

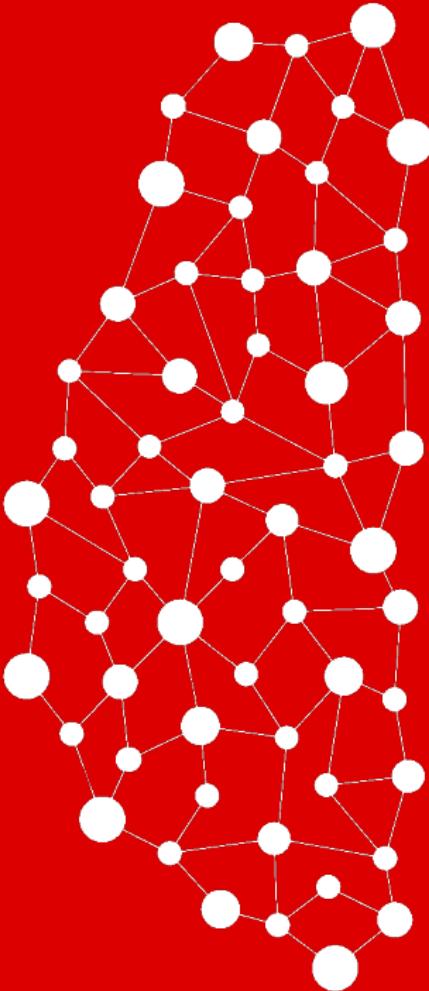
Exploring brain heterogeneity in neuropsychiatric patients with explainable artificial intelligence



Esten H. Leonardsen

Post-doc at the Department of Psychology,
University of Oslo

Chief Scientific Officer, baba.vision



UNIVERSITY
OF OSLO

Outline

1. Background and motivation
2. Explainable dementia classification
3. Explainable brain age predictions



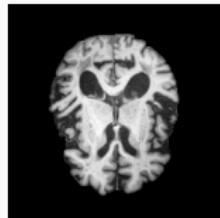
Background: Predictive clinical neuroscience



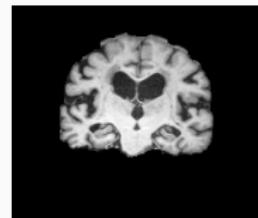
Background: Predictive clinical neuroscience



Side

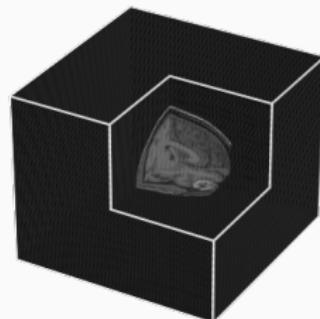


Above



Front

Structural Magnetic
Resonance Imaging (MRI) scans



Background: Predictive clinical neuroscience



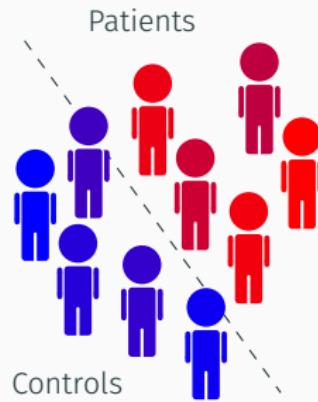
Background: Predictive clinical neuroscience



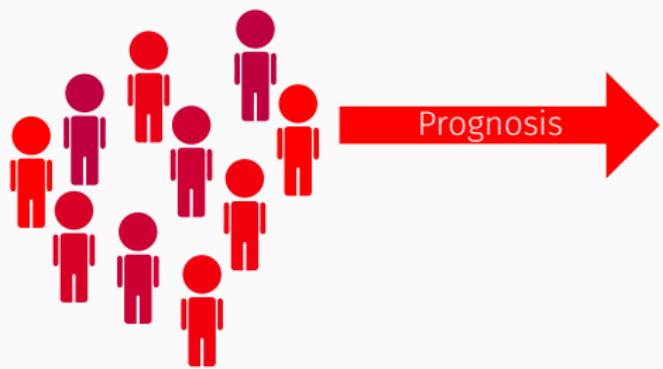
Background: Predictive clinical neuroscience



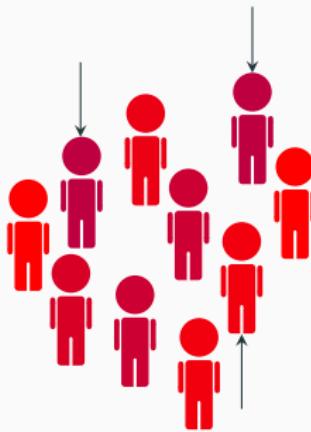
Background: Predictive clinical neuroscience



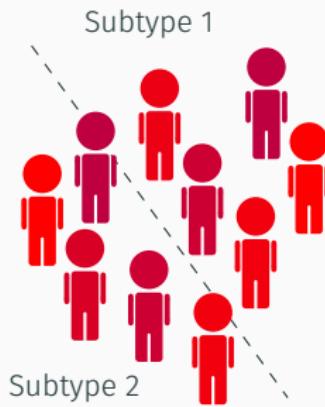
Background: Predictive clinical neuroscience



Background: Predictive clinical neuroscience



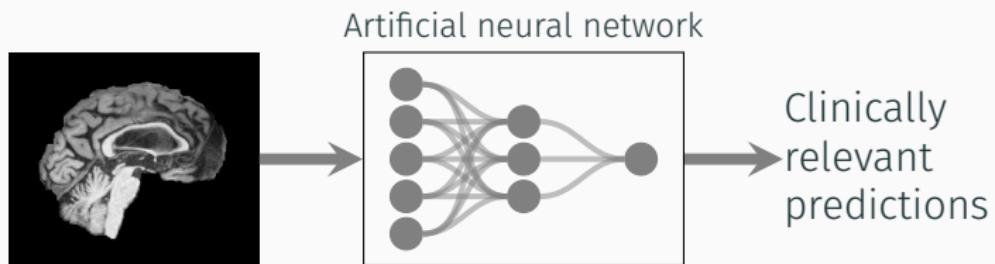
Background: Predictive clinical neuroscience



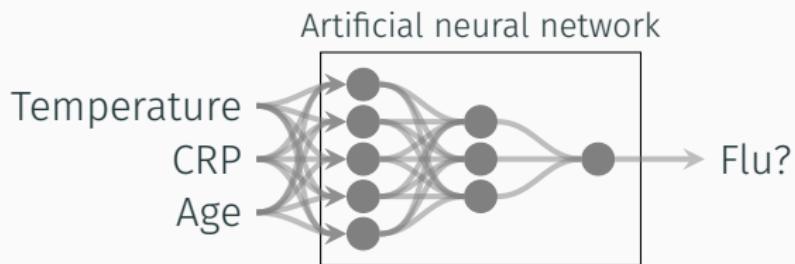
Background: Explainable artificial intelligence



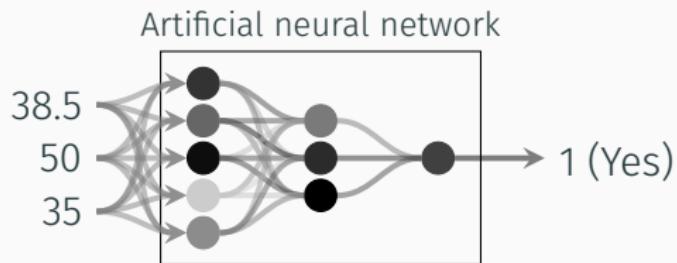
Background: Explainable artificial intelligence



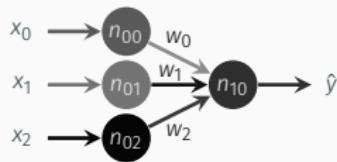
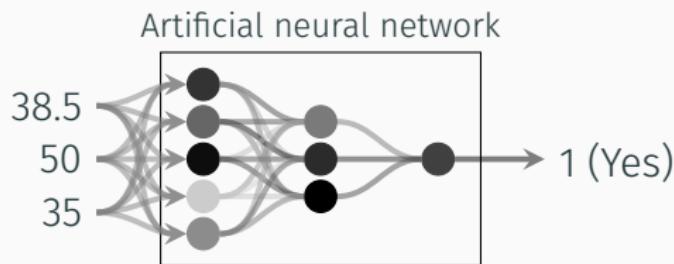
Background: Explainable artificial intelligence



Background: Explainable artificial intelligence



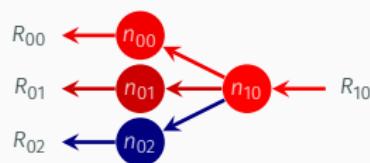
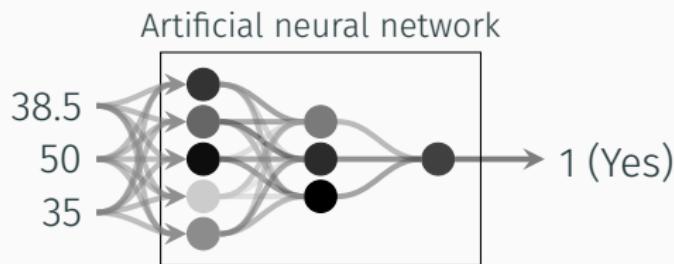
Background: Explainable artificial intelligence



$$\hat{y} = f \left(\sum_i^N w_i \cdot n_{0i} \right)$$



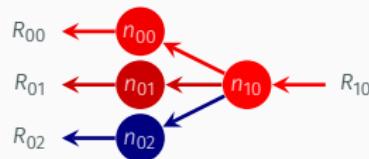
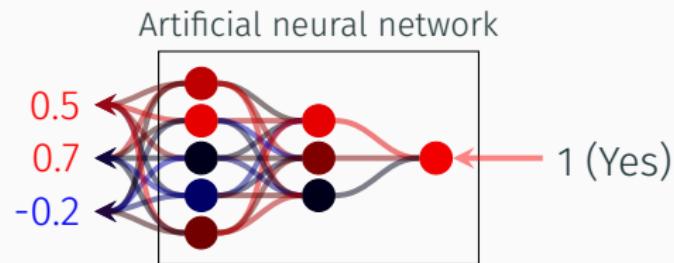
Background: Explainable artificial intelligence



$$R_{0i} = \sum_j \frac{n_{0i} w_i}{\sum_k n_{0k} w_k} R_{1j}$$



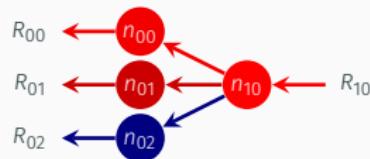
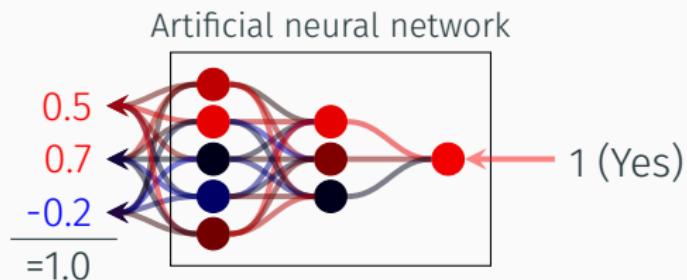
Background: Explainable artificial intelligence



$$R_{0i} = \sum_j \frac{n_{0i} w_i}{\sum_k n_{0k} w_k} R_{1j}$$



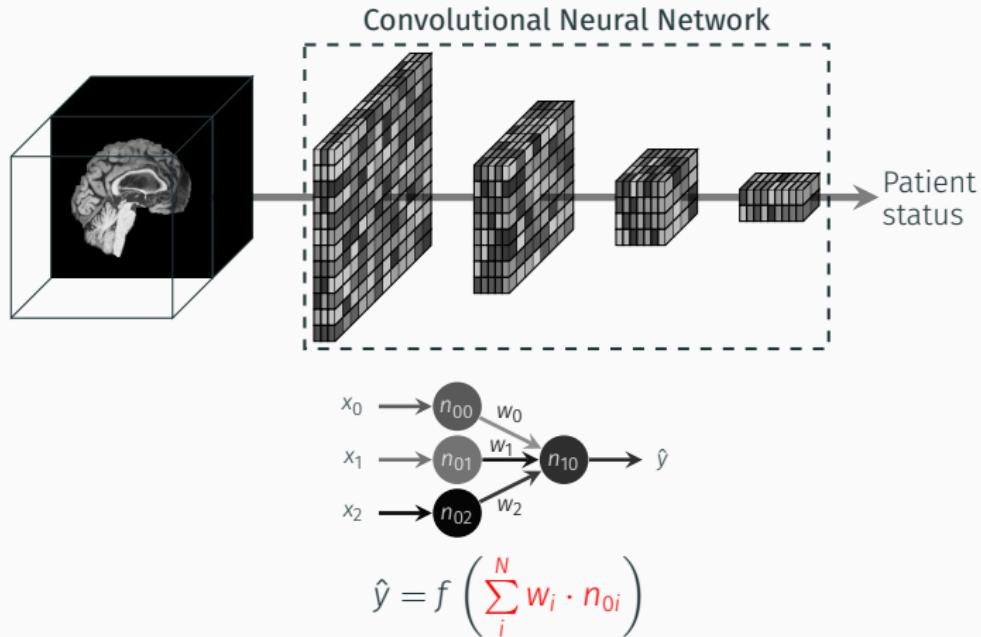
Background: Explainable artificial intelligence



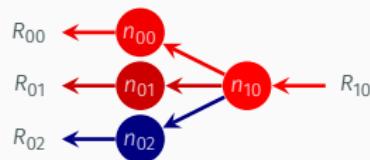
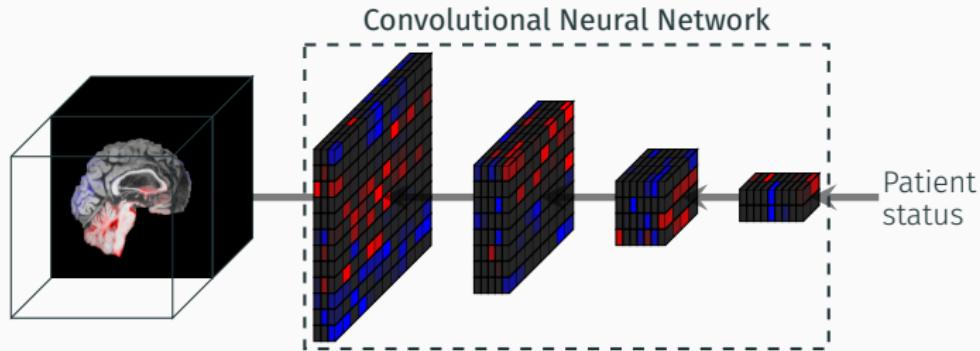
$$R_{0i} = \sum_j \frac{n_{0i} w_i}{\sum_k n_{0k} w_k} R_{1j}$$



Background: Explainable artificial intelligence



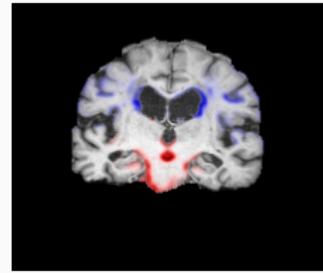
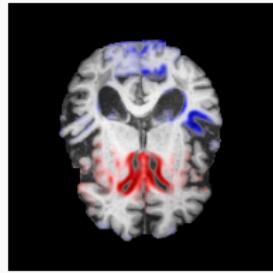
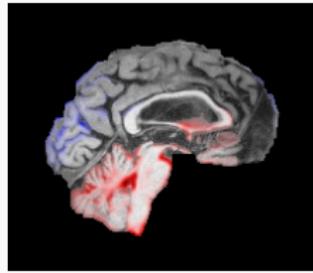
Background: Explainable artificial intelligence



$$R_{0i} = \sum_j \frac{n_{0i} w_i}{\sum_k n_{0k} w_k} R_{1j}$$



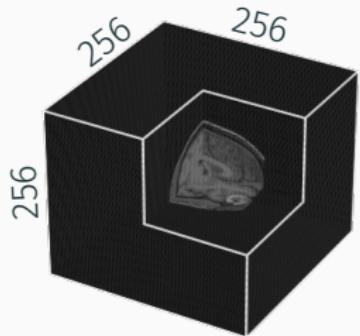
Background: Explainable artificial intelligence



Background: Conceptualization



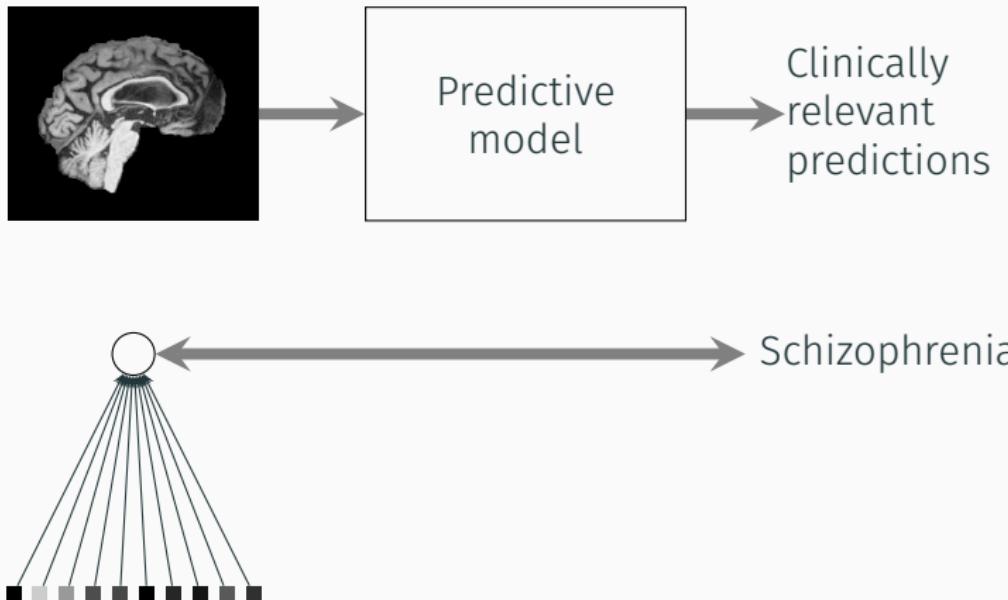
Background: Conceptualization



Background: Conceptualization



Background: Conceptualization



Background: Conceptualization

"Despite considerable heterogeneity at the regional level, deviations were often coupled to common functional circuits and networks"

Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders
Segal, A. et al., *Nature Neuroscience* (2023)



Background: Conceptualization



Background: Conceptualization



Background: Conceptualization



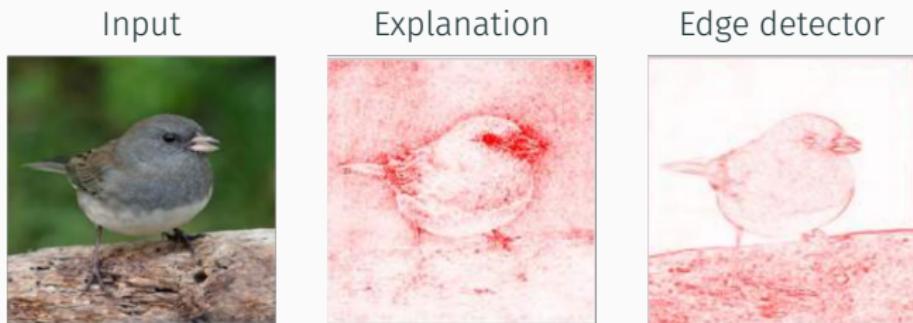
Background: Conceptualization

"We propose explanation space as a powerful embedding to facilitate detection of latent disease subtypes, [...]. Transforming from feature space to explanation space should not be seen in competition to dimensionality reduction, but rather as a complimentary processing step"

Inferring disease subtypes from clusters in explanation space
Schulz, MA. et al., *Scientific Reports* (2020)



Background: Conceptualization



Sanity Checks for Saliency Maps
Adebayo, J. et al., *preprint at arXiv* (2018)



Explainable dementia classification



UNIVERSITY
OF OSLO

Explainable dementia classification: Dataset



FEMALE

MALE

55

60

65

70

75

80

85

90

95

n=1708

?? Controls
?? Patients



Explainable dementia classification: Dataset



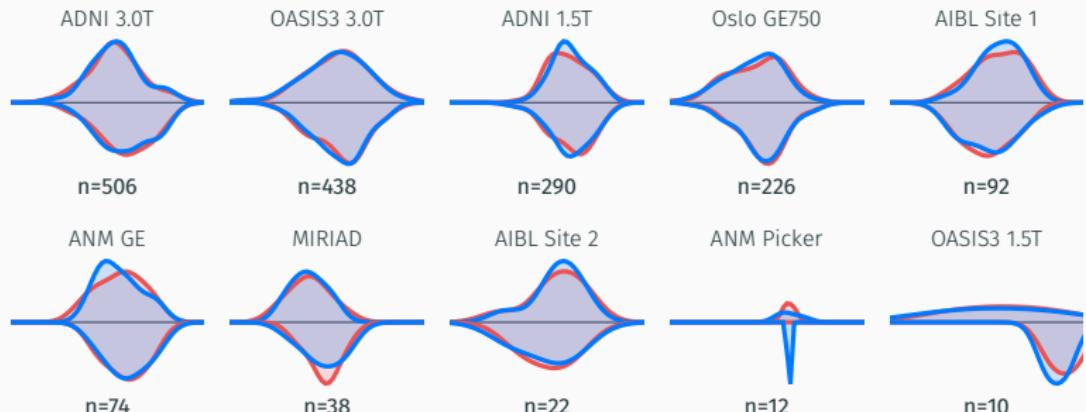
FEMALE

MALE

55 60 65 70 75 80 85 90 95

n=1708

?? Controls
?? Patients



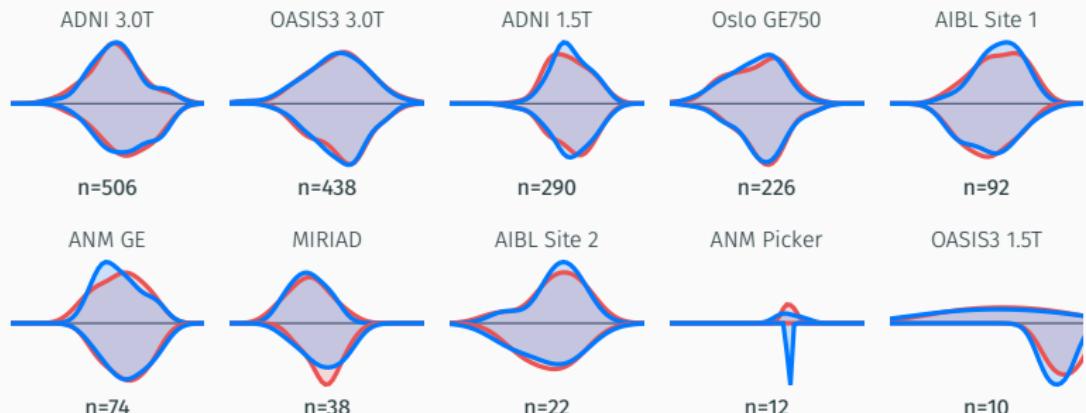
Explainable dementia classification: Dataset



ADNI:
Probable
Alzheimer's disease

OASIS3:
Dementia

Oslo:
Alzheimer's disease,
Vascular dementia,
Unspecified dementia



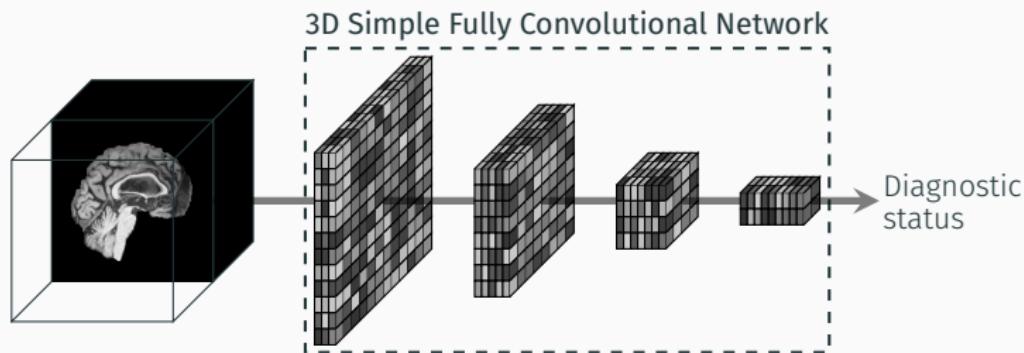
Explainable dementia classification: Modelling



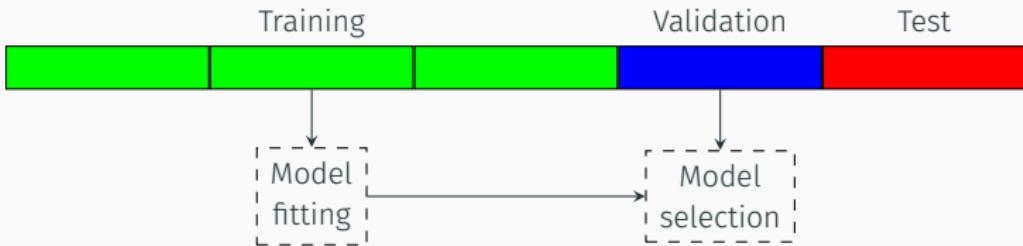
Explainable dementia classification: Modelling



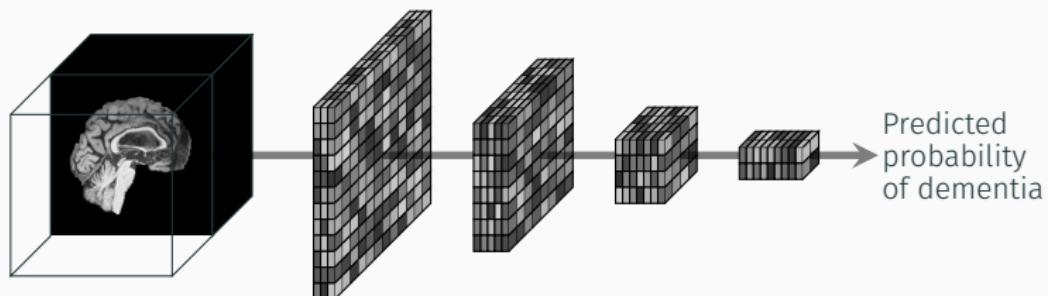
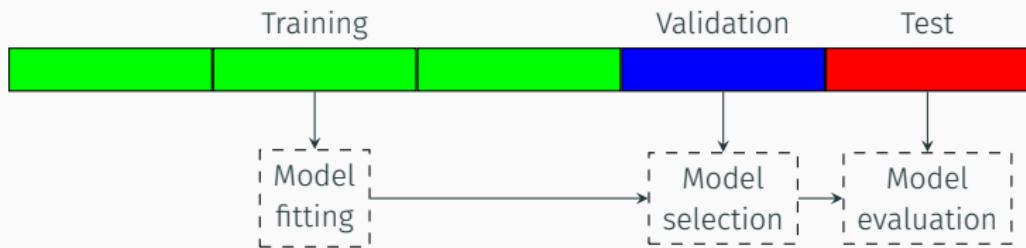
Explainable dementia classification: Modelling



Explainable dementia classification: Modelling



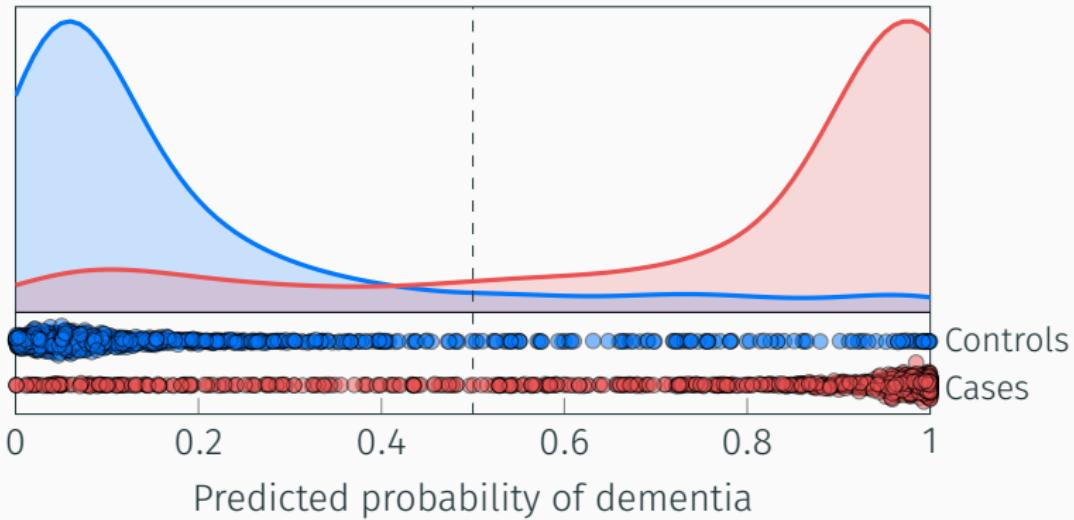
Explainable dementia classification: Modelling



Explainable dementia classification: Modelling



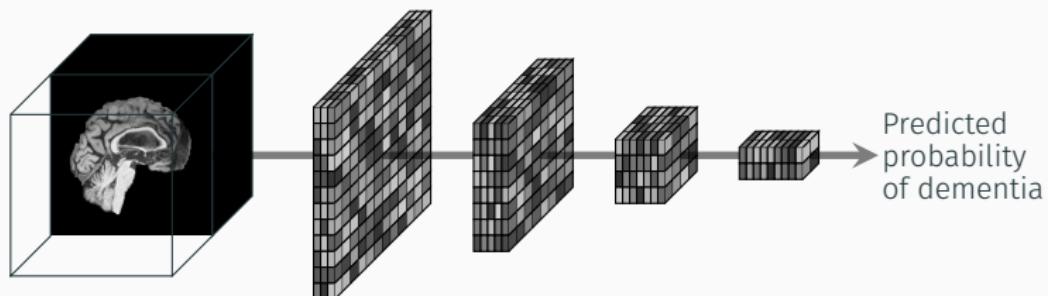
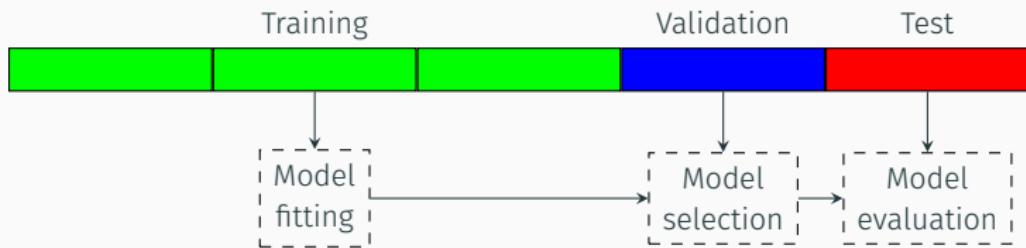
Explainable dementia classification: Modelling



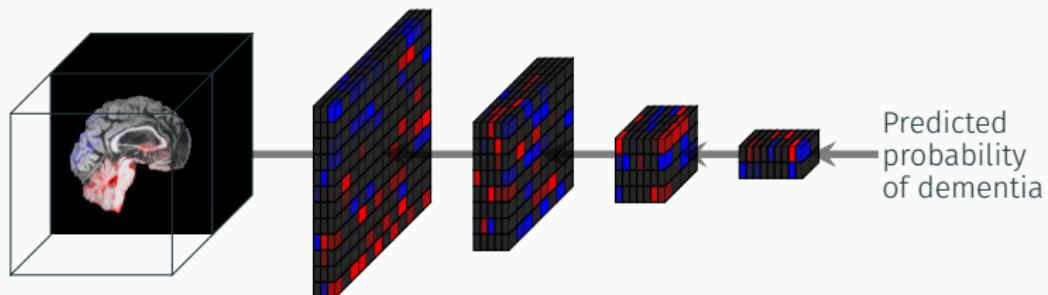
AUC=0.91, balanced accuracy=85%



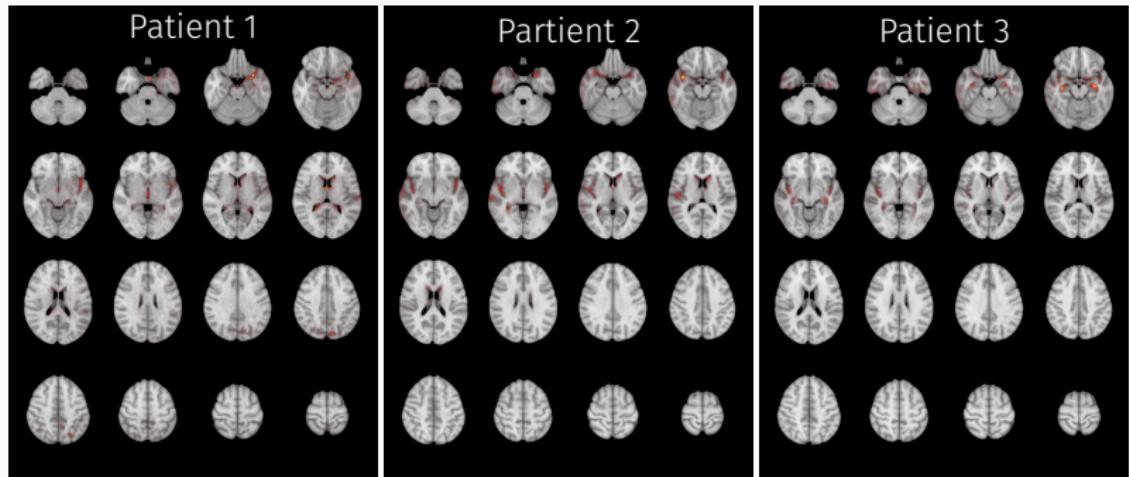
Explainable dementia classification: Modelling



Explainable dementia classification: Modelling



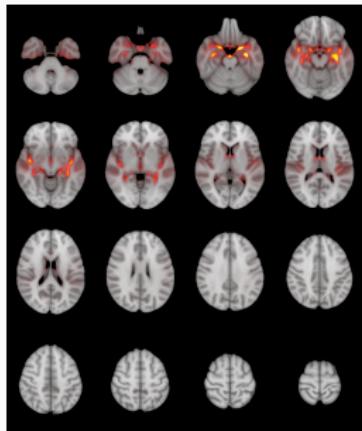
Explainable dementia classification: Validation



Explainable dementia classification: Validation



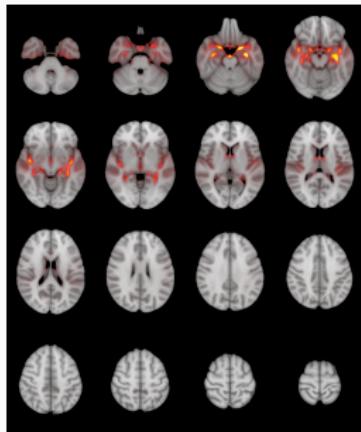
LRP



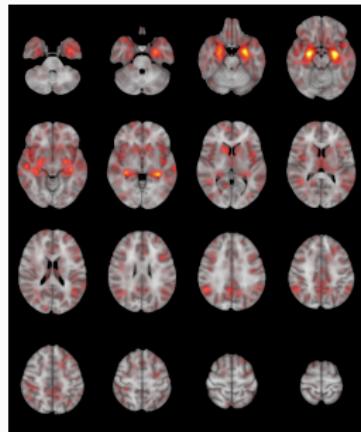
Explainable dementia classification: Validation



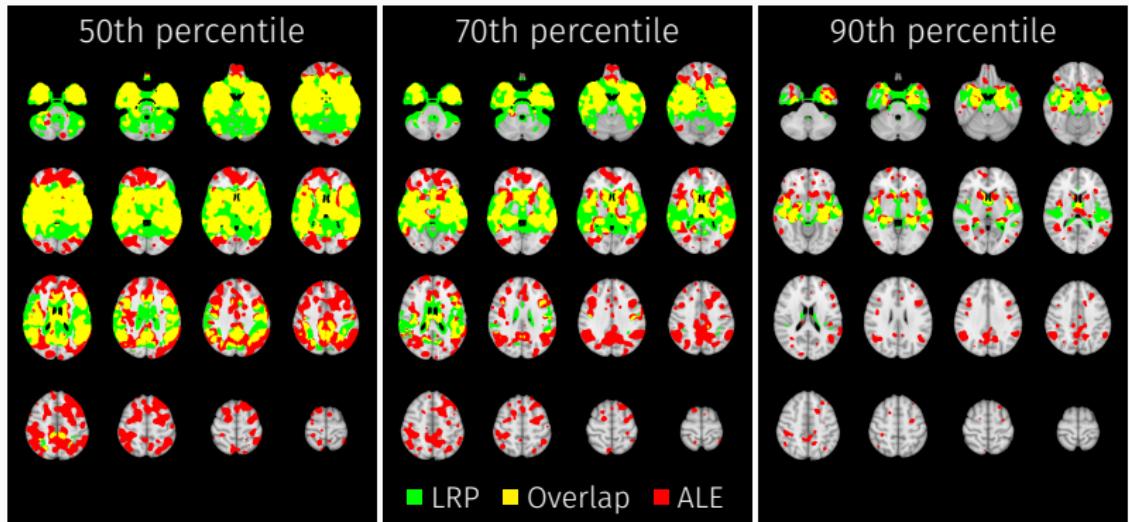
LRP



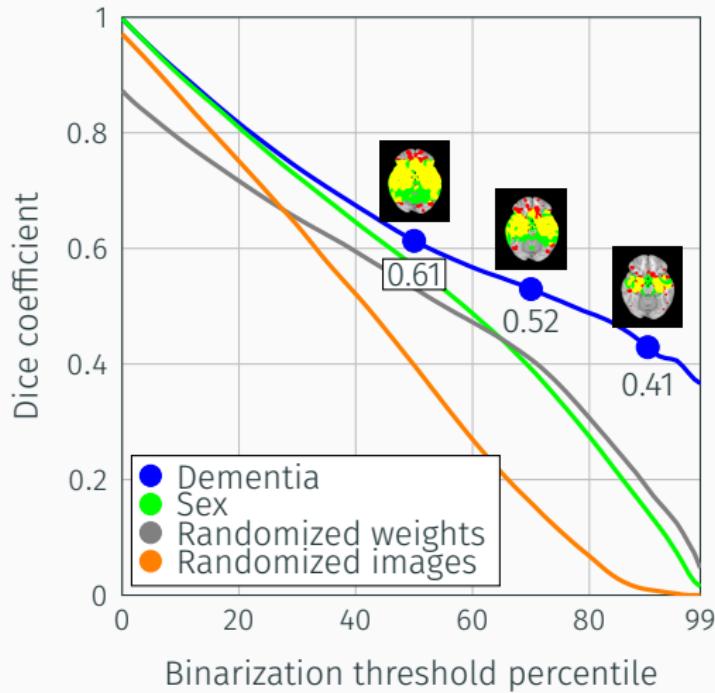
GingerALE



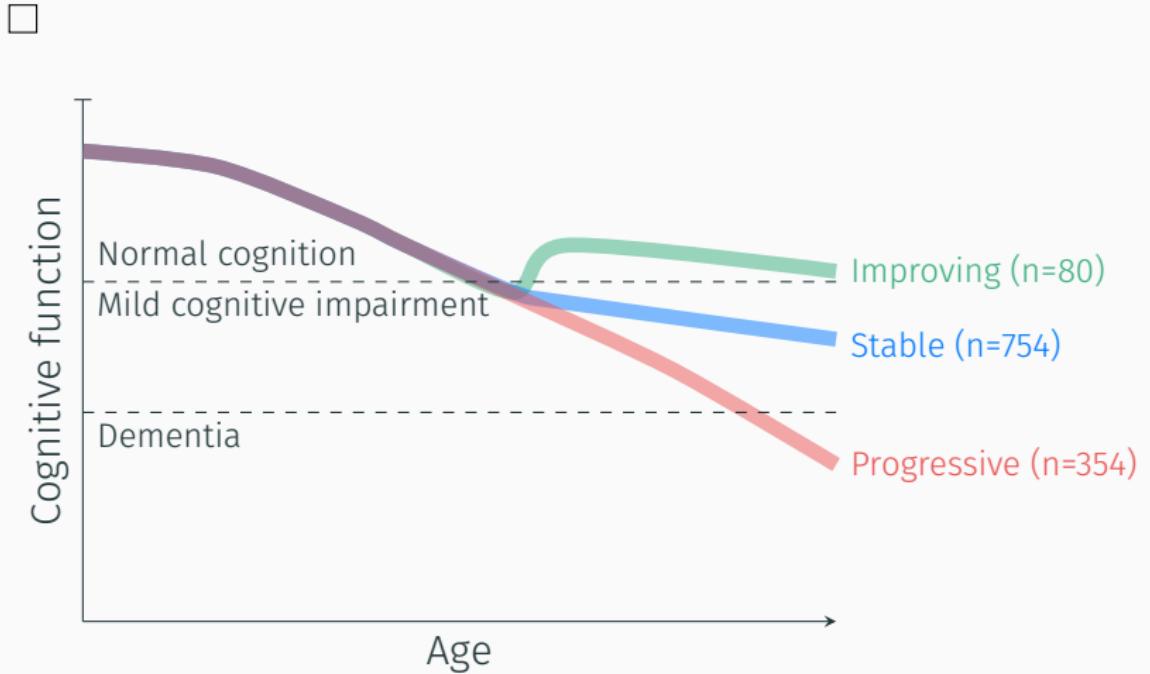
Explainable dementia classification: Validation



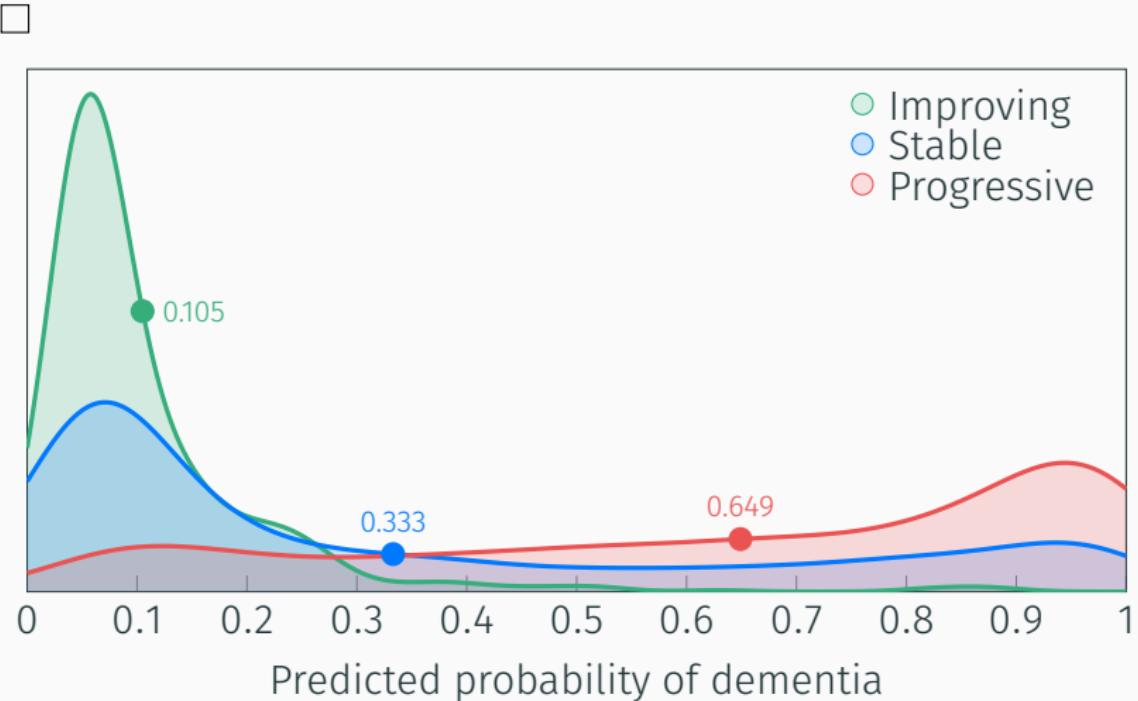
Explainable dementia classification: Validation



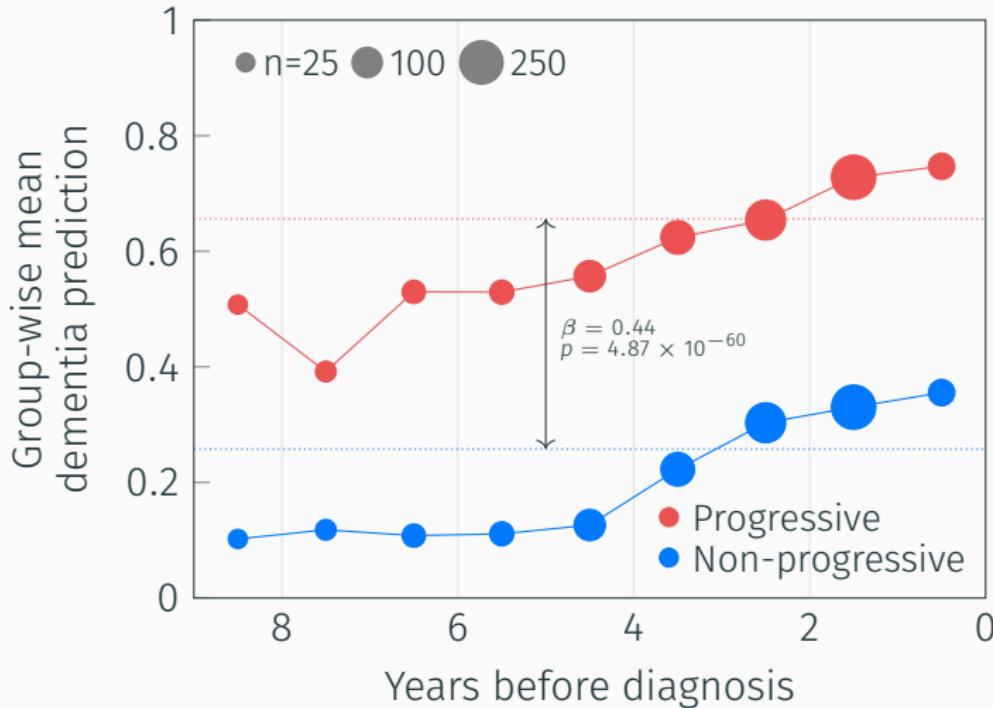
Explainable dementia classification: Application



Explainable dementia classification: Application



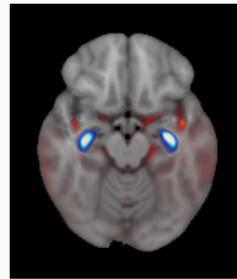
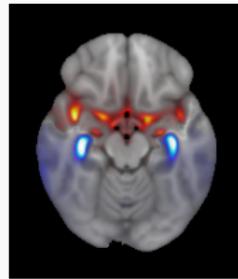
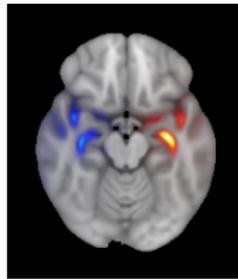
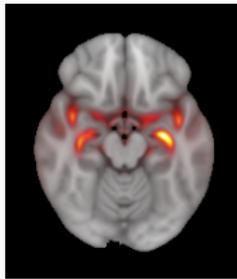
Explainable dementia classification: Application



Explainable dementia classification: Application



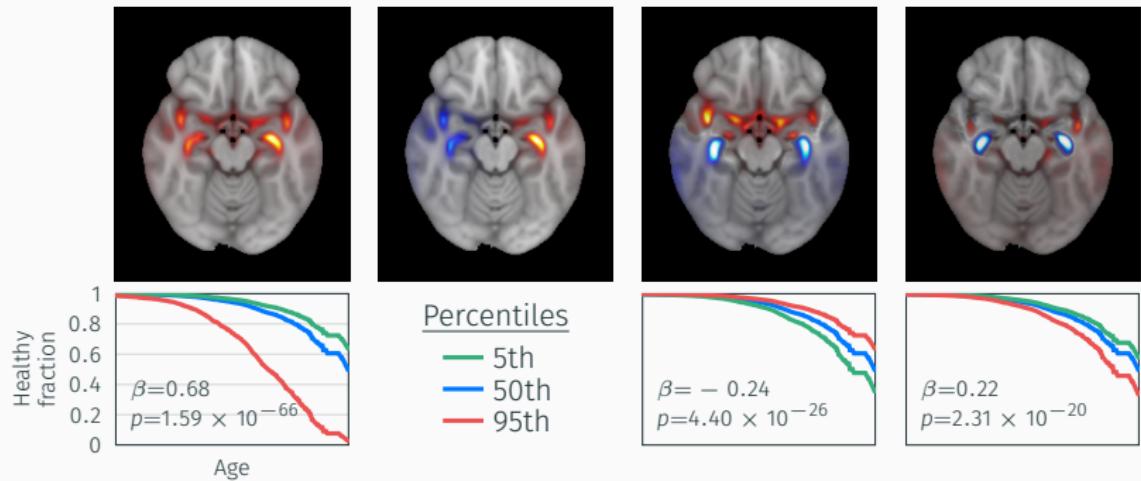
Component 0 Component 1 Component 2 Component 3



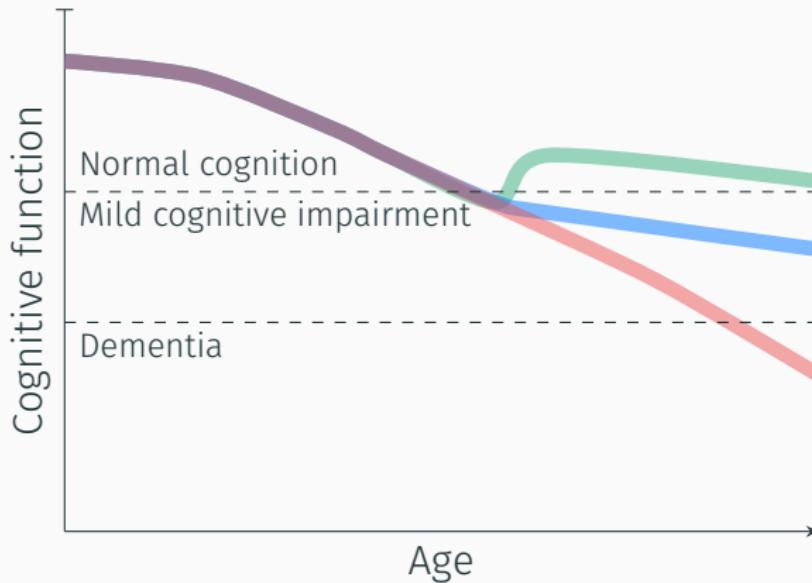
Explainable dementia classification: Application



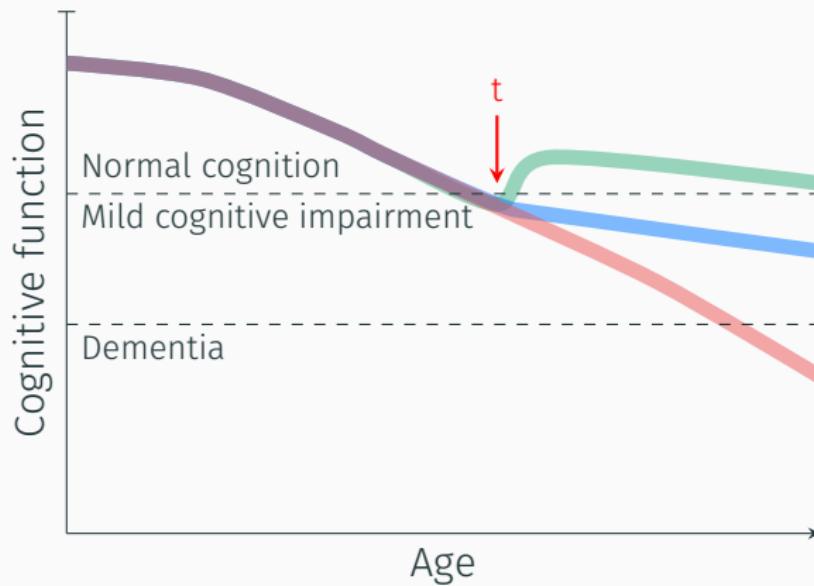
Component 0 Component 1 Component 2 Component 3



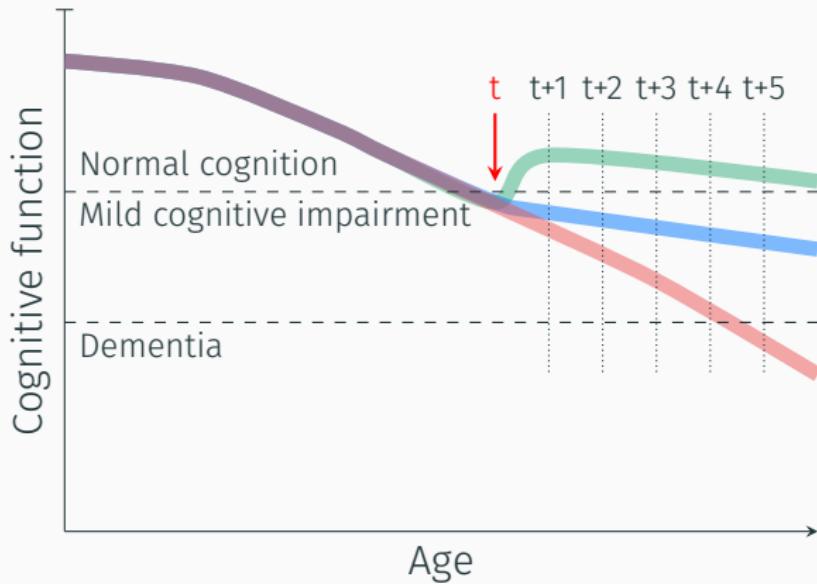
Explainable dementia classification: Application



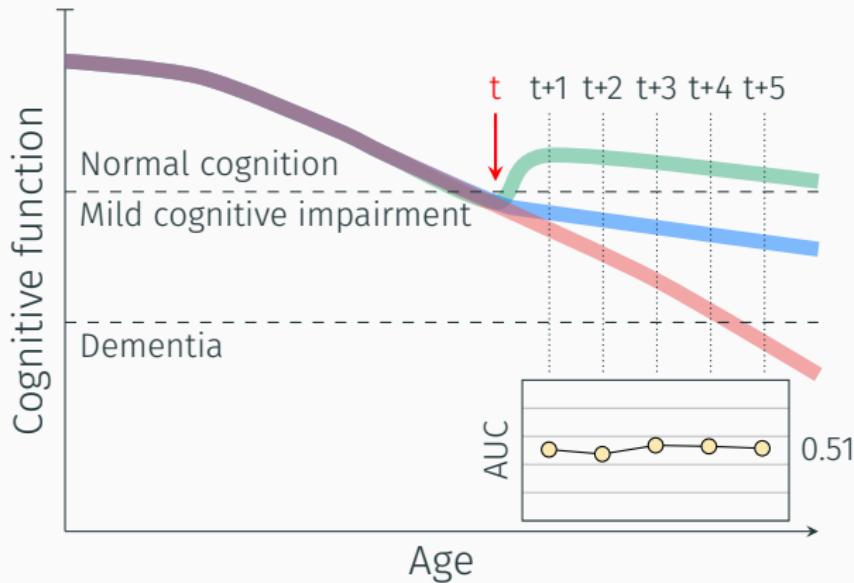
Explainable dementia classification: Application



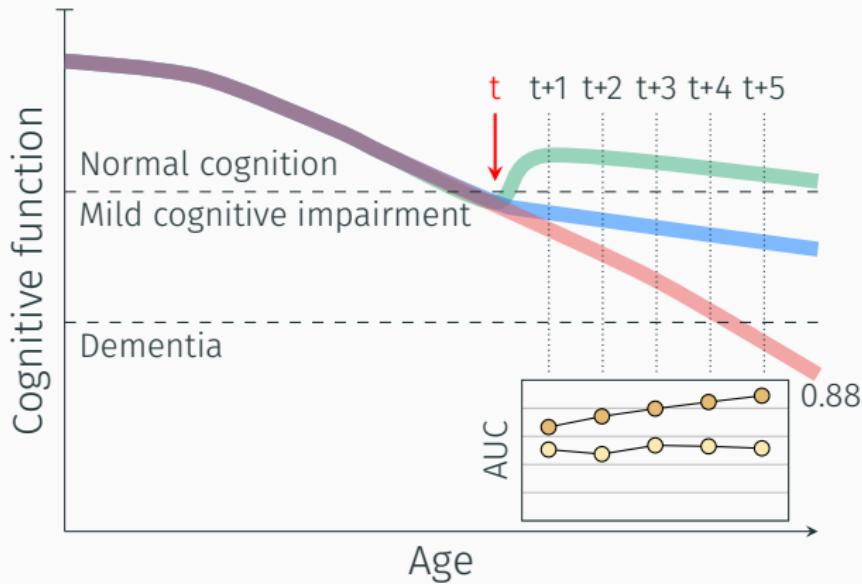
Explainable dementia classification: Application



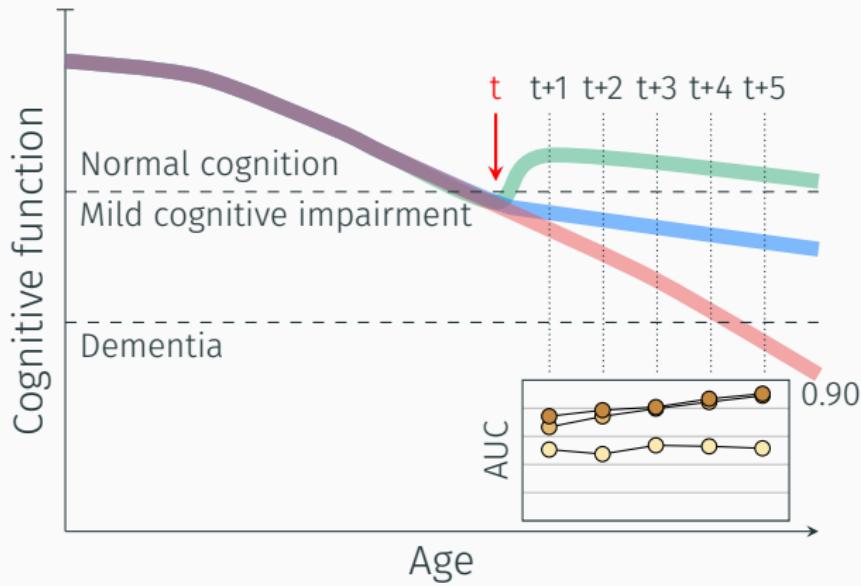
Explainable dementia classification: Application



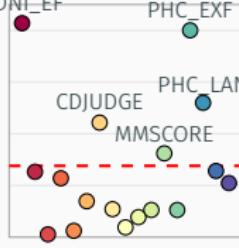
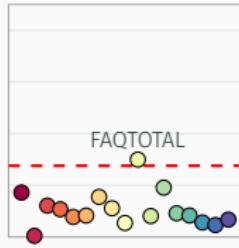
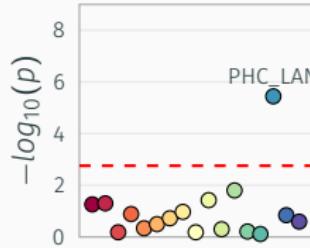
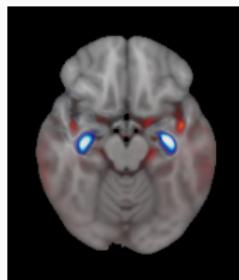
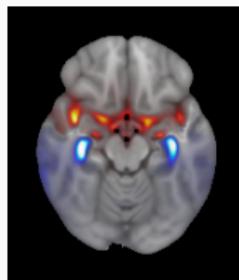
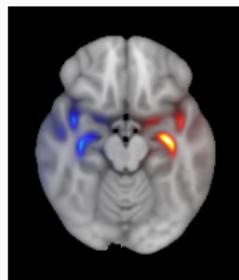
Explainable dementia classification: Application



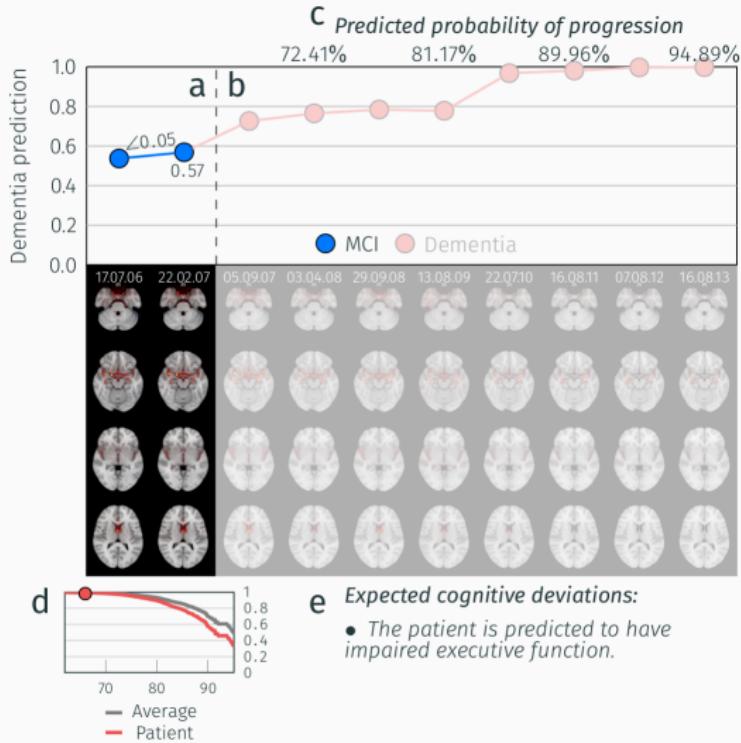
Explainable dementia classification: Application



Explainable dementia classification: Application



Explainable dementia classification: Application

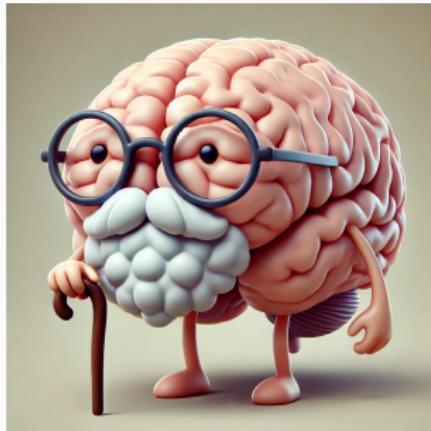


Explainable brain age



UNIVERSITY
OF OSLO

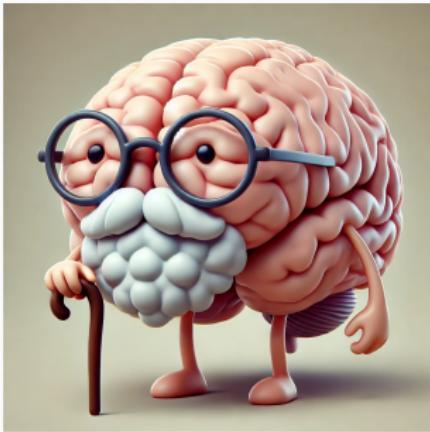
Explainable brain age: Motivation



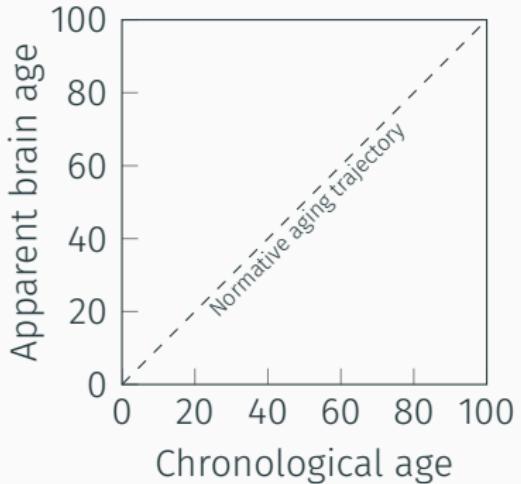
Generated by Dall-E 3



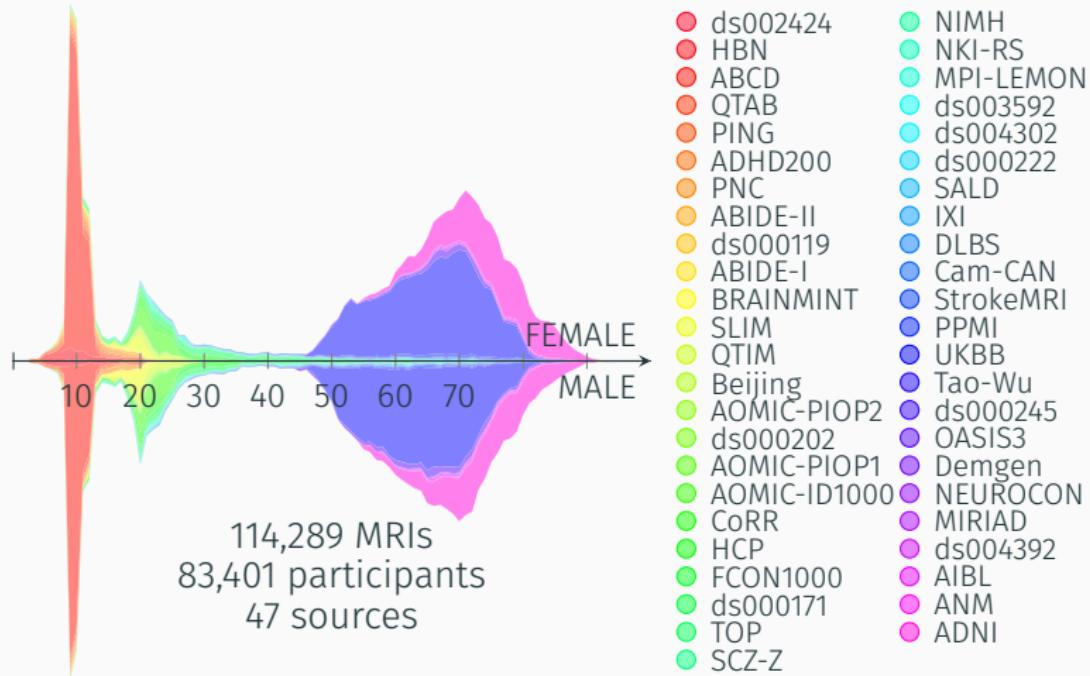
Explainable brain age: Motivation



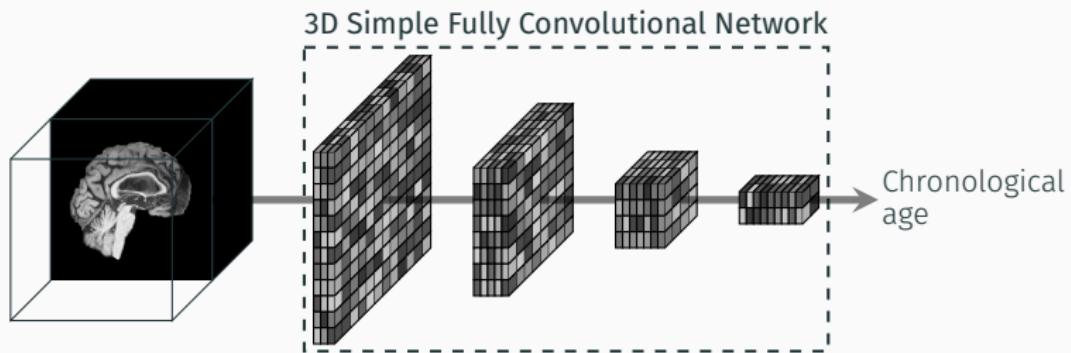
Generated by Dall-E 3



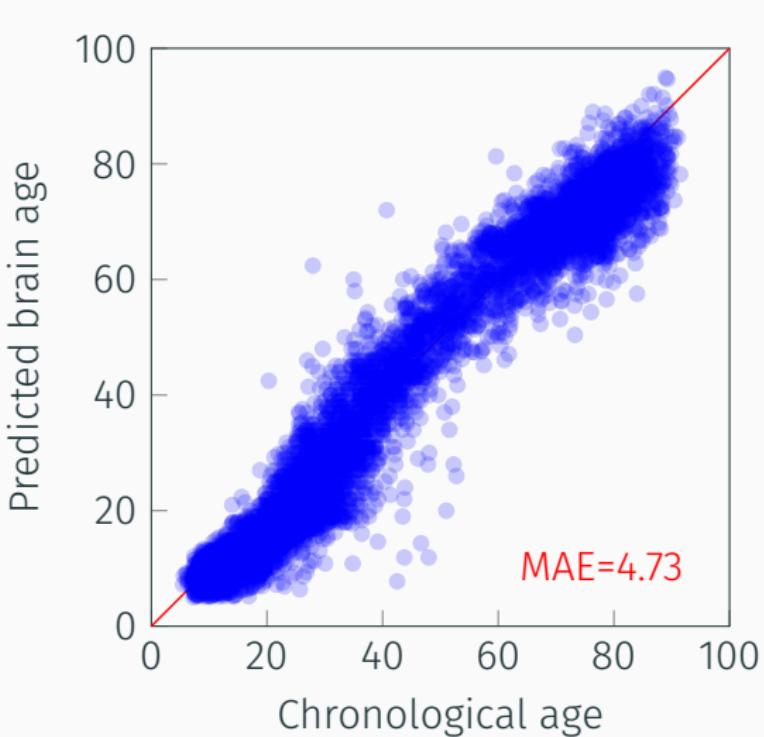
Explainable brain age: Motivation



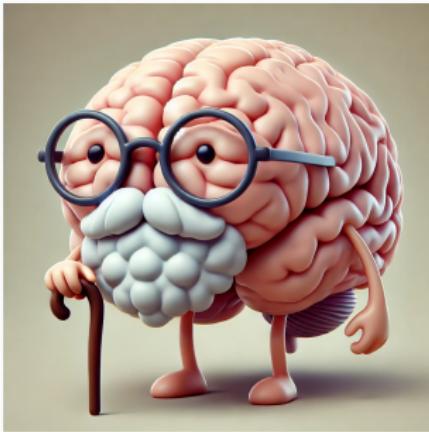
Explainable brain age: Motivation



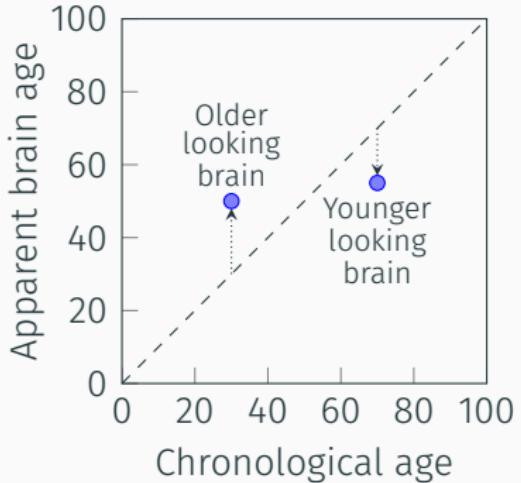
Explainable brain age: Motivation



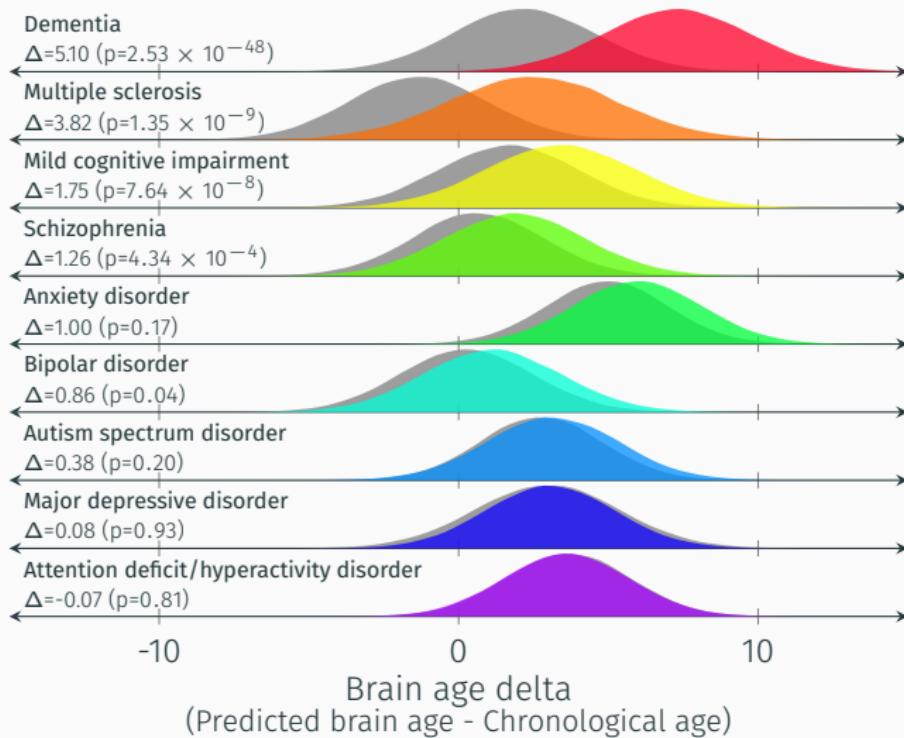
Explainable brain age: Motivation



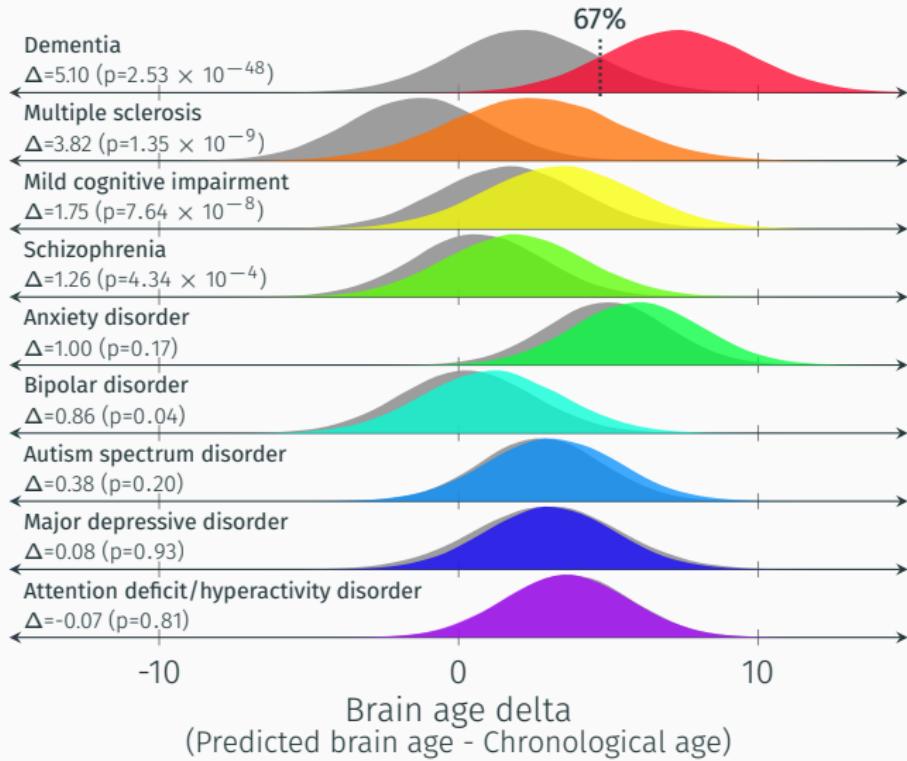
Generated by Dall-E 3



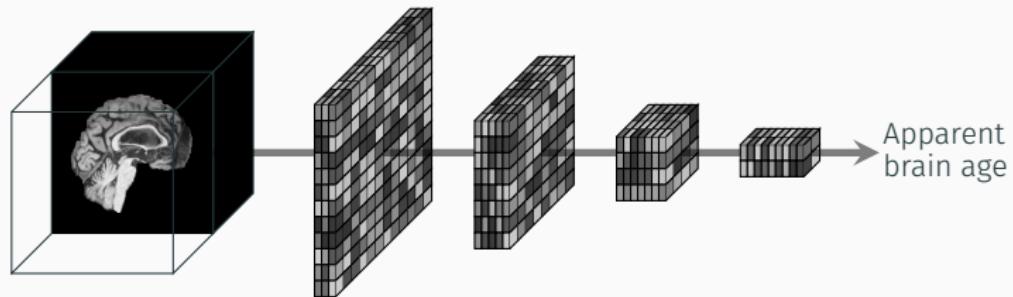
Explainable brain age: Motivation



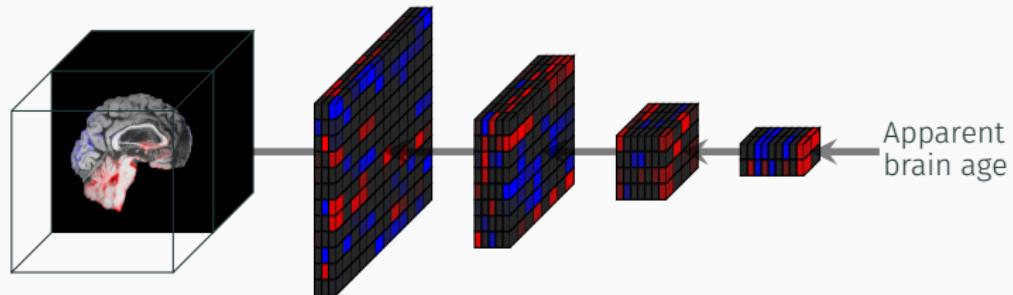
Explainable brain age: Motivation



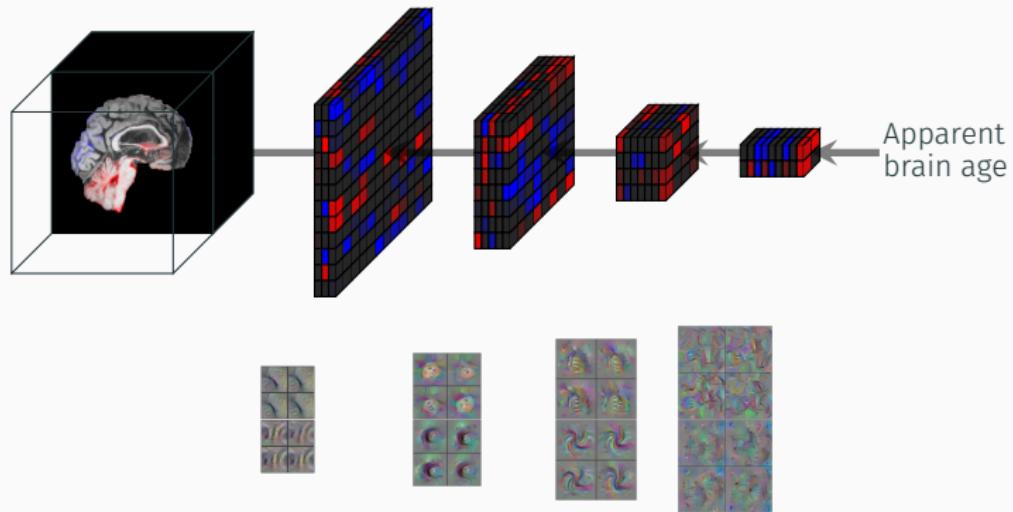
Explainable brain age: Methods



Explainable brain age: Methods



Explainable brain age: Methods



Understanding Neural Networks Through Deep Visualization
Yosinski, J. et al., preprint at arXiv (2015)



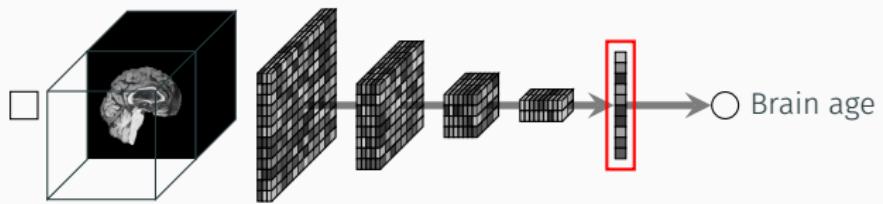
Explainable brain age: Methods



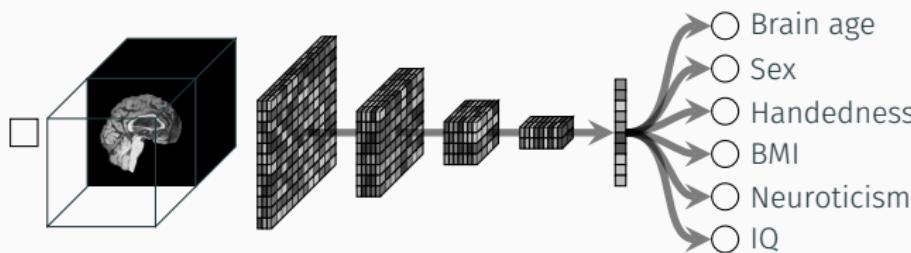
Feature correlation



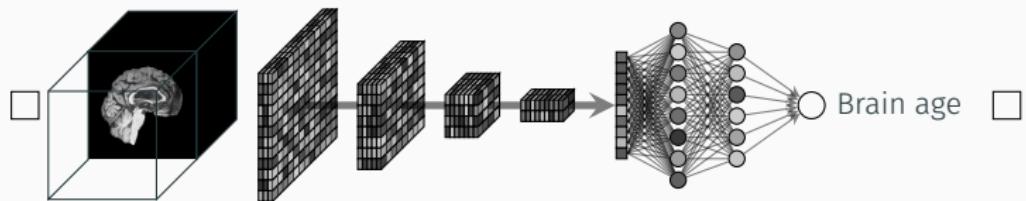
Explainable brain age: Methods



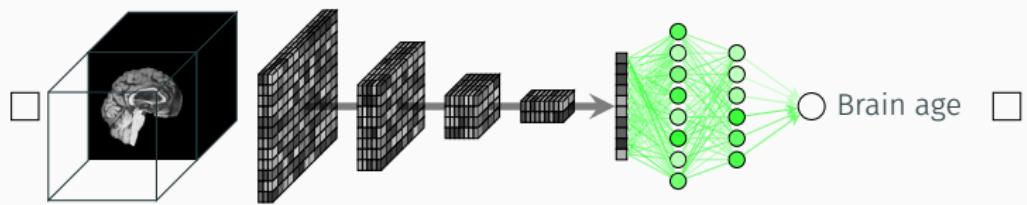
Explainable brain age: Methods



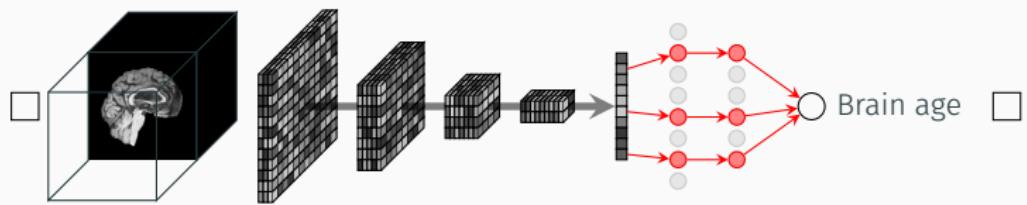
Explainable brain age: Methods



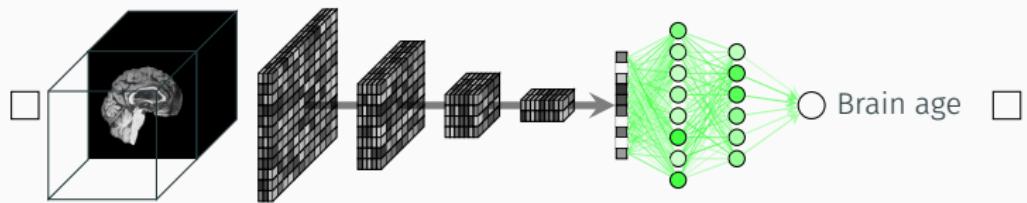
Explainable brain age: Methods



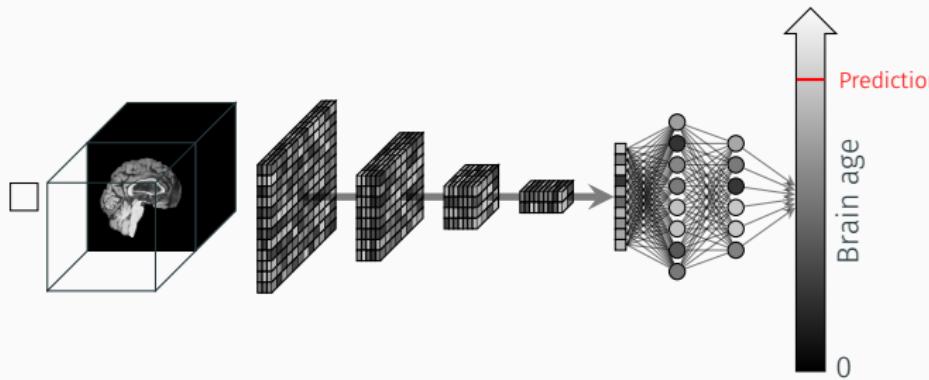
Explainable brain age: Methods



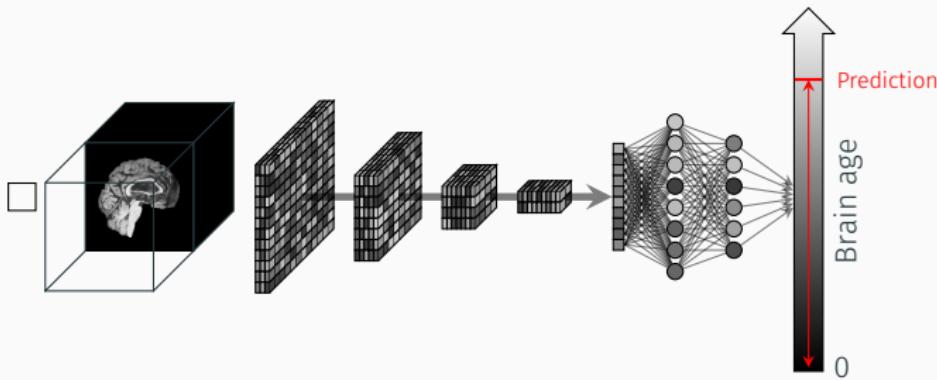
Explainable brain age: Methods



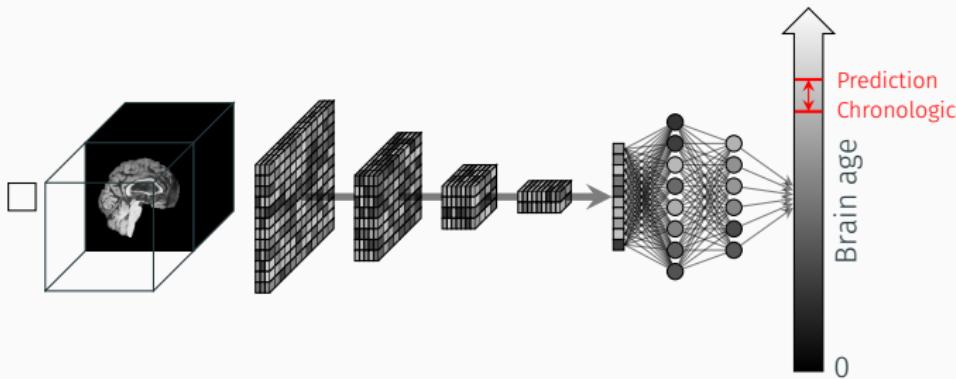
Explainable brain age: Methods



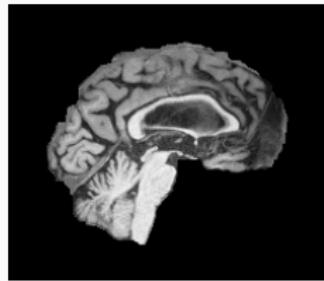
Explainable brain age: Methods



Explainable brain age: Methods



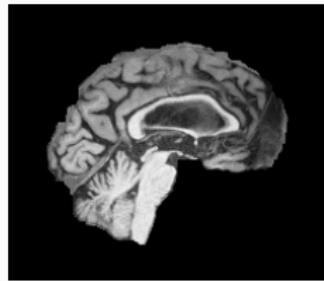
Explainable brain age: Methods



Age
62



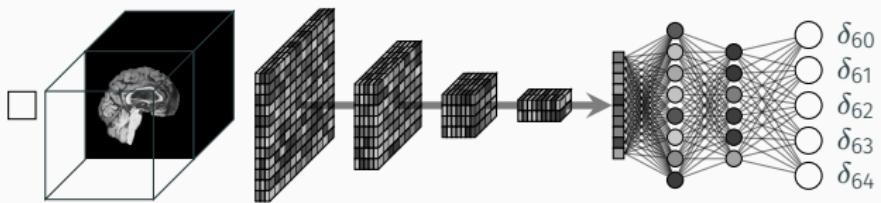
Explainable brain age: Methods



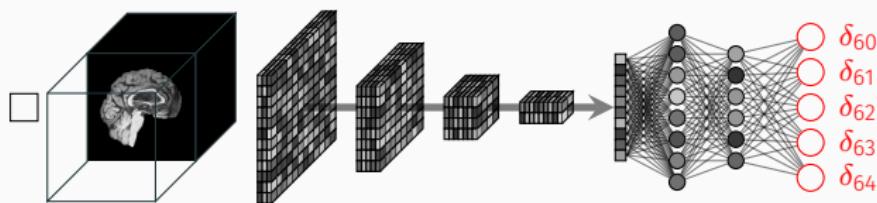
δ_{60}	δ_{61}	δ_{62}	δ_{63}	δ_{64}
2	1	0	-1	-2



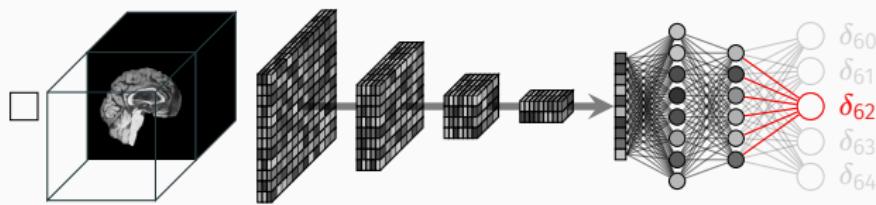
Explainable brain age: Methods



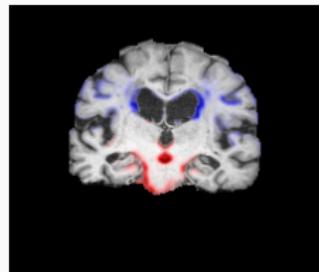
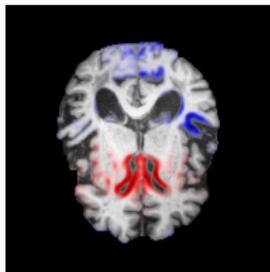
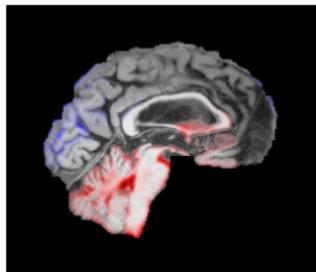
Explainable brain age: Methods



Explainable brain age: Methods



Explainable brain age: Results



Younger
appearing

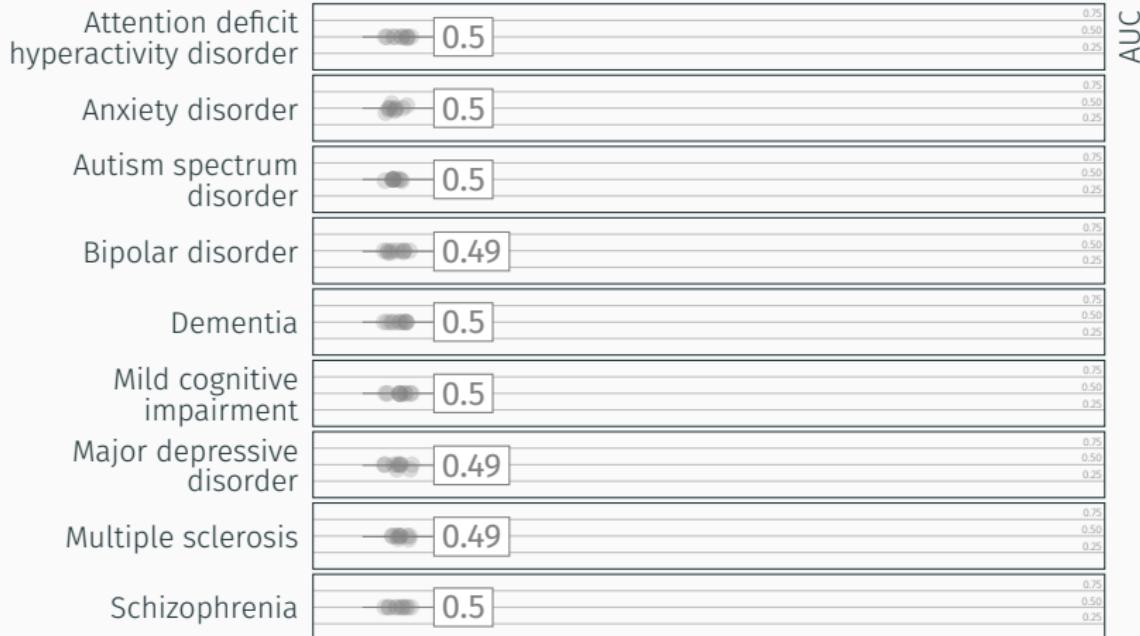
Older
appearing



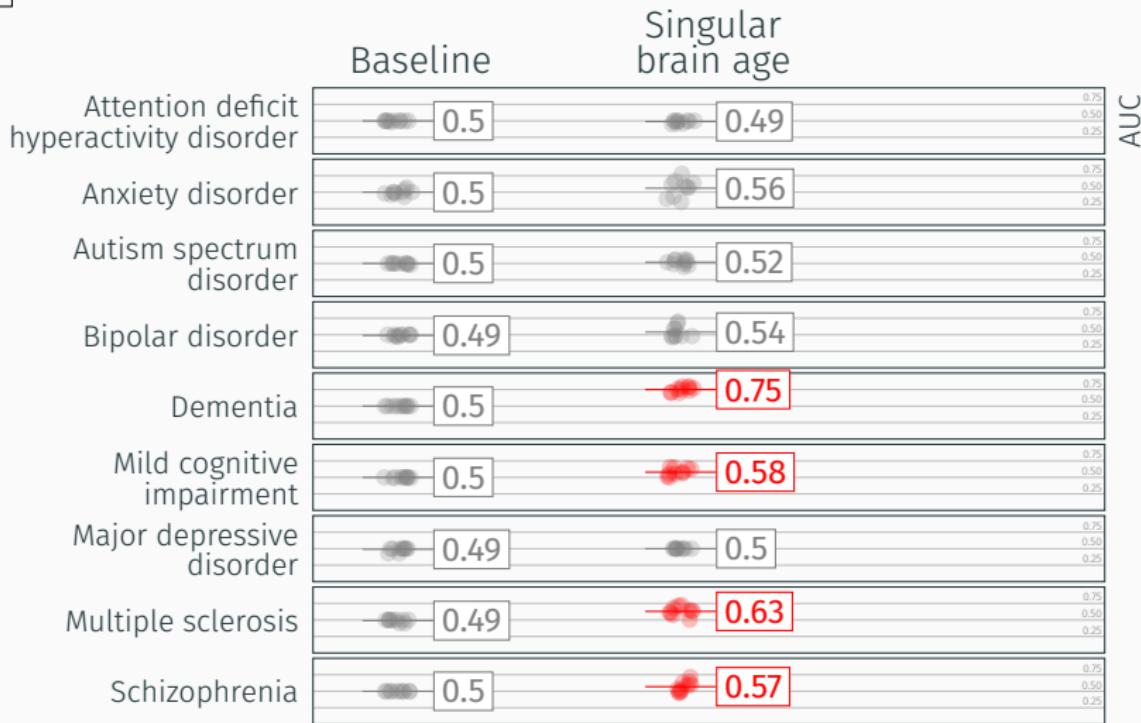
Explainable brain age: Results



Baseline



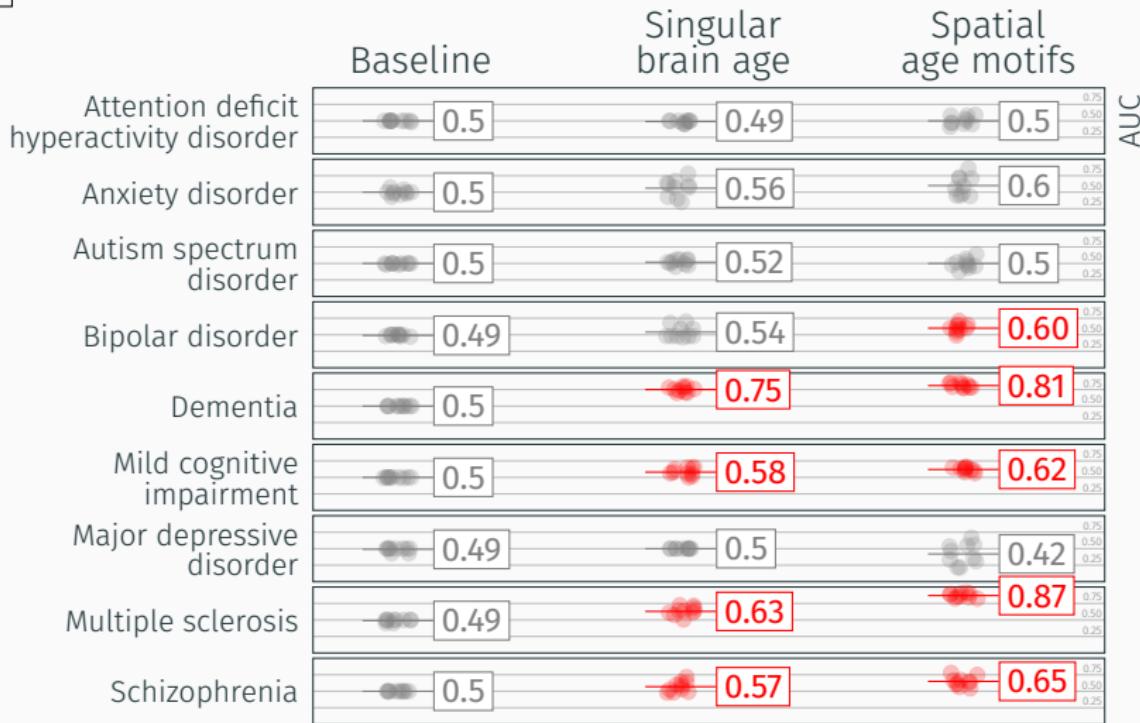
Explainable brain age: Results



AUC



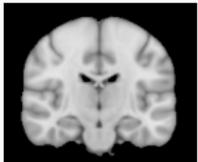
Explainable brain age: Results



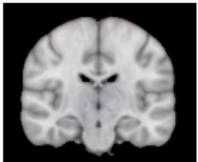
AUC



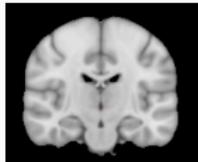
Explainable brain age: Results



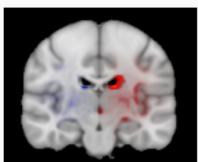
Attention deficit
hyperactivity disorder



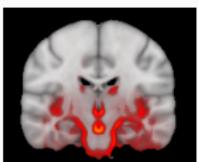
Anxiety
disorder



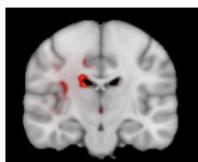
Autism spectrum
disorder



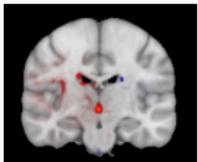
Bipolar
disorder



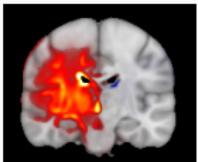
Dementia



Schizophrenia



Major depressive
disorder



Multiple
sclerosis



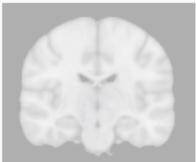
Mild cognitive
impairment



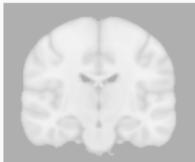
Explainable brain age: Results



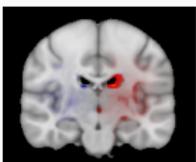
Attention deficit
hyperactivity disorder



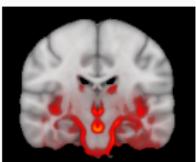
Anxiety
disorder



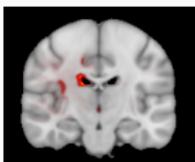
Autism spectrum
disorder



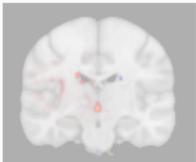
Bipolar
disorder



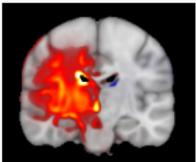
Dementia



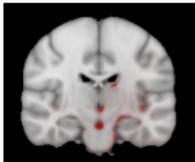
Schizophrenia



Major depressive
disorder



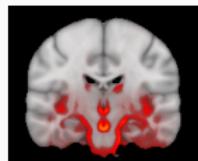
Multiple
sclerosis



Mild cognitive
impairment



Explainable brain age: Results

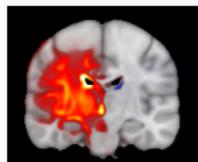


Dementia

- Temporal pole
- Left amygdala
- Parahippocampal gyrus



Explainable brain age: Results

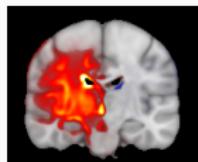


Multiple
sclerosis

- Right pallidum
- Right thalamus
- Right putamen



Explainable brain age: Results

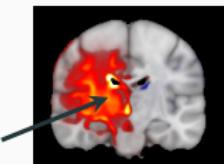


Multiple
sclerosis

- Right pallidum
- Right thalamus
- Right putamen



Explainable brain age: Results

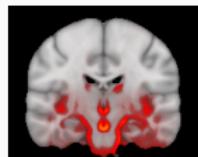


Multiple
sclerosis

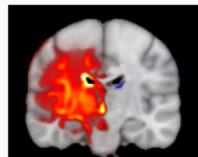
- Right pallidum
- Right thalamus
- Right putamen



Explainable brain age: Results



Dementia



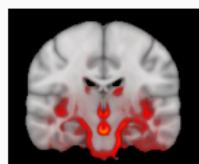
Multiple
sclerosis



Explainable brain age: Results

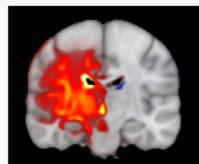


AUC



Dementia

Singular brain age	Spatial age motifs
0.75	0.81



Multiple
sclerosis

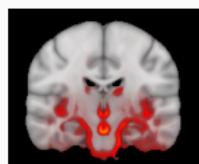
0.63	0.87
------	------



Explainable brain age: Results

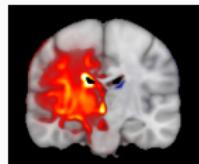


AUC



Singular brain age	Spatial age motifs
0.75	0.06

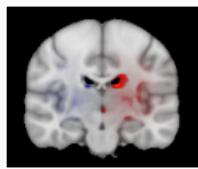
0.75 → 0.81



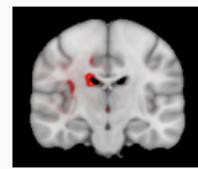
0.63 → 0.87
0.24



Explainable brain age: Results



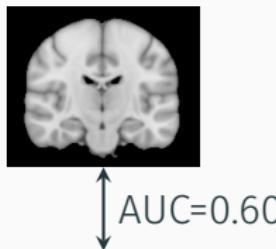
Bipolar
disorder



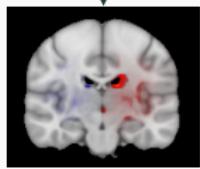
Schizophrenia



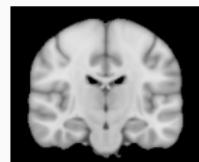
Explainable brain age: Results



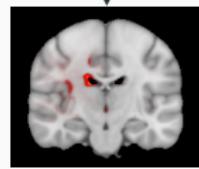
AUC=0.60



Bipolar
disorder



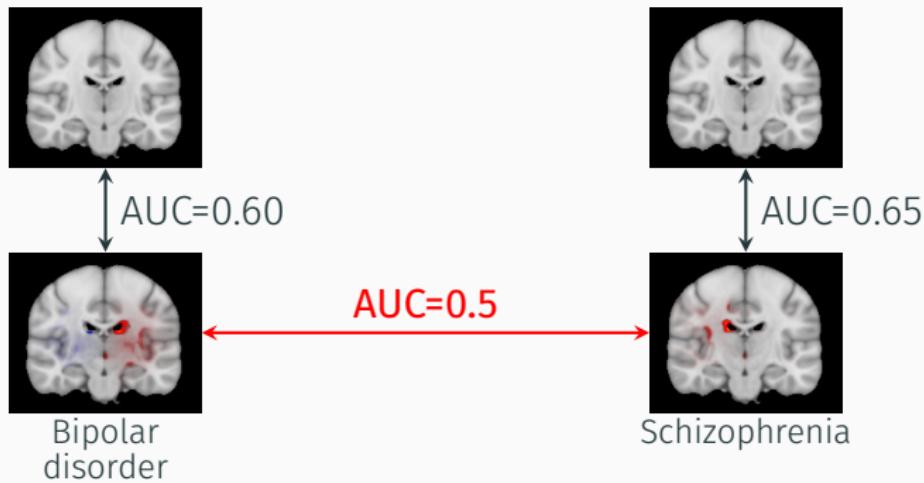
AUC=0.65



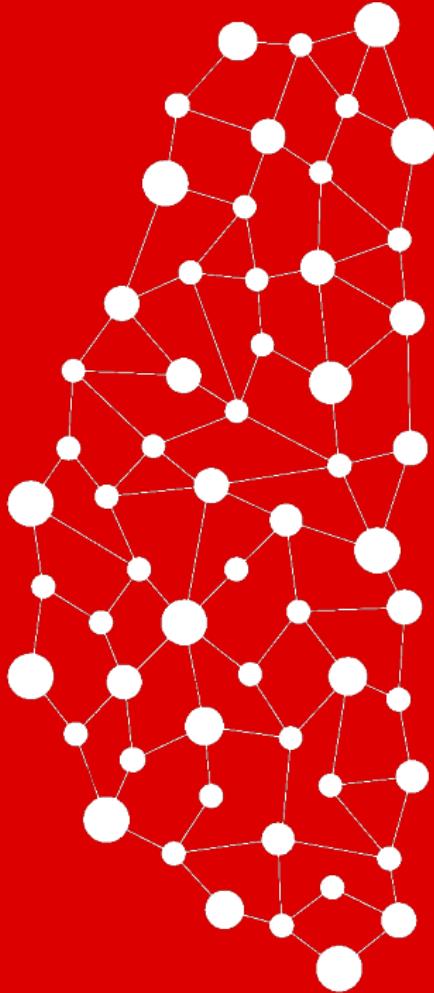
Schizophrenia



Explainable brain age: Results



Thank you for your attention!
estenhl@ui.no



UNIVERSITY
OF OSLO