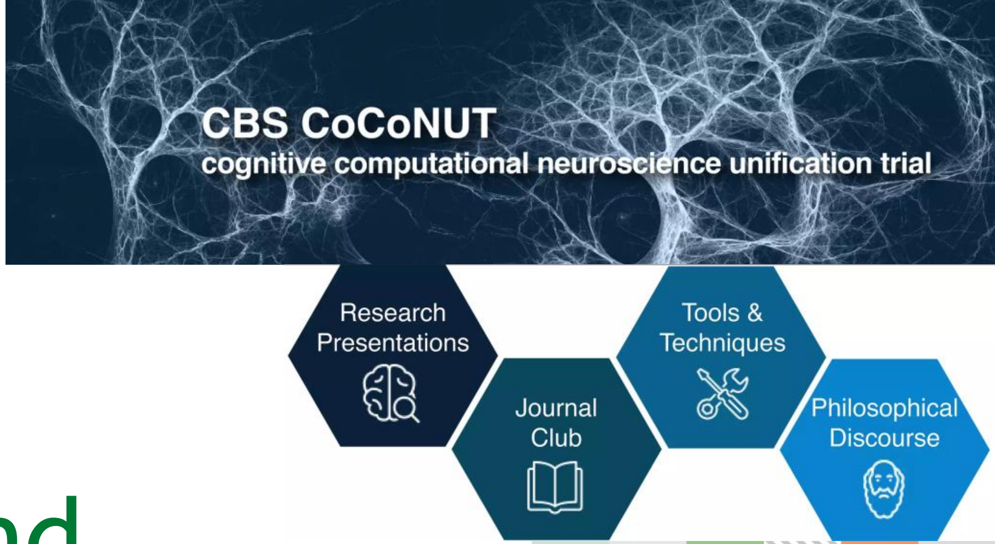
Medical University of Graz

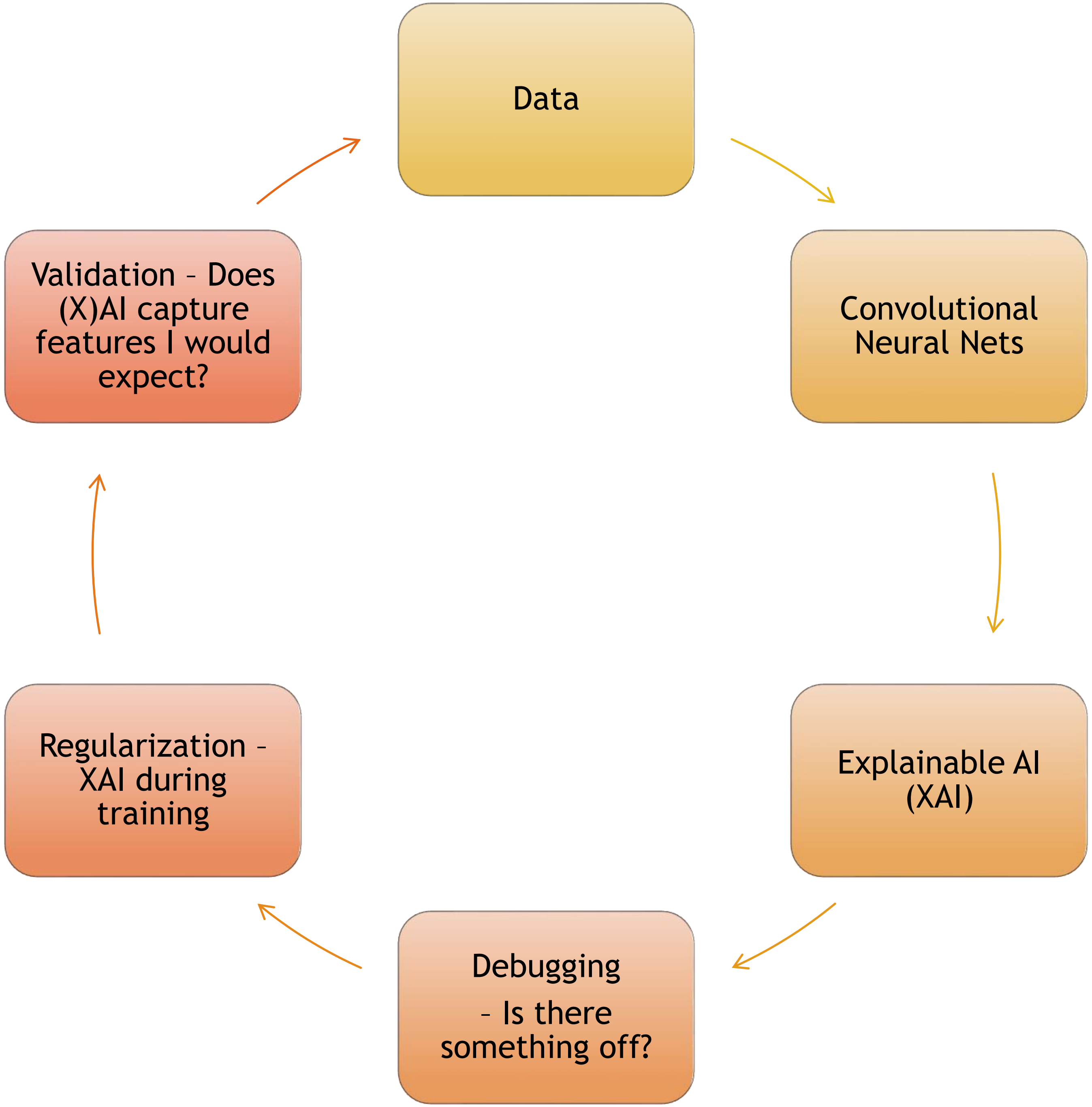


Debugging, regularization and **validation** - explainable deep learning as a means to enhance brain disease classification models using MRI data

Christian Tinauer[1]
[1]Department of Neurology, Medical University of Graz, Graz, Austria

CBS-CoCoNUT Talk, 25th October 2024

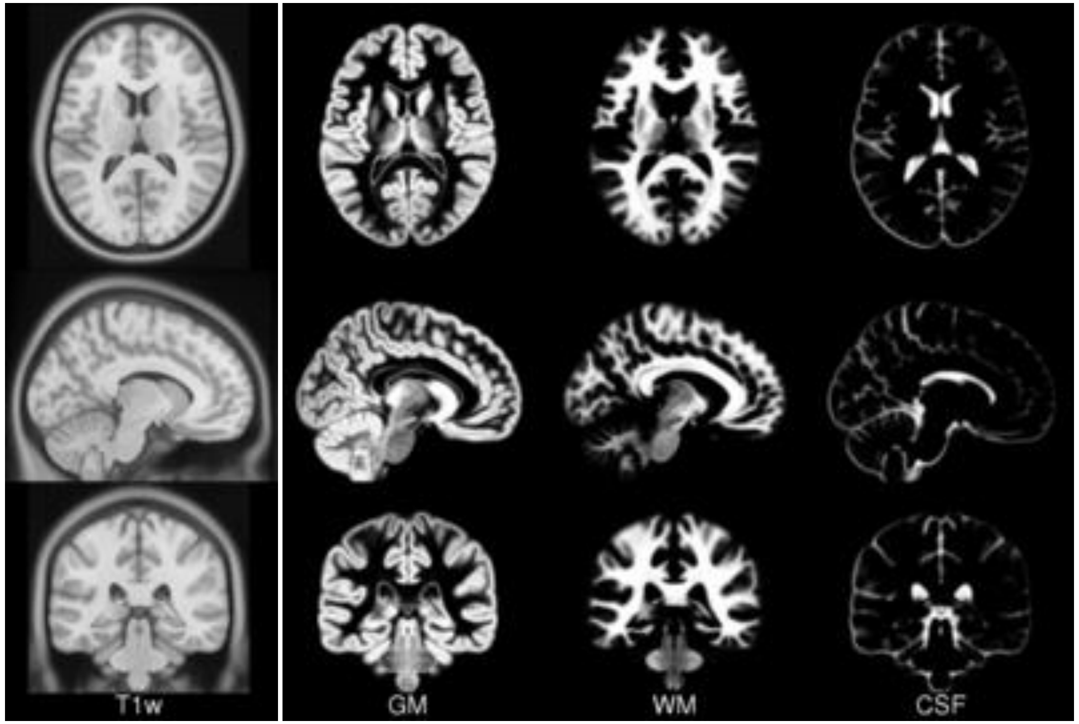
Outline


Data
Convolutional Neural Nets
Explainable AI (XAI)
Debugging – Is there something off?
Regularization – XAI during training
Validation – Does (X)AI capture features I would expect?

Introduction – Disease and MRI data

- Alzheimer’s disease (AD) is the most common form of dementia
- ~55 million living with AD worldwide

Structural MRI



R2* maps (Quantitative MRI)

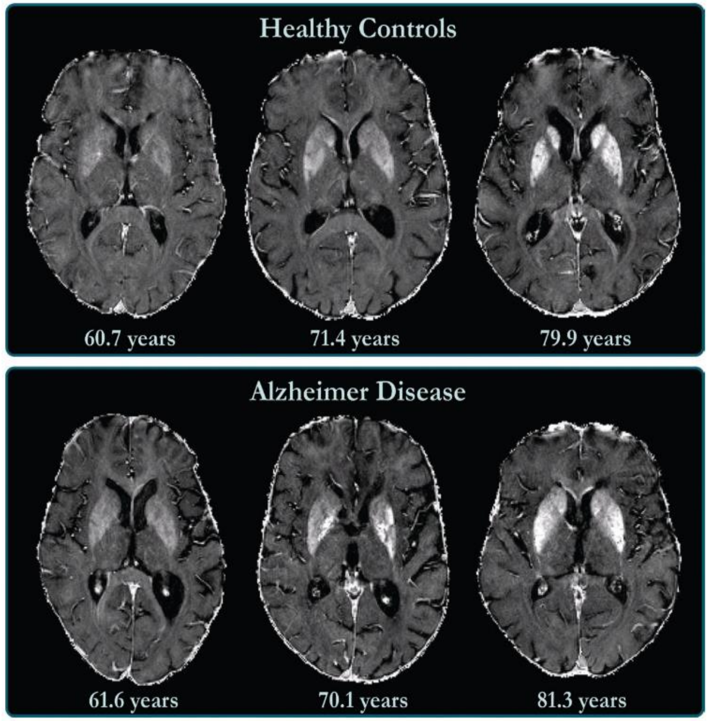
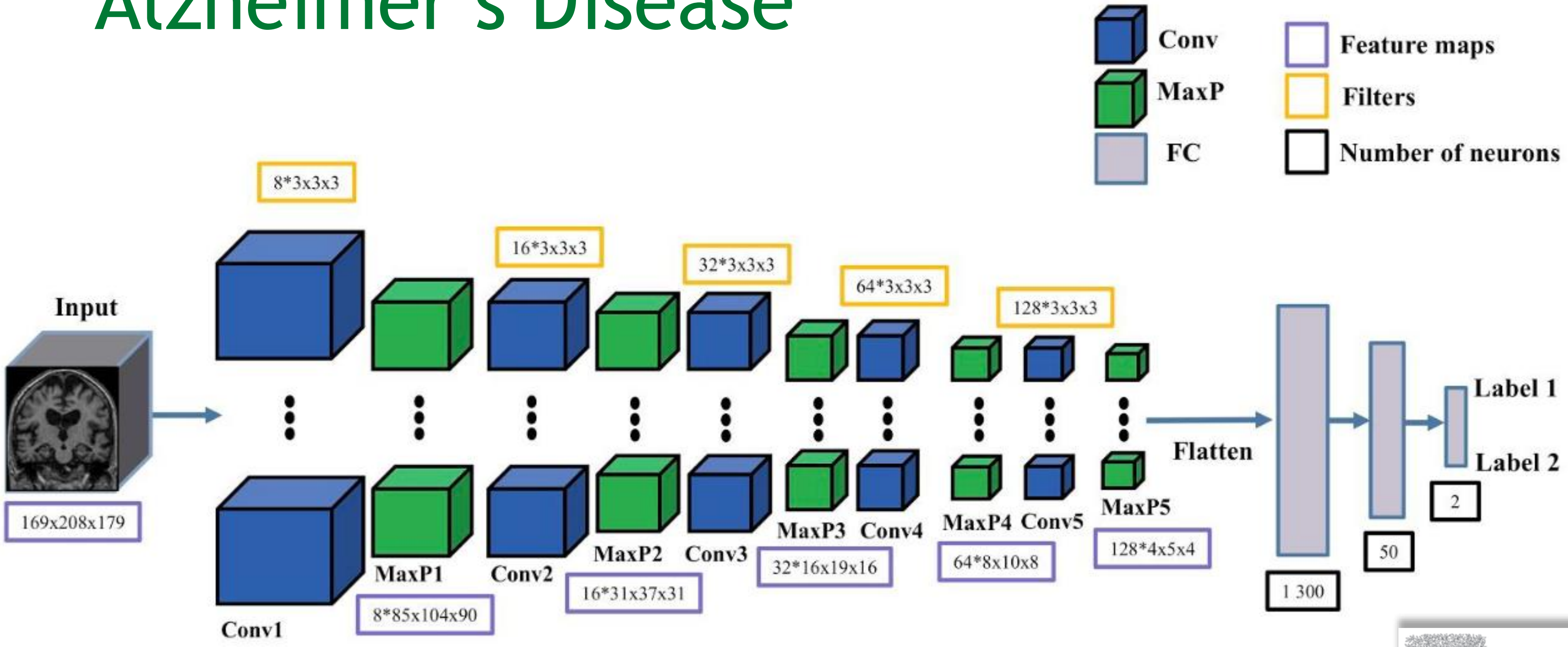


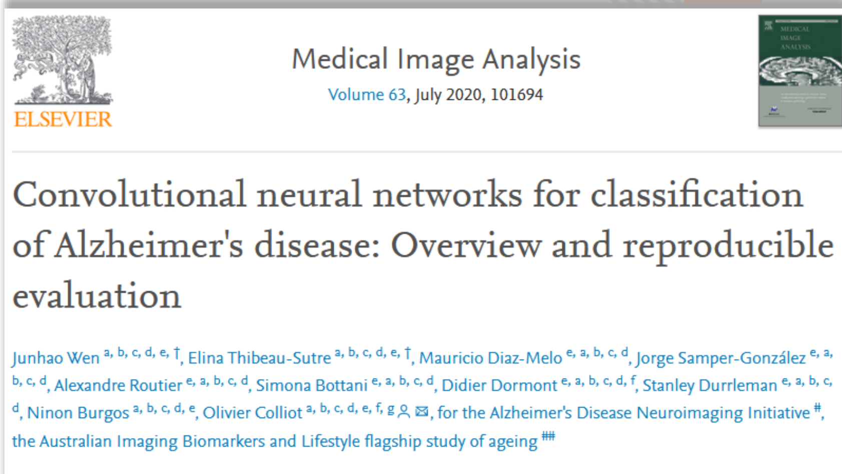
Figure 2: R2* maps of healthy control participants and participants with Alzheimer disease. R2* maps are windowed between 10 and 50 sec^{-1}.

[1] ICBM 2009a Nonlinear Asymmetric 1×1x1mm template (modified)
[2] Damulina et al, Radiology, 2021

Introduction – Deep Learning in Alzheimer's Disease



Medical University of Graz



ELSEVIER

Medical Image Analysis

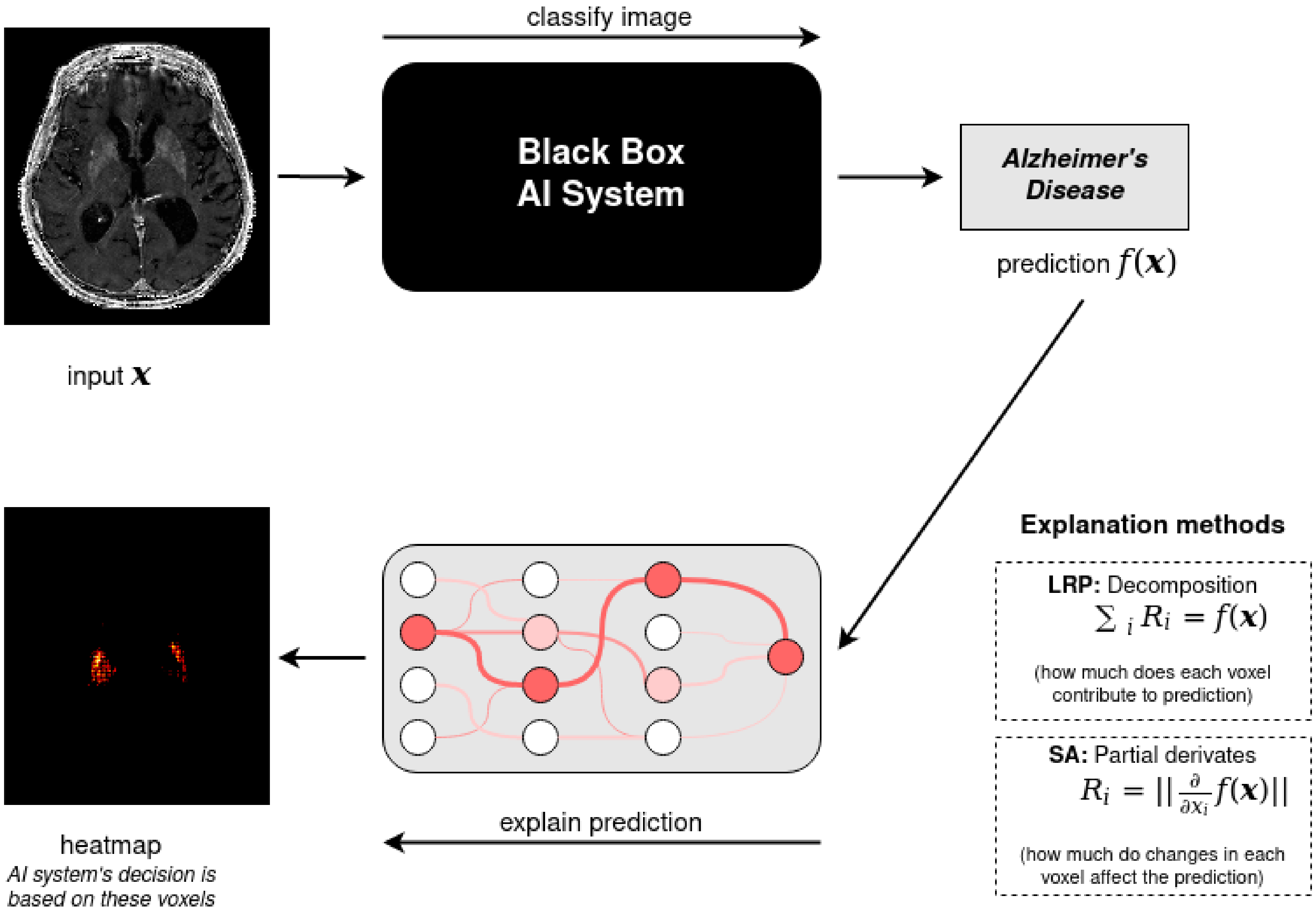
Volume 63, July 2020, 101694

Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation

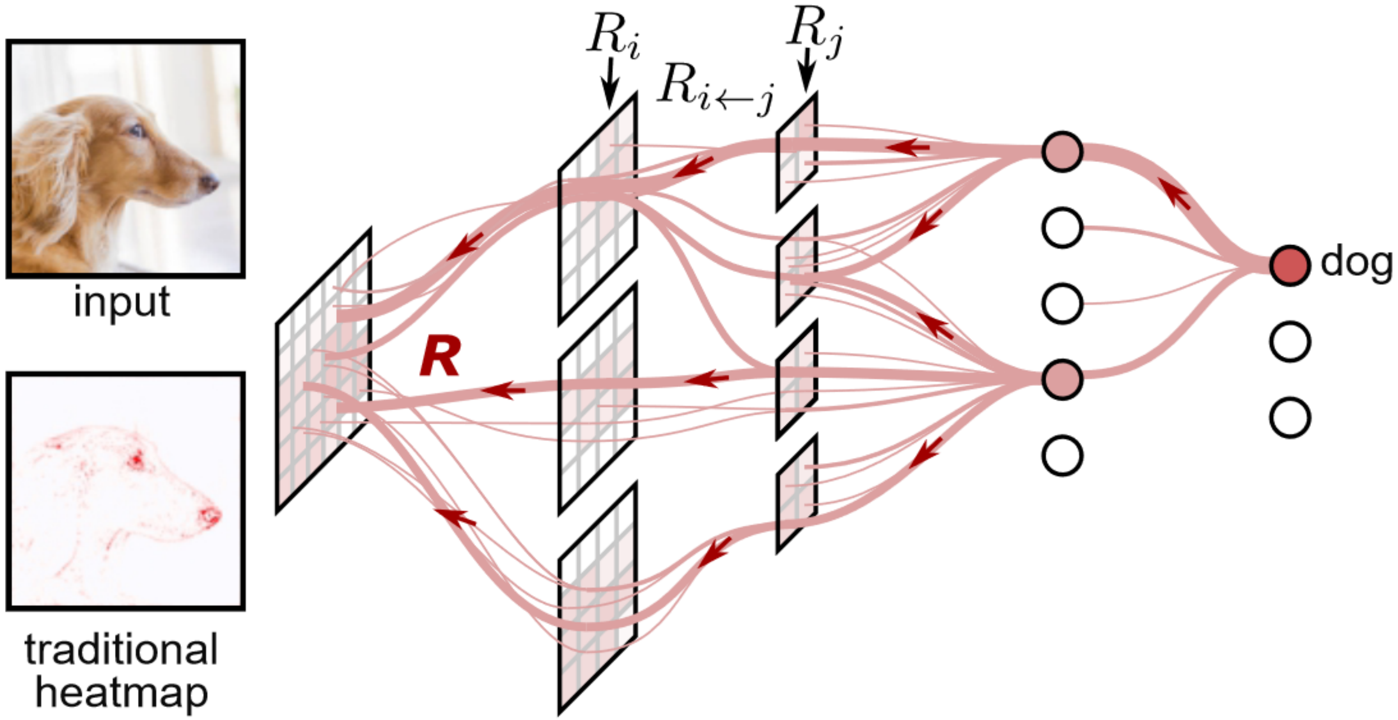
Junhao Wen [a, b, c, d, e, †], Elina Thibeau-Sutre [a, b, c, d, e, †], Mauricio Diaz-Melo [e, a, b, c, d], Jorge Samper-González [e, a, b, c, d], Alexandre Routier [e, a, b, c, d], Simona Bottani [e, a, b, c, d], Didier Dormont [e, a, b, c, d, f], Stanley Durrleman [e, a, b, c, d], Ninon Burgos [a, b, c, d, e], Olivier Colliot [a, b, c, d, e, f, g], for the Alzheimer's Disease Neuroimaging Initiative [#], the Australian Imaging Biomarkers and Lifestyle flagship study of ageing [##]

[1] Wen et al, Medical Image Analysis, 2020

Introduction – Explainable Deep Learning



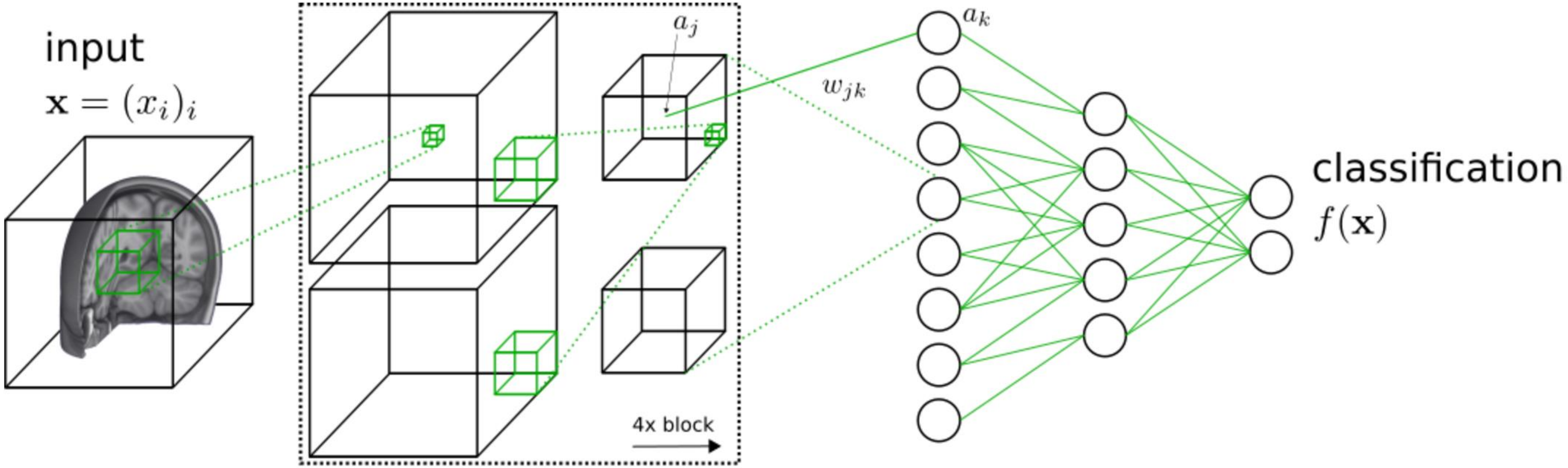
Introduction – Layer-wise Relevance Propagation



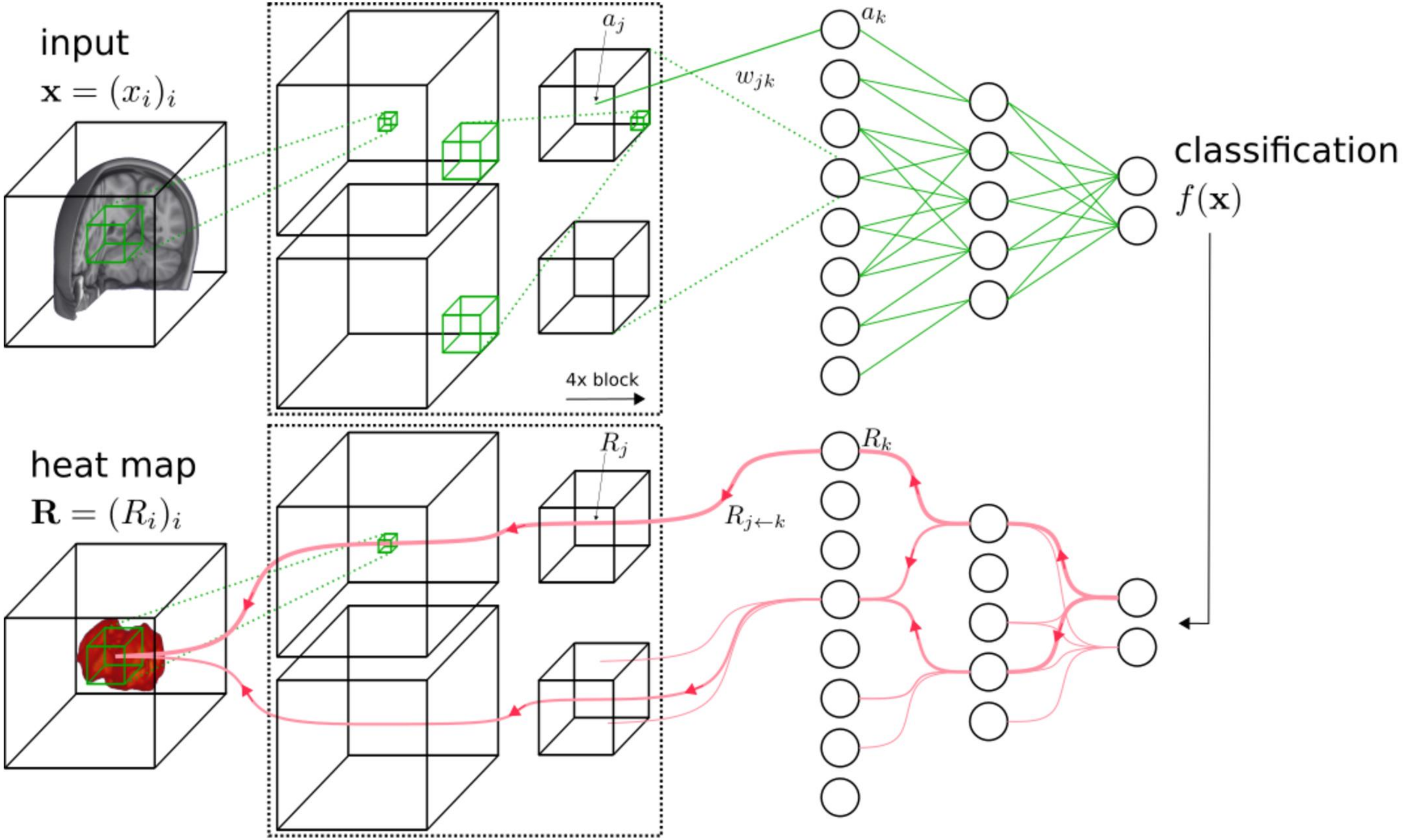
$$R_{i \leftarrow j}^{(l-1,l)}(\mathbf{x}) = \frac{z_{ij}}{z_j} R_j^l(\mathbf{x})$$

[1] Achtibat et al, Nature Machine Intelligence, 2023

Introduction – Explainable Deep Learning in AD

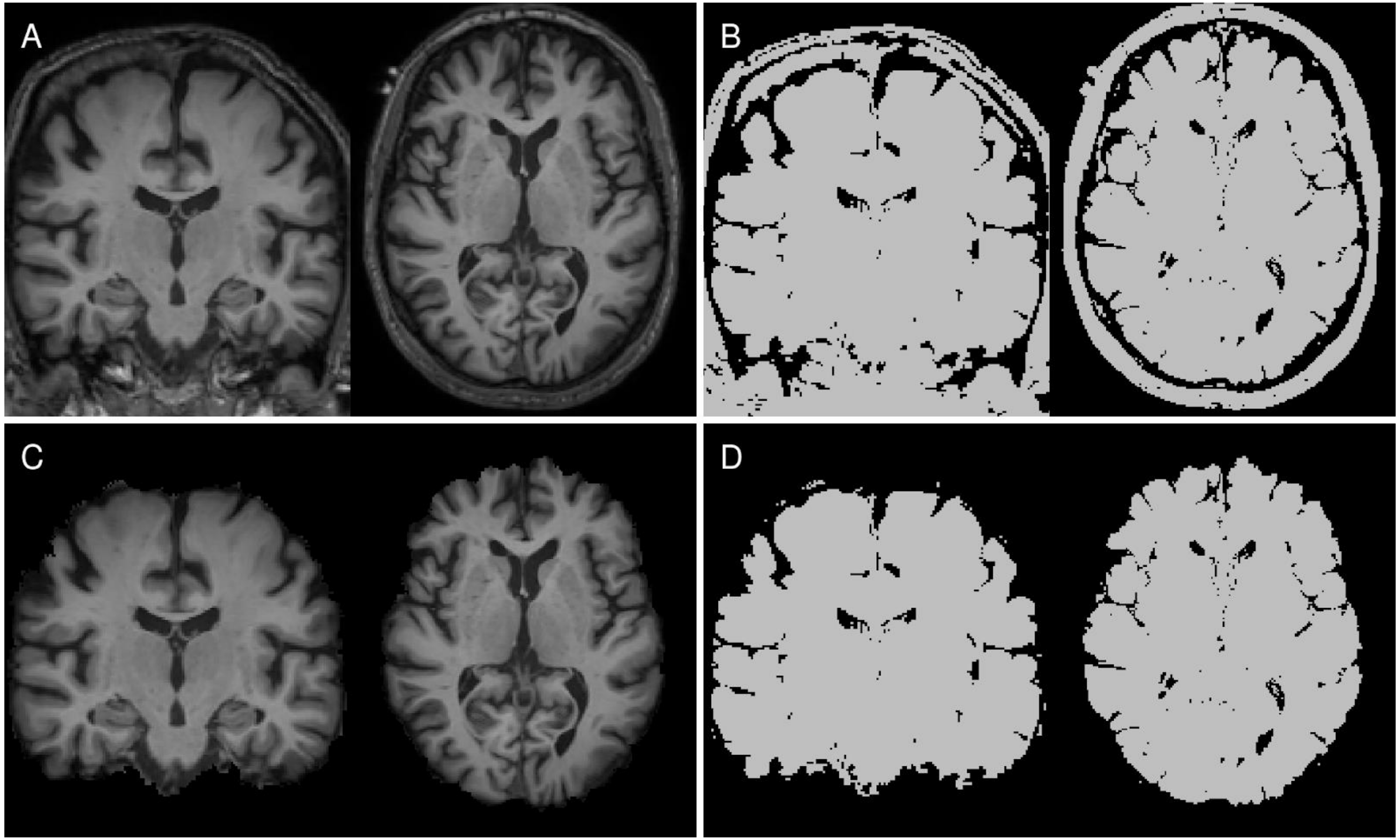


Introduction – Explainable Deep Learning in AD



1st Research Question

How much does T1w image texture influence deep learning Alzheimer's disease classifier?



Medical University of Graz

Medical University of Graz

Debugging:

When texture does not matter: Misinterpretation of deep learning-based Alzheimer's disease classification

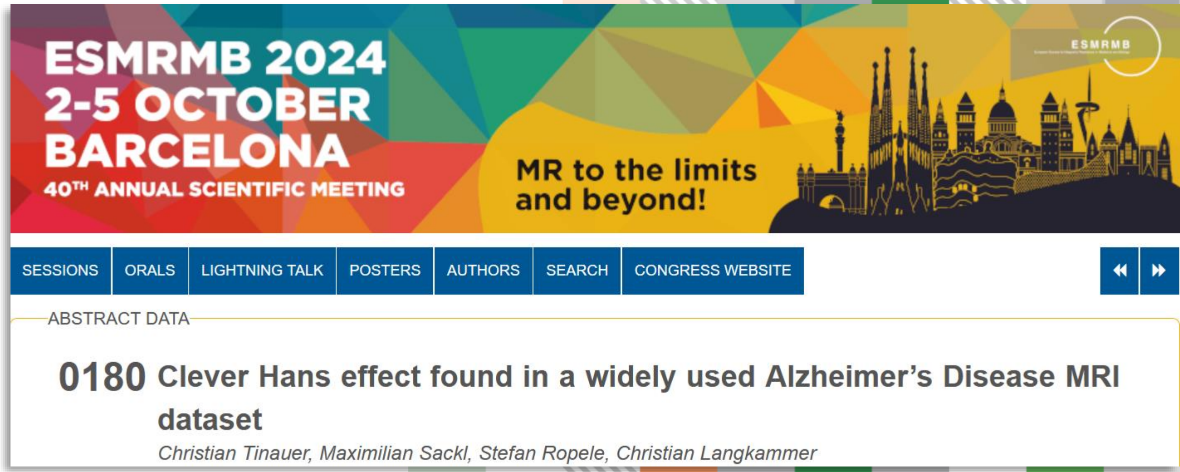
and

Clever Hans effect found in a widely used Alzheimer's Disease MRI dataset

Abstract #3046

When texture does not matter: Misinterpretation of deep learning-based Alzheimer's disease classification

Christian Tinauer[1], Stefan Ropele[1], and Christian Langkammer[1]
[1]Medical University of Graz, Graz, Austria



ABSTRACT DATA

0180 Clever Hans effect found in a widely used Alzheimer's Disease MRI dataset

Christian Tinauer, Maximilian Sackl, Stefan Ropele, Christian Langkammer

Methods – Dataset and Preprocessing

- 990 MRIs from 201 patients with probable AD;
  mean age=75.1±7.1 years, m/f=102/99

…propensity-logit-matched (covariates: age, sex) with…

- 990 MRIs from 159 normal controls;
  mean age=75.3±7.9 years, m/f=91/68

- ADNI database; https://adni.loni.usc.edu/

- FSL BET on T1w in native space and non-linear registration to MNI152 space

- Intensity rescaling based on white matter peak(s) in the brain tissue histogram per image

Medical University of Graz

Results - Performance

Input images	Binarizer	Accuracy	Sensitivity	Specificity	AUC
Native T1w	13,75%	62.51±5.45% [53.39%, 72.09%]	62.26±9.48% [47.70%, 84.93%]	62.79±8.35% [46.56%, 74.42%]	0.63±0.054 [0.53, 0.72]
	27,50%	72.74±5.49% [61.41%, 82.11%]	71.37±9.10% [56.86%, 88.87%]	74.15±9.94% [51.67%, 88.37%]	0.73±0.055 [0.61, 0.82]
	41,25%	77.95±4.57% [70.90%, 86.34%]	76.74±9.41% [60.10%, 94.56%]	79.15±6.80% [64.54%, 89.56%]	0.78±0.045 [0.71, 0.86
	None	71.12±5.01% [61.34%, 82.52%]	67.47±9.90% [51.66%, 85.94%]	74.76±7.07% [62.03%, 85.45%]	0.71±0.05 [0.62, 0.83]
Skull-stripped T1w	13,75%	78.12±4.63% [70.79%, 85.79%]	76.83±7.03% [62.65%, 87.05%]	79.40±6.76% [65.53%, 89.91%]	0.78±0.046 [0.71, 0.86]
	27,50%	79.57±3.92% [73.46%, 86.45%]	78.32±7.74% [66.79%, 93.87%]	80.92±7.71% [67.53%, 91.95%]	0.80±0.039 [0.74, 0.86]
	41,25%	81.56±4.63% [72.31%, 88.67%]	79.69±9.42% [62.59%, 96.48%]	**83.50±6.77%** [72.48%, 96.29%]	0.82±0.046 [0.72, 0.89]
	None	**81.63±3.77%** [74.36%, 88.01%]	**81.22±6.94%** [69.59%, 93.15%]	82.11±7.92% [65.50%, 93.60%]	**0.82±0.037** [0.74, 0.88]

Medical University of Graz

Related - Performance

Medical University of Graz

Classification architectures	Training data	Image preprocessing	Intensity rescaling	Data split	Training approach	Transfer learning	Task	Validation balanced accuracy	Exp #
3D subject-level CNN	Baseline	Minimal	None	subject-level	single-CNN	None	AD vs CN	0.50 ± 0.00 [0.50, 0.50, 0.50, 0.50, 0.50]	1
			MinMax					0.80 ± 0.05 [0.76, 0.86, 0.81, 0.85, 0.74]	2
						AE pre-training		0.82 ± 0.05 [0.74, 0.90, 0.83, 0.77, 0.83]	3
	Longitudinal	Minimal	MinMax	subject-level	single-CNN	AE pre-training		0.85 ± 0.04 [0.88, 0.88, 0.84, 0.85, 0.78]	4
		Extensive						0.86 ± 0.06 [0.88, 0.94, 0.85, 0.85, 0.76]	5

- Minimal preprocessing: No skull-stripping!
- Intensity rescaling is important!

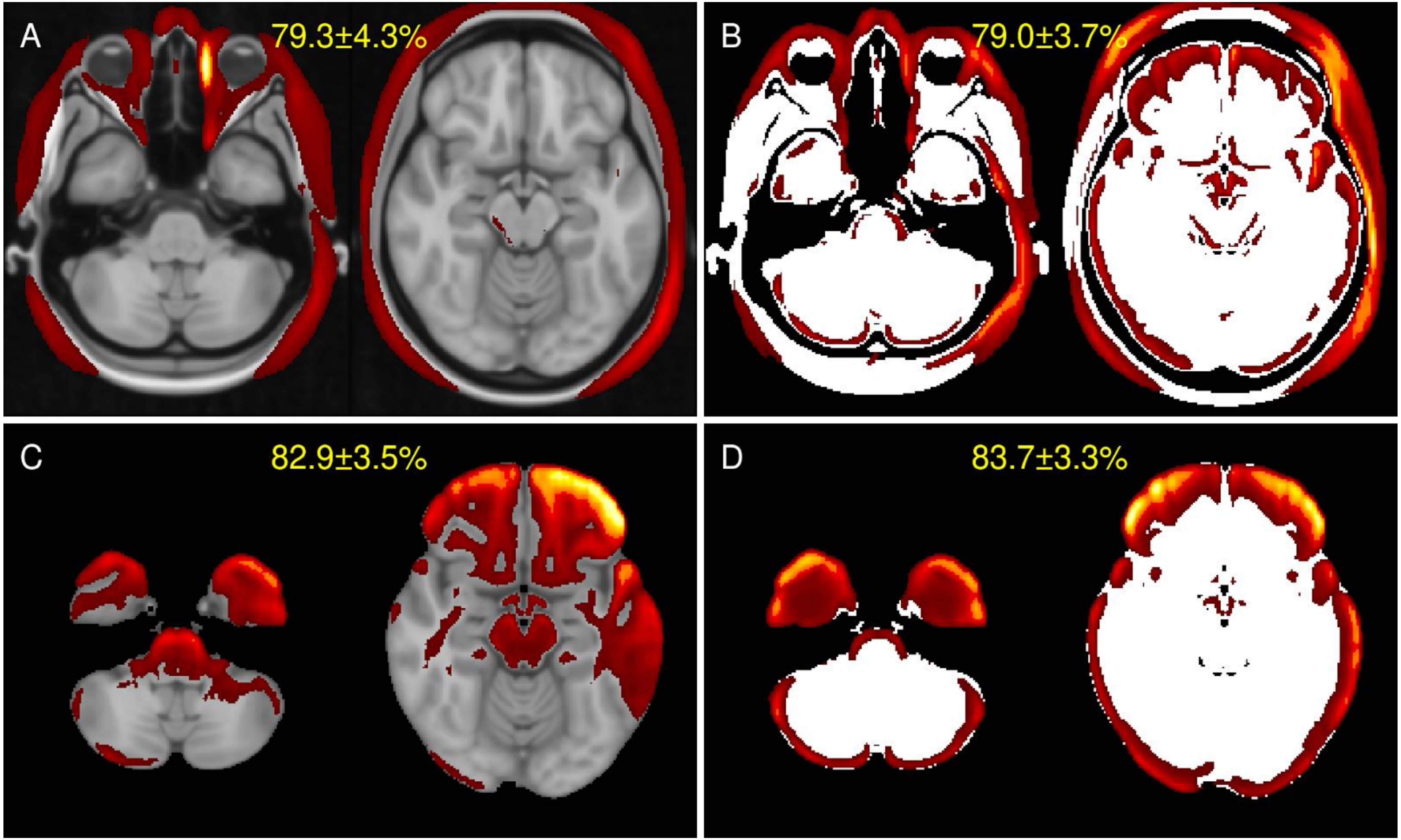
ELSEVIER
Medical Image Analysis
Volume 63, July 2020, 101694

Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation

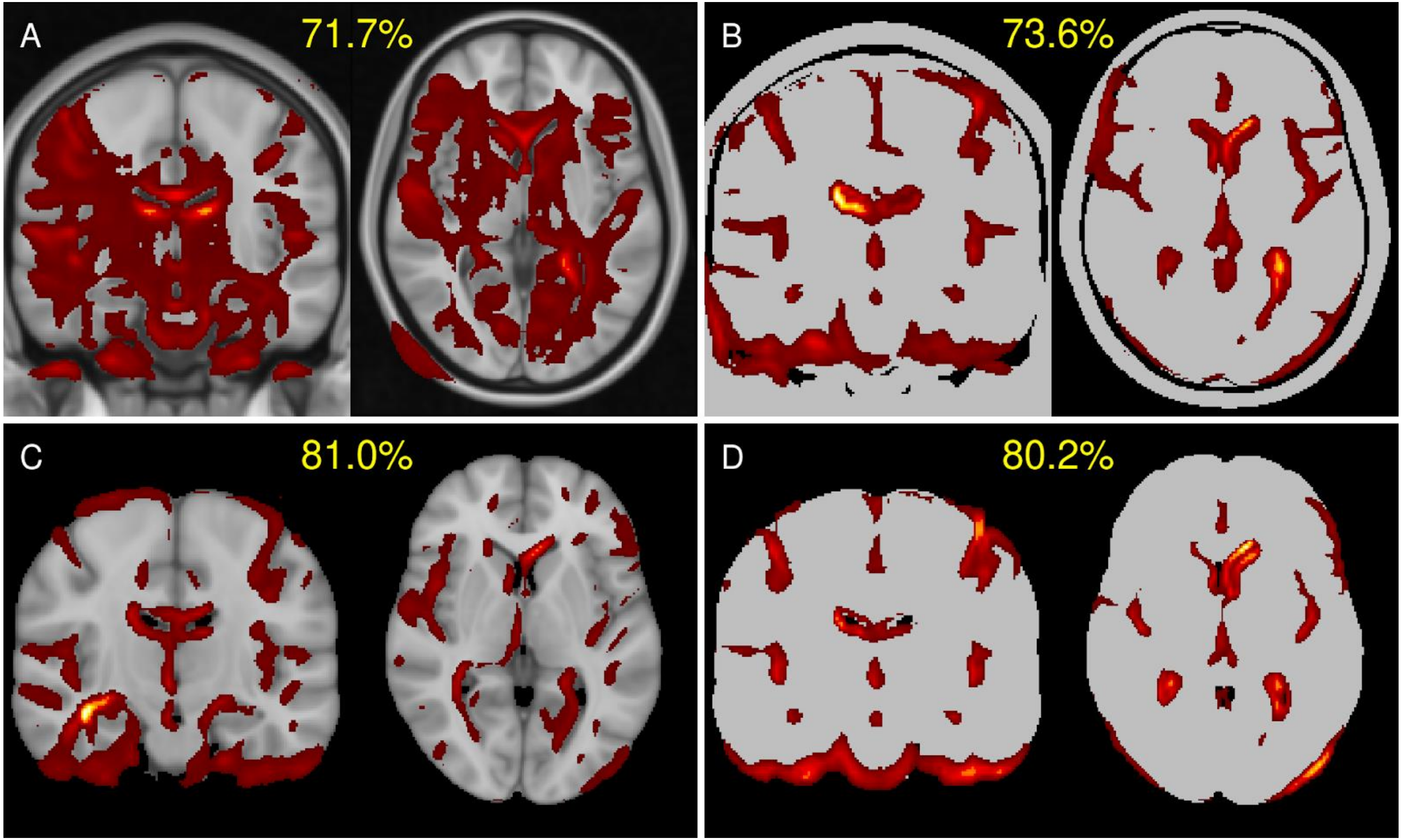
Junhao Wen [a, b, c, d, e, †], Elina Thibeau-Sutre [a, b, c, d, e, †], Mauricio Diaz-Melo [e, a, b, c, d], Jorge Samper-González [e, a, b, c, d], Alexandre Routier [e, a, b, c, d], Simona Bottani [e, a, b, c, d], Didier Dormont [e, a, b, c, d, f], Stanley Durrleman [e, a, b, c, d], Ninon Burgos [a, b, c, d, e], Olivier Colliot [a, b, c, d, e, f, g], for the Alzheimer's Disease Neuroimaging Initiative [#], the Australian Imaging Biomarkers and Lifestyle flagship study of ageing [##]

[a] Institut du Cerveau et de la Moelleépinière, ICM, Paris F-75013, France
[b] SorbonneUniversité, ParisF-75013,France
[c] Inserm, U 1127, Paris F-75013, France
[d] CNRS, UMR 7225, Paris F-75013, France
[e] Inria, Aramis project-team, Paris F-75013, France
[f] Department of Neuroradiology, AP-HP, Hôpital de la PitiéSalpêtrière, Paris F-75013, France
[g] Department of Neurology, AP-HP, Hôpital de la PitiéSalpêtrière, Paris F-75013, France

Results – Heatmaps, Unmatched-Data

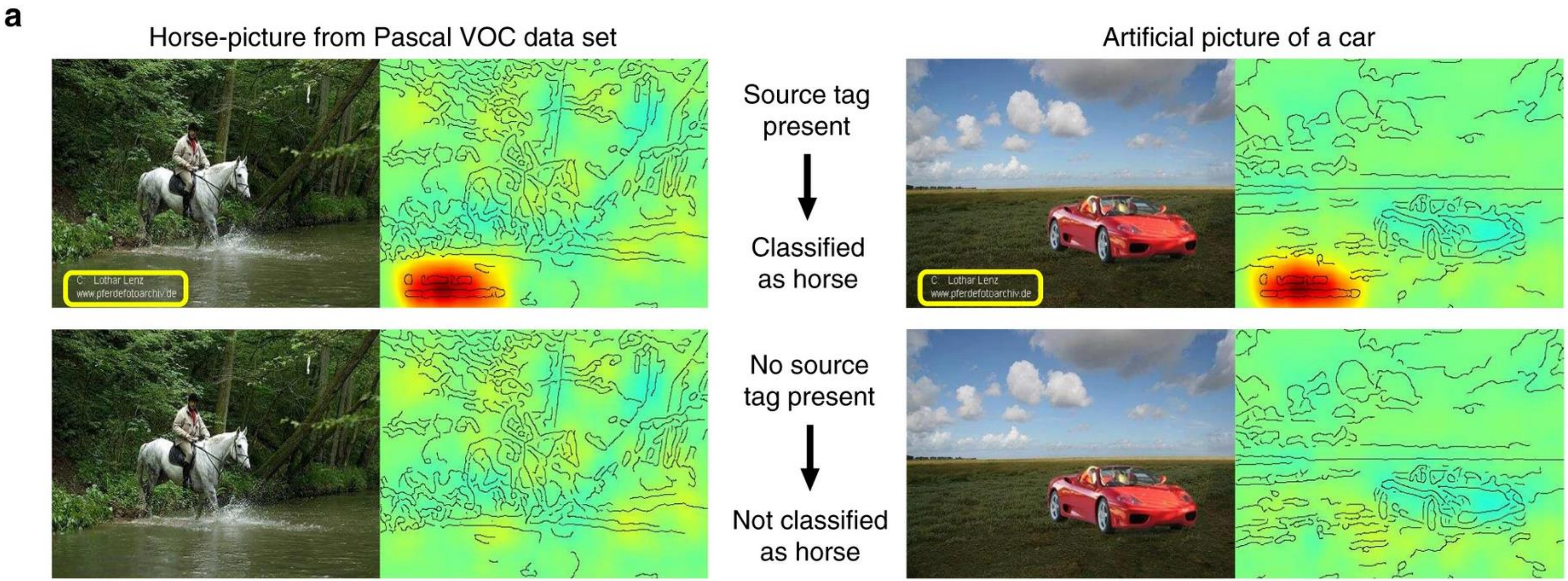


Results – Heatmaps, Propensity-Score-Matched



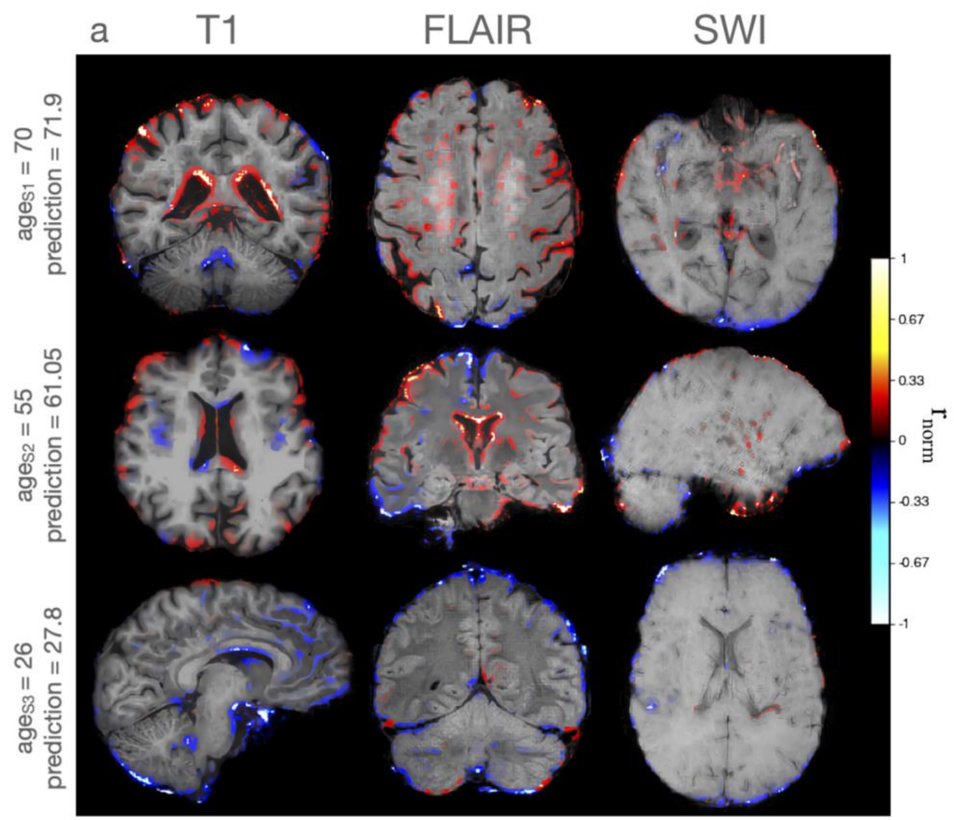
Medical University of Graz

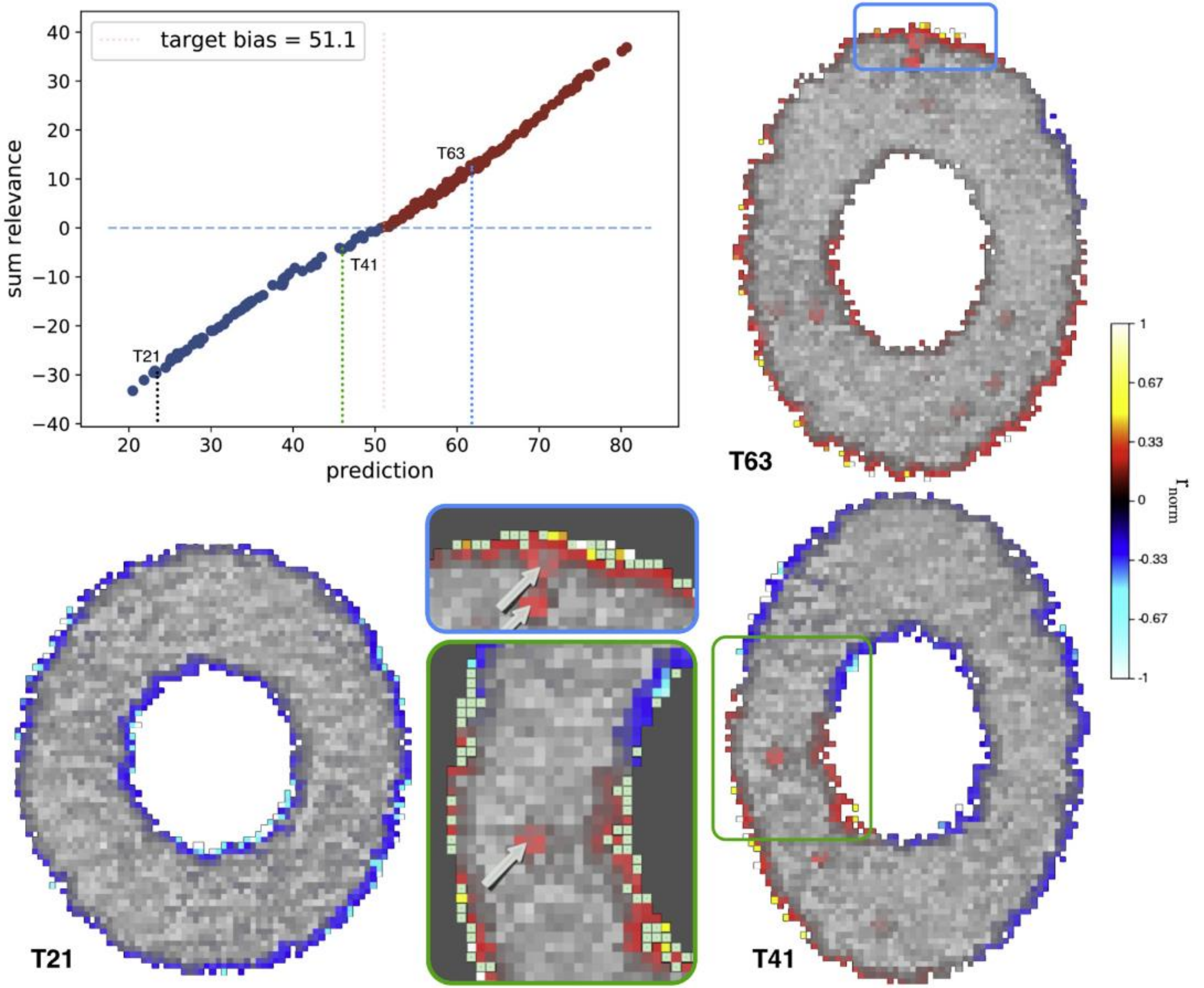
Related – Explainability



[1] Lapuschkin et al, Nature Communications, 2019

Related – (Simulated) Brain Aging





[1] Hofmann et al, NeuroImage, 2022

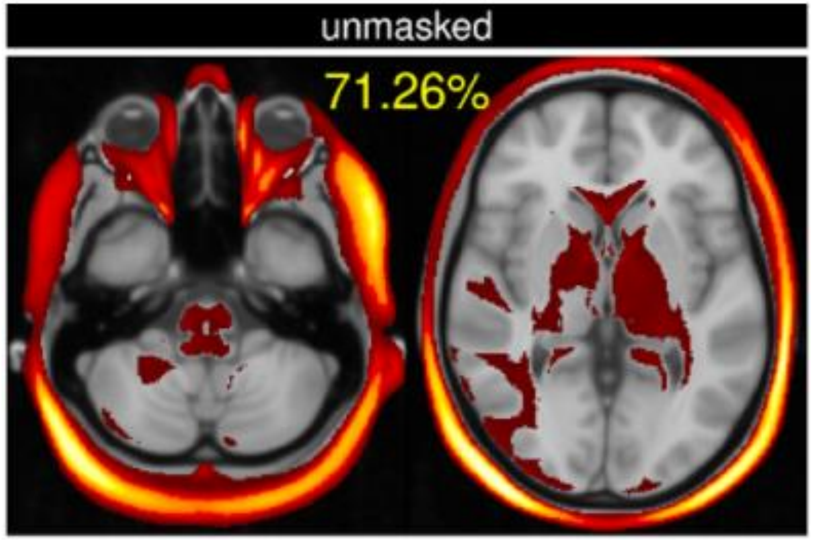
Discussion & Conclusion

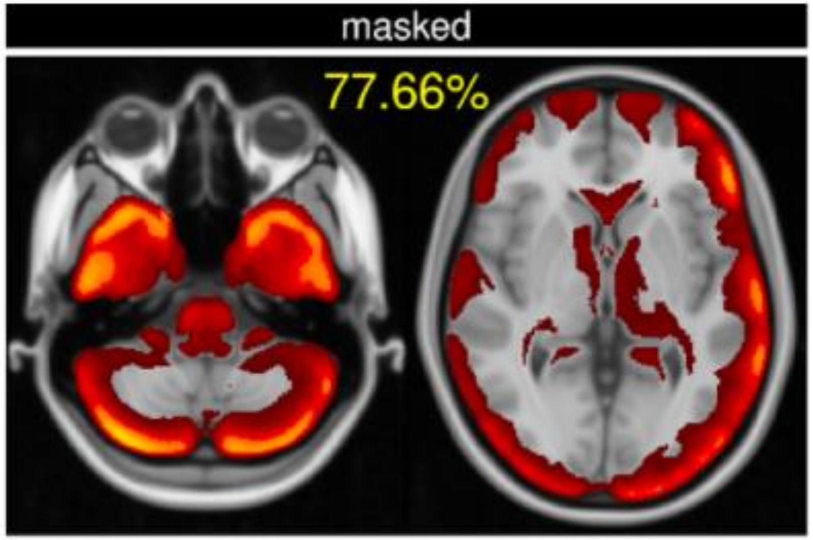
- Deep learning AD classification is strongly driven by volumetric features.
- Gray-white matter texture does not improve classification performance.

[1] https://en.wikipedia.org/wiki/Clever_Hans#/media/File:Osten_und_Hans.jpg

2nd Research Question

Is it possible to integrate (problem specific) a priori information into the deep learning training process using explainability?





Medical University of Graz

scientific reports

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > scientific reports > articles > article

Article | Open access | Published: 24 November 2022

Interpretable brain disease classification and relevance-guided deep learning

Christian Tinauer, Stefan Heber, Lukas Pirpamer, Anna Damulina, Reinhold Schmidt, Rudolf Stollberger, Stefan Ropele & Christian Langkammer

Scientific Reports **12**, Article number: 20254 (2022) | Cite this article

2905 Accesses | **8** Altmetric | Metrics

Regularization:

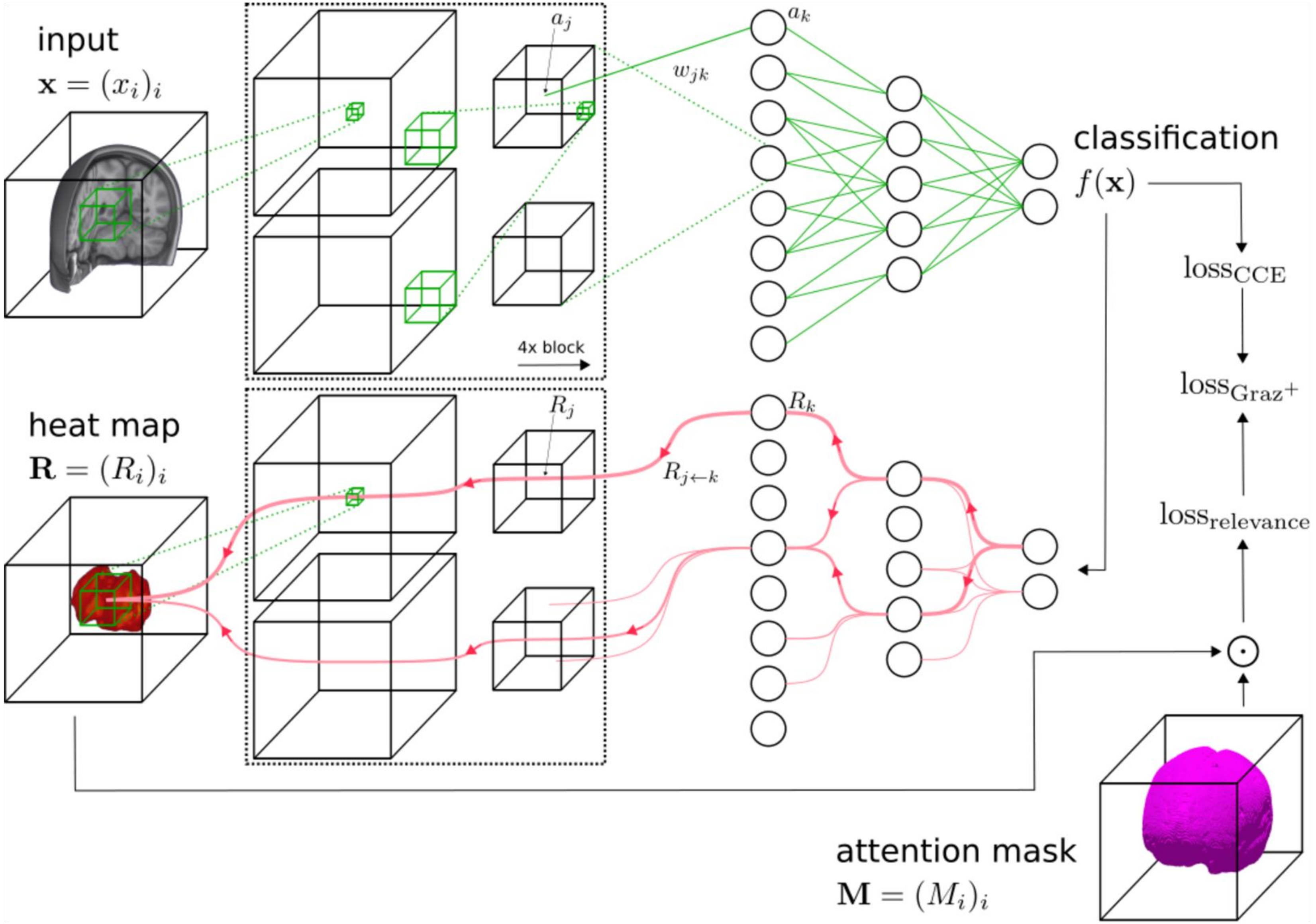
Interpretable brain disease classification and relevance-guided deep learning

Methods – Dataset

- 264 MRIs from 128 patients with probable AD;
  mean age=71.9±8.5 years, ProDem study
- 378 MRIs from 290 normal controls;
  mean age=71.3±6.4 years, community-dwelling study
- Skull stripping: FSL BET for brain masks
- Registrations: none (native space), linear to MNI152, non-linear to MNI152
- Intensity rescaling with one fixed value for all images

Medical University of Graz

Methods – Network



Guided
relevance
by
adaptive
z$^+$-rule

[1] Tinauer et al, Scientific Reports 2022

Methods – Loss function

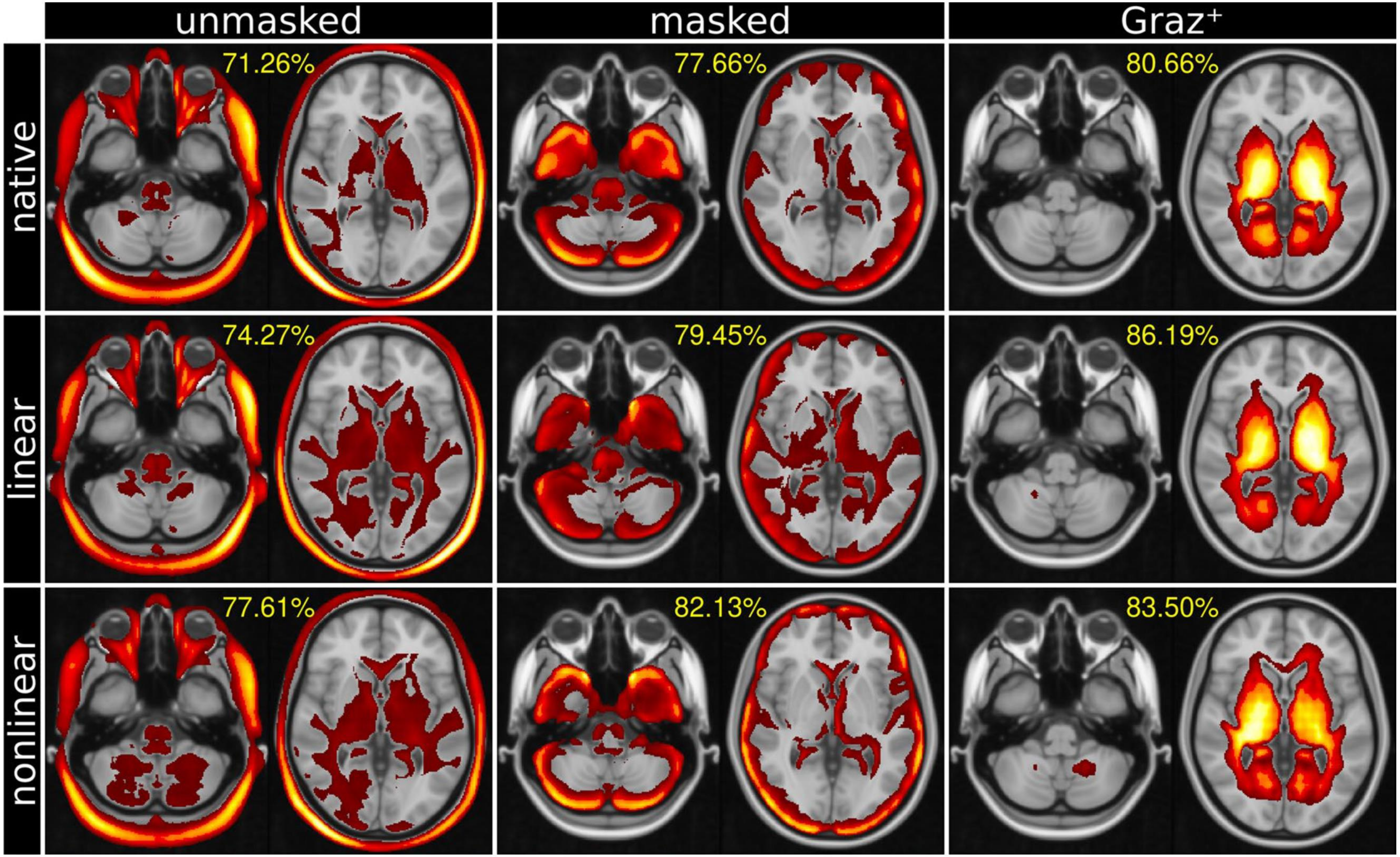
$$\text{loss}_{\text{relevance}}(\mathbf{R}, \mathbf{M}) = -\mathbf{1}^T \text{vec}(\mathbf{R} \odot \mathbf{M})$$

$$\begin{aligned}\text{loss}_{\text{Graz}^+} &= \text{loss}_{\text{relevance}} + \text{loss}_{\text{CCE}} \\ &= -\mathbf{1}^T \text{vec}(\mathbf{R} \odot \mathbf{M}) - \sum_{i=1}^{\text{outputs}} y_i \cdot \log(\hat{y}_i)\end{aligned}$$

Results - Performance

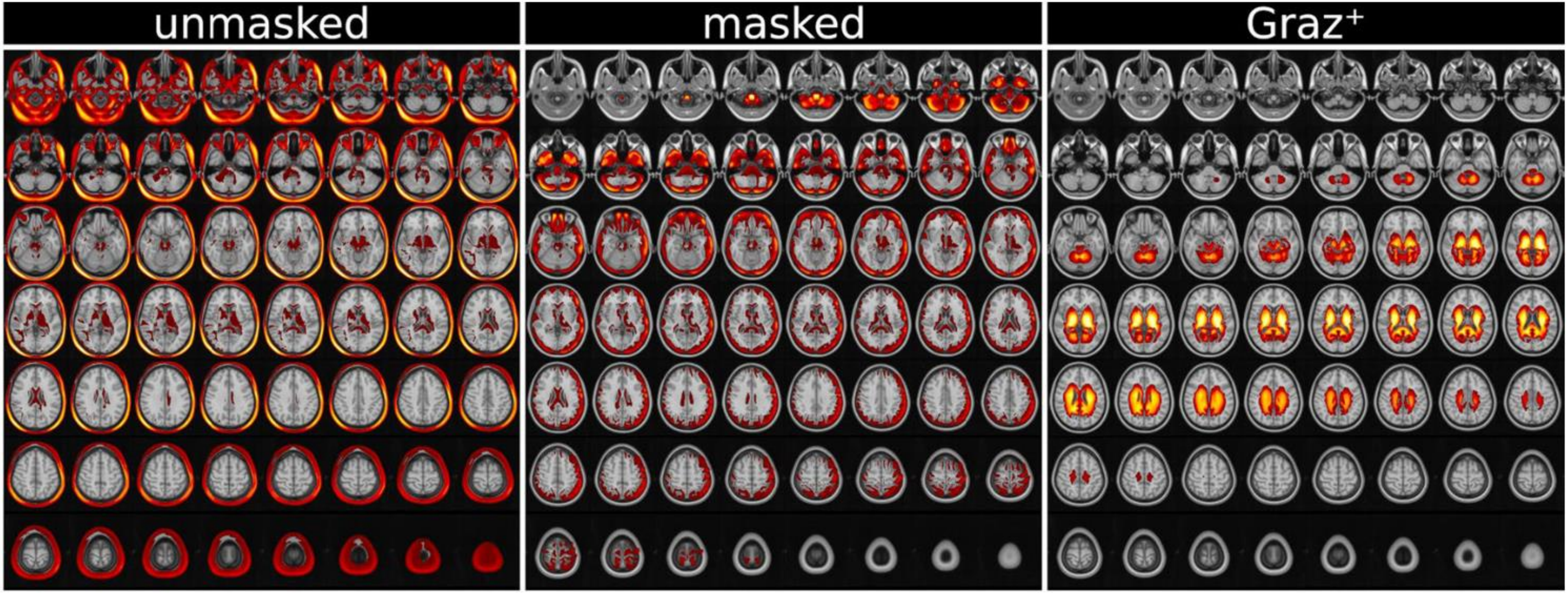
Classifier	Skull stripping	Registration	Balanced accuracy	Sensitivity	Specificity	AUC
CNN	No	–	71.26 ± 2.86%	55.55 ± 7.51%	86.96 ± 3.95%	0.75 ± 0.02
		Lin.	74.27 ± 3.83%	63.13 ± 9.05%	85.40 ± 6.45%	0.80 ± 0.05
		Nonlin.	77.61 ± 4.44%	64.79 ± 5.02%	90.43 ± 5.19%	0.85 ± 0.06
CNN	Yes	–	77.66 ± 4.39%	69.70 ± 7.65%	85.63 ± 4.06%	0.83 ± 0.05
		Lin.	79.45 ± 3.34%	76.87 ± 4.81%	82.03 ± 6.23%	0.86 ± 0.05
		Nonlin.	82.13 ± 5.08%	73.47 ± 7.89%	90.78 ± 4.92%	0.88 ± 0.05
CNN+Graz$^+$	No	–	80.66 ± 4.80%	74.95 ± 7.85%	86.36 ± 2.85%	0.88 ± 0.04
		Lin.	**86.19 ± 6.01%**	79.73 ± 10.72%	**92.66 ± 3.73%**	**0.92 ± 0.04**
		Nonlin.	83.50 ± 5.90%	77.16 ± 8.95%	89.83 ± 4.49%	0.90 ± 0.04
Logistic regression*	Yes	Lin.**	82.00 ± 4.25%	**80.57 ± 7.16%**	83.43 ± 2.45%	0.90 ± 0.04

Results – Heatmaps I

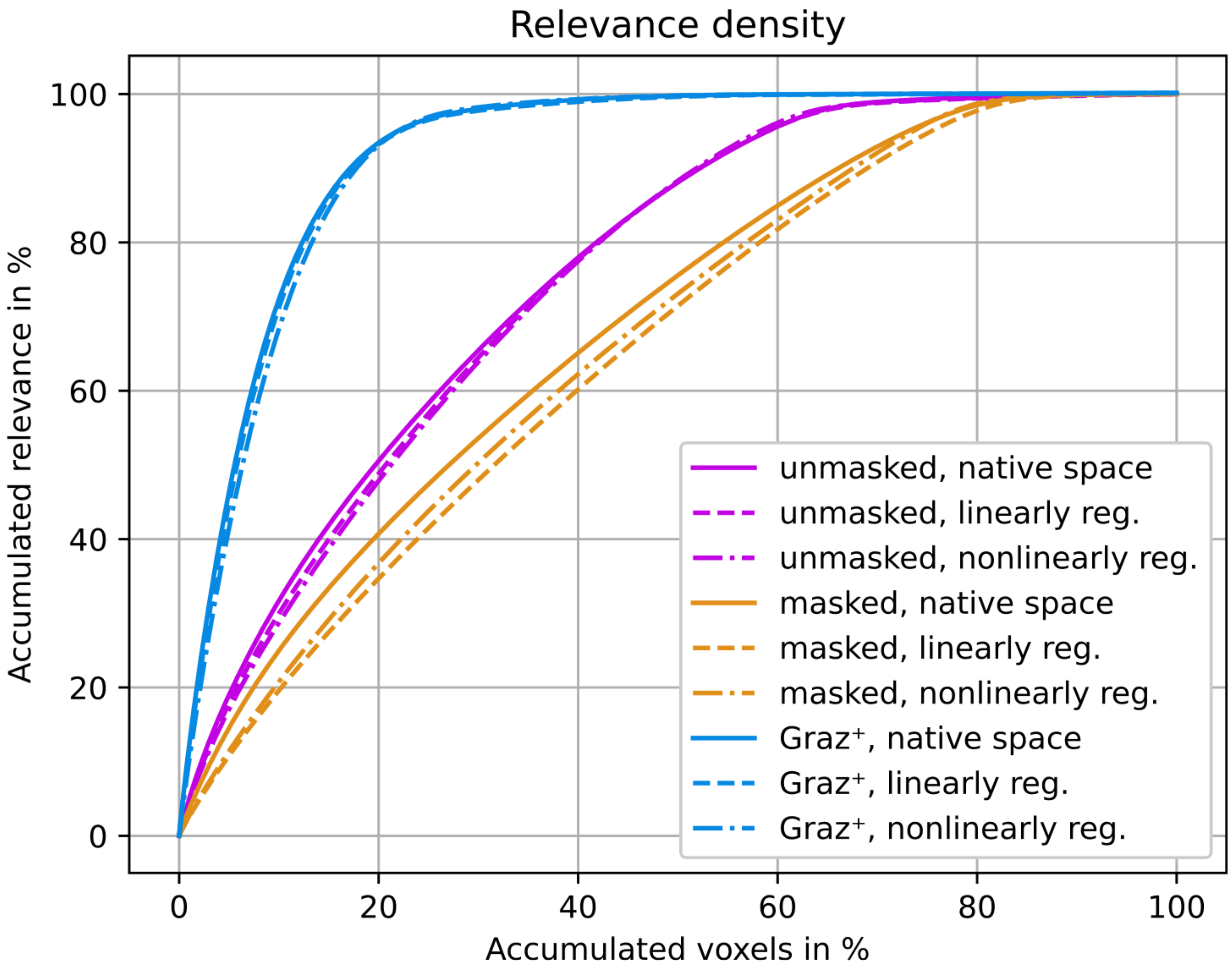


Medical University of Graz

Results – Heatmaps II



Results



Medical University of Graz

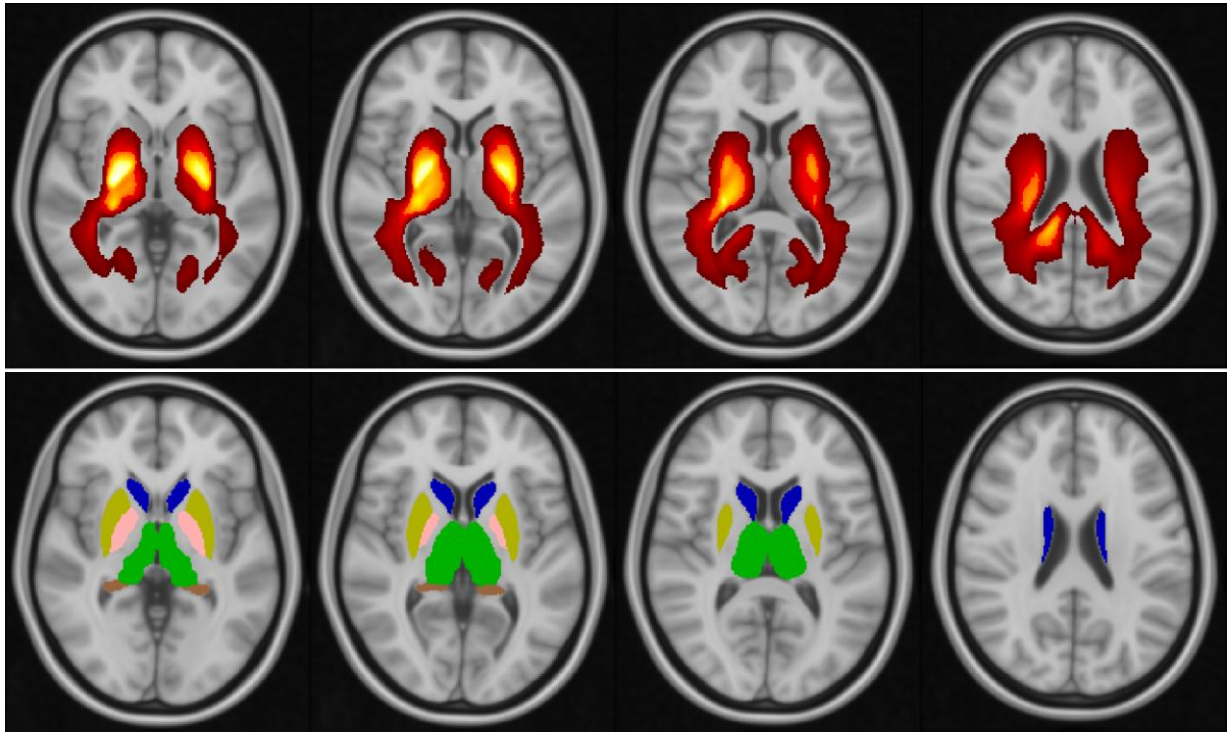
Discussion & Conclusion

Medical University of Graz

- Preprocessing of MR images is crucial for the feature identification by DNNs.
- The proposed relevance-guided approach identified regions with highest relevance in brain tissue located adjacent to the ventricles.
- Increased feature sparsity improves the classification accuracy.

3rd Research Question

Does explainable deep learning capture pathological tissue changes?



Medical University of Graz

Home > Explainable Artificial Intelligence > Conference paper

Explainable Concept Mappings of MRI: Revealing the Mechanisms Underlying Deep Learning-Based Brain Disease Classification

Conference paper | First Online: 10 July 2024

pp 202–216 | Cite this conference paper

Explainable Artificial Intelligence

(xAI 2024)

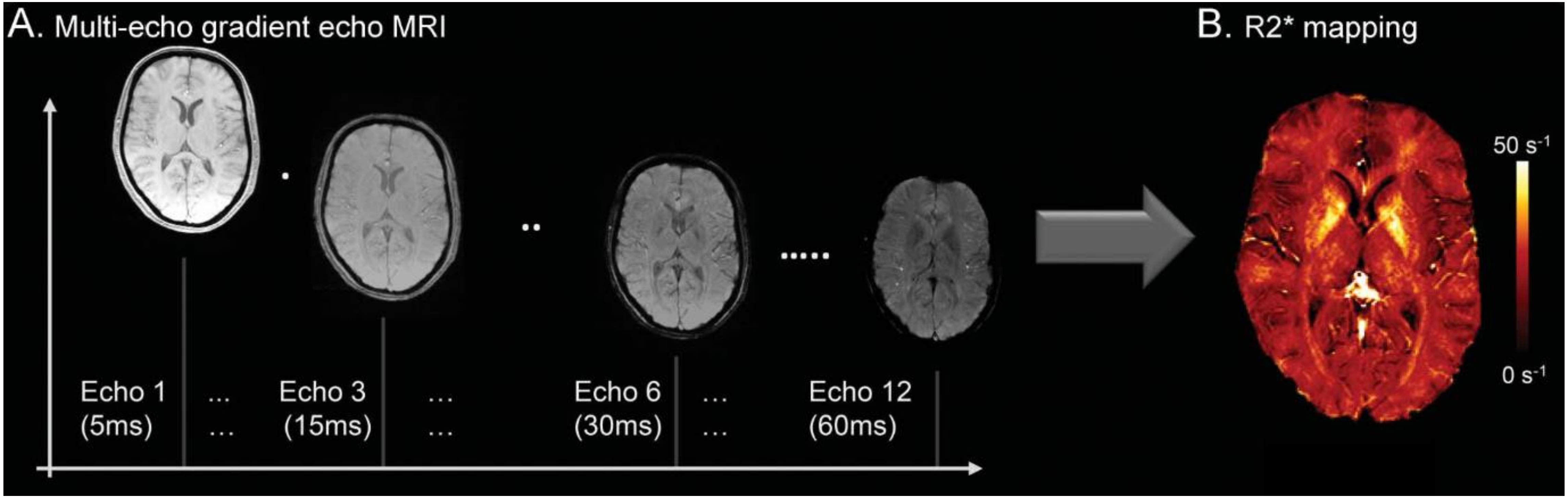
Christian Tinauer, Anna Damulina, Maximilian Sackl, Martin Soellradl, Reduan Achtibat, Maximilian Dreyer, Frederik Pahde, Sebastian Lapuschkin, Reinhold Schmidt, Stefan Ropele, Wojciech Samek & Christian Langkammer

Validation:

Explainable concept mappings of MRI: Revealing the mechanisms underlying deep learning-based brain disease classification

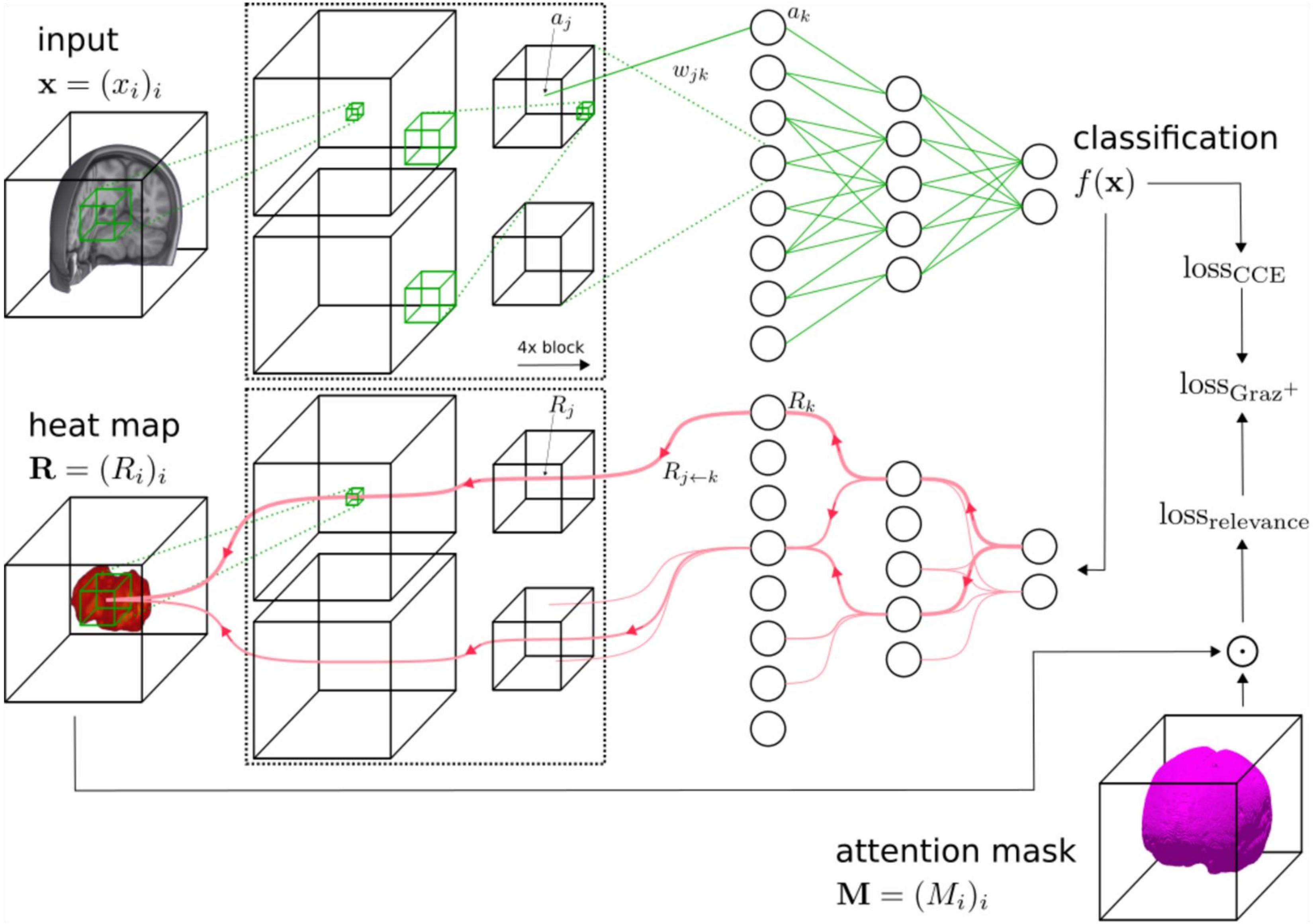
Methods – Dataset

- Brain masks from T1-weighted MPRAGE images
- R2* maps from images of spoiled FLASH sequence using a monoexponential model
- 226 R2* maps from 117 patients
- 226 R2* maps from 219 controls



[1] Khalil et al, Neurology, 2015

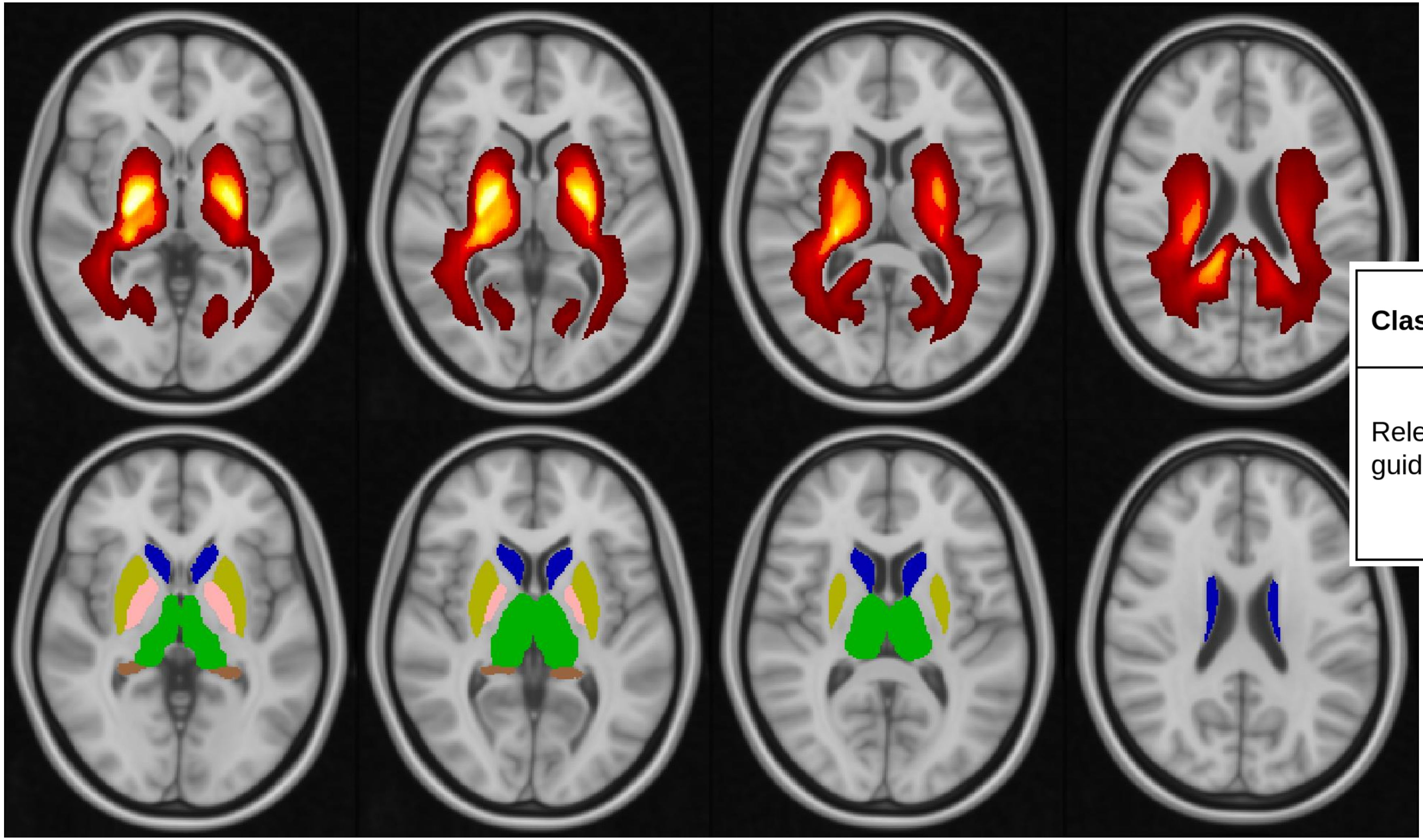
Methods – Network



Guided
relevance
by
adaptive
z$^+$-rule

[1] Tinauer et al, Scientific Reports 2022

Results I – Heat maps from R2* maps input

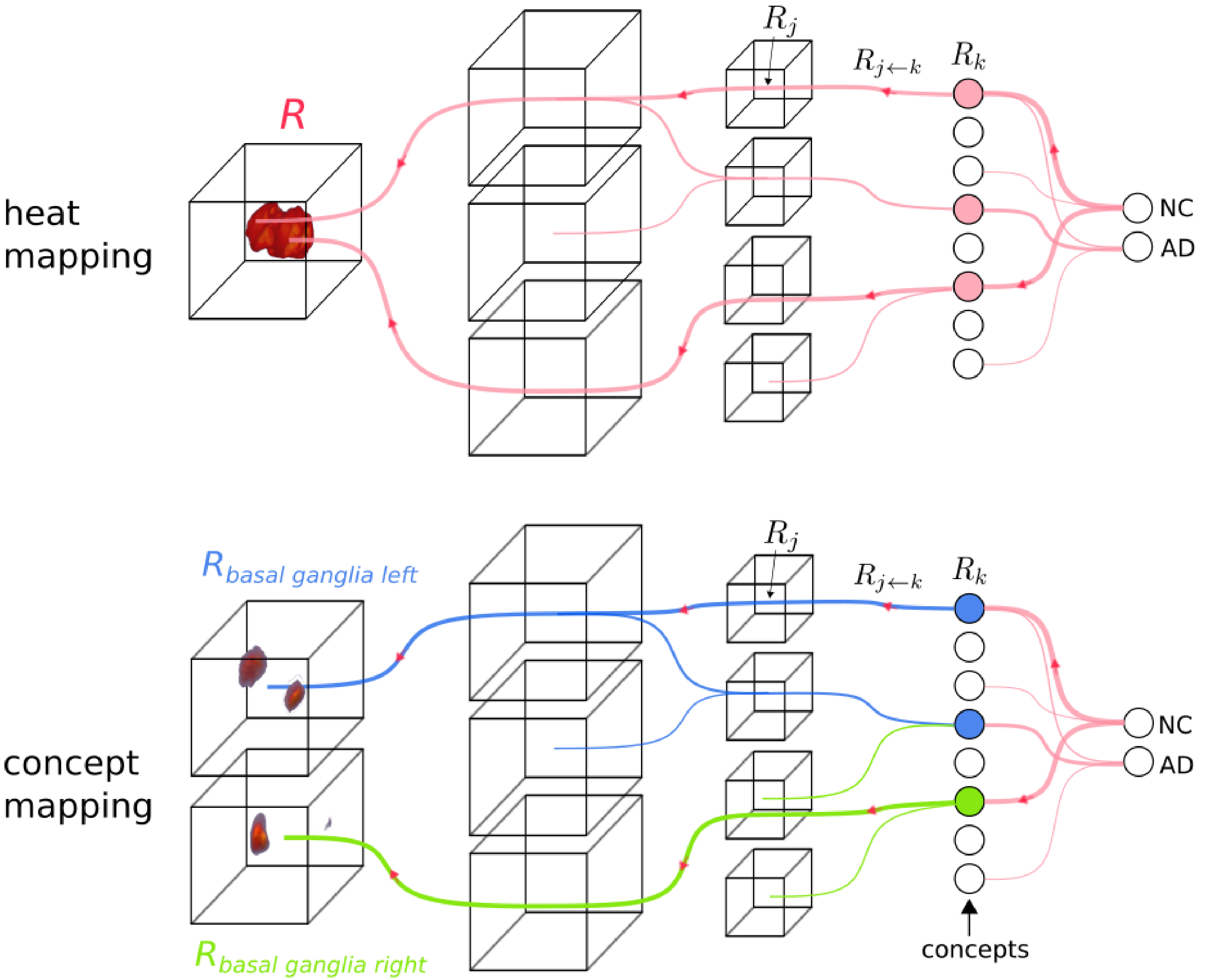


thalami - caudate nuclei - putamen - pallidum

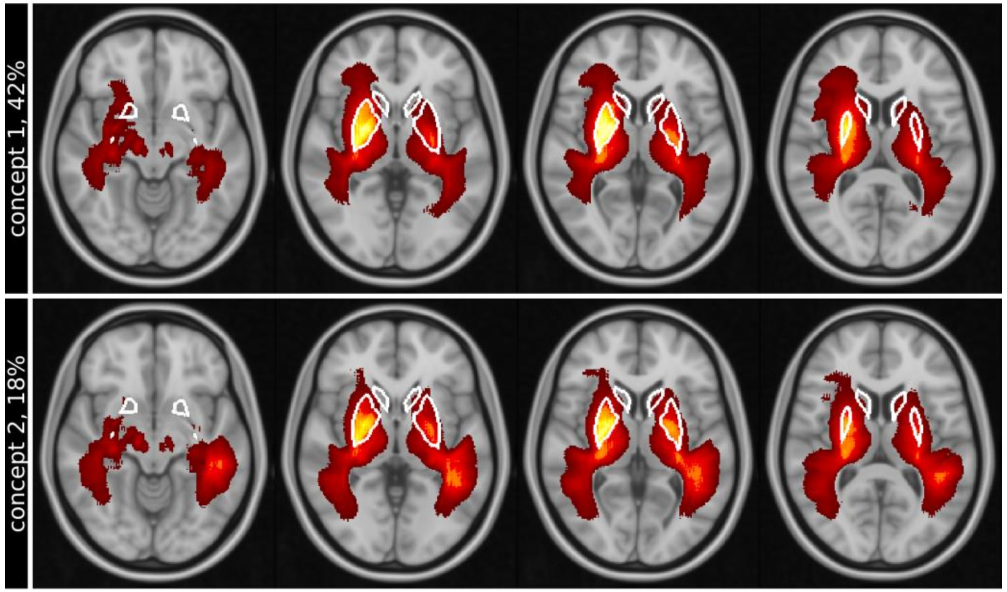
Classifier	Registration	Balanced accuracy
Relevance-guided CNN	-	**80.36±2.43%**
	lin.	77.51±3.27%
	nonlin.	76.27±3.89%

Medical University of Graz

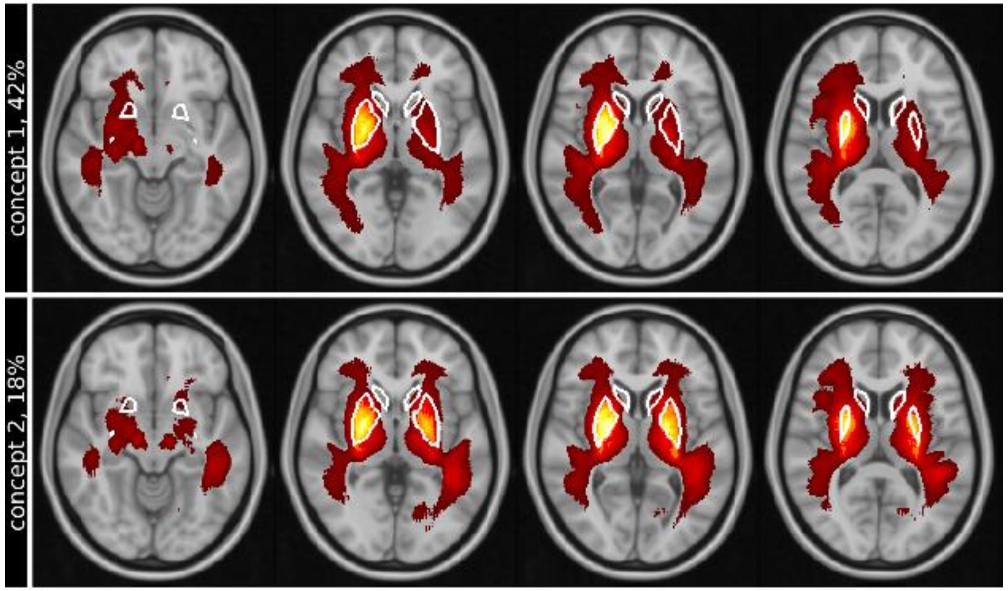
Methods – Heat- vs Concept-Mapping



Results II – Concept maps from R2* maps input

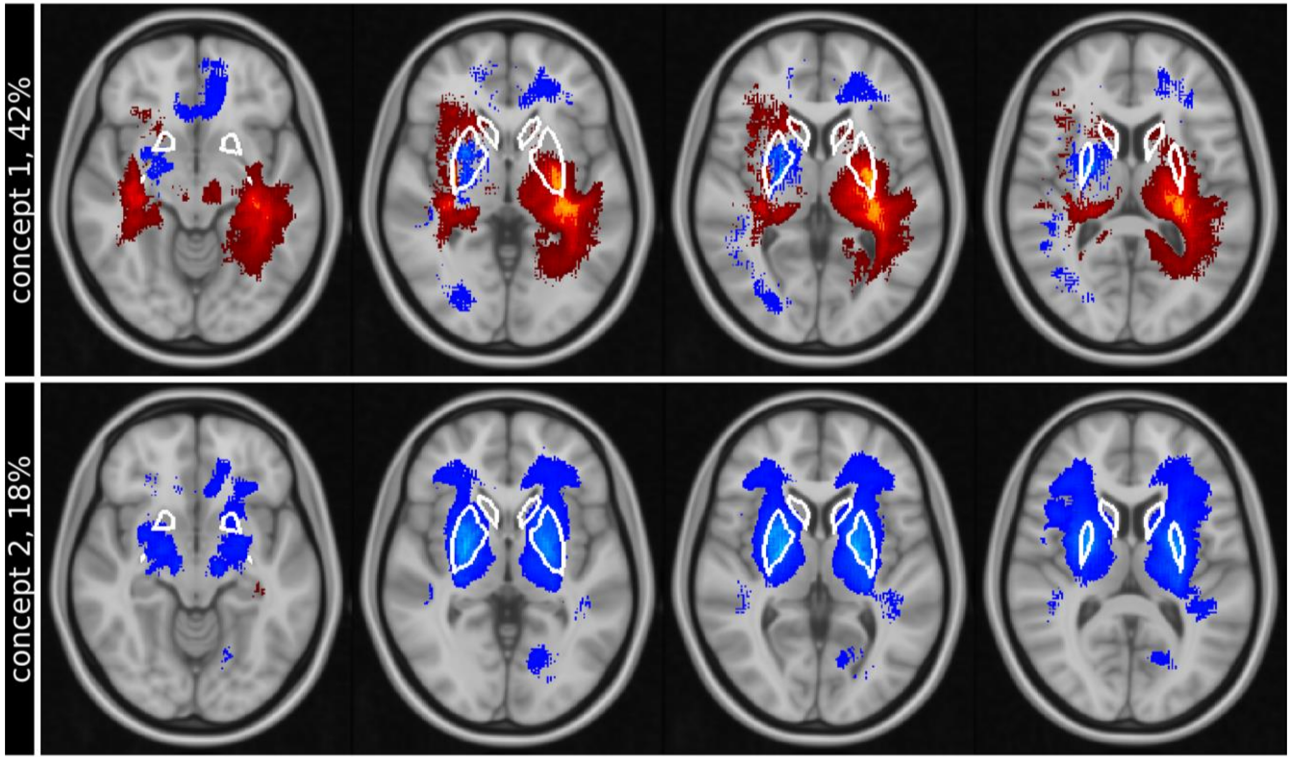


$R_{NC,mean}$



$R_{AD,mean}$

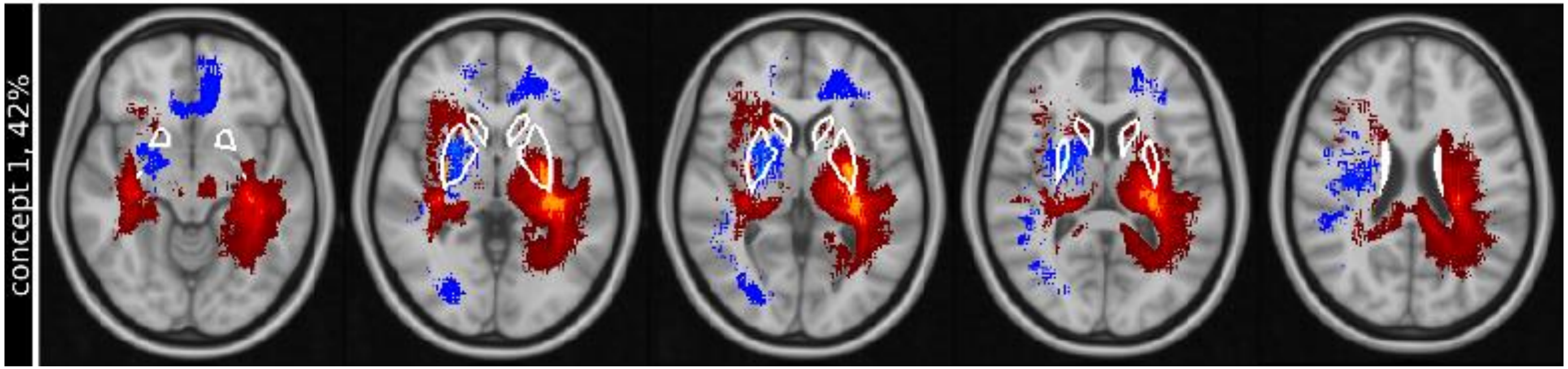
$R_{diff} = R_{NC,mean} - R_{AD,mean}$



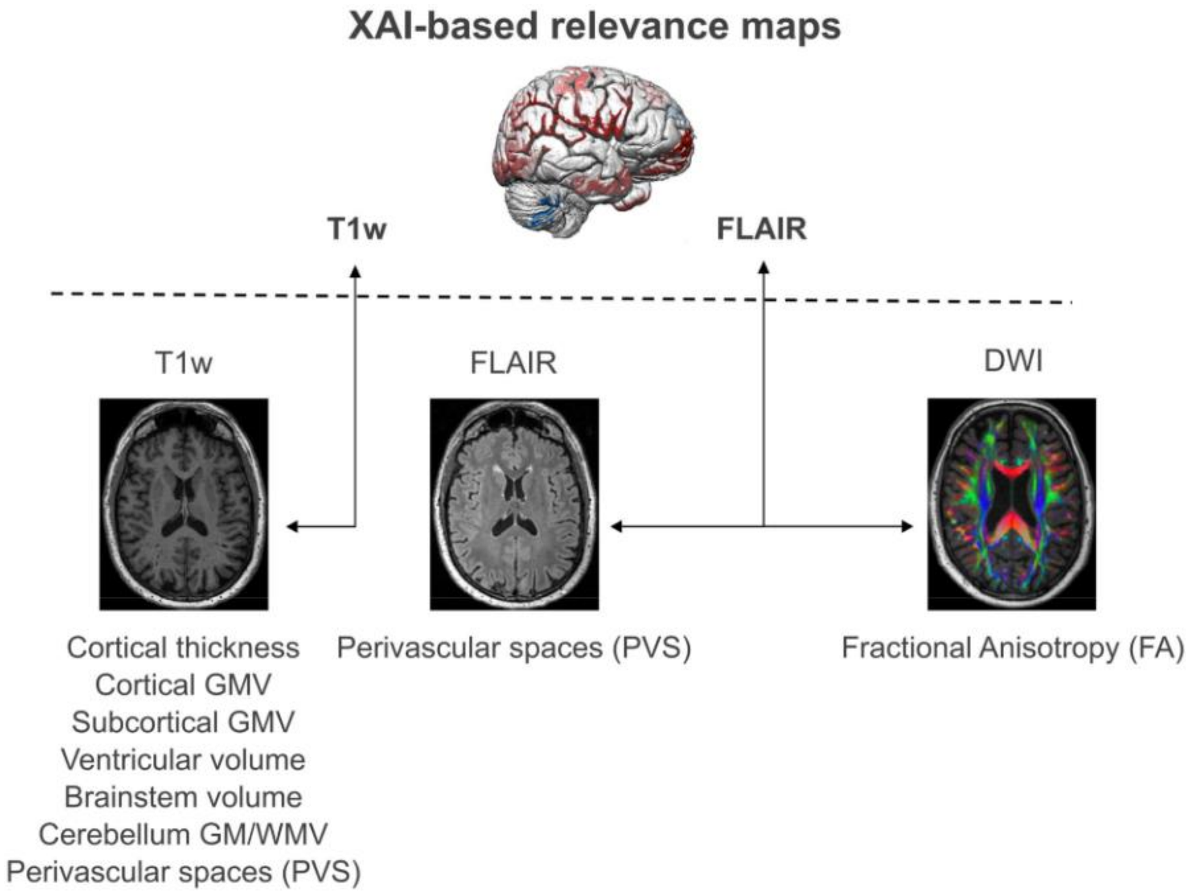
Medical University of Graz

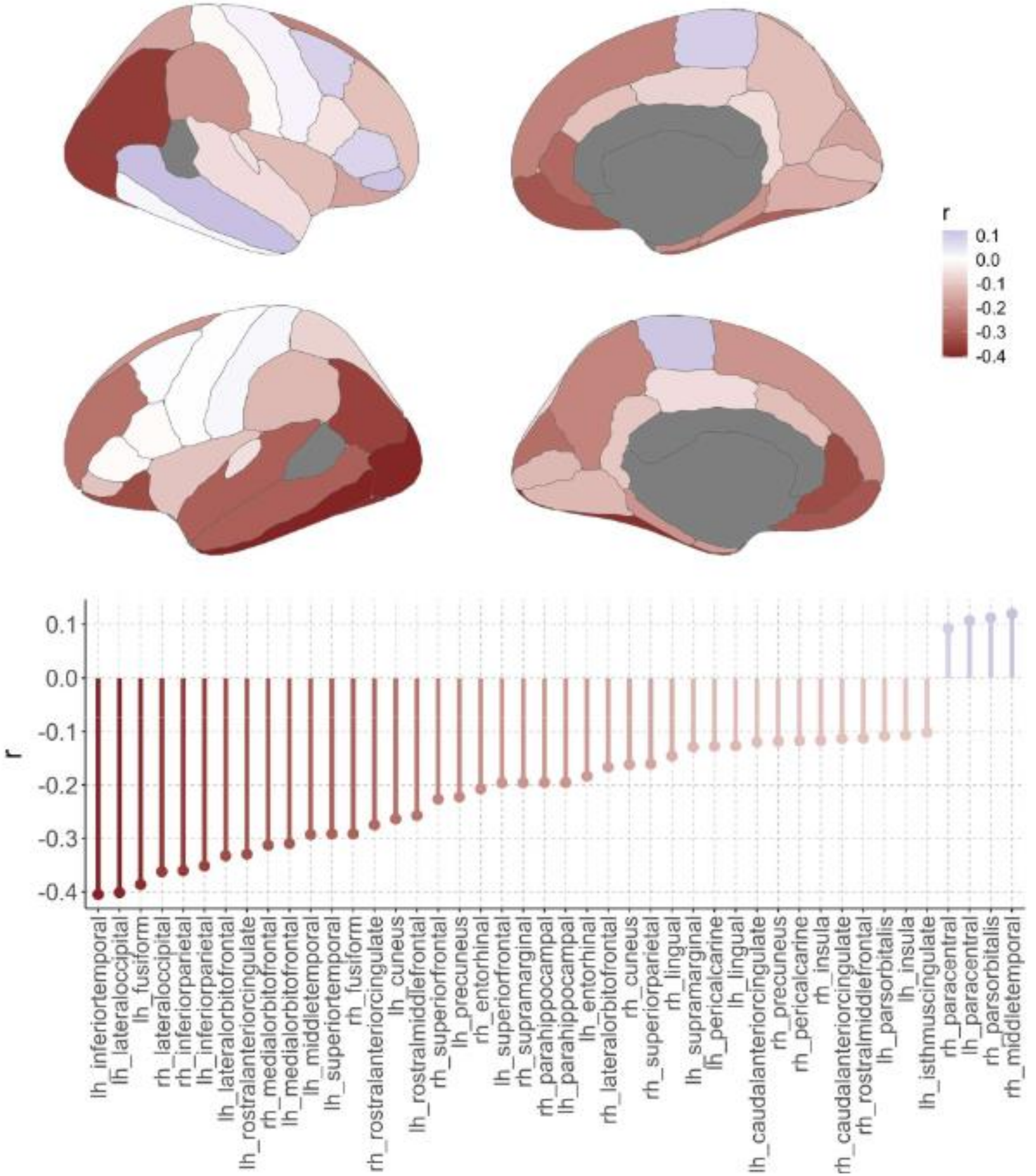
Results – ROI-based t-test

Region	Median R_2^* NC (sec^{-1})	Median R_2^* AD (sec^{-1})	p Value
Left basal ganglia	29.00 (27.24-32.66)	30.76 (28.15-33.60)	**0.006***
Left caudate nucleus	23.37 (21.43-25.70)	24.28 (22.54-26.71)	**0.004***
Left pallidum	36.52 (33.17-40.58)	36.43 (33.61-41.30)	**0.401**
Left putamen	27.48 (25.15-31.75)	30.11 (27.27-33.94)	**0.000***
Left hippocampus	16.40 (15.34-17.49)	15.88 (14.78-17.06)	**0.003***
Left thalamus	19.70 (18.77-20.89)	19.71 (18.85-20.92)	**0.972**
Right basal ganglia	29.21 (26.74-32.14)	30.62 (28.09-32.91)	**0.002***
Right caudate nucleus	23.06 (21.35-25.17)	24.19 (22.11-25.94)	**0.002***
Right pallidum	36.25 (32.86-40.61)	36.76 (33.68-41.54)	**0.094**
Right putamen	28.08 (25.07-31.76)	30.17 (27.08-34.11)	**0.000***
Right hippocampus	16.44 (15.54-17.38)	16.02 (14.88-17.19)	**0.008***
Right thalamus	19.97 (18.82-20.71)	19.76 (18.90-20.62)	**0.497**



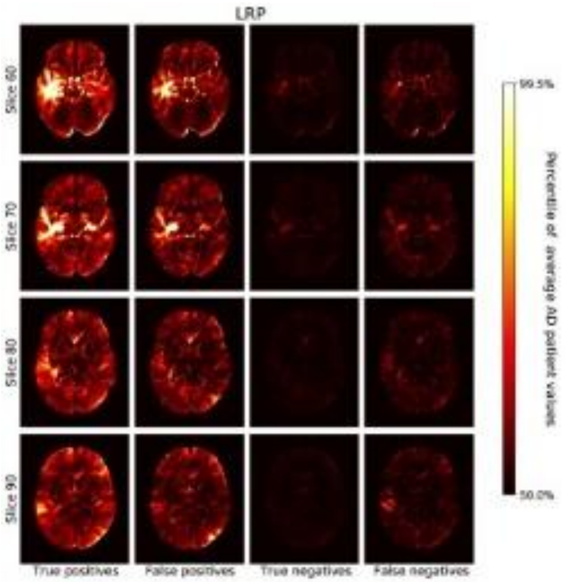
Related

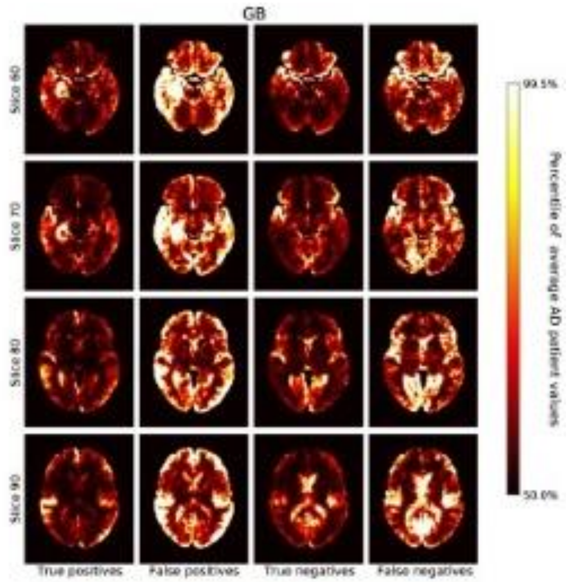


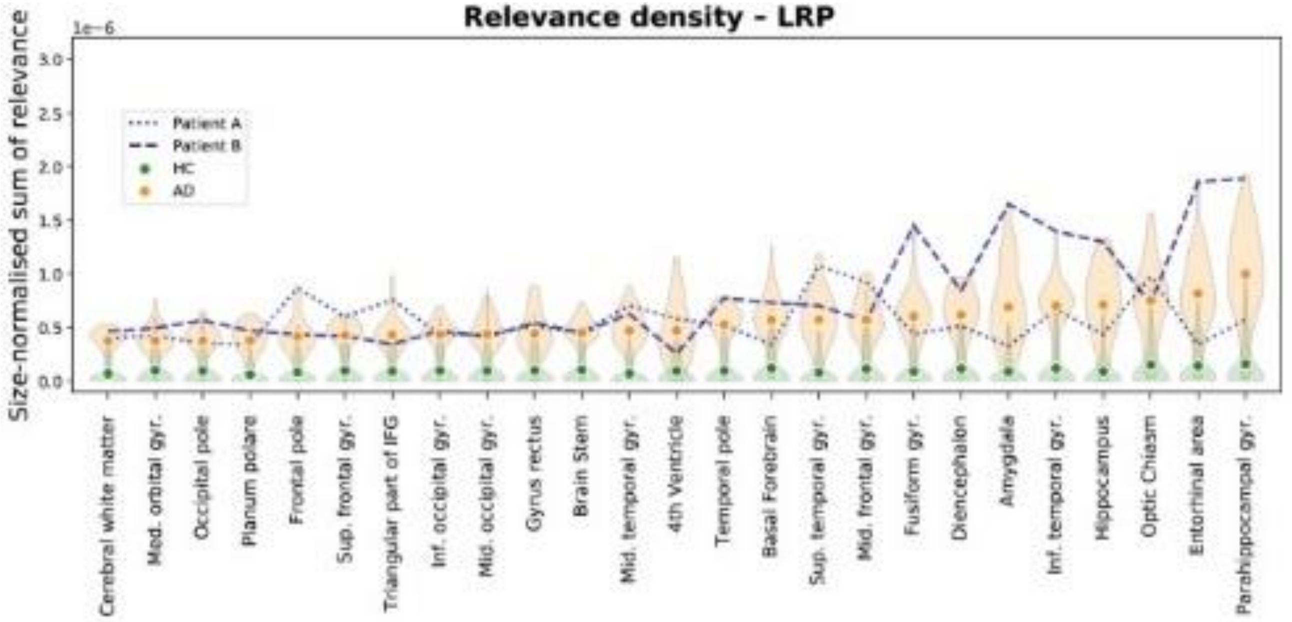


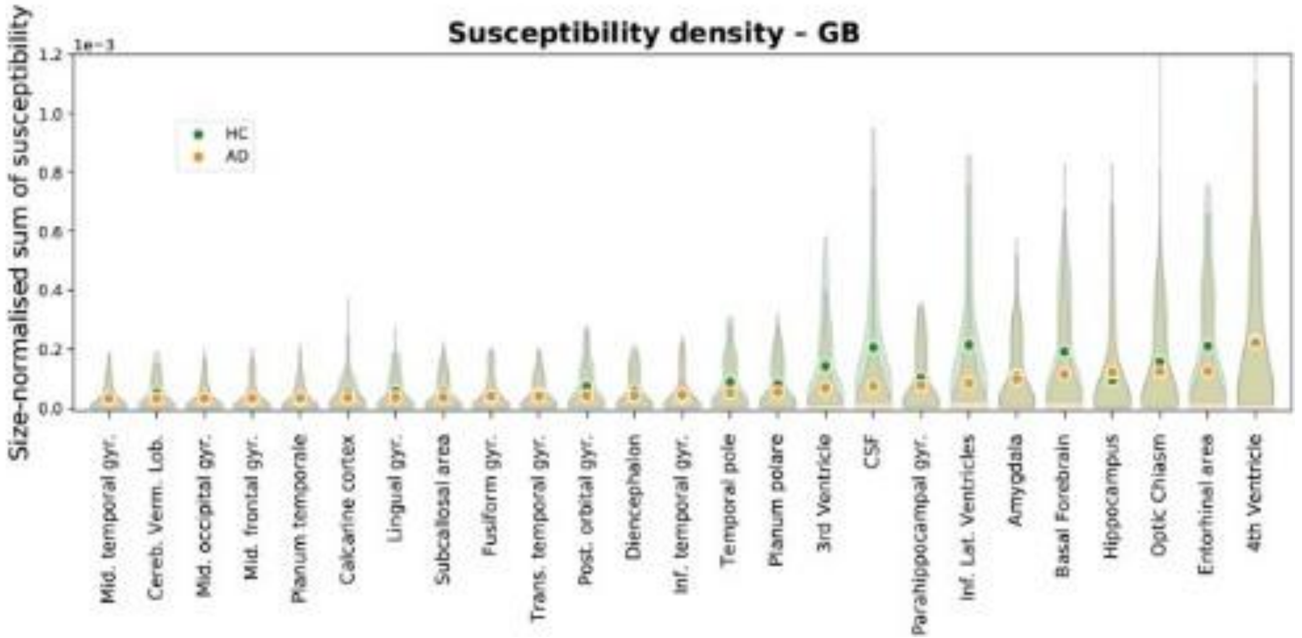
[1] Hofmann et al, Preprint bioRxiv, 2024

Related







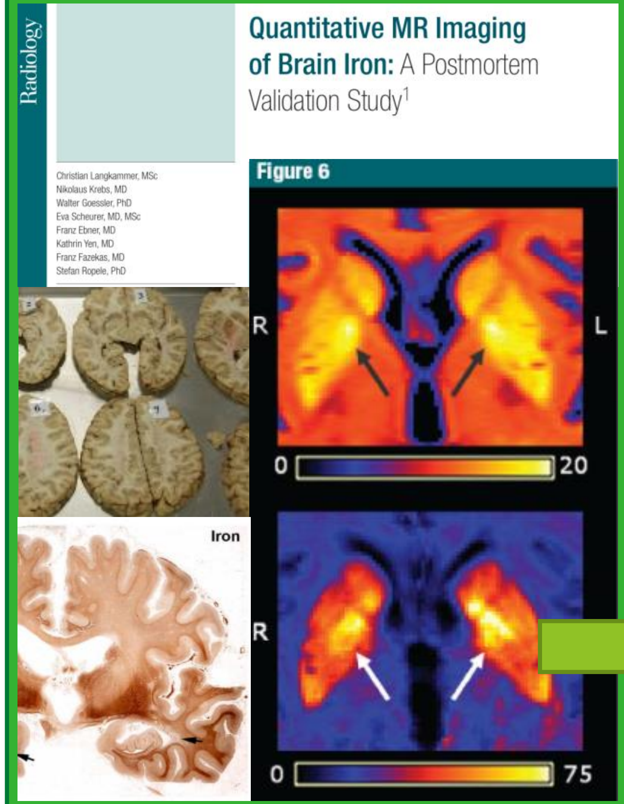


[1] Böhle et al, Frontiers in Aging Neuroscience, 2019

Discussion & Conclusion

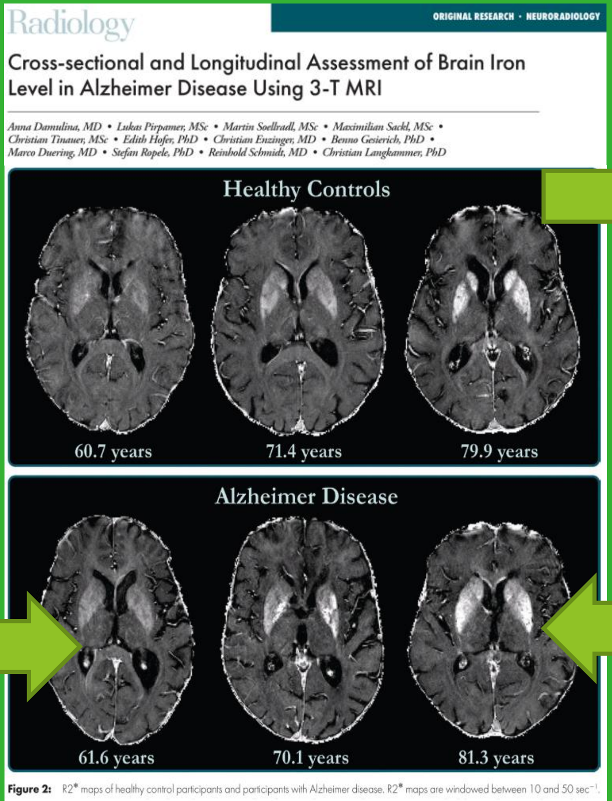
Medical University of Graz

Postmortem MRI



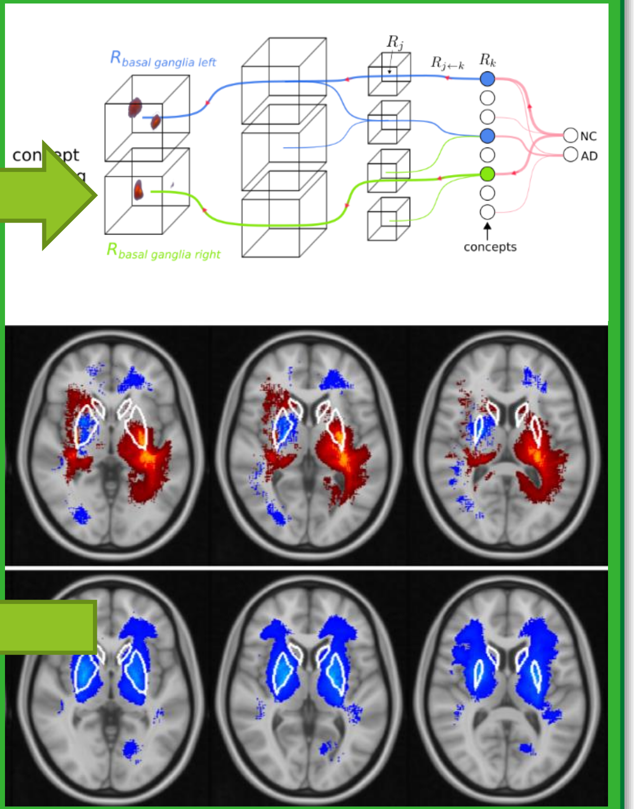
Langkammer, 2010, Radiology

Iron in AD



Damulina, 2021, Radiology

CNN based AD class



Tinauer, 2024, XAI-2024

Summary and take home message

- Data
- Convolutional Neural Nets
- Explainable AI (XAI)
- Debugging – Is there something off?
- Regularization – XAI during training
- Validation – Does (X)AI capture features I would expect?

Medical University of Graz

Thank you!

Christian Tinauer
Department of Neurology
Auenbruggerplatz 22
8036 Graz, Austria

christian.tinauer@medunigraz.at
www.neuroimaging.at
https://github.com/christiantinauer