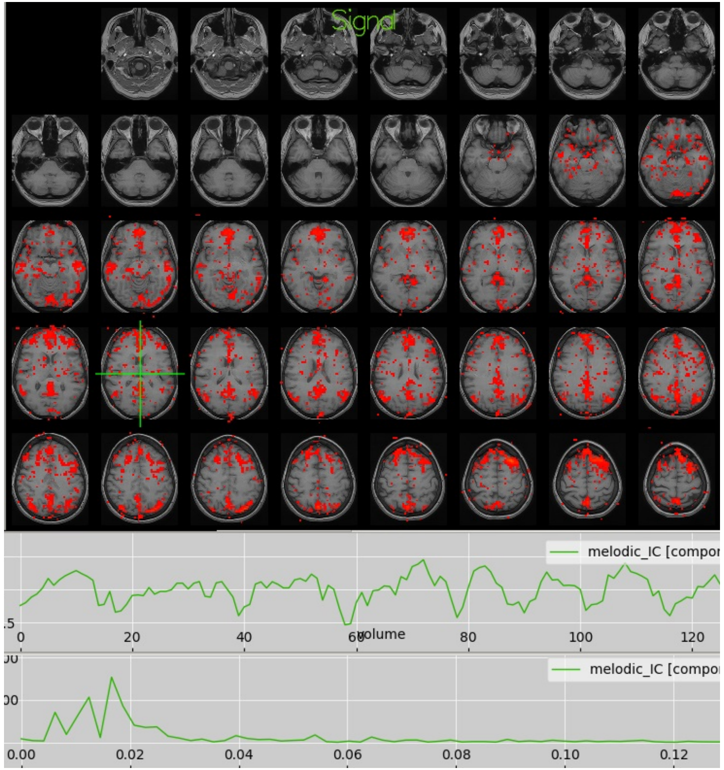
- **Neuroimaging data sources**
- **Neuroimaging data peculiarities**
- **Neuroimaging data analysis**
  - Noise
  - Connectivity features and interpretable models
- **Biomedical tasks**
- **Educational tasks**
- **Neuroimaging in eSport**
- **Conclusions**

Independent Component Analysis: recap



$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)$$
$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)$$
$$x(t) = A \cdot s(t)$$

Signal: Default Mode Network



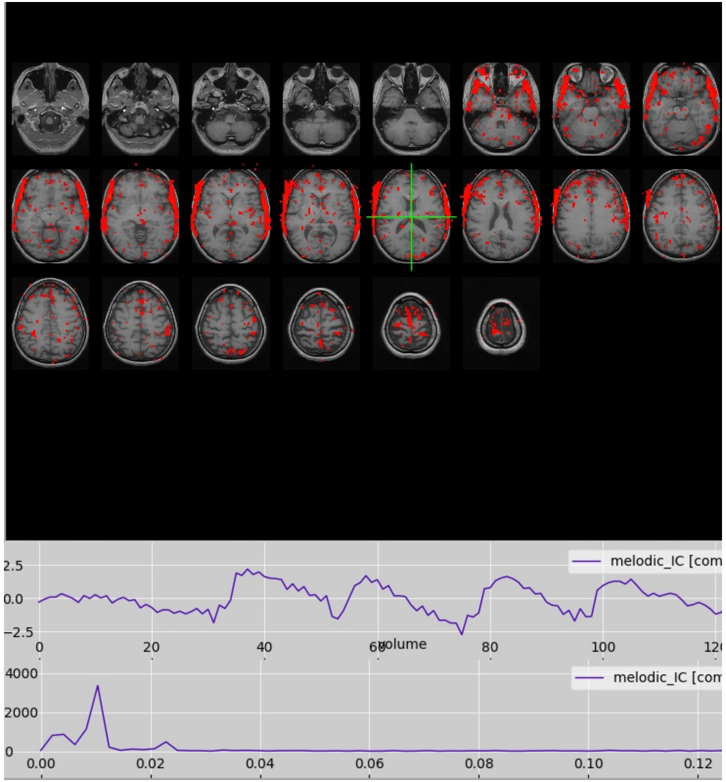
“Fingerprints”:

ICA Spatial maps

ICA time-series

ICA spectrum

Noise: motion



ICA Spatial maps

EDGE fraction

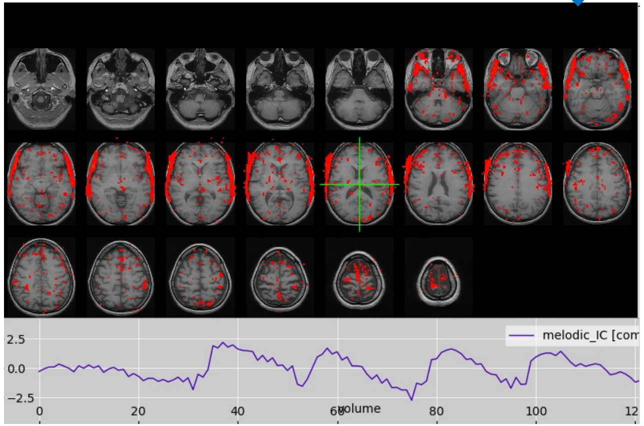
ICA time-series

ICA spectrum

Skoltech
Skolkovo Institute of Science and Technology

Noise detection and elimination

- ❑ ICA decomposition
- ❑ Noise differ from neuronal signal in spatial and temporal patterns of ICs

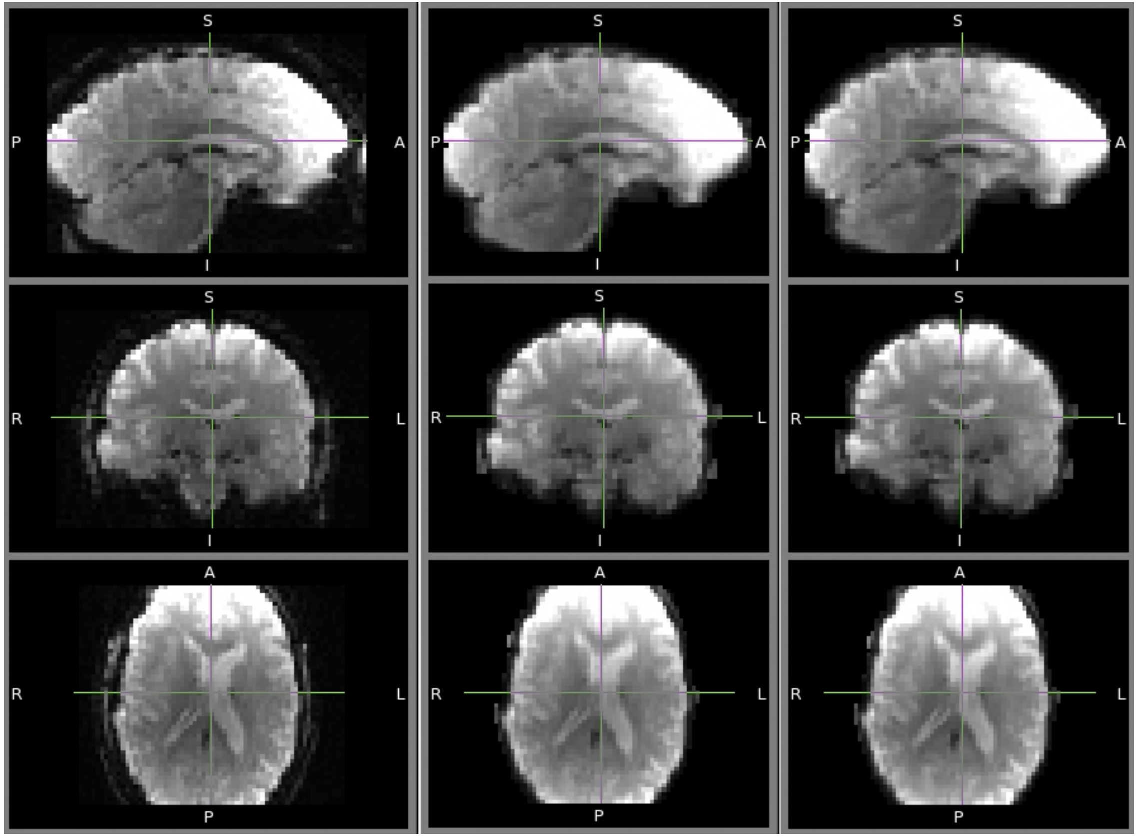


Now IC classification is performed **manually** by experts or **by trained classifier** (FSL FIX) – need for labelled data, highly dependent on specific scanner and parameters.

We introduce metrics based on physics of process and work on making it less dependent on a particular scanner (HF component, Correlation with motion, EDGE fraction, CSF fraction)

Skoltech
Skolkovo Institute of Science and Technology

Noise detection and elimination: example



Result: 5-20% increase of AUC

Summary: neuroimaging data

1) Different sources, data with different properties

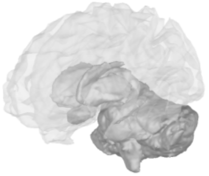
2) Data is noisy

3) All data modalities reflect brain structure/function from different prospetives and are usually considered together by doctor/researcher (example – epilepsy diagnostics: MRI/fMRI/EEG/iEEG)

- **Neuroimaging data sources**
- **Neuroimaging data peculiarities**
- **Neuroimaging data analysis**
  - Noise
  - Connectivity features and interpretable models
- **Biomedical tasks**
- **Educational tasks**
- **Neuroimaging in eSport**
- **Conclusions**

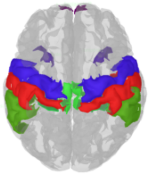
General pipeline for data analysis

1. Data acquisition



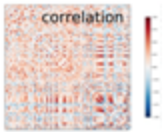
sMRI/fMRI

2. Data processing

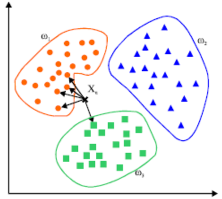


Data pre-processing,
Noise removal

3. Feature engineering



4. Machine Learning



5. Interpretation

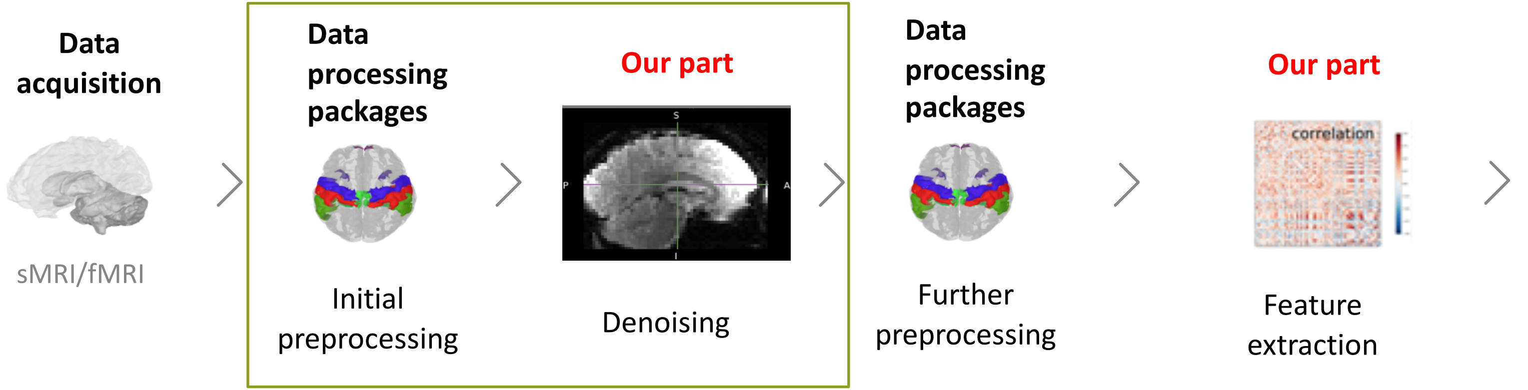
Model interpretation:

Potential biomarkers

Neuroimaging data analysis: low-level, preprocessing

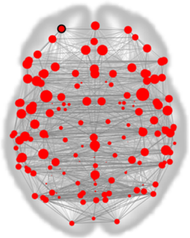
As an example, steps for fMRI data:

1) Filtering
2) Temporal corrections (i.e. different time for slice acquisition)
3) Coregistration, normalization, …
4) **Denoising**
5) Spatial smoothing, …

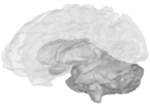


Dimensionality reduction and feature extraction

- Statistical methods: PCA, ICA, manifold learning, and etc. – problems with interpretation, not always capture clinically meaningful information
- Domain-specific methods:
    - vectorized features (volumes, curvatures, thicknesses from MRI - Freesurfer),
    - **matrix features** (correlation matrices from fMRI/EEG data) also could be considered as connectivity graphs – CONN, Nilearn

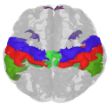


Data acquisition



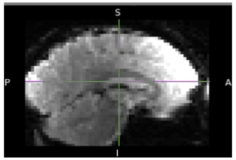
sMRI/fMRI

Data processing packages



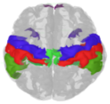
Initial preprocessing

Our part



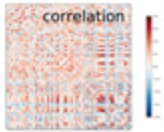
Denoising

Data processing packages



Further preprocessing

Our part



Feature extraction

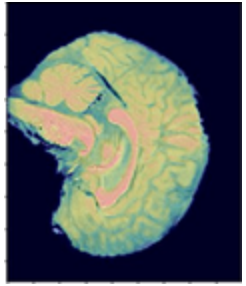
High-level tasks

Methods: general flow + denoise

MRI/fMRI/EEG

Denoising

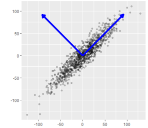
Deep Learning



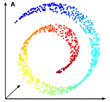
Low-dim representation

Domain-specific methods

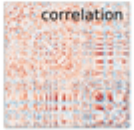
Standard methods



Vector



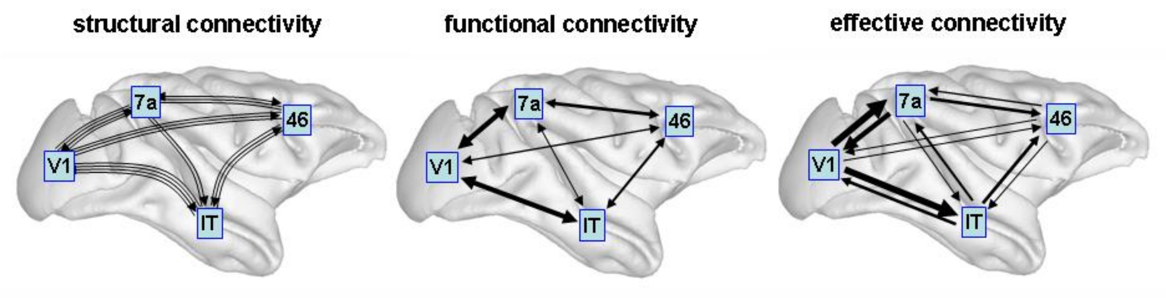
Vector/matrix



Transform to vector

Work with matrix (graph)

Networks and connectivity types



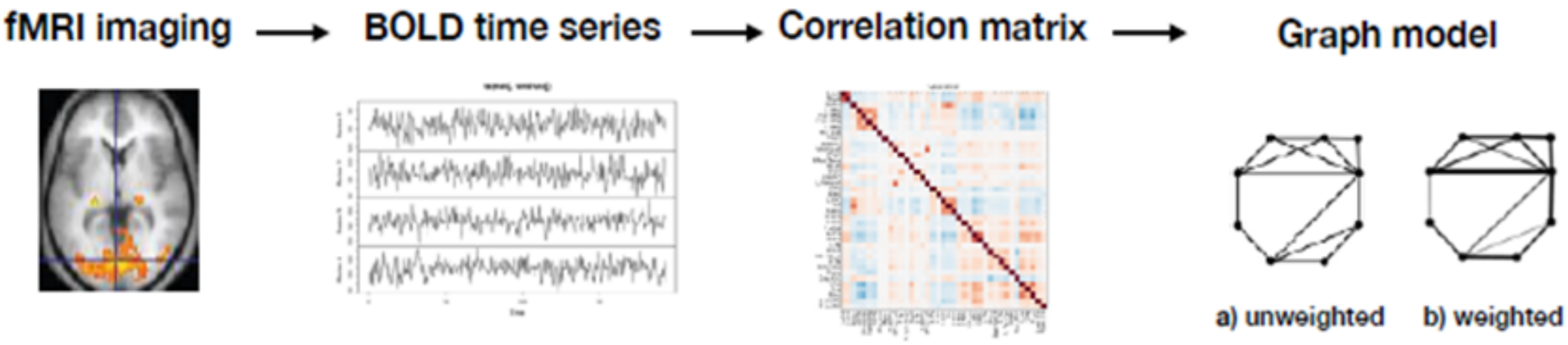
Sporns (2007)

- **anatomical/structural connectivity**
  = presence of axonal connections.

- **functional connectivity**
  = statistical dependencies between regional time series.

- **effective connectivity**
  = causal (directed) influences between neurons or neuronal populations.

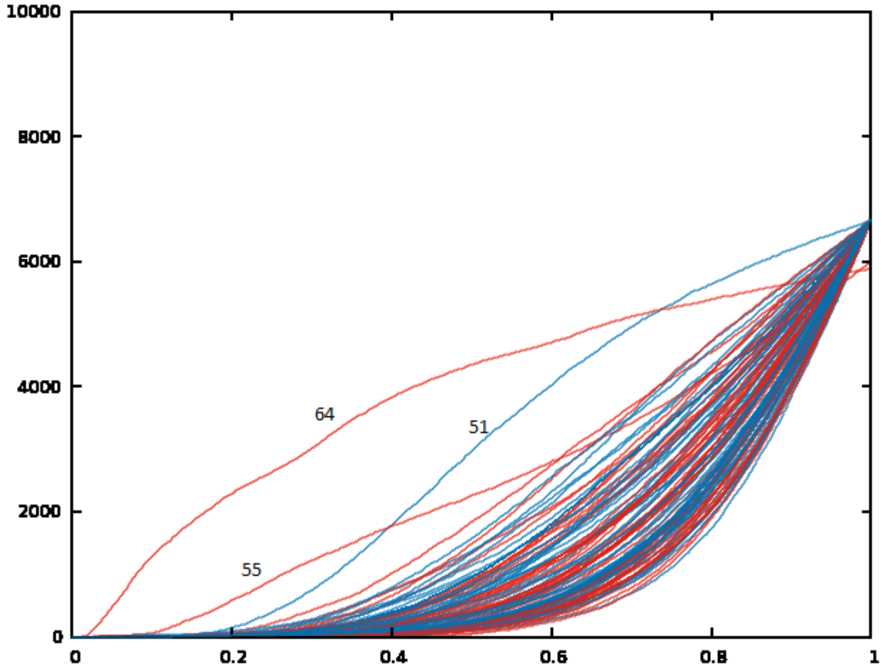
Connectivity analysis and feature extraction



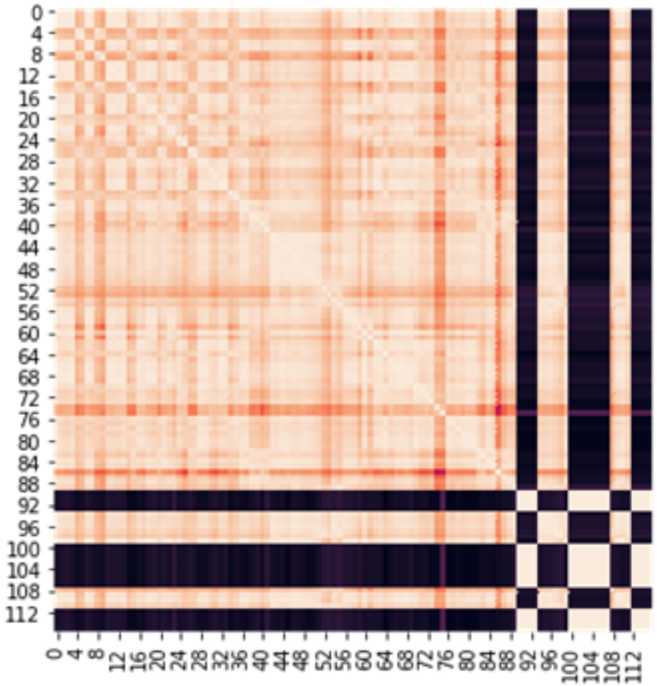
"Descriptor" of brain networks

Anomalies detection in data: Betti numbers

Topological Data Analysis



Patient 64



Correlational connectivity

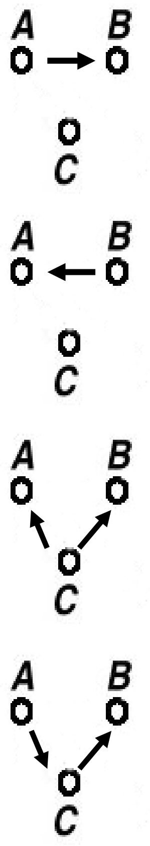
** Pros:

- Easy to compute;

** Cons:

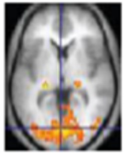
- no mechanistic insight
- interpretation of resulting patterns is difficult / arbitrary, i.e. false correlations;

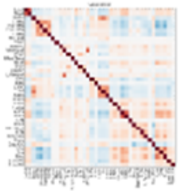
→ Causal (Effective) connectivity



Methods to assess effective connectivity

fMRI imaging → ?? → causation matrix





Causality "Descriptor" of brain networks

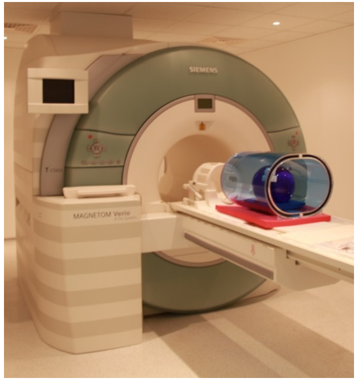
- Time-series analysis
  - Can dynamics of region A be predicted better using past values of region A and region B as opposed to using past values of region A alone, e.g. **Granger Causality, Transfer Entropy**
- Methods based on (non)linear dynamic models
  - **Dynamic Causal Modelling (DCM)**

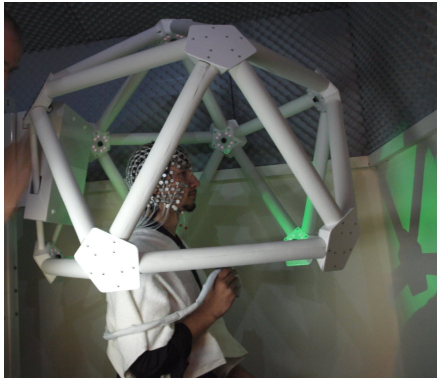
Dynamic Causal Modelling

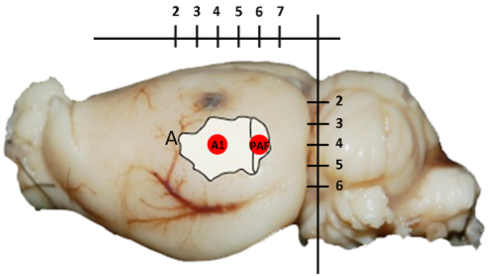
Generative model

$$\dot{x} = f(x, u, \theta) + \omega$$
$$y = g(x, u, \theta) + v$$

Experimental data







= Silverball electrode, diameter: 1 mm

Skoltech
Skolkovo Institute of Science and Technology

Resting-state fMRI modelling

➜ $\dot{\boldsymbol{x}}(\boldsymbol{t}) = \boldsymbol{A}\boldsymbol{x}(\boldsymbol{t}) + \boldsymbol{v}(\boldsymbol{t})$

➜ Coupled dynamical systems with no external input

➜ $$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

n regions

Aij - Extrinsic effective connectivity parameters

DCM parameters = rate constants

Integration of a first-order linear differential equation gives an exponential function:

$$\frac{dx}{dt} = ax \quad \Rightarrow \quad x(t) = x_0 \exp(at)$$

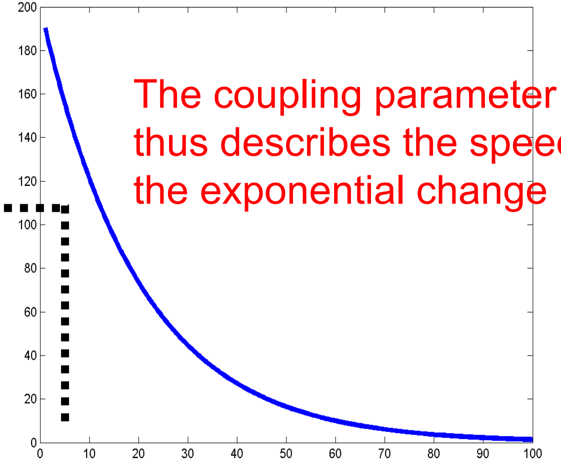
Coupling parameter a is inversely proportional to the half life τ of x(t):

$$x(\tau) = 0.5x_0$$

$$= x_0 \exp(a\tau)$$

$$\Rightarrow \quad \boxed{a} = \ln 2 / \tau$$

$0.5x_0$



$$\tau = \ln 2 / a$$

The coupling parameter a thus describes the speed of the exponential change in x(t)

DCM: pros&cons

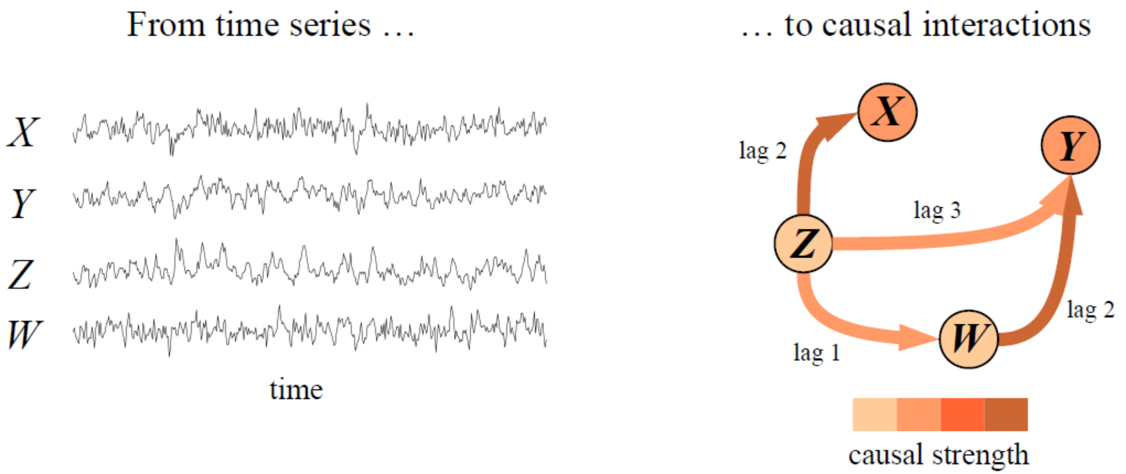
PROS:

1. Neurobiologically plausible
2. Easy to interpret parameter values
3. Parameter priors could be derived from anatomical data, tractography, invasive study, laser Doppler flowmetry and etc.

CONS:

1. Model-based
2. Parameter estimation procedures

Transfer entropy: model-free approach to assess causality



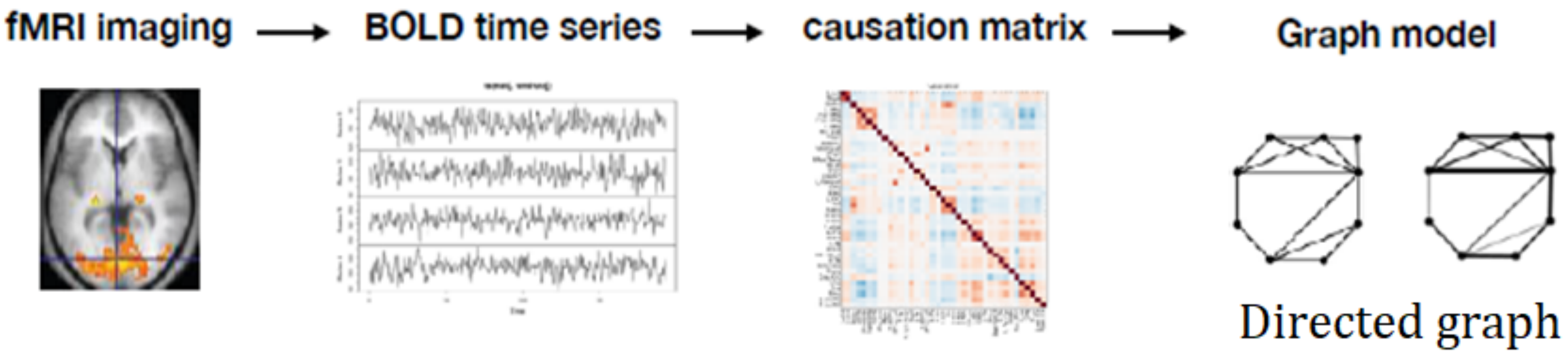
$$p(Y_n|Y_n^-, X_n^-) = p(Y_n|Y_n^-)$$

Markovian independence

$$TE_{X\to Y} = \sum p(Y_n, Y_n^-, X_n^-) \log \frac{p(Y_n|Y_n^-, X_n^-)}{p(Y_n|Y_n^-)}$$

Transfer Entropy (TE) – deviation from equality (KL distance)

Causal connectivity analysis and feature extraction

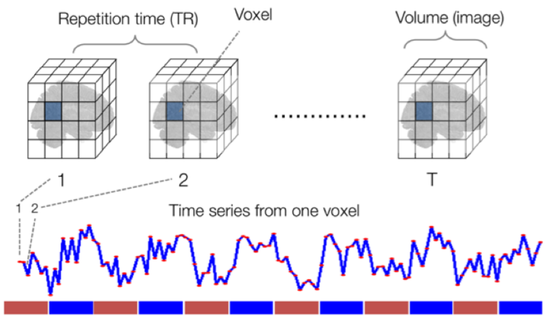


fMRI models interpretation

Recap:

functional MRI (fMRI)

4D (3D time-dependent data)
1.5×10^6-dimensional MRI-measurements
received per a few seconds
>100 Mb

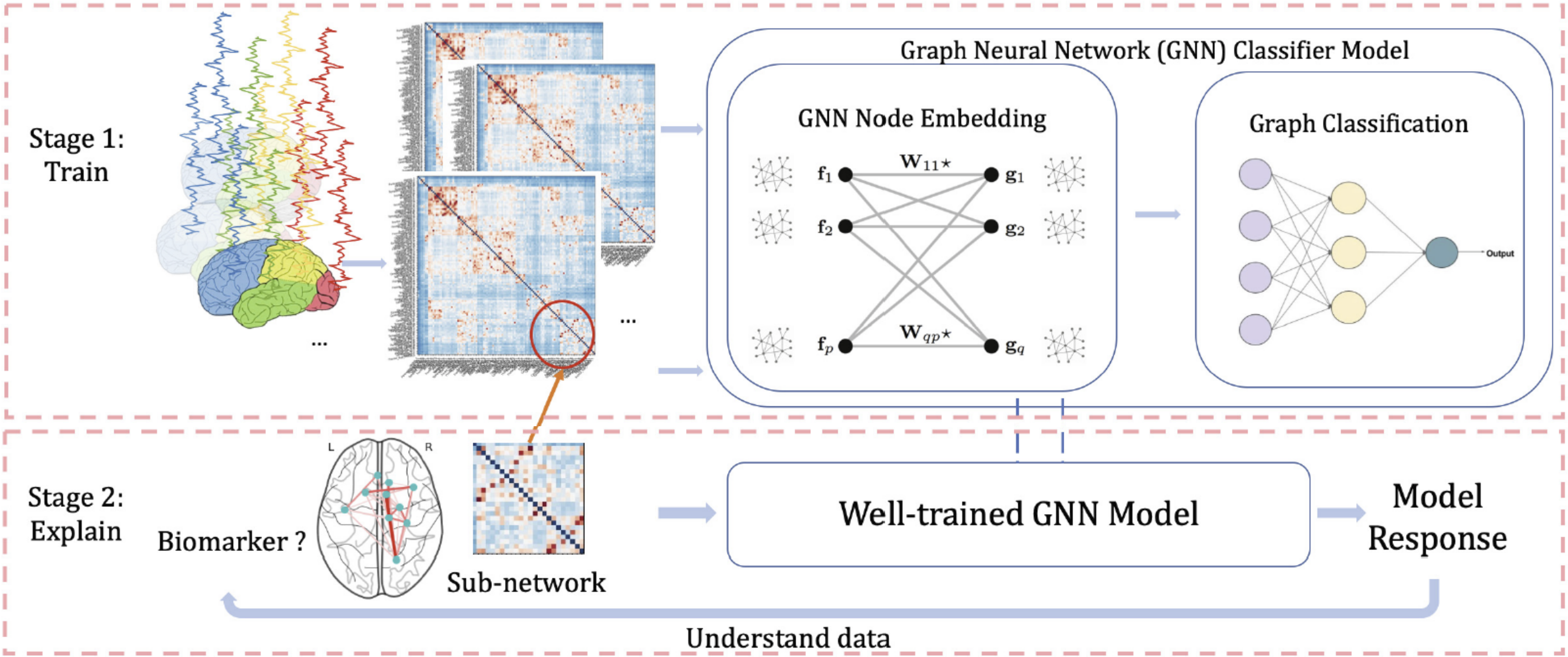


Why interpretable models are so important?

Approaches to interpretability: connectomes

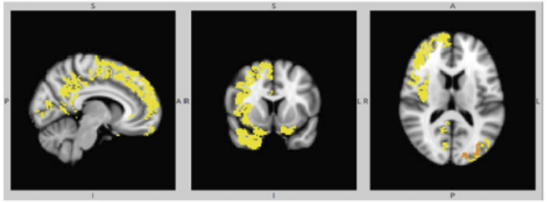
“Graph Neural Network for Interpreting Task-fMRI Biomarkers” MICCAI 2019

- **75 ASD** children and **43 healthy controls** (age and IQ-matched)
- the **“biopoint” task**, viewing point light animations of coherent and scrambled biological motion in a block design
- **ROI-based** connectome
- **Node attributes**: degree of connectivity, General Linear Model (GLM) coefficients, mean, standard deviation of task-fMRI, and ROI center coordinates
- **Edge attributes**: Pearson correlation, partial correlation, geometric distance

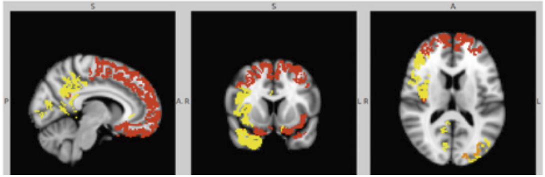


- cluster the whole brain graph into subgraphs
- investigate the predictive power of each sub-graph
- assign importance score to each ROI

(a)



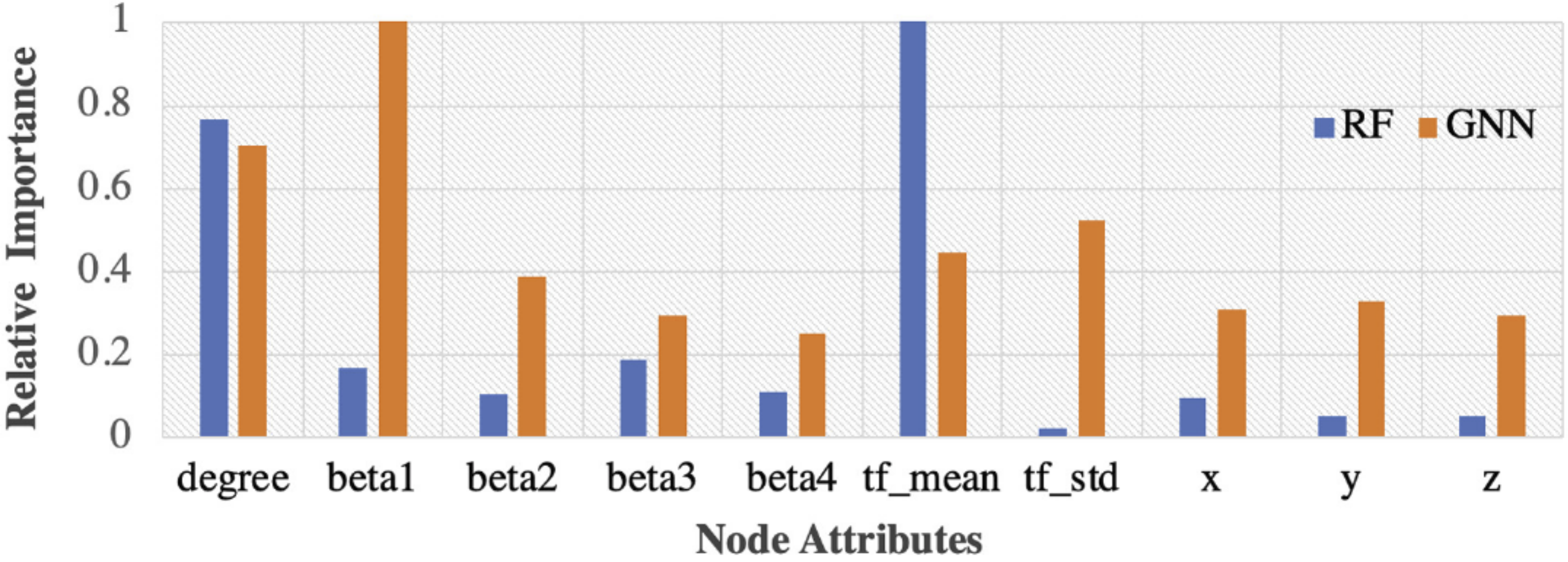
(b)

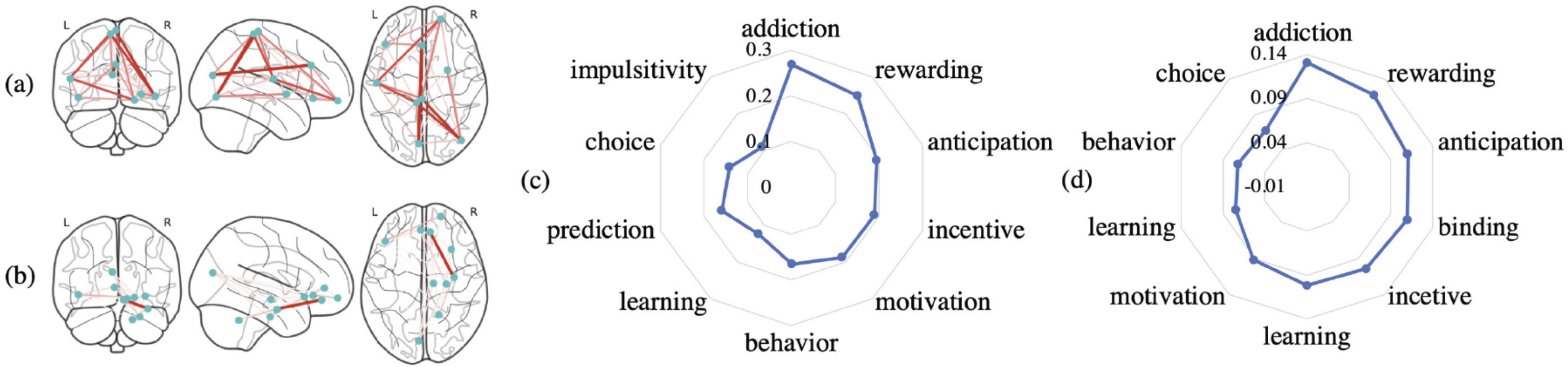


Top 30 important ROIs
a – RF
b – GNN (red) laying over a

Model	GNN(r = 0.5)
Accuracy	**0.76 ± 0.06**
F-score	**0.79 ± 0.08**
Precision	**0.76 ± 0.12**
Recall	**0.82 ± 0.06**

(c)

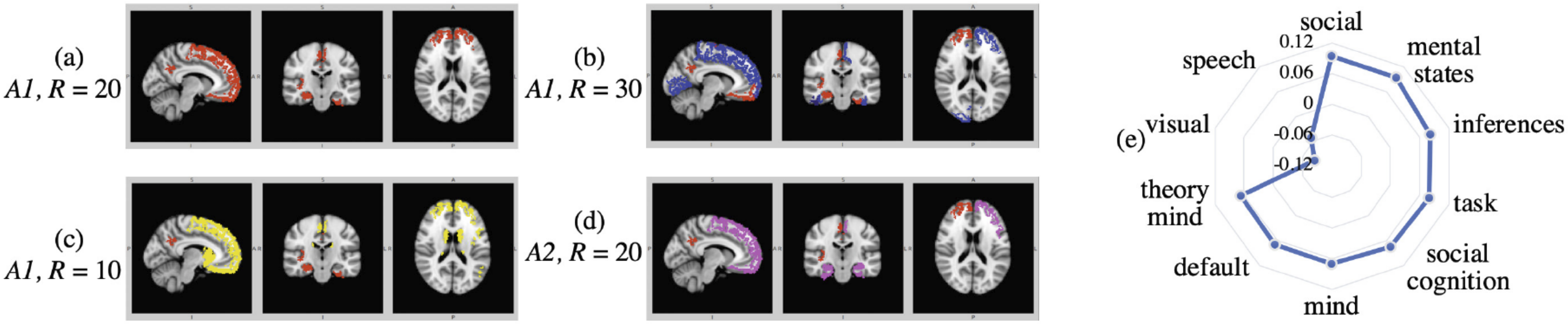




(a), (c) Top scoring sub-graph and corresponding functional decoding keywords and coefficients (Neurosynth)

(b), (d) The 2nd high scoring sub-graph and corresponding functional decoding keywords and coefficients

Robustness



(a) The biomarkers (red) interpreted on A1 (**Destrieux atlas**) with 20 clusters;
(b)–(d) The biomarkers interpreted by different R (number of clusters) and atlas (**Desikan-Killiany**) laying over on (a) with different colors;
(e) The correlation between overlapped ROIs and functional keywords.

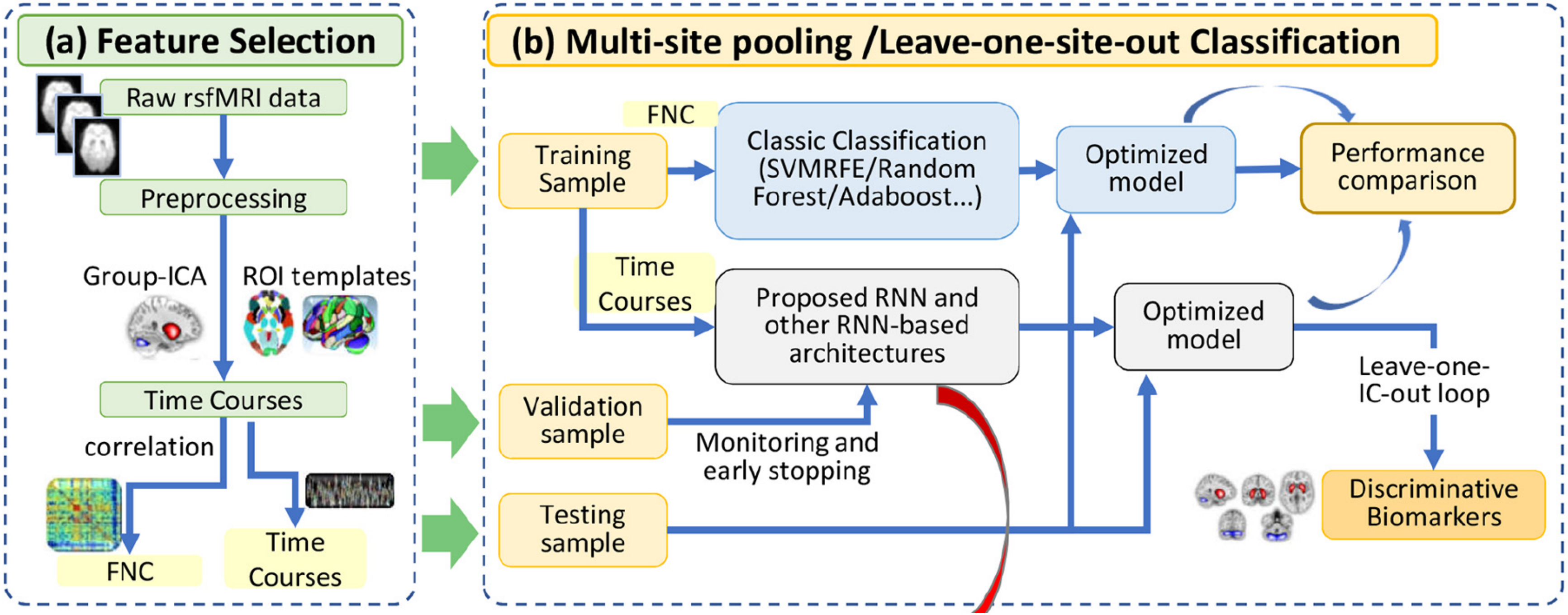
Approaches to interpretability: full-size fMRI

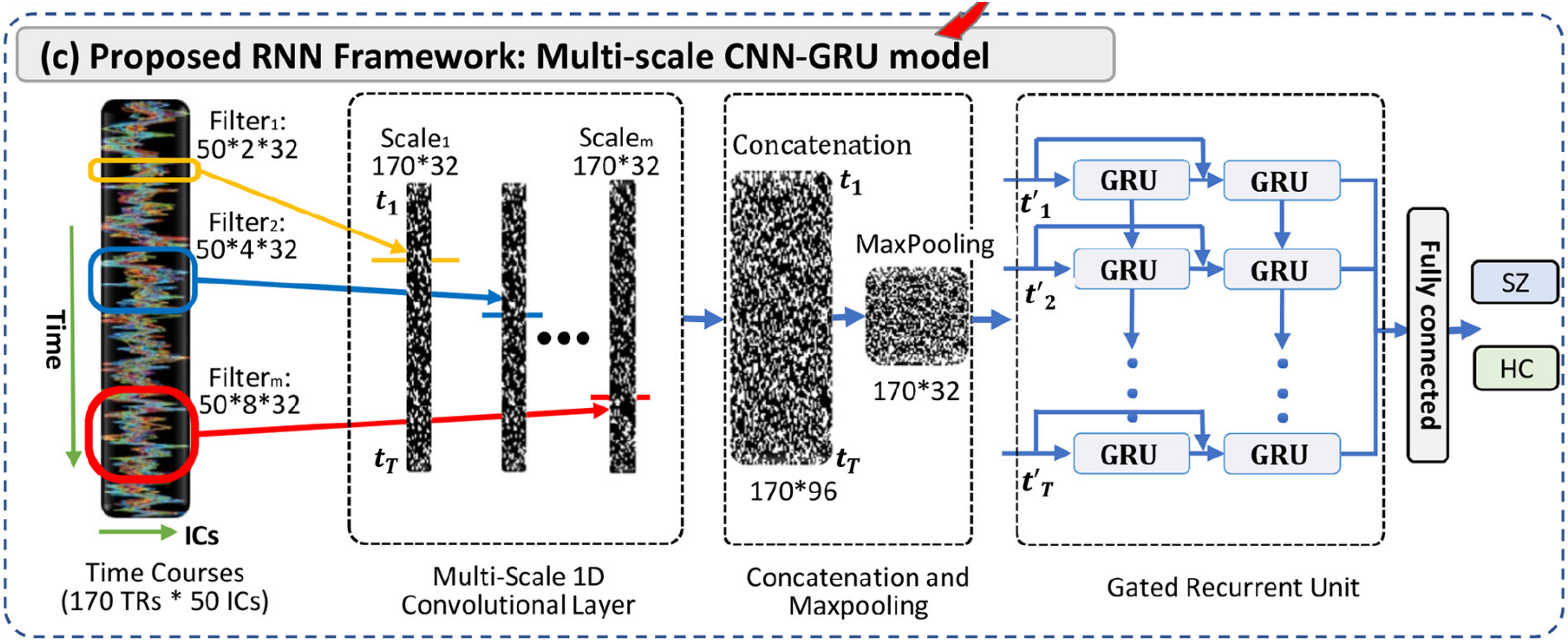
"Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data" EBioMedicine 2019

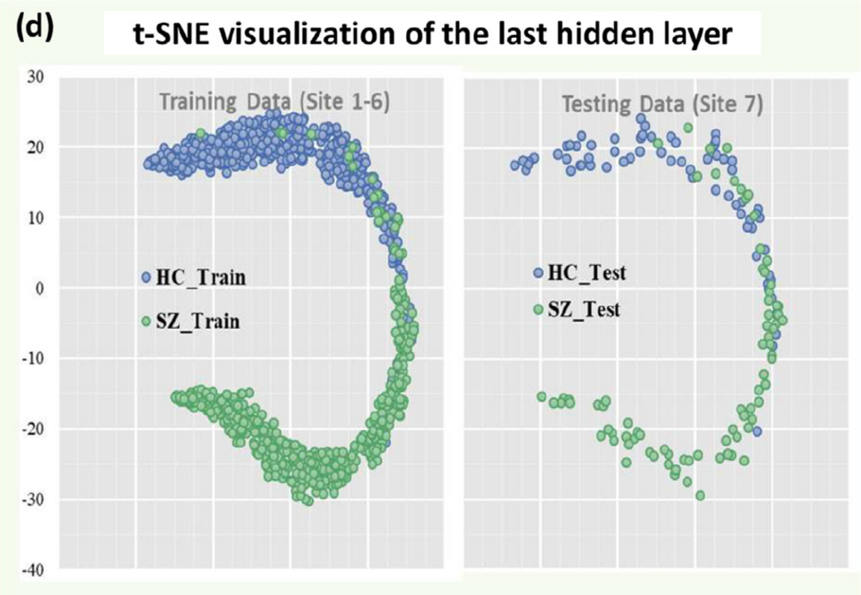
➜ classification between **558 schizophrenia and 542 healthy controls**, **7 sites** (rs-fMRI)

➜ Time courses were extracted using group ICA

➜ FNC matrices as baseline


(a) Feature Selection
Raw rsfMRI data
Preprocessing
Group-ICA
ROI templates
Time Courses
correlation
FNC
Time Courses
(b) Multi-site pooling /Leave-one-site-out Classification
FNC
Training Sample
Classic Classification (SVMRFE/Random Forest/Adaboost...)
Optimized model
Performance comparison
Time Courses
Proposed RNN and other RNN-based architectures
Optimized model
Leave-one-IC-out loop
Validation sample
Monitoring and early stopping
Testing sample
Discriminative Biomarkers


(c) Proposed RNN Framework: Multi-scale CNN-GRU model
Filter1: 50*2*32
Filter2: 50*4*32
Filterm: 50*8*32
Time
ICs
Time Courses (170 TRs * 50 ICs)
Scale1 170*32
Scalem 170*32
t1
tT
Multi-Scale 1D Convolutional Layer
Concatenation
MaxPooling
170*96
170*32
Concatenation and Maxpooling
GRU
t'1
t'2
t'T
Gated Recurrent Unit
Fully connected
SZ
HC

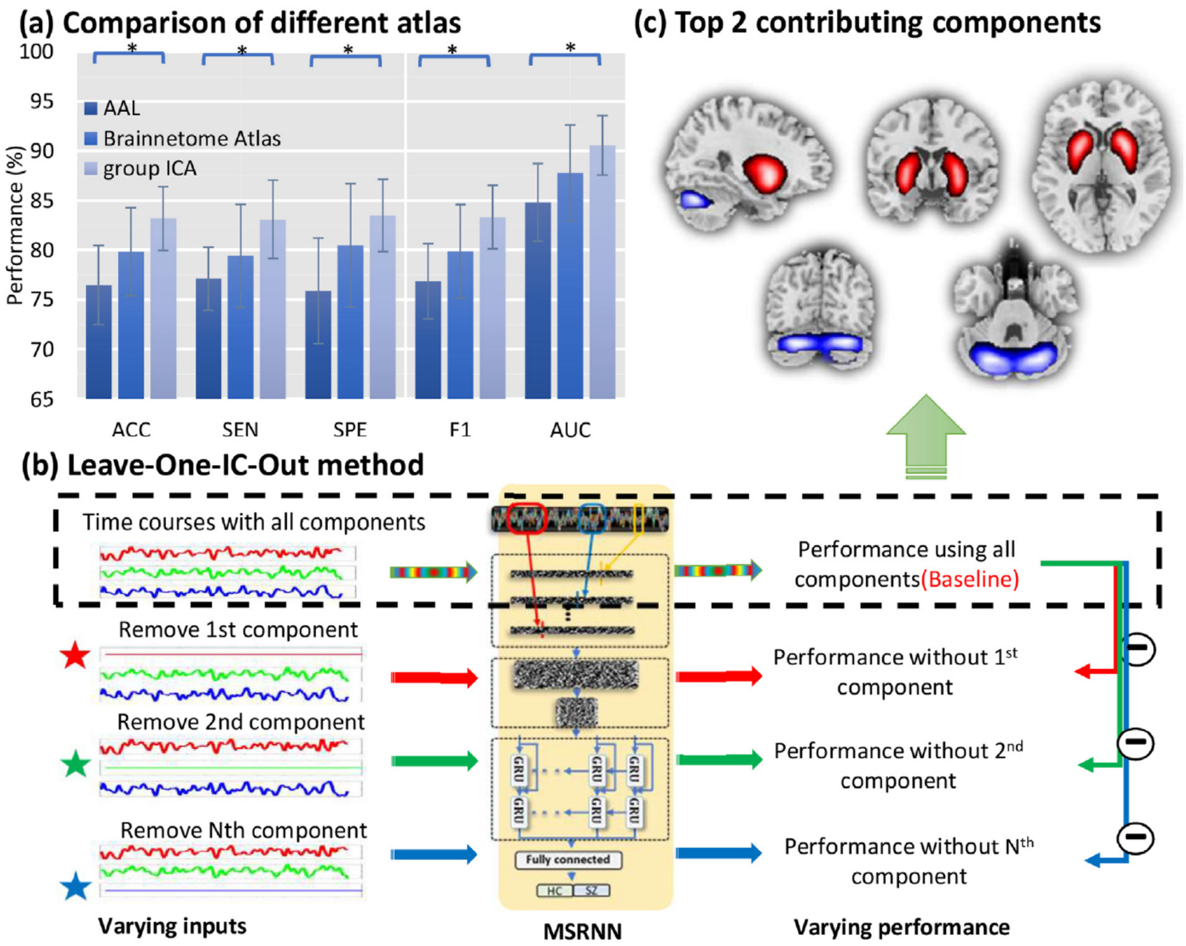


Performance comparison in multi-site pooling classification.

Methods		ACC
CON	Adaboost	75.6(3.8)**
CON	Random Forest	76.0(3.5)**
CON	SVM	79.4(3.1)*
RNN	GRU_1_last	51.6(3.6)**
RNN	GRU_1_ave	77.8(3.4)**
RNN	GRU_2_ave	78.0(3.9)**
CMLP	Multi_CNN_MLP	77.8(3.4)**
CRNN	Simple_CNN_GRU_2_ave	80.8(3.0)○
CRNN	Multi_CNN_GRU_1_ave	80.6(3.5)○
CRNN	Multi_CNN_GRU_2_ave	81.2(3.4)○
CRNN	Multi_CNN_LSTM_2_ave	81.6(2.9)○
CRNN	MsRNN(Proposed)	83.2(3.2)

Performance comparison in leave-one-site-out classification.

Methods		ACC
CON	Adaboost	72.9(3.0)**
CON	Random Forest	72.6(4.4)**
CON	SVM	76.0(3.1)*
RNN	GRU_1_last	47.7(3.2)**
RNN	GRU_1_ave	78.7(2.8)○
RNN	GRU_2_ave	77.9(3.9)○
CMLP	Multi_CNN_MLP	76.1(3.2)*
CRNN	Simple_CNN_GRU_2_ave	79.1(3.7)○
CRNN	Multi_CNN_GRU_1_ave	80.3(3.0)○
CRNN	Multi_CNN_GRU_2_ave	79.7(3.0)○
CRNN	Multi_CNN_LSTM_2_ave	78.7(3.9)○
CRNN	MsRNN(Proposed)	80.2(3.0)



1. **putamen** and **caudate** which - parts of striatum;
2. **declive** and **uvula**: parts of the cerebellum

1. Train on 880 samples
2. 220 subjects without removing any component to obtain a baseline
3. Same 220 subjects with one IC removed
4. Repeat step 3 until each IC has been removed once

Open questions

1. How to obtain purely temporal biomarkers (raw data)?
2. How to work with a single subject?

➜ Treat prediction as true

➜ Follow 1st or second scheme

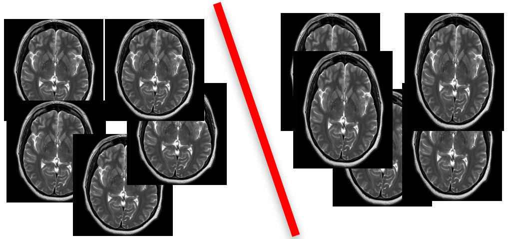
Skoltech
Skolkovo Institute of Science and Technology

Questions to Part 1?

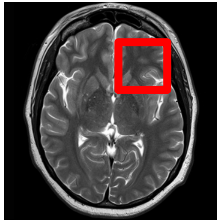
- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- Educational tasks
- Neuroimaging in eSport
- Conclusions

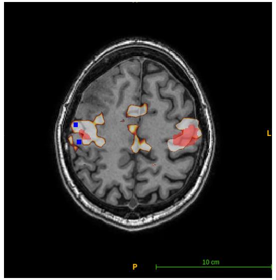
AI systems in medicine: tasks

- ❑ **Classification and regression tasks: diagnostics and prognosis**

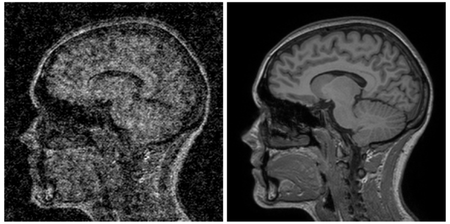


- ❑ **Detection tasks: disease patterns**





- ❑ **More specific tasks: segmentation, reconstruction, enhancement, retrieval, …**



NeuroML Group: partners

НИИ НЕЙРОХИРУРГИИ
имени академика Н.Н.Бурденко

ЦЕНТР ПАТОЛОГИИ РЕЧИ
И НЕЙРОРЕАБИЛИТАЦИИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

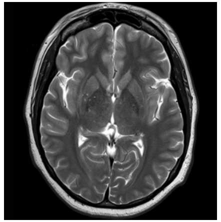
федеральное государственное бюджетное учреждение
Национальный медико-хирургический Центр
имени Н.И. Пирогова
Министерства здравоохранения Российской Федерации

AI systems in medicine: flow

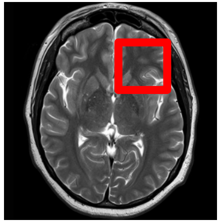
Data-driven **Medical Decision Support System Software (DSS SW)**

Data (training dataset) → **AI technologies** → **Predictive model** → **Medical DSS SW**

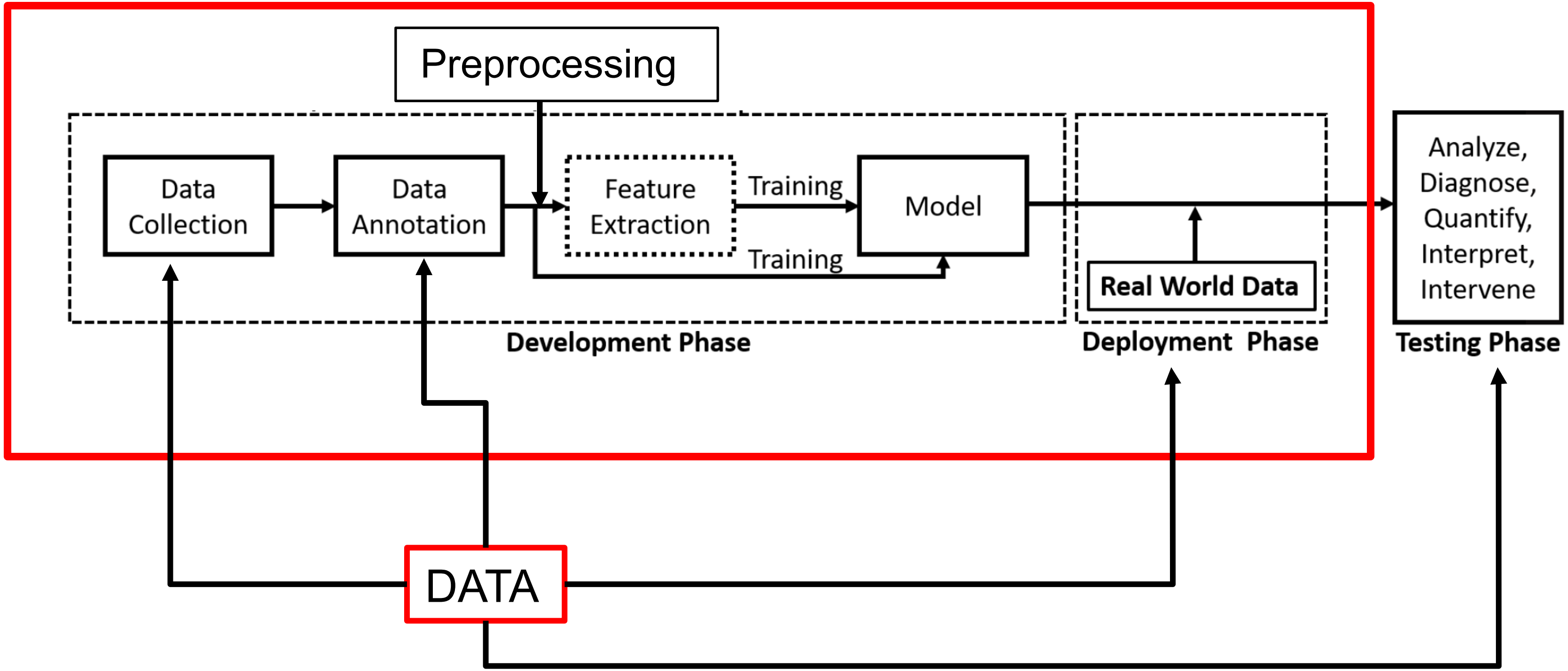
Example: based on structural MRI image detect epileptogenic lesions in brain cortex



→ **Medical DSS SW** →



AI system creation pipeline



* from (Qayyum et al., 2020)

Vulnerabilities in data acquisition/collection

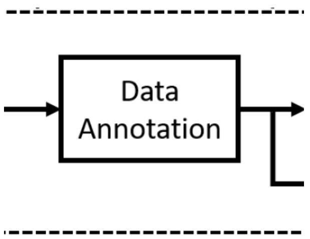
Data acquisition → Data Collection →

Data for medicine (one patient) → Data for AI (dataset)

- ❑ **Instrumental and environmental noise**
- ❑ **Missing guidelines**

Possible solution: develop protocols and universal formats, data engineer working with medical center

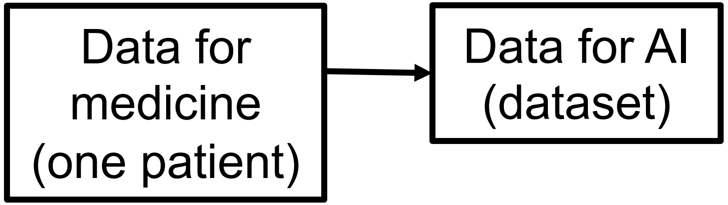
Vulnerabilities in data annotation



- ❑ **Ambiguous Ground Truth**

(e.g. GT is unknown or there is no agreement)

- ❑ **Improper Annotation**
  - Labels and strata for diagnostics
  - Images annotation for detection



A D A S E

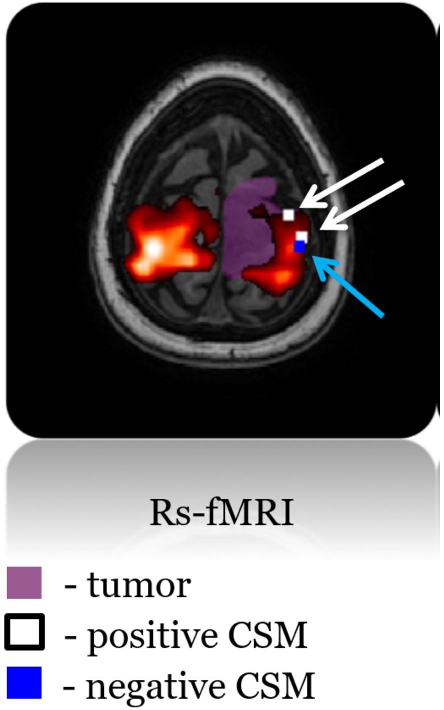
Skoltech
Skolkovo Institute of Science and Technology

Secure and robust ML for healthcare: Possible solutions

Solutions to Address Labelling

- ❑ **Know the origin of ground truth**: imaging alone/clinical confirmation
- ❑ **Confirmatory labelling** from different modalities (MRI-CT-PET, …)
- ❑ **Labelling assistance**:
  - Text analysis for retrospective labelling (NLP)
  - Active learning systems
- ❑ **Assume noisy labels and levels of trust**

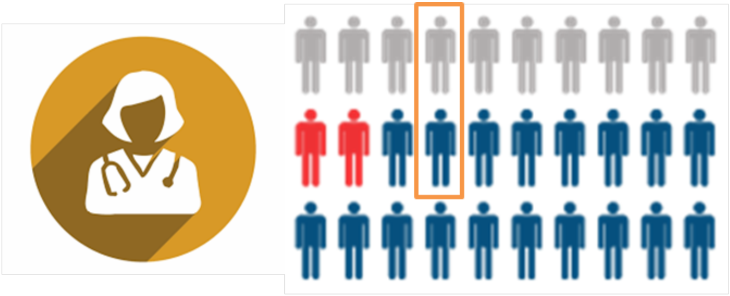
Example: labels adjustment



Vulnerabilities in sample

- ❑ **Limited and Imbalanced Datasets**
- ❑ **Class Imbalance and Bias**
- ❑ **Incomplete data**

Data for medicine (one patient) → Data for AI (dataset)



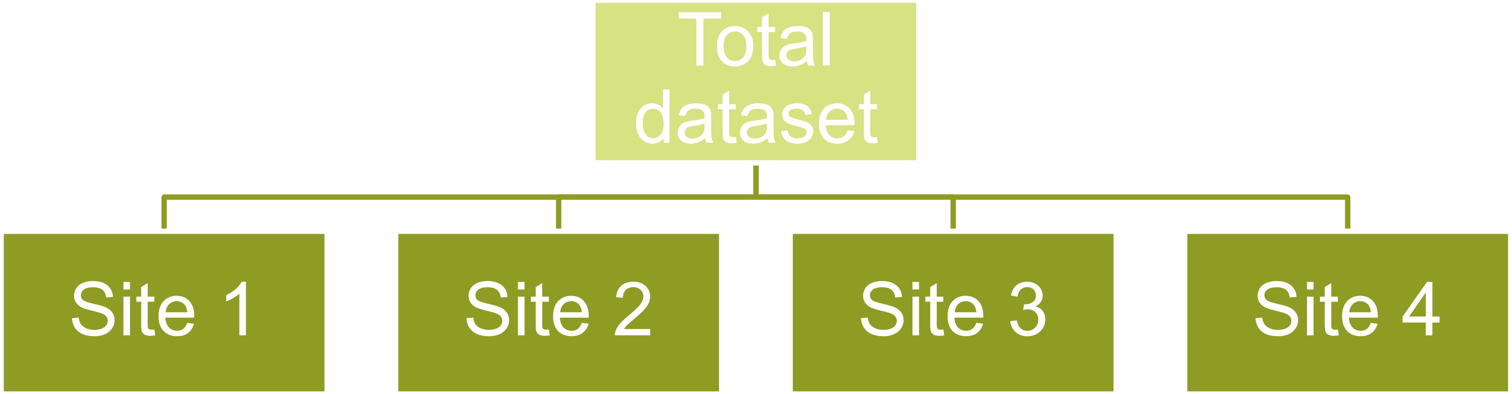
* from (Oakden-Rayner, 2019)

Vulnerabilities in sample: distribution shift

Distribution Shift:

- ✓ equipment
- ✓ scanning protocols
- ✓ personnel
- ✓ patients
- ✓ … etc.

Vulnerabilities in sample: distribution shift

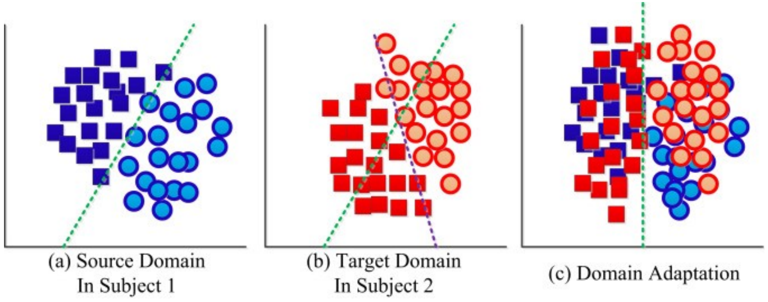


- ✓ Conventional ML is not robust to site
- ✓ Separate model for each site – small sample size

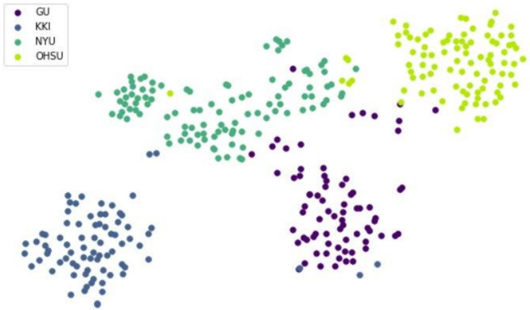
Secure and robust ML for healthcare: Possible solutions

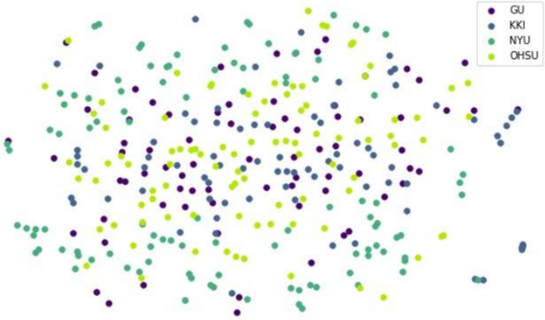
Solutions to Address Distribution Shifts

- ❑ Transfer Learning
- ❑ Domain Adaptation
  - Supervised
  - Unsupervised
  - Semi-supervised
- ❑ Data normalization (e.g. with nonlinear ICA)



Example: ABIDE dataset





* from (Pominova et al., 2020)

Summary

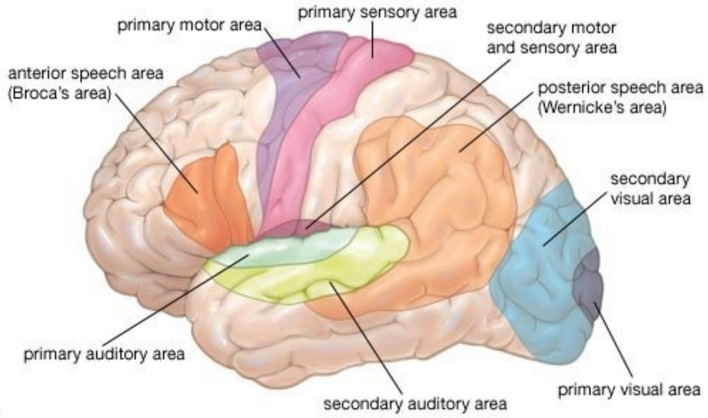
Step	Problem	Possible solution
Data acquisition	Instrumental and environmental noise	AI-based preprocessing technologies: denoising, artifact removal, etc.
Data collection	Missing guidelines	Data collection technology (data formats, storage protocols, common diagnosis codes, etc.)
Data annotation	Ambiguous Ground Truth	ML for data with noisy labels and different levels of confidence (levels of trust)
	Improper Annotation: Labels and strata for diagnostics	AI-based assistance for labeling: text analysis for retrospective labelling (NLP), active learning systems, etc.
	Improper Annotation: Images annotation for detection	AI-based assistance for automated anomaly detection and labeling
		Confirmatory labelling from different modalities
Training datasets	Limited and Imbalanced Datasets	ML methods for Limited and Imbalanced Datasets: data augmentation, sampling techniques,…
	Class Imbalance and Bias	ML methods for Imbalanced classification
	Incomplete data	ML methods for reconstruction of missing data
	Distribution shift	Transfer Learning
		Domain Adaptation (supervised, unsupervised, semi-supervised)
		Data normalization (signal separation)

Skoltech
Skolkovo Institute of Science and Technology

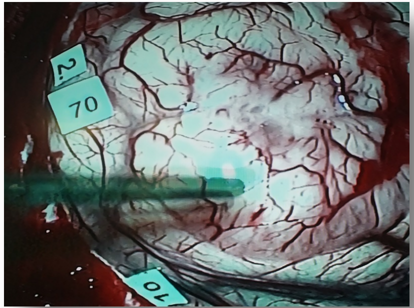
- **Neuroimaging data sources**
- **Neuroimaging data peculiarities**
- **Neuroimaging data analysis**
- **Biomedical tasks**
    - Functional brain areas mapping
    - Epileptogenic foci localization
    - Depression diagnostics
- **Educational tasks**
- **Neuroimaging in eSport**
- **Conclusions**

What should neurosurgeons plan?

- Locations of functional areas in healthy people and patients differ
- Before operation - pre-operative mapping
- During the operation - precise cortical stimulation mapping (CSM)



*from https://www.britannica.com



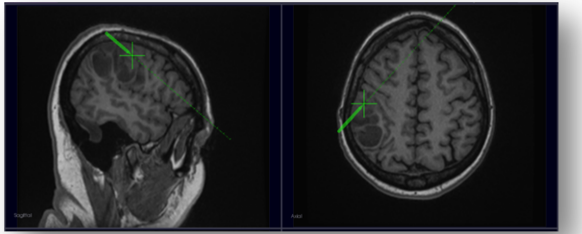
Skoltech
Skolkovo Institute of Science and Technology

Current mapping techniques and problems

- ❑ Invasive

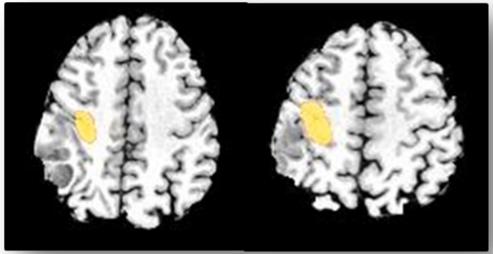
  intraoperative cortical

  stimulation mapping (CSM)

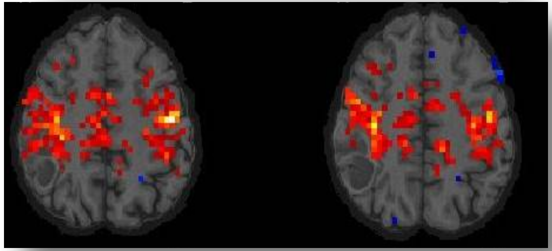


- ❑ Non-invasive

  task-based functional MRI



  resting-state functional MRI

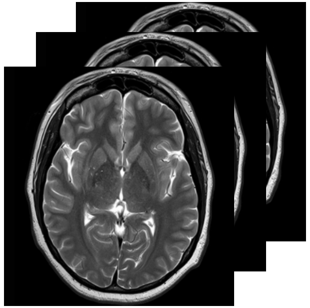


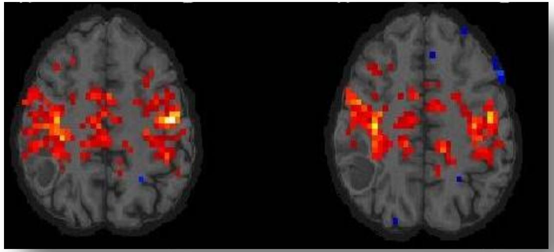
Resting-state fMRI

Shows:

- All **mixed** resting-state brain activity
- Reflects brain functional organization

fMRI





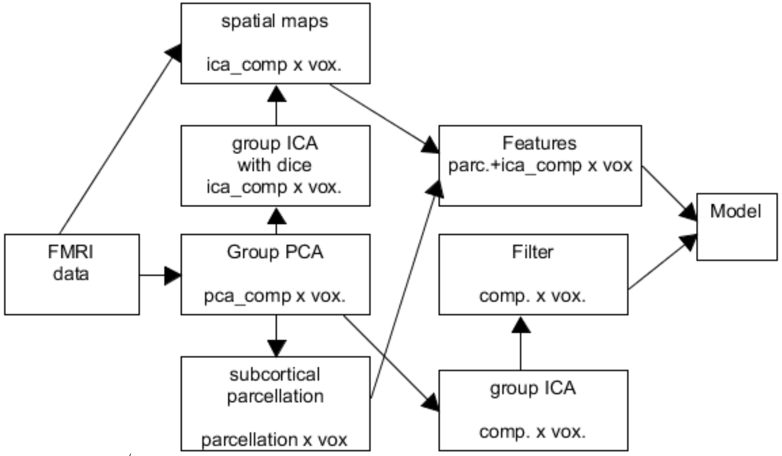
We need:

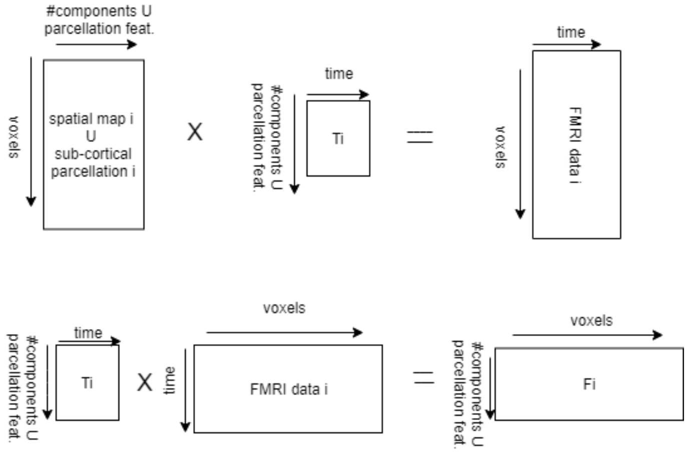
- One **particular individual brain** network
- Localization and properties are known in healthy
- This information could be used for patients

Open databases and model pre-training

Human Connectome Project: ~1200 healthy subjects with **both** resting-state and task fMRI data https://www.humanconnectome.org/

4D (3D time-dependent data) 1.5×10^6-dimensional MRI-measurements received per a few seconds
>100 Mb





*Tavor, 2016

Open databases and model pre-training

Input:

F_i – Extracted features with dim. (voxels x features)

y_i - true activation (dim. voxels)

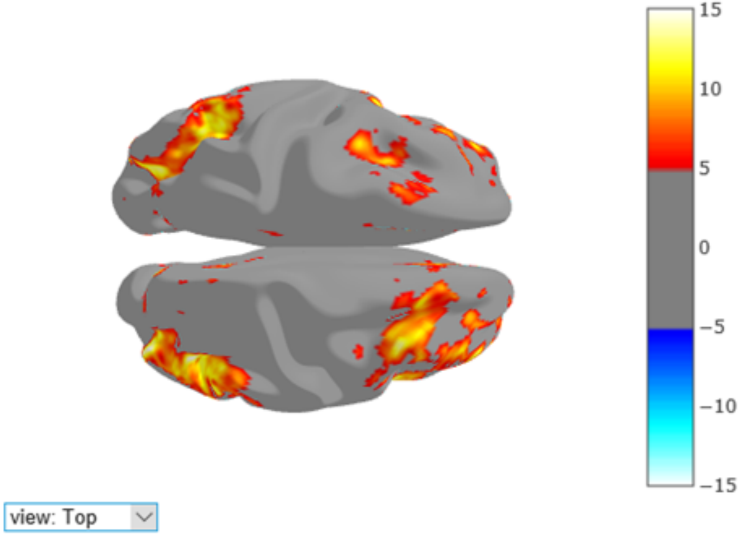
where *i* stands for subject *i*.

For each subject determine β_i by minimizing MSE:

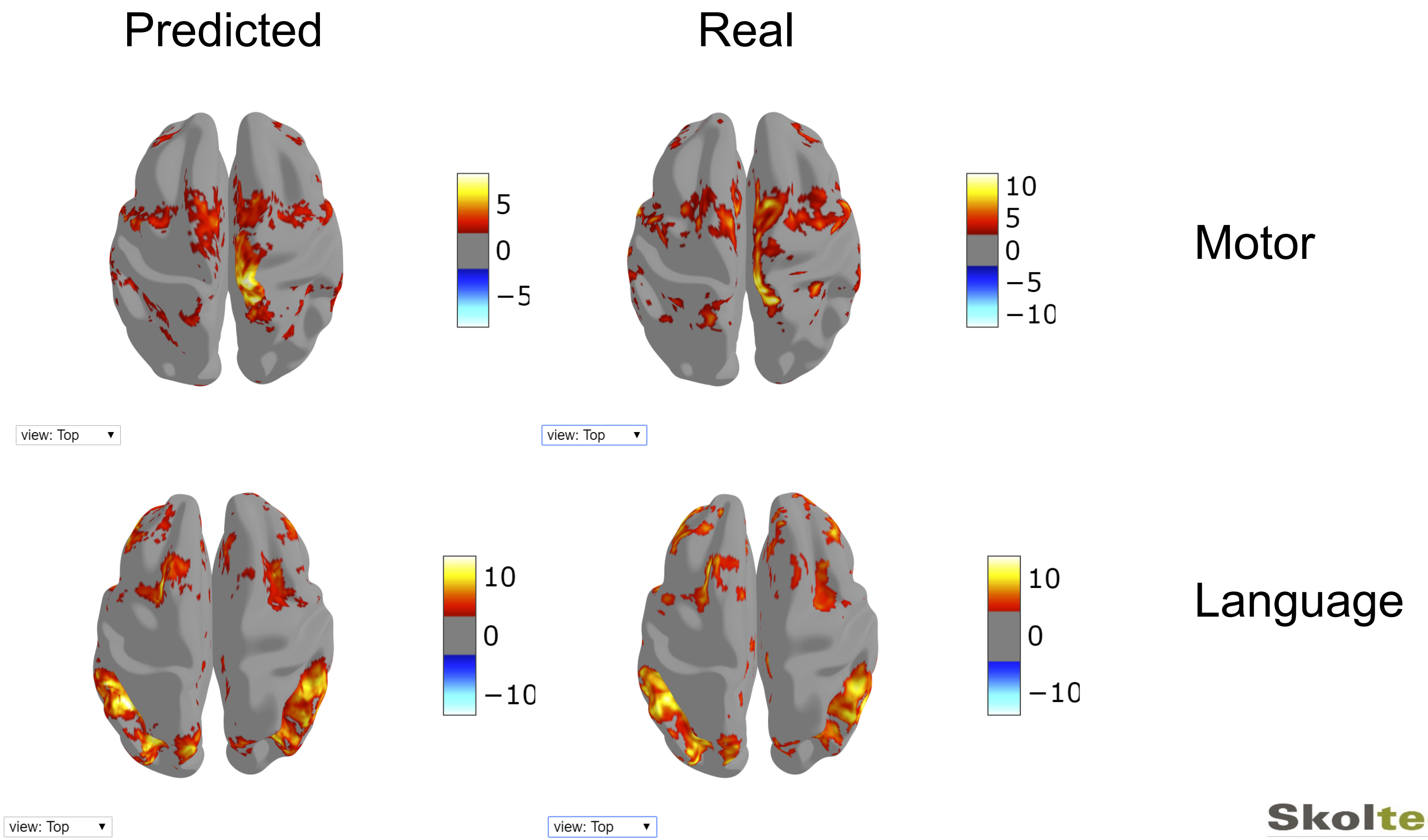
$$y_i = F_i \beta_i$$

Average β_i over all *N* training subjects :

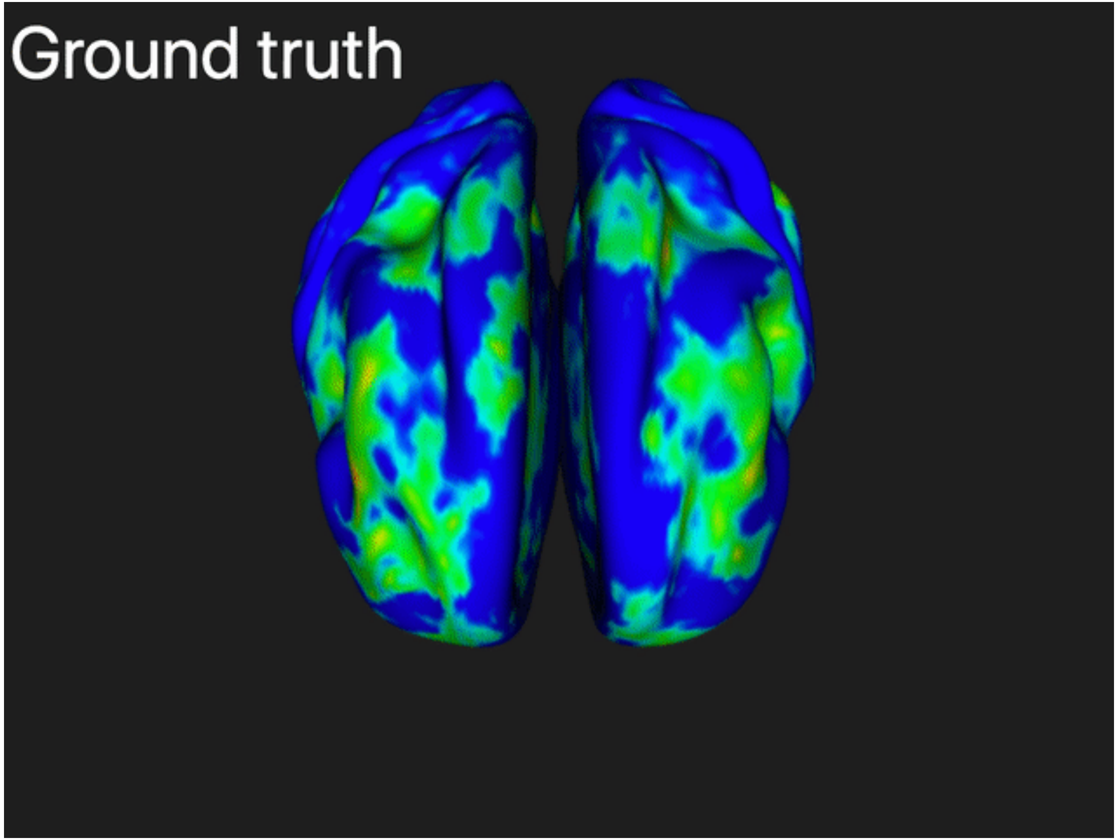
$$\beta = \frac{1}{N} \sum_{i=1}^{N} \beta_i$$

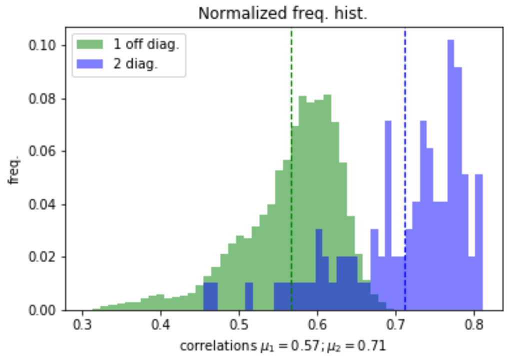


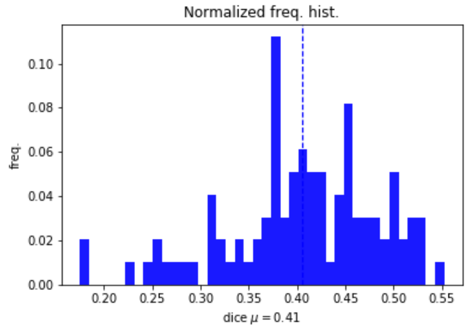
Open databases and model pre-training



Open databases and model pre-training





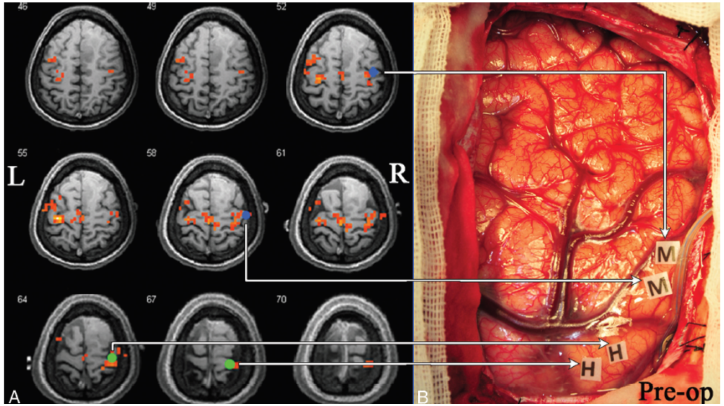


Skoltech
Skolkovo Institute of Science and Technology

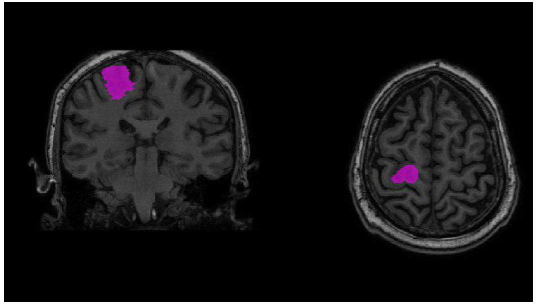
Additional imaging and intraoperative data

- Patient imaging data

- CSM from previous patients: both true and false activations



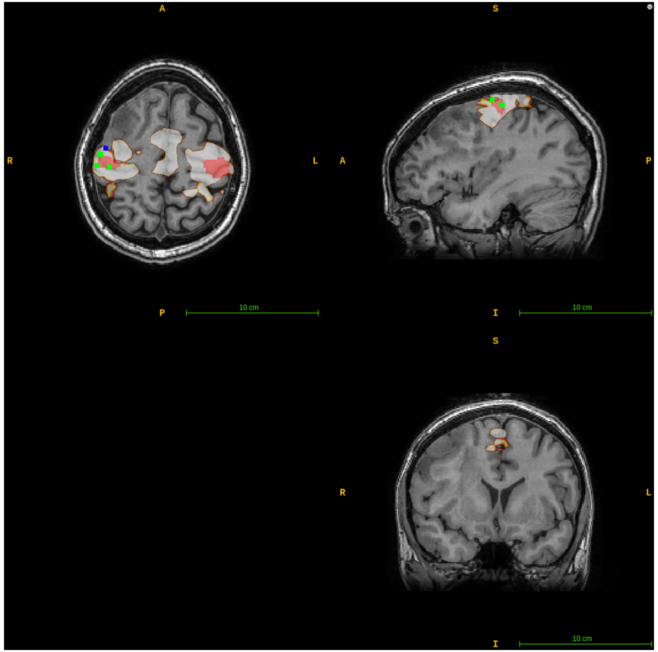
- Tumor mask

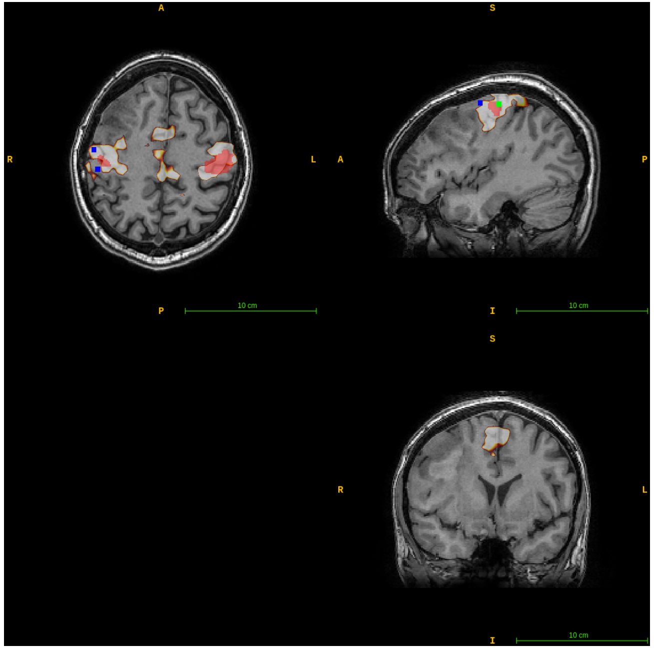


ADASE

Skoltech
Skolkovo Institute of Science and Technology

Showcases

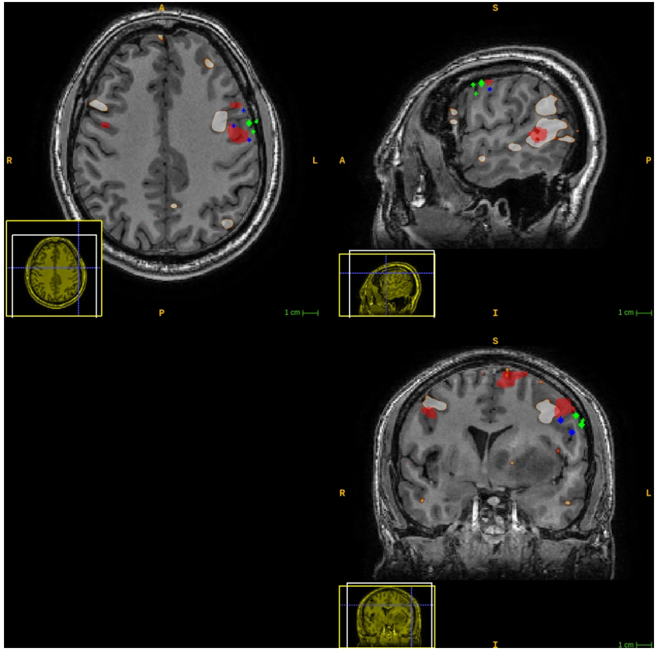


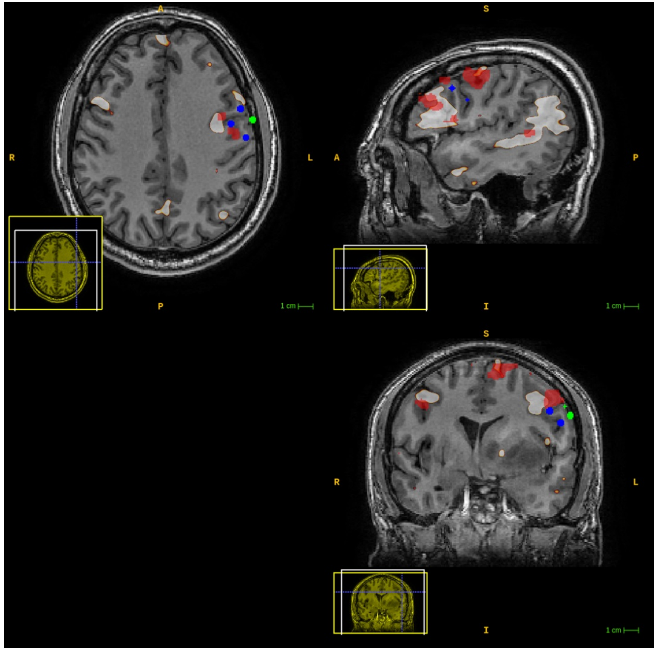


Patient A
Glioma grade III
Motor network

Method	Positive	Negative
Tb-fmri	100%	25%
Rs-fmri	100%	0%

Showcases





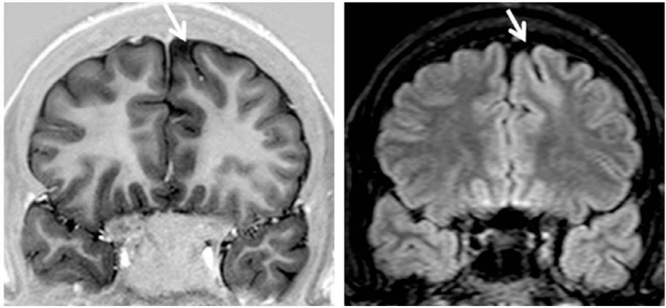
Patient B
Glioma grade II
Language network

Method	Positive	Negative
Tb-fmri	100%	50%
Rs-fmri	100%	0%

- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
  - Functional brain areas mapping
  - Epileptogenic foci localization
  - Depression diagnostics
- Educational tasks
- Neuroimaging in eSport
- Conclusions

Foci of epileptogenic activity localization

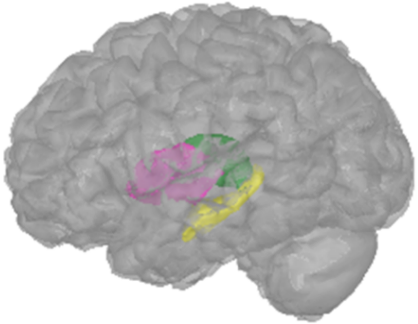
- **Medical partner**: V.I. Kulakov Research Center for Obstetrics, Gynecology and Perinatology
- Often treatment with drugs is ineffective
- Removal of foci of epileptogenic activity leads to recovery
- MRI positive/negative is subjective
- Radiologists with different level of qualification

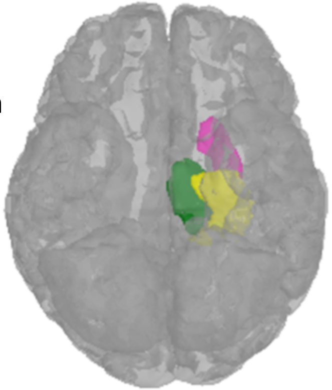


⇩

Segmented and labelled data to train models

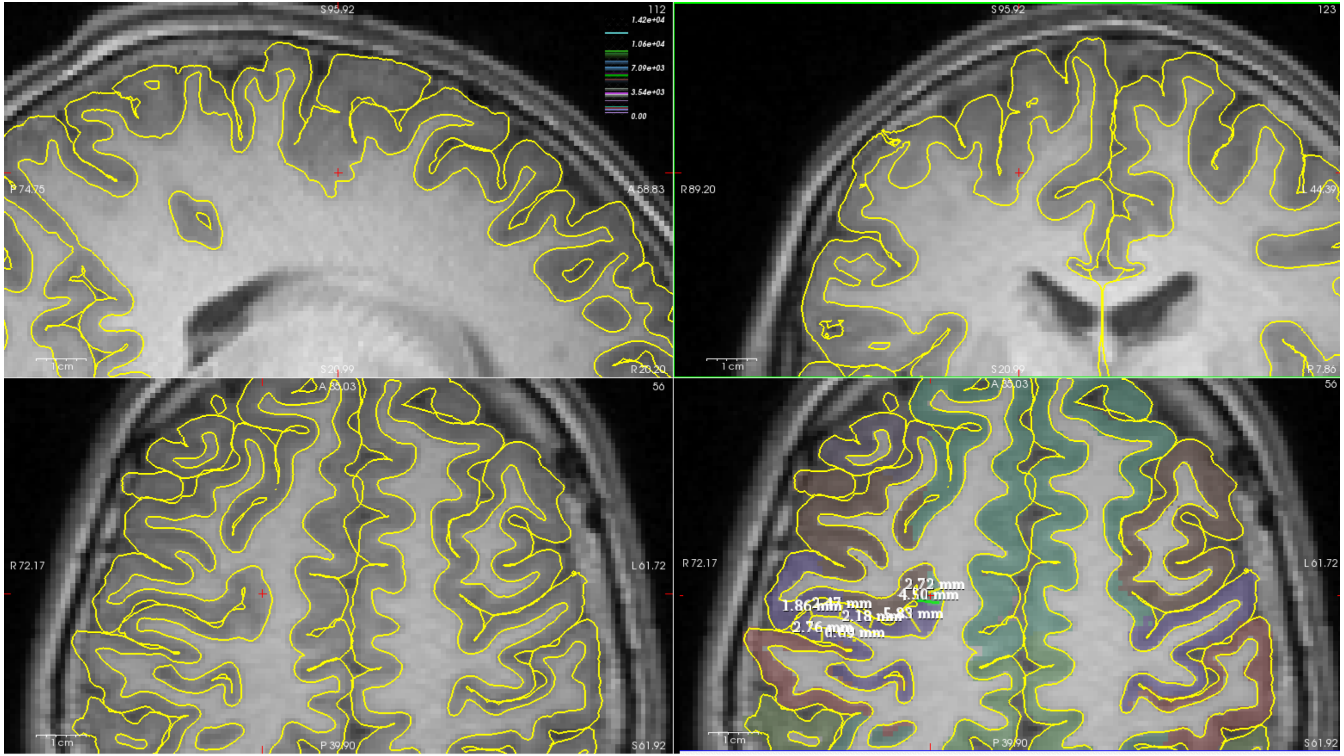
Left Hippocampus, Thalamus, Putamen



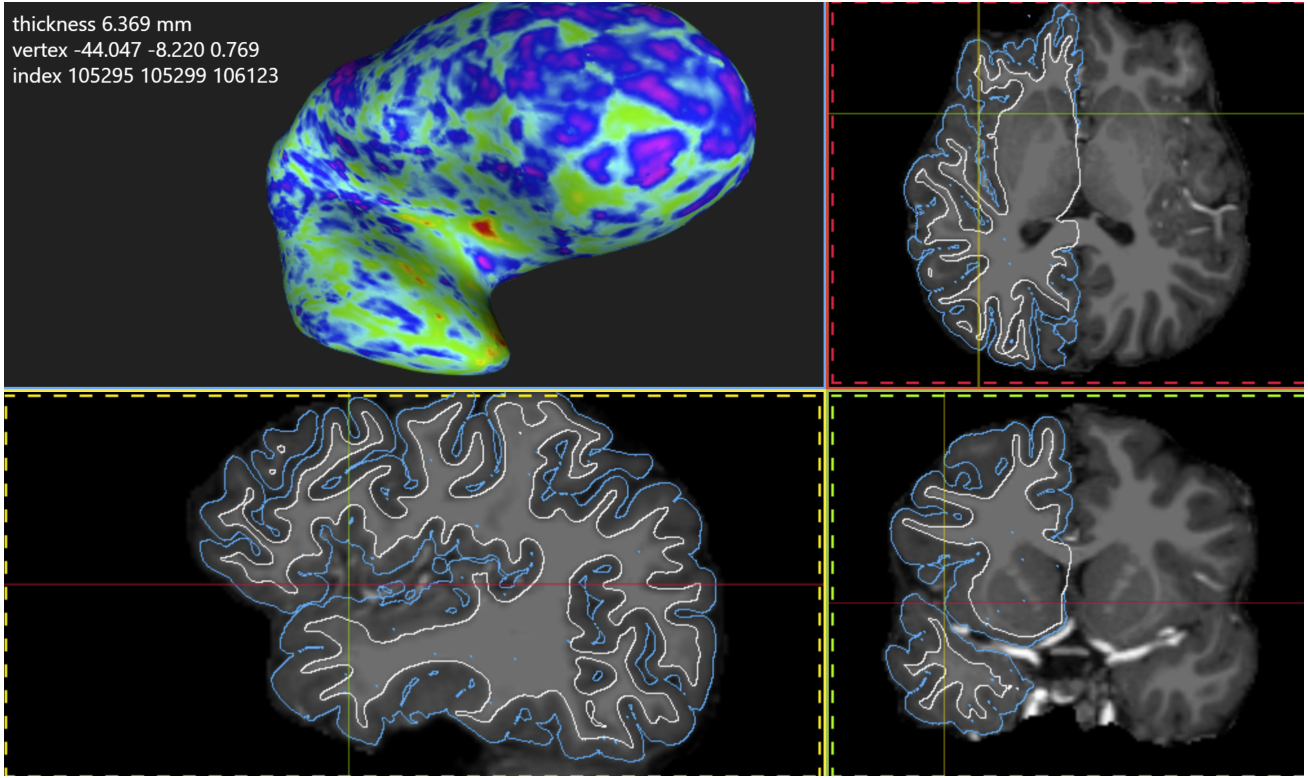


Skoltech
Skolkovo Institute of Science and Technology

Brain tissue segmentation: epileptogenic focus



Visualization system - 1: detection of lesions

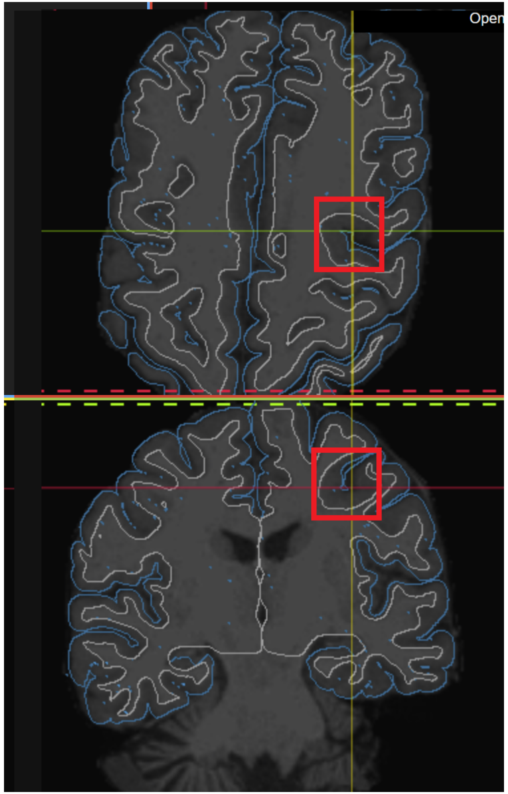


Raw values
(not normalized)

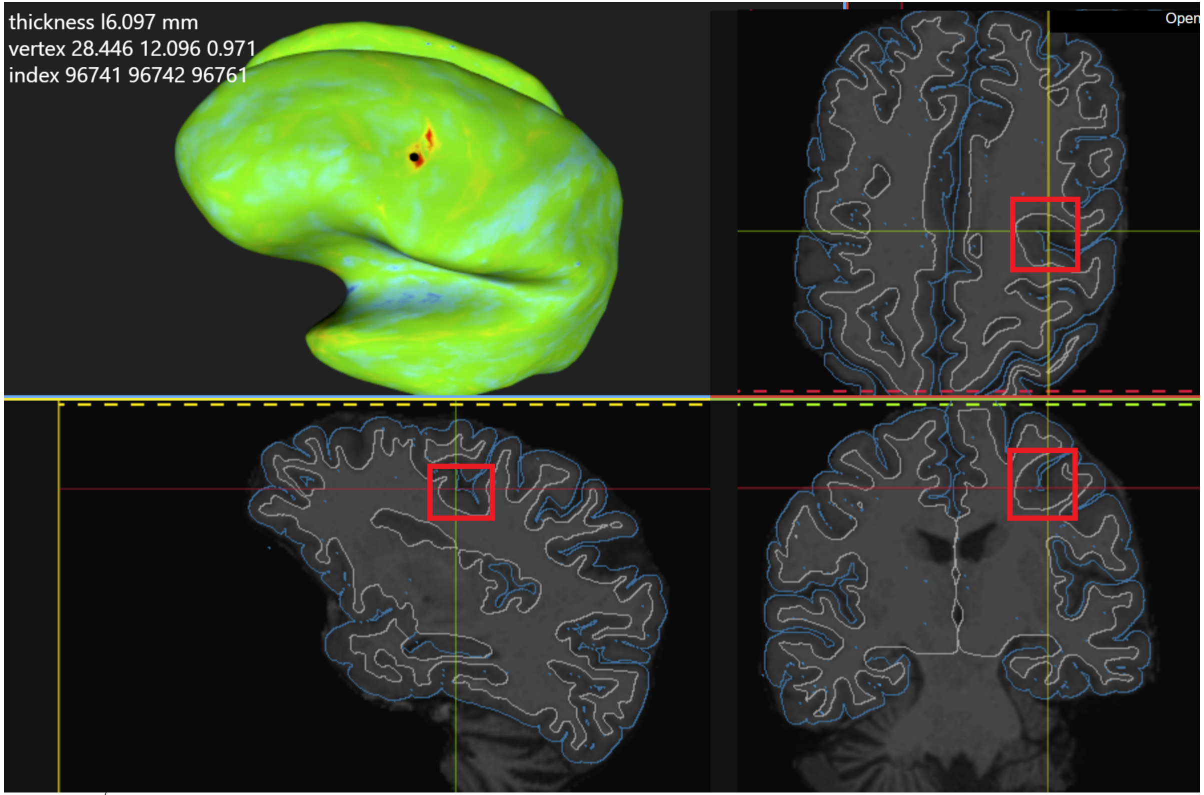
What information is available?

Kulakov NMRC:
- lots of MR images (retrospective study)
- all with text reports
- none properly annotated

Data for medicine (one patient) → Data for AI (dataset)

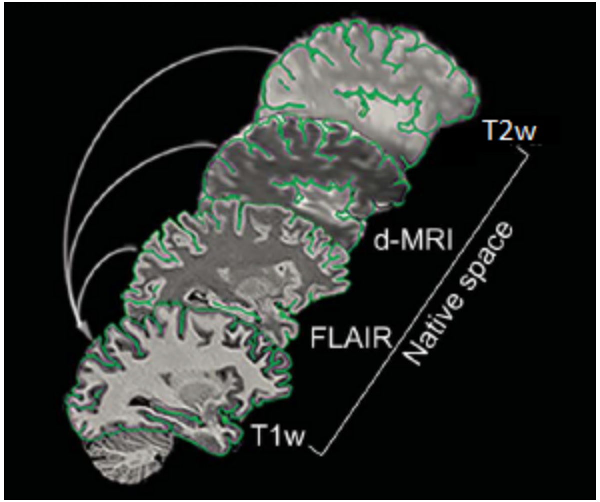


Visualization system - 2: detection of lesions



- Manual labeling assistance (normalized values)
- Fully-automated labeling

Data labelling and weakly supervised learning



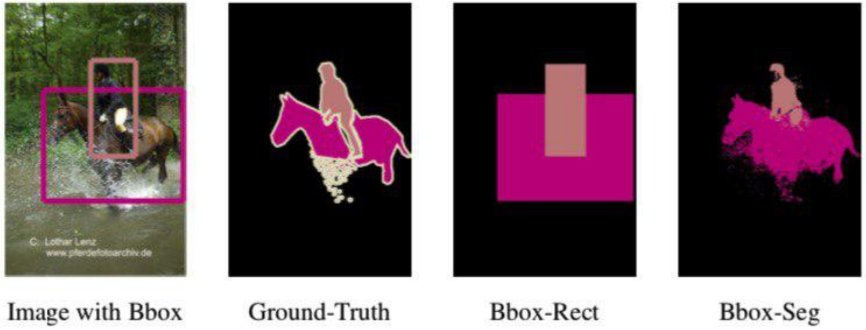
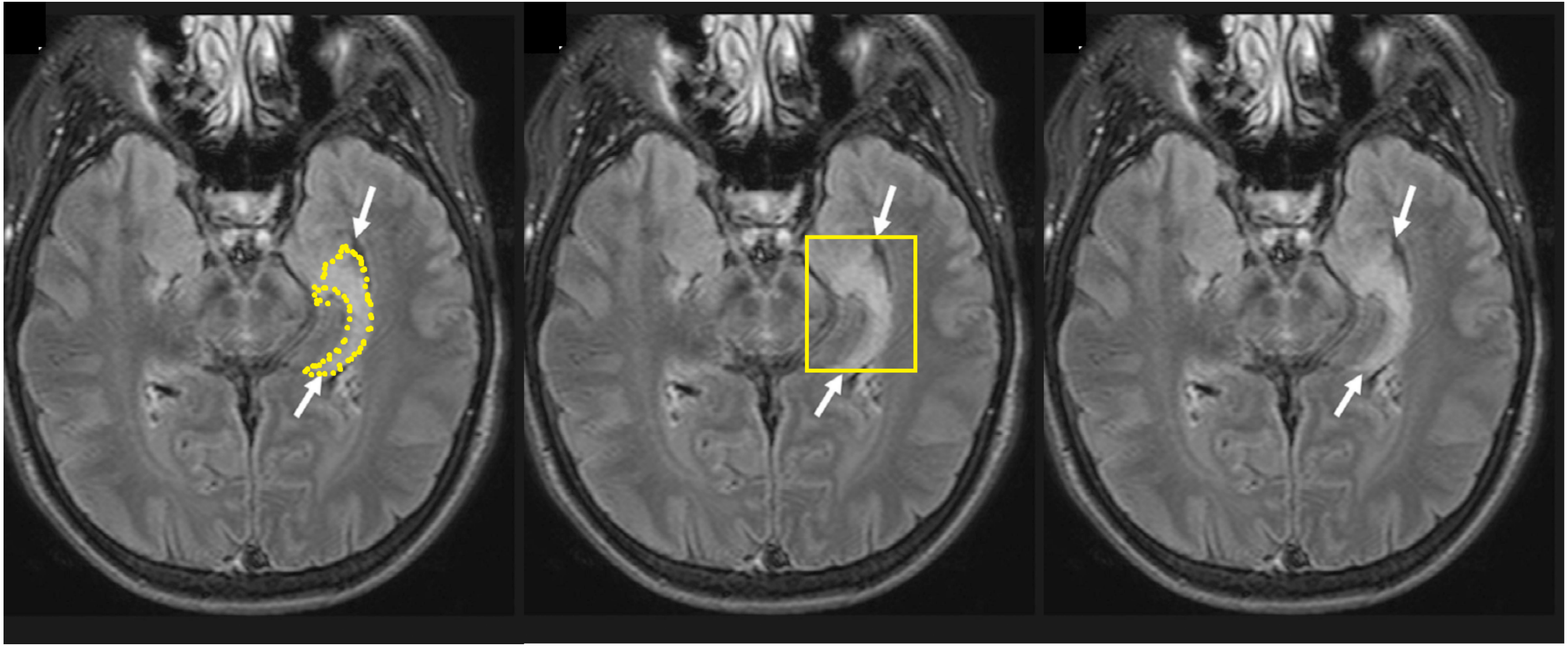


Figure 4. Estimated segmentation from bounding box annotation.

*Weakly- and Semi-Supervised Learning of a Deep Convolutional Network for Semantic Image Segmentation, 2015

Data labelling and weakly supervised learning

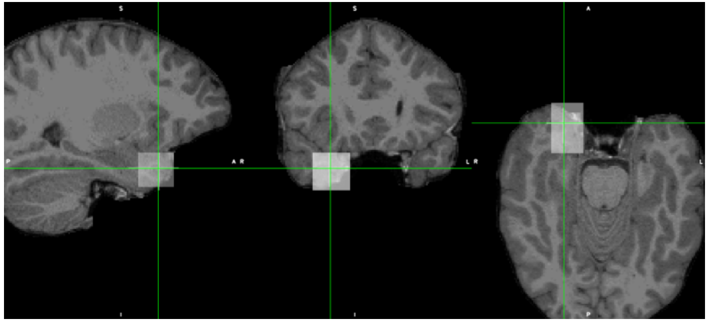


After we have labelled data

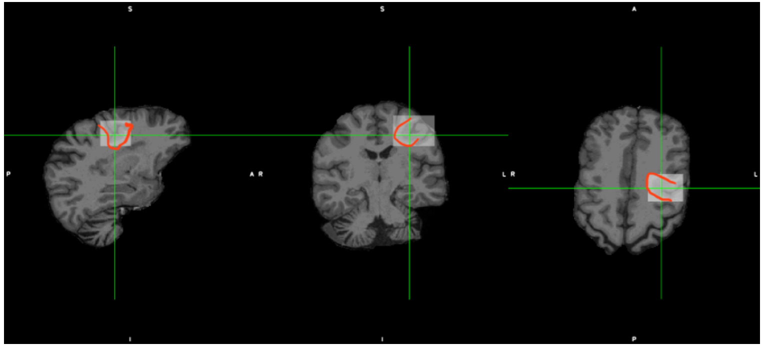
30 FCD subjects
- T1w images
- 256x256x256, 1mm

15 labeled FCD subjects
- weak labeling
- not orthonormalized
- each view (axial, sagital, coronal) is labeled with 2d rectangle by specialist
- we, as non-specialists, could observe FCD lesions only on two subjects (FCD5 and FCD8)

15 unlabeled FCD subjects
- specialists could not detect FCD lesion
- from the same domain as labeled subjects
- can be used for training autoencoder



Subject FCD0

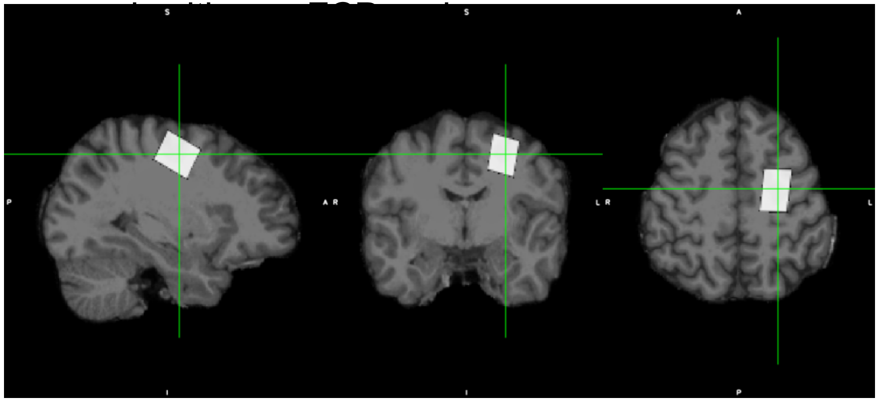


Subject FCD5

Dataset

Rectangular mask. Issues:

- After MRI alignment, rectangle squeezes and changes angle
- Mask becomes too big, overlapping too



*Rectangular mask, subject FCD1, after alignment**

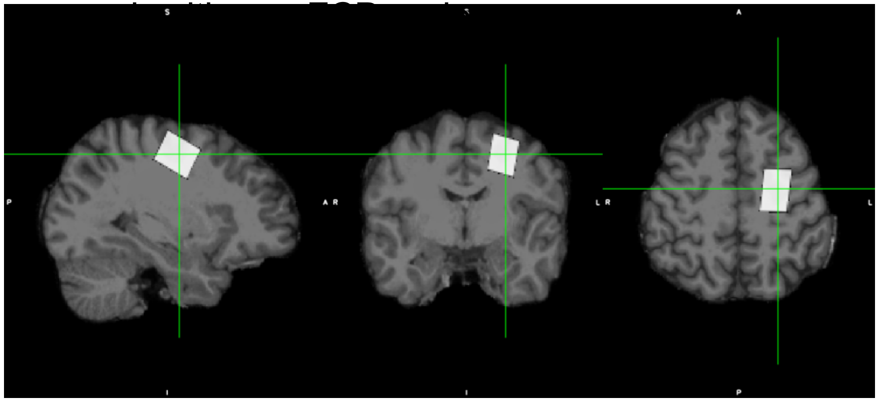
* For alignment we use FSL FLIRT algorithm with MNI152 template

Dataset

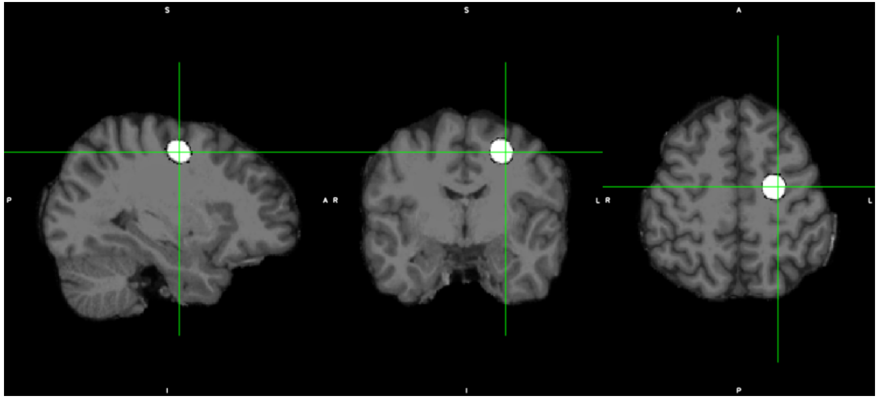
Rectangular mask. Issues:

- After MRI alignment, rectangle squeezes and changes angle
- Mask becomes too big, overlapping too

Use **ellipse** masks



*Rectangular mask, subject FCD1, after alignment**



*Ellipse mask, subject FCD1, after alignment**

* For alignment we use FSL FLIRT algorithm with MNI152 template

Preprocessing

- Bias field correction
- Skull stripping
- Image registration
  (FLIRT, MNI152 1mm template)
- Histogram Standardization (see Fig. 1)
- Z-normalization (mean=0, std=1)

FSL, FreeSurfer software, TorchIO library

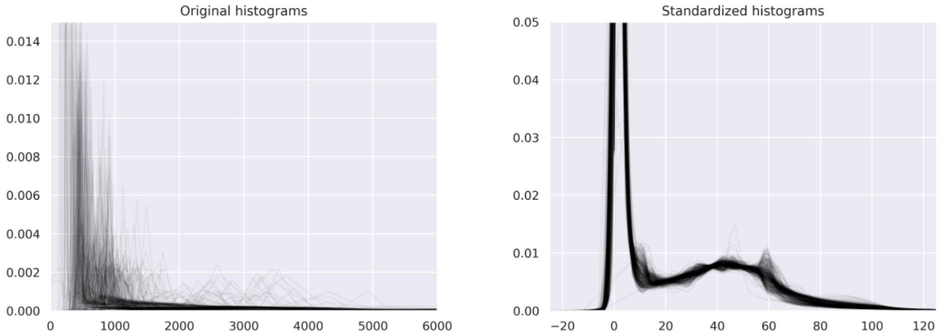
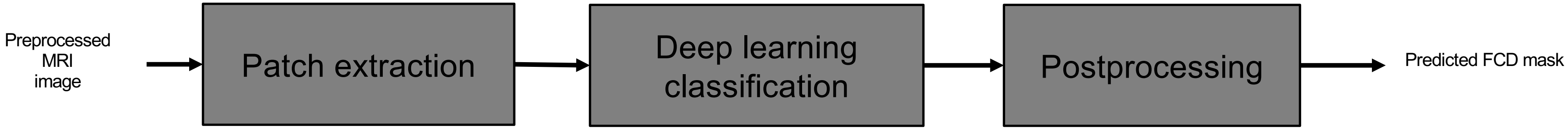


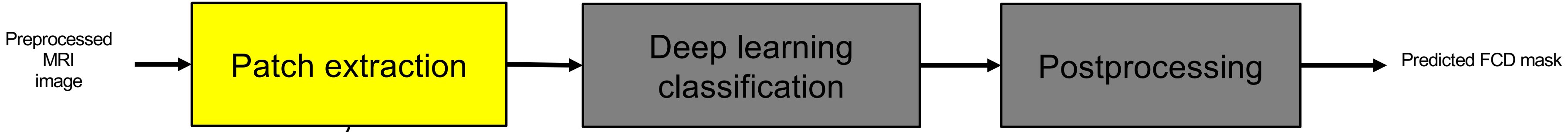
Fig. 1. Histograms of intensities before and after standardization

Baseline algorithm

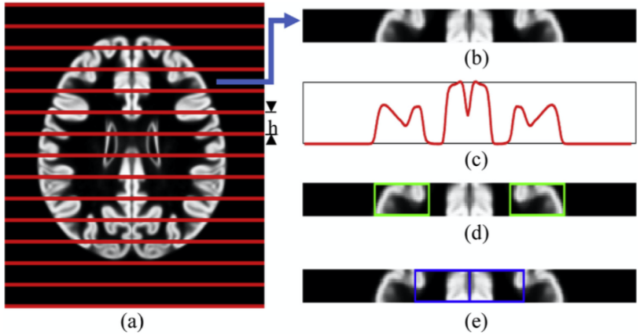


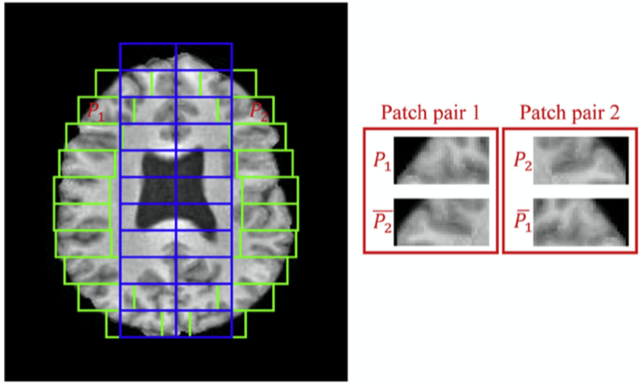
*From Automated detection of focal cortical dysplasia using a deep convolutional neural networks, 2020, Wang et al.

Baseline algorithm



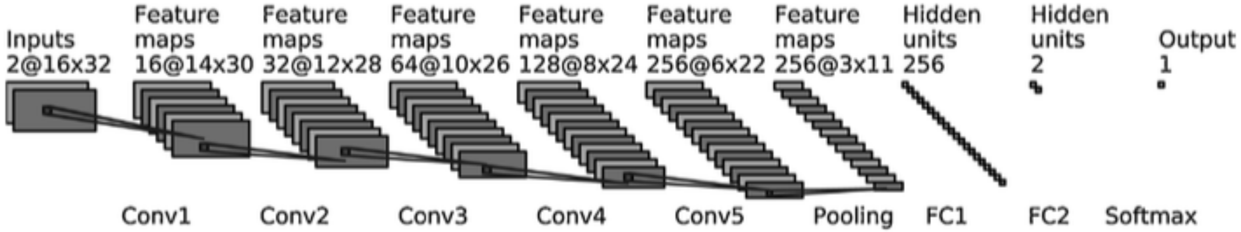
- Taking small patches along cortical regions and inter hemisphere tissue
- Each patch is stacked with its mirror reflection about the middle sagital line
- If a patch has any overlap with FCD mask, it is labeled as 'FCD'
- Upsampling via reextracting shifted versions of FCD patches
- Result: ~216000 patches, 30% of them are FCD





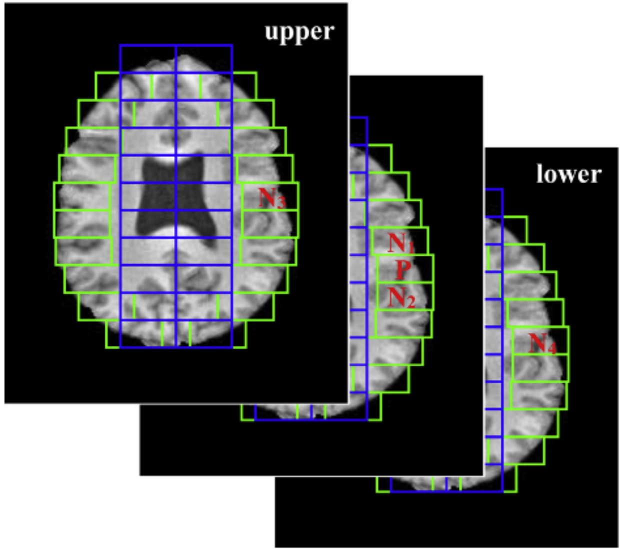
Baseline algorithm

- Standard CNN for binary classification:
- 4 conv layers
- 2 fc layer
- Dropout



Baseline algorithm

Preprocessed MRI image → Patch extraction → Deep learning classification → Postprocessing → Predicted FCD mask



- Choosing optimum threshold for patch classification
- If non-FCD patch's neighbors are FCD -> change to FCD
- If FCD patch's neighbors are non-FCD -> change to FCD

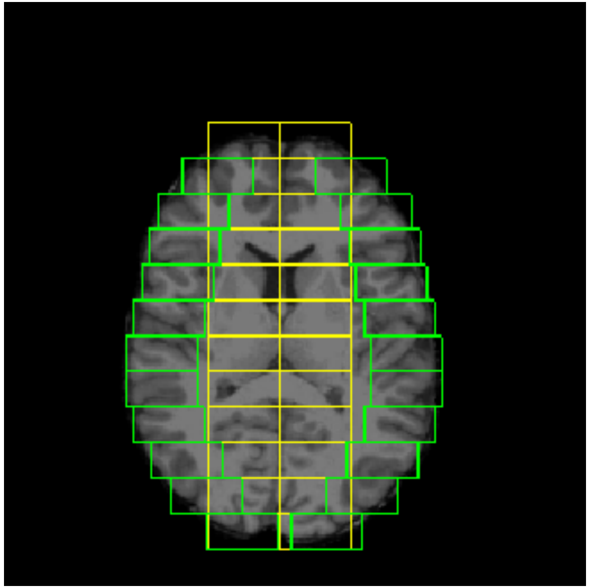
top-k score* = 0.20

*average number of top-20 by probability patches intersecting ground-truth FCD region

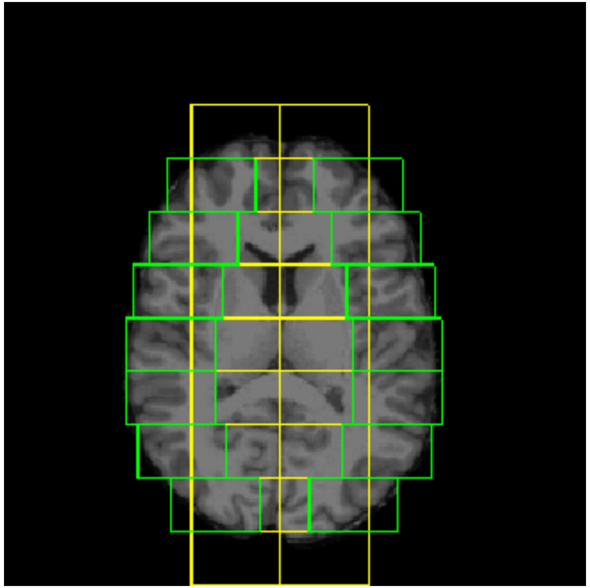
Improvements

- Increasing size of patches (16x32 -> 24x40)
  - With using small size of patches, there are some missing regions

Increasing size of patches



16x32 patches. Subject FCD1

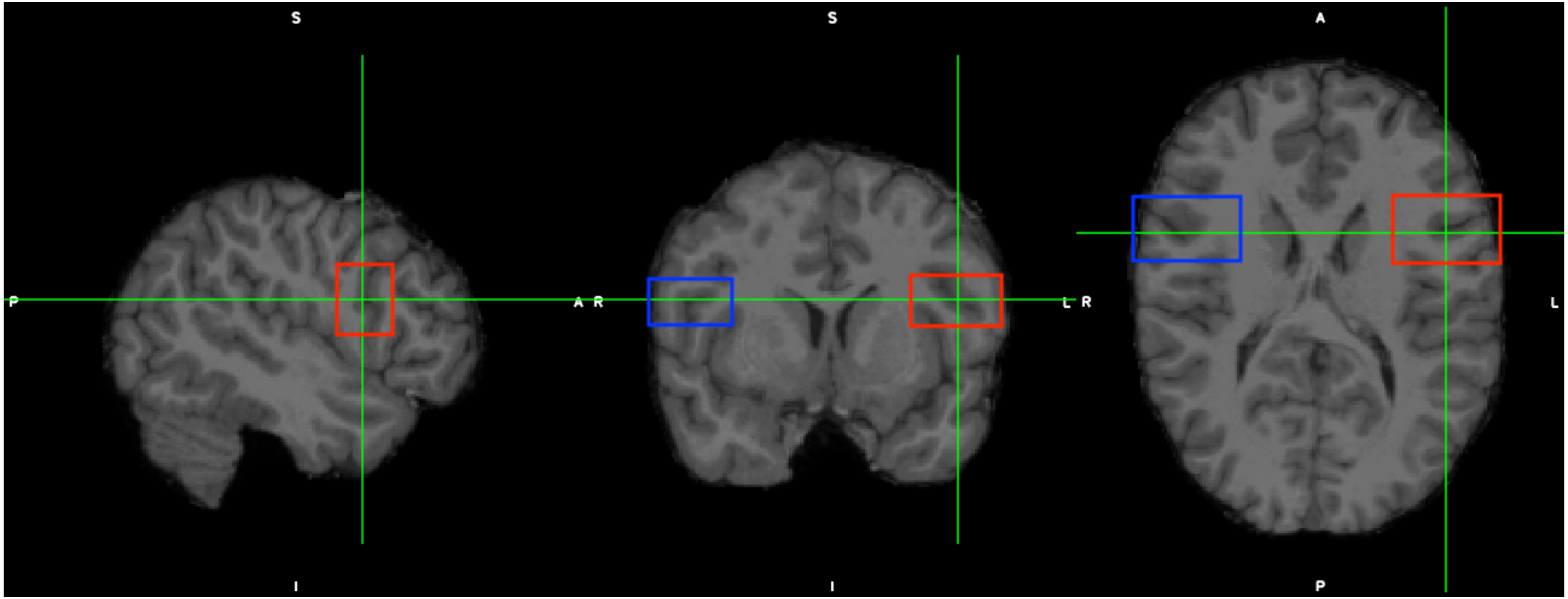


24x40 patches. Subject FCD1

Improvements

- Increasing size of patches (16x32 -> 24x40)
  - With using small size of patches, there are some missing regions
- Coronal, sagital slices
  - Why should we ignore 2 other dimensions?
  - To each axial patch concatenate coronal and sagital patches, orthonormal to it
  - Each input is now tensor of shape 6x24x40

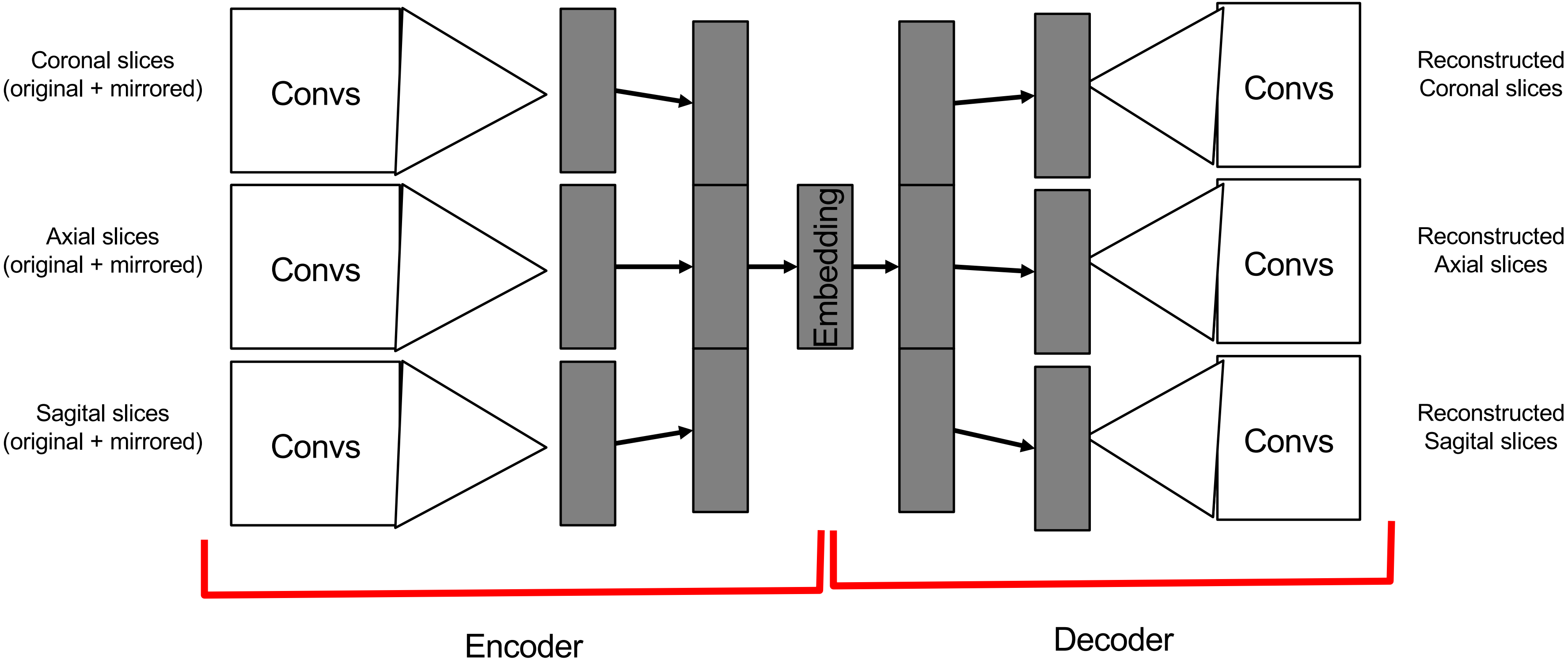
Improvements



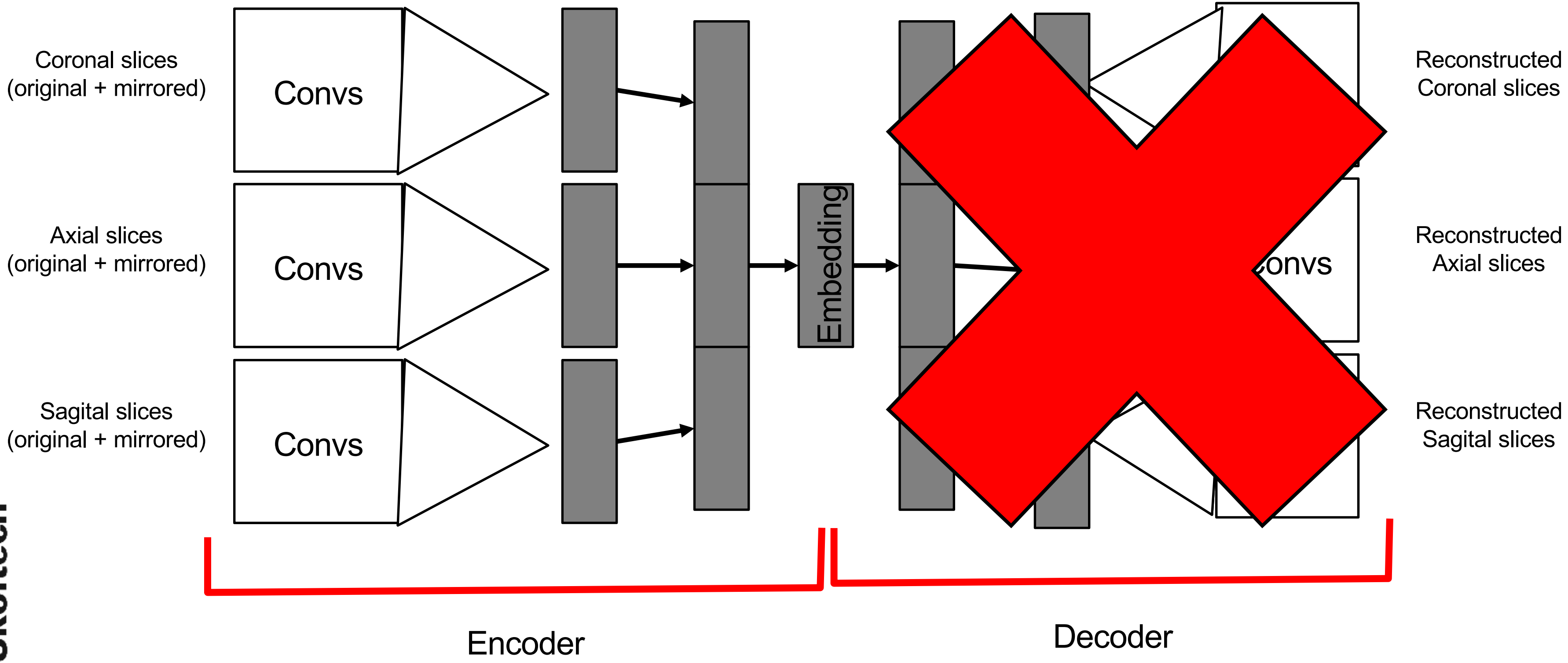
Improvements

- Increasing size of patches (16x32 -> 24x40)
  - With using small size of patches, there are some missing regions
- Coronal, sagital slices
  - Why should we ignore 2 other dimensions?
  - To each axial patch concatenate coronal and sagital patches, orthonormal to it
  - Each input is now tensor of shape 6x24x40
- Autoencoder pretraining
  - We can use unlabeled FCD subjects to pretrain model, using autoencoder

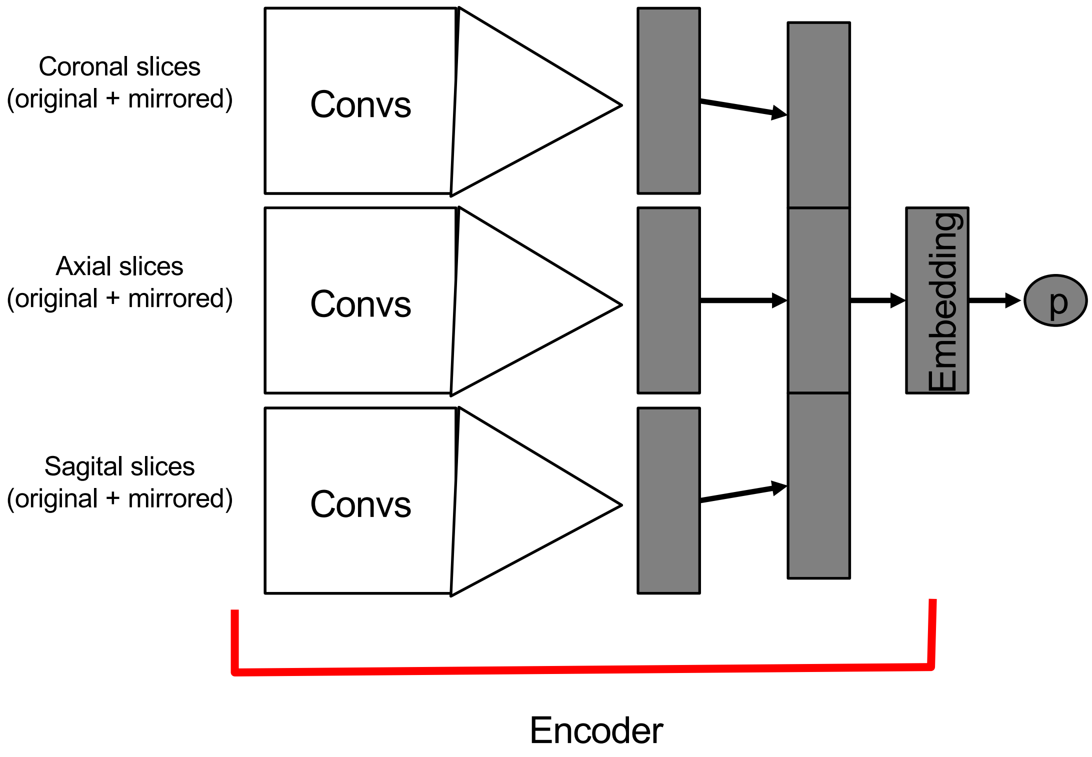
Autoencoder



Autoencoder



Autoencoder



Improvements

- Increasing size of patches (16x32 -> 24x40)
  - With using small size of patches, there are some missing regions
- Coronal, sagital slices
  - Why should we ignore 2 other dimensions?
  - To each axial patch concatenate coronal and sagital patches, orthonormal to it
  - Each input is now tensor of shape 6x24x40
- Autoencoder pretraining
  - We can use unlabeled FCD subjects to pretrain model, using autoencoder
- Patch labeling
  - Hard labeling: any overlap with mask -> 100 % FCD
  - Soft labeling:
    large overlap with mask -> 90 % FCD (as example)
    small overlap with mask -> 10 % FCD (as example)

Improvements

- Increasing size of patches (16x32 -> 24x40)
  - With using small size of patches, there are some missing regions
- Coronal, sagital slices
  - Why should we ignore 2 other dimensions?
  - To each axial patch concatenate coronal and sagital patches, orthonormal to it
  - Each input is now tensor of shape 6x24x40
- Autoencoder pretraining
  - We can use unlabeled FCD subjects to pretrain model, using autoencoder
- Patch labeling
  - Hard labeling: any overlap with mask -> 100 % FCD
  - Soft labeling:
    large overlap with mask -> 90 % FCD (as example)
    small overlap with mask -> 10 % FCD (as example)
- Ensembling specific-to-localization models
  - Temporal and non-temporal FCD regions are very different by their nature.
  - Separate them to enhance model performance

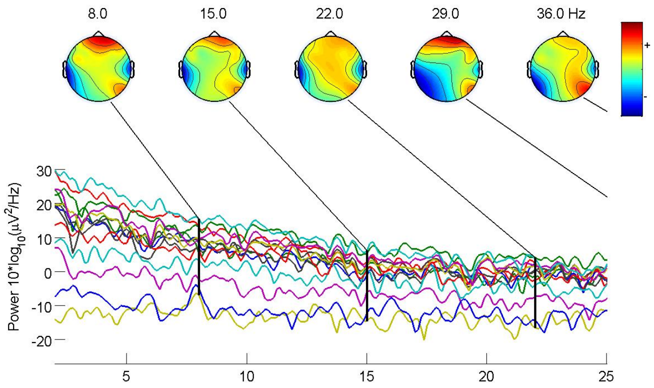
Ablation study

Notation	Model	Top-20 score
A	Baseline	0.200
B	+ soft patch labeling	0.267
C	+ increased patch size	0.400
D	+ autoencoder pretraining	0.533
E	+ temporal separation	**0.733**
F	+ coronal slices	0.667
G	+ sagittal slices	**0.733**

- **Neuroimaging data sources**
- **Neuroimaging data peculiarities**
- **Neuroimaging data analysis**
- **Biomedical tasks**
  - Functional brain areas mapping
  - Epileptogenic foci localization
  - Depression diagnostics
- **Educational tasks**
- **Neuroimaging in eSport**
- **Conclusions**

Depression types diagnostics

- **Medical partner**: V.P. Serbsky Moscow Research Institute of Psychiatry
- 128-channel EEG data on depression
- Aims:
  - to develop an algorithm that could be used as additional source of information at initial depression diagnostics
  - to develop an algorithm that could be used to distinguish between different types of depression



Used features

- Magnitudes of rhythm bands – mediocre results
- Coherence – good results

$$Coh_{xy}(f) = \frac{S_{xy}(f)}{\sqrt{S_x(f)S_y(f)}}$$

- Imaginary Coherence – poor results

$$ImCoh_{xy}(f) = Im(Coh_{xy}(f))$$

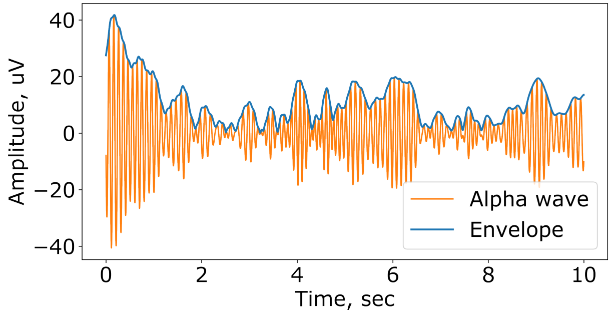
- Phase Lag Index – poor results

$$PLI_{xy}(f) = \left|\frac{1}{n}\sum sign(\varphi_x(f) - \varphi_y(f))\right|$$

- Phase Slope Index – poor results

- Alpha/beta band envelopes – good results

Alpha/beta band filtering
Hilbert transform and envelope extraction
Calculation of correlation between channels



Selected results

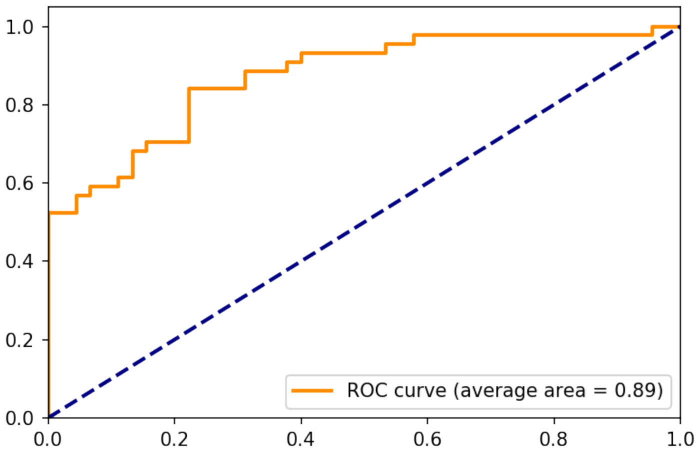
Model	ROC-AUC (Repeated 10-fold CV)
Rhythm features, best linear model	0.72
Coherence, best linear model	**0.88**
Coherence, best XGB model	0.83
Envelopes, best linear model	**0.89**
Envelopes, best XGB model	0.80
All features, final model	**0.98**
Neural Nets on raw signal	0.5

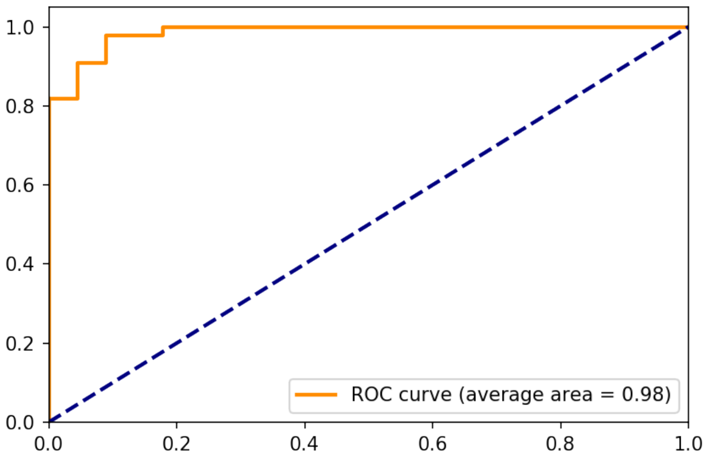
Receiver Operator Characteristic

Only envelopes correlation (alpha)

All features

Sensitivity
= Recall
= TP /
(TP + FN)

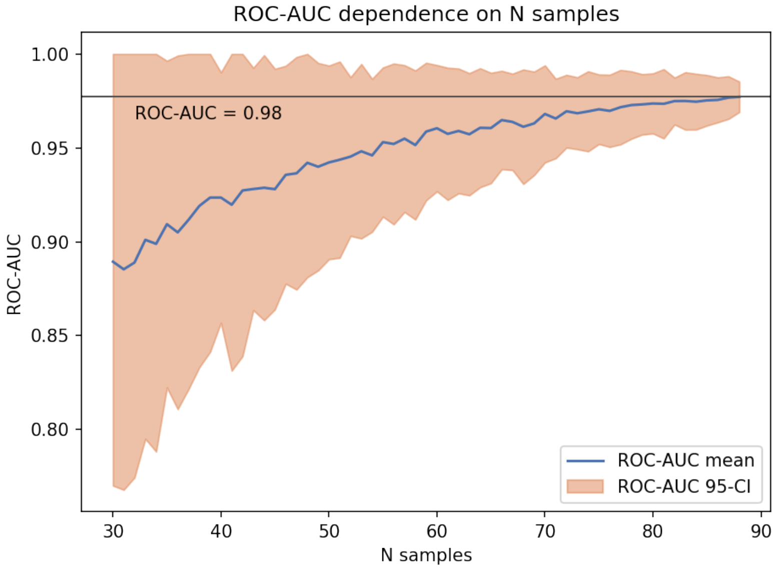


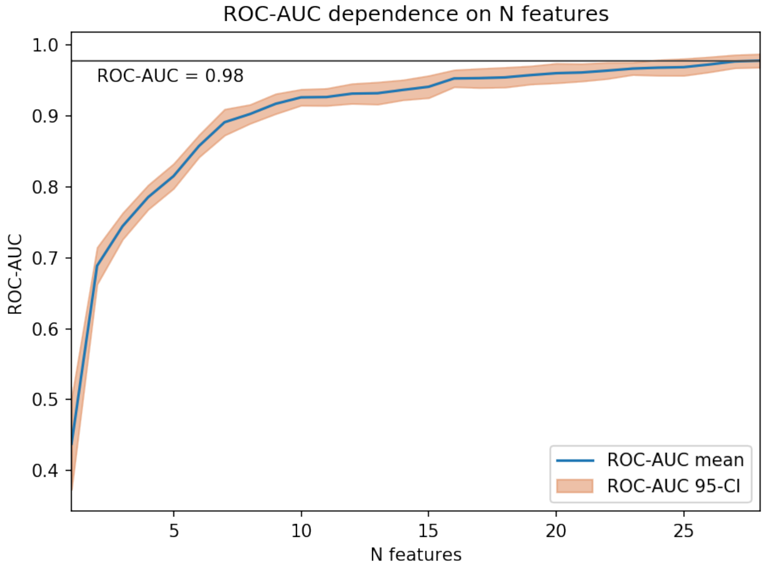


1 - Specificity
= False Negative Rate
= FN / (TP + FN)

Skoltech
Skolkovo Institute of Science and Technology

Dependence on sample size and N features





Next steps: MDD subtypes classification, source reconstruction, electrode clustering, applying neural networks and graph classification approaches. **Genomic/transcriptomic data**

- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- **Educational tasks**
- Neuroimaging in eSport
- Conclusions

Neuroeducation: motivation

“Adaptive Learning Startup Squirrel AI Raises CN¥1B (US$150 million)”

“Squirrel AI uses customized resources and learning activities to identify and address the **unique needs** of each learner based on their **profiles, learning level, strengths and weaknesses**—an educational method is known as AI Adaptive Learning”

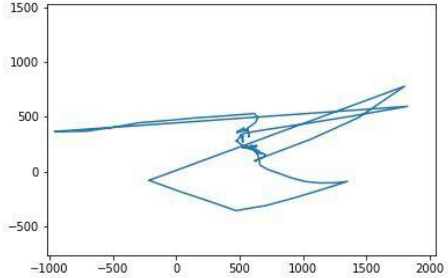
*from https://medium.com

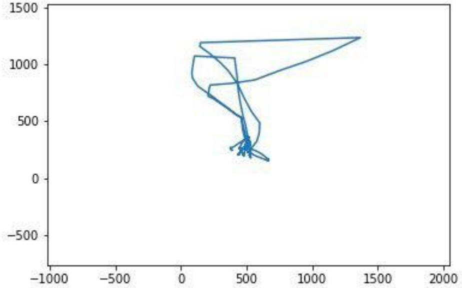
Neuroeducation

- Behavioral indexes (like accuracy and reaction time) and neuroimaging markers
- Are objective markers
- We aim at finding these correlates as well as their dependencies on study programs

Neuroeducation

Test performance prediction based on eye-tracking data





Skoltech
Skolkovo Institute of Science and Technology

Predicting cognitive performance using eye-movements, reaction time and difficulty level.

Marie Arsalidou[1,2,3], Valentina Bachurina[1,2], Svetlana Suchchinskaya[4], Maxim Sharaev[4], Evgeny Burnaev[4]

[1] HSE University, Moscow, Russian Federation, [2] Sirius University of Science and Technology, Moscow, Russian

Background

Cognitively challenging tasks require complex coordination of information beyond visual input. Mental attention capacity corresponds to the number of items one can attend to a process simultaneously[1]. Task that assess mental attention have multiple levels of difficulty[2].

Goal: Evaluate machine learning models that predict accuracy by reaction time, eye-movements and difficulty level.

References

1. Pascual-Leone (1970). Acta Psychologica, 32, 301-345
2. Arsalidou et al., (2010). Cognitive Development, 25, 262–277.

Российский научный фонд

Method

Participants: N = 57 (34 females; 20-30 years). EyeLink Portable Duo SR Research eye-tracker with 1ms temporal resolution (at 1000 Hz frequency) in remote head-free-to-move mode.

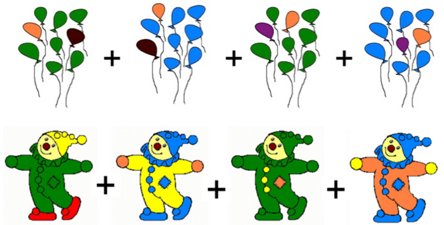


Figure 1
Color matching task. Participants indicate whether relevant colors are same or different Difficulty levels range from 1 to 6.

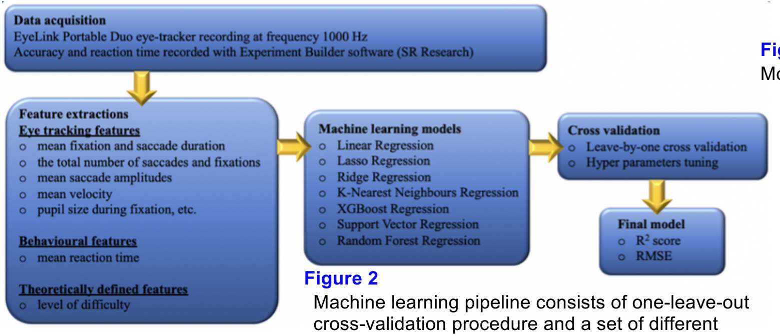


Figure 2
Machine learning pipeline consists of one-leave-out cross-validation procedure and a set of different machine learning algorithms

Results

All data: R^2 = 0.80; Reaction time R^2 = 0.73; Difficulty level: R^2 = 0.61; Eye-metrics: R^2 = 0.36

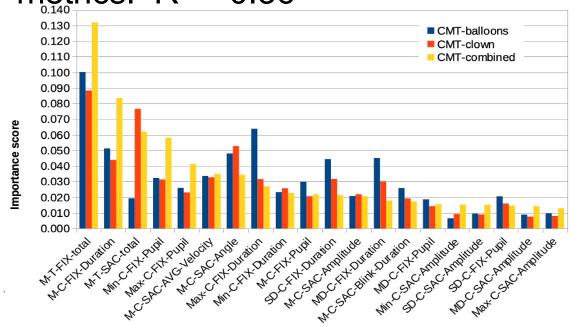


Figure 3
Most informative features based on the model on eye-tracking metrics

Conclusion

- Reaction time, difficulty level and eye tracking metrics are effective independent predictors of accuracy.
- Eye-tracking indices with the most importance include the number of fixations, number of saccades, duration of the current fixation and pupil size.

Machine learning, eye movements and mathematical problem solving

Maxim Sharaev[1], Svetlana Sushchinskaya[1], Valentina Bachurina[2,3], George Taranov[1], Evgeny Burnaev[1], Marie Arsalidou[2,3,4];
[1] Skoltech, [2] National Research University Higher School of Economics, [3] Sirius University of Science and Technology, [4] York University

Introduction

Major discoveries in technology and science often rely on mathematical skills. Mathematical knowledge is founded on basic math problem solving such as addition, subtraction, multiplication, and division. Research shows that problem solving is associated with eye movements that index allocation of attention.

Goal: Use machine learning on eye-tracking metrics dataset to predict performance on real-life user efficiency in mathematical task solving.

Methods

Participants & experiment design

Participants (n = 26, 18-30 years, 17 females and 9 males) viewed mathematical problems in three levels of difficulty indexed by 1-, 2-, and 3-digit problems along with four possible answers, while their eye movements were recorded. The experiment for each participant was divided in 3 runs with 15 blocks (same difficulty for the block) of math tasks (Fig. 1).

Data acquisition & pre-processing

Eye-tracking data was acquired with EyeLink Portable Duo SR Research eye-tracker with 1ms temporal resolution (at 1000 Hz frequency) in remote head-free-to-move mode (Fig. 1). Events and their characteristics such as saccades and fixations were obtained from EyeLink Data Viewer.

Acknowledgements: Support is gratefully acknowledged from the Russian Basic Research Foundation (#19-313-51010)

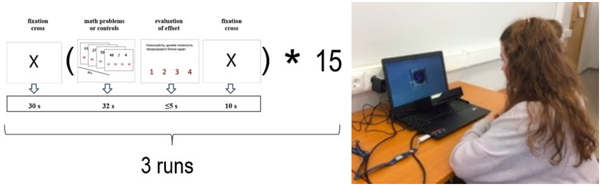


Fig. 1. Participant mathematical tasks pipeline eye-tracking system

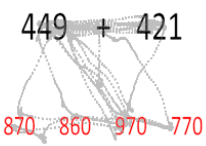


Fig. 2. Sample task lvl-3 with eye-tracking path

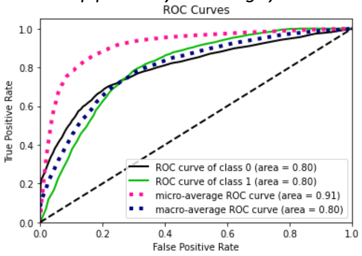


Fig. 3. Random Forest ROC Curves

Modelling

Dataset - 9200 images and results (Fig.2).
Numerous features were drawn from events data - mean, median, standard deviation, maximum and minimum values per trial.
5 fold cross-validation (divided by participants) was used in modelling. Different types of algorithms were used - Logistic Regression, Random Forests, Boosting Trees.

Results

Results show that trial correctness can be classified with a 0.80 ROC AUC score based (Fig. 3).
Predicting the task difficulty level of each trial was attained with 62% accuracy, which is significantly better than the random prediction (i.e., 51% - all predictions are 1-difficulty).

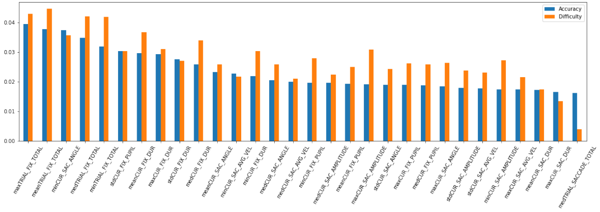


Fig. 4. Most informative features for Random Forest classifier for Accuracy and Difficulty prediction

Conclusion

Critically, no study to date has evaluated eye-tracking metrics associated with mathematical operations using machine learning approaches to classify trial correctness and predict task difficulty level.

- The most important features include metrics associated with current pupil fixation, current saccade amplitudes, and current fixation duration (Fig. 4)
- Theoretically, findings contribute to theories of mathematical cognition
- Practically, algorithms can contribute to further research in ML use for mathematical problems, which potentially has applications in education in terms of assessment and personalized learning

- Neuroimaging data sources
- Neuroimaging data peculiarities
- Neuroimaging data analysis
- Biomedical tasks
- Educational tasks
- Neuroimaging in eSport
- Conclusions

$612M
2015 eSPORTS Spent

134M
Watches

$309
Average annual cost per member

TWITCH
> 70% TV traffic USA

7.5M
Players in LoL

12M
Attended competitions in the USA and Europe

Neurotechnologies in cybersport: motivation

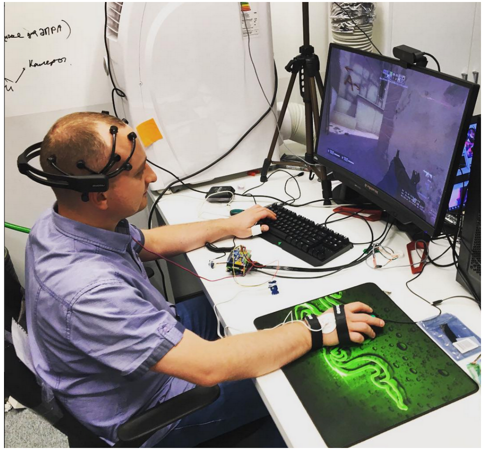
- ➜ Professional trainer Tsagolov: "We have no tools for a sportsman **training mental assessment**…"
- ➜ Irina Semenova, e-sport coach, Virtus Pro founder: "Professional trainers rely on their **own experience** and **visual feedback** rather than on **science**".

We address the following problems:

Quick reaction Under pressure	Sustained concentration over time	Mental resources allocation

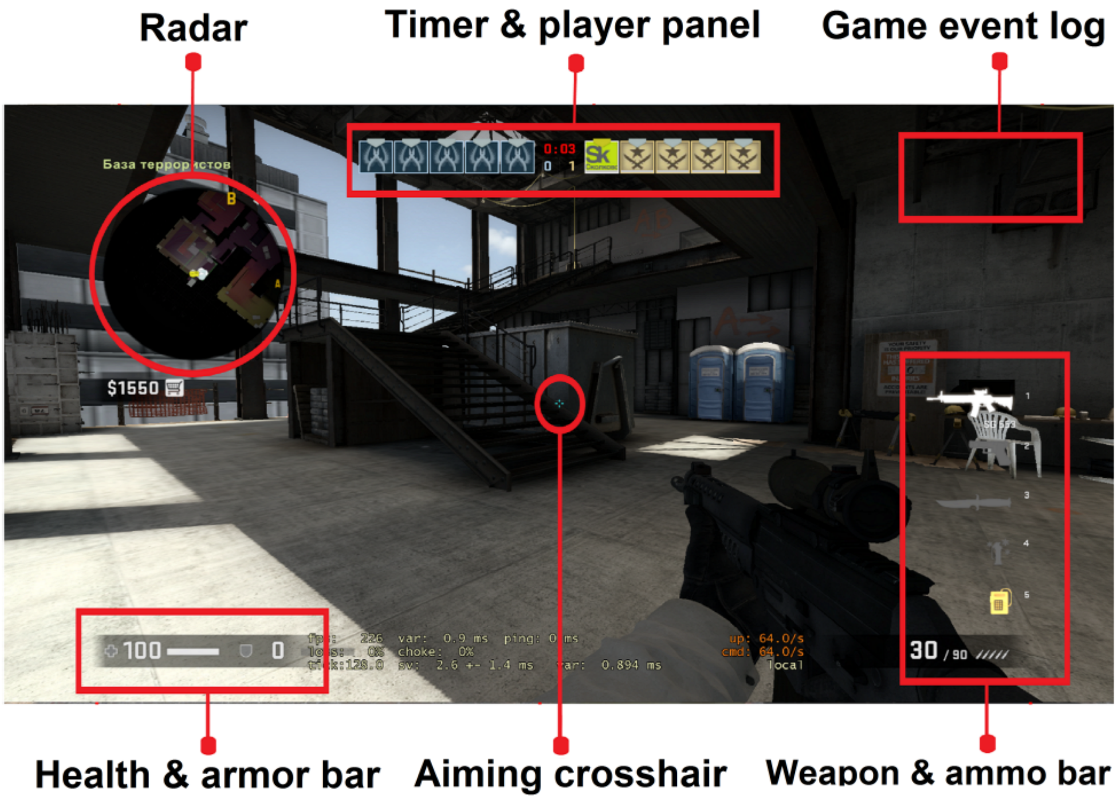
Neurotechnologies in cybersport

- **Partner**: Skoltech Cyberacademy
- Aims:
  - find and explore biomarkers of performance in games, make predictions and recommendations
  - use marked data as a starting point for other tasks like monitoring operator’s fatigue



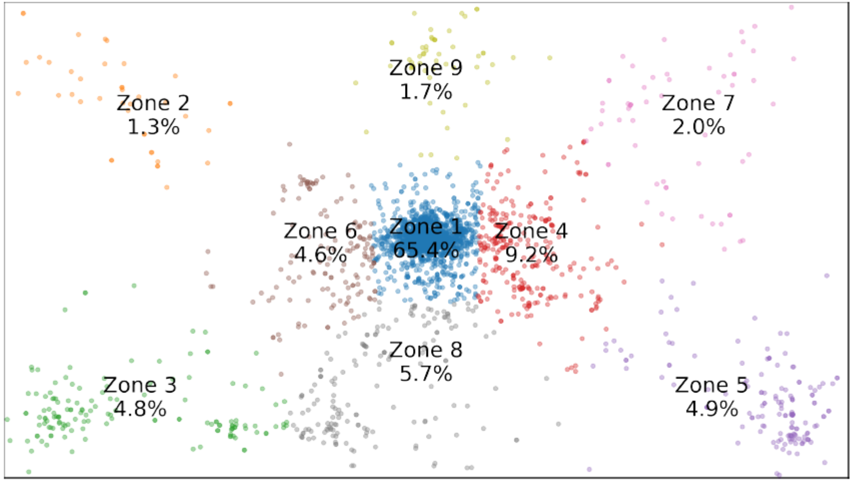

29:15
1 0
ВЫ ИГРАЕТЕ ЗА КОМАНДУ ТЕРРОРИСТОВ
Дальнобой
Skoltech Experimental R...
eau
B
AK-47
Игрок BATTERS123 покидает игру [connect R.ub1que.ru]
73
95
fps: 199 var: 1.0 ms ping: 4 ms
loss: 0% choke: 0%
tick:128.0 sv: 3.9 +- 0.6 ms var: 0.015 ms
up:128.0/s
cmd:128.0/s
online
28 / 90

eSport: eye-tracking

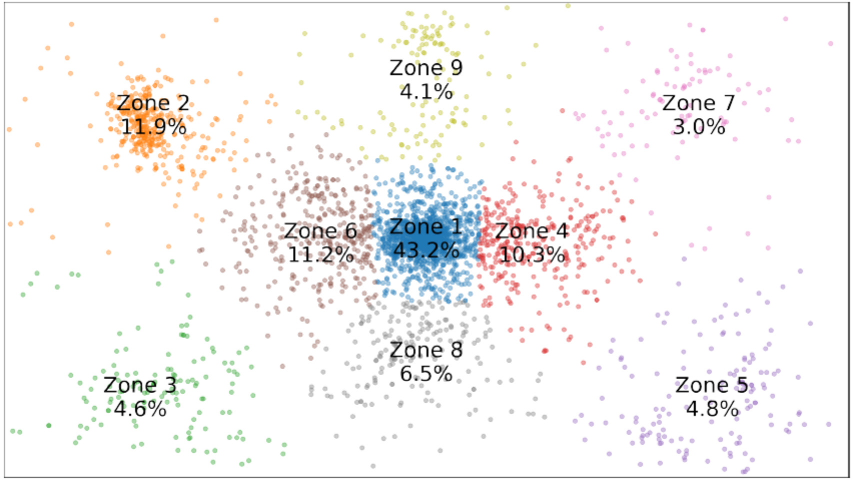


Skoltech
Skolkovo Institute of Science and Technology

eSport: eye-tracking

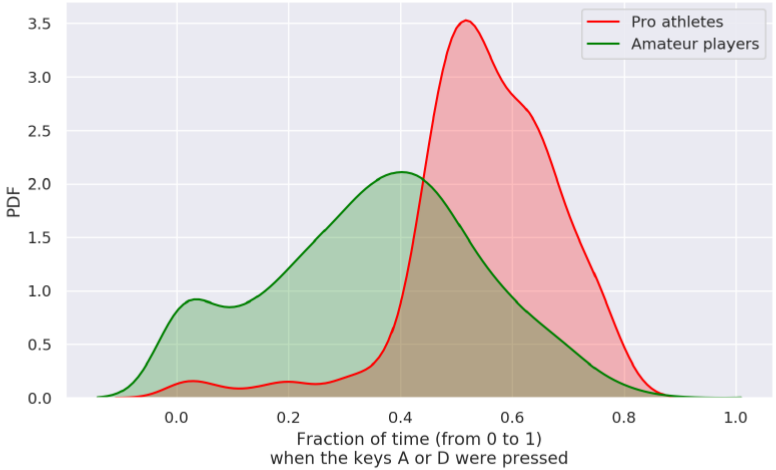


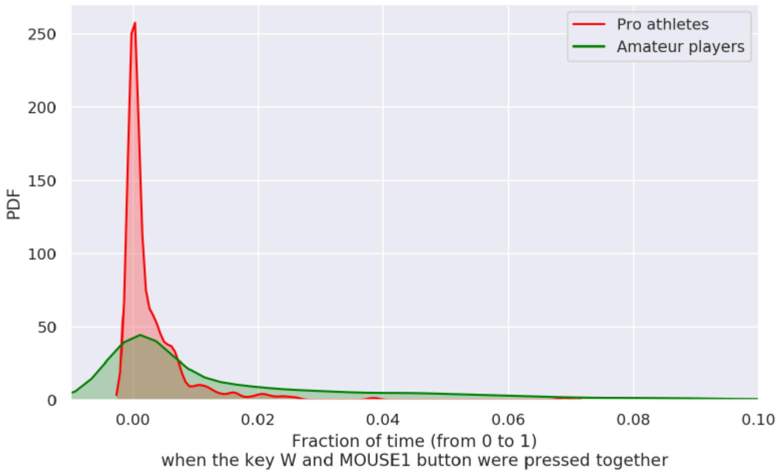
(a) Professional player's gaze.



(b) Amateur player's gaze.

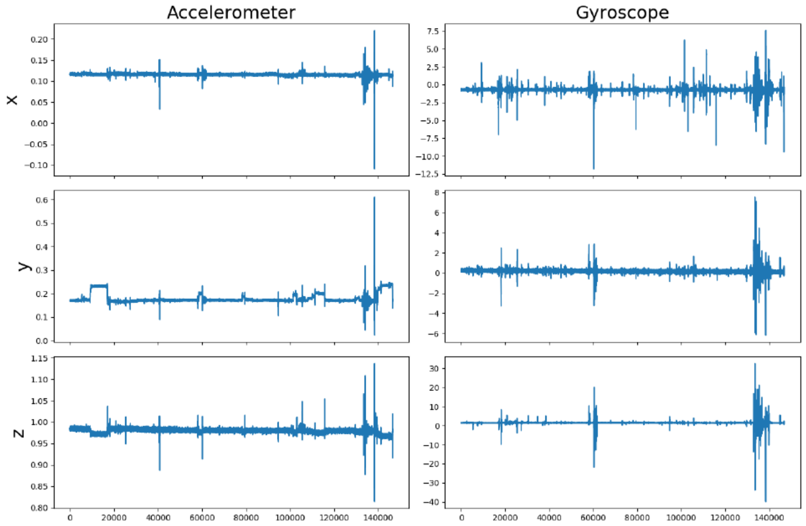
eSport: keyboard/mouse

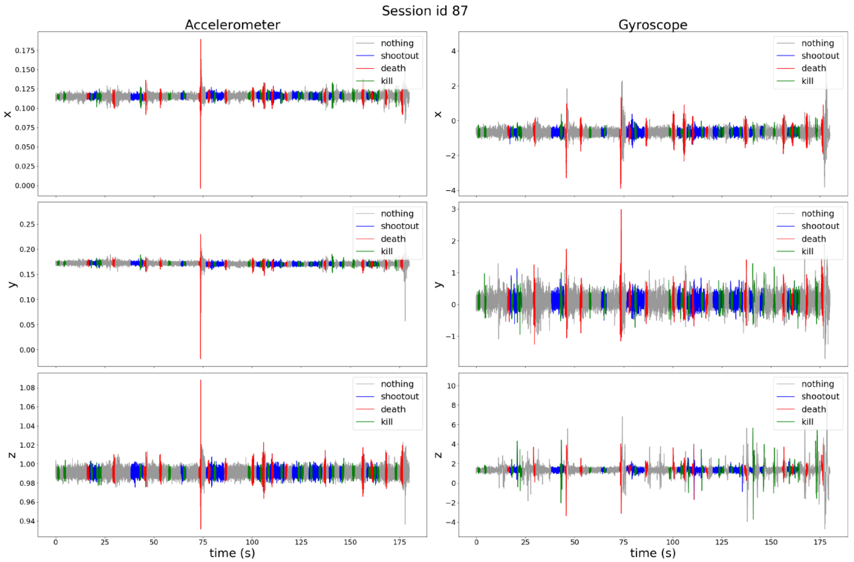




Skoltech
Skolkovo Institute of Science and Technology

eSport: Smart Chair

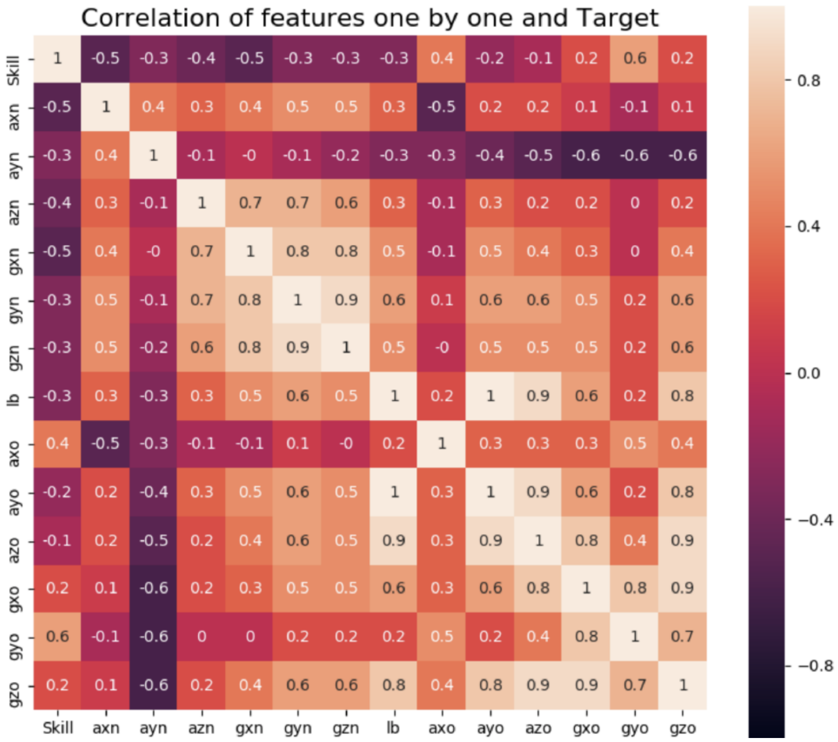




Skoltech
Skolkovo Institute of Science and Technology

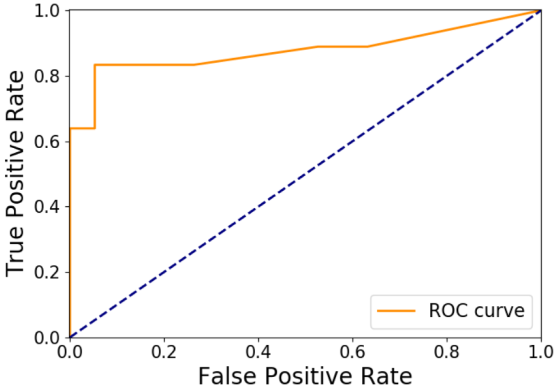
eSport: Smart Chair

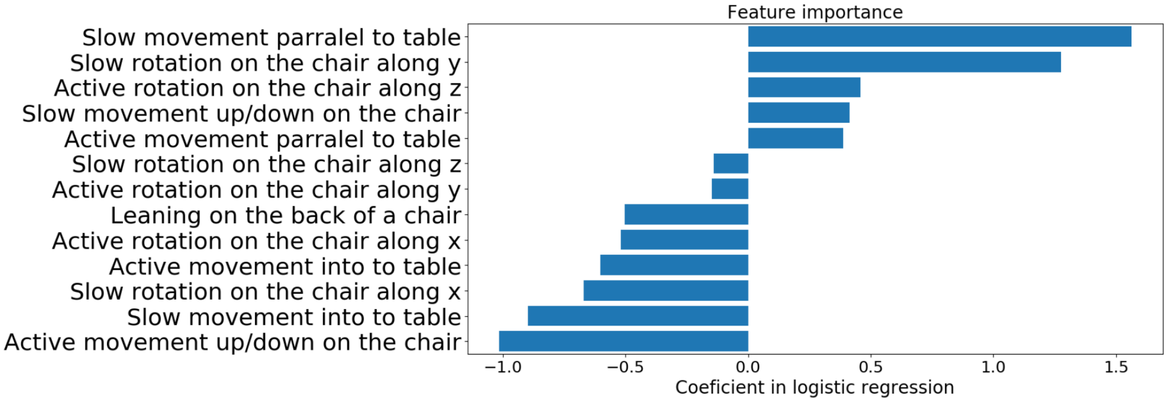
Name	Description
Skill	Players skill in CS:GO. 0 or 1 (low or high)
axn	Active movement along x-axis. Intensity of rectilinear motion to the right/left.
ayn	Active movement along y-axis.Intensity of rectilinear motion to/from table.
azn	Active movement along z-axis. Intensity of rectilinear motion up/down.
gxn	Active rotation on x-axis. Frequency of approaching to/distancing from a monitor.
gyn	Active rotation on y-axis. Intensity of swaying to the right/left side of a chair.
gzn	Active rotation on z-axis. Intensity of rotations on a vertical axis.
lb	Portion of time when player leans to the back of a chair.
axo	Intensity of subtle rectilinear oscillations parallel to table.
ayo	Intensity of subtle rectilinear oscillations to/from table.
azo	Intensity of subtle rectilinear oscillations up/down.
gxo	Intensity of subtle approaching to/distancing from a monitor.
gyo	Intensity of subtle swaying to the right/left side of a chair.
gzo	Intensity of subtle rotations on a vertical axis.



eSport: Smart Chair

Method	AUC, mean	AUC, std
Logistic Regression	0.85	0.14
Support Vector Machine	**0.86**	0.13
KNN, 5 neighbours	0.80	0.13
Random Forest, depth 4	0.82	0.16





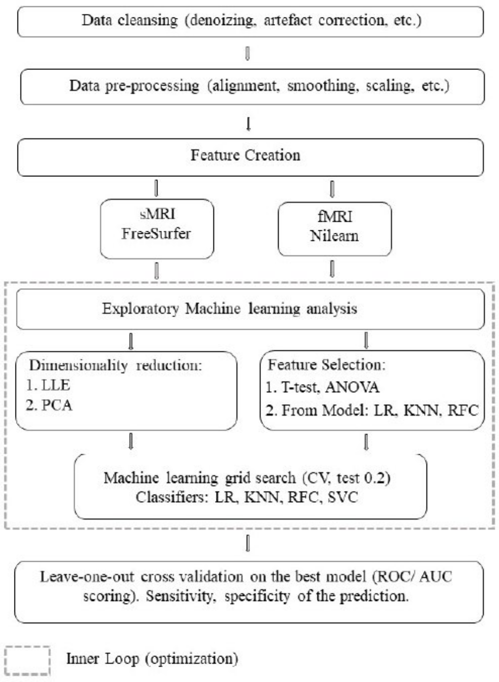
We need a Pipeline!

- ❑ Lots of multimodal datasets
- ❑ Similar preprocessing and analysis methods

- ❑ Routine work should be automatized (less labor, less errors)

- ❑ We introduce a block-pipeline, meta-data is saved on each step

Multimodal data input ⇨ Data preprocessing ⇨ Analysis and inference ⇨ Results visualization



Questions to Part 2?

Conclusions

- Neuroimaging data are the basis for solving applied problems and fundamental tasks of AI
- Neuroimaging data have peculiarities, due to both the complexity of the object (brain) and data acquisition methods
- Standard data processing methods, including denoising and feature extraction should use domain knowledge and models; CDISE's experience has demonstrated this
- To work in this area people need new specialization
- Open Data and collaboration is beneficial!

Thanks for attention!

Visit us

http://adase.group/neuro/

Skoltech
Skolkovo Institute of Science and Technology