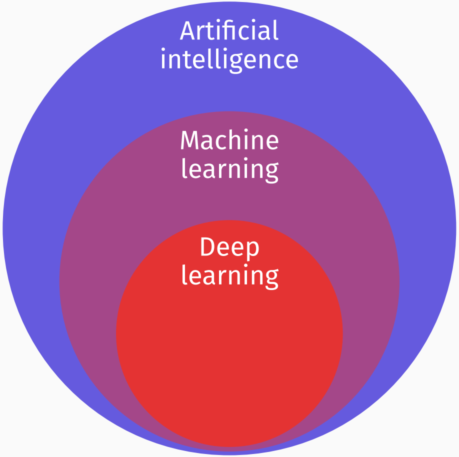
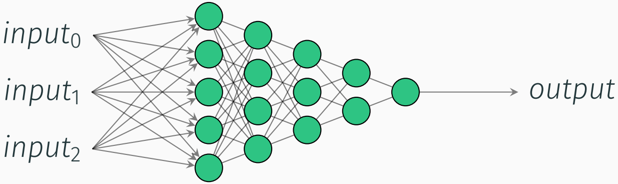
Enabling deep learning in small MRI datasets with transfer learning

UNIVERSITY OF OSLO

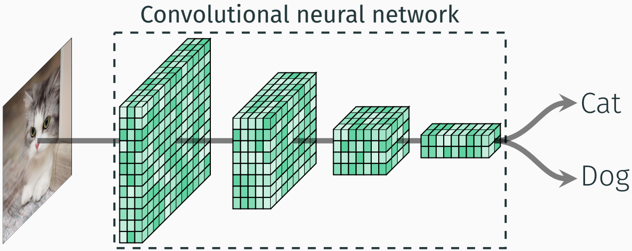
Deep learning on MRI data



Deep learning on MRI data



Deep learning on MRI data

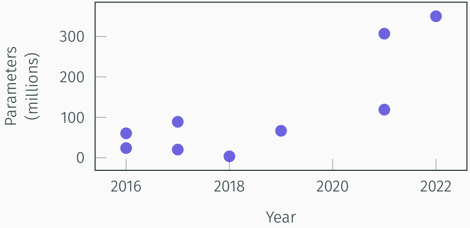


Deep learning on MRI data

Model	Year	Imagenet top-1 accuracy	Parameters
InceptionV3	2016	77.9%	23.9M
ResNet152V2	2016	78.0%	60.4M
DenseNet201	2017	77.3%	20.2M
NASNetLarge	2017	82.5%	88.9M
MobileNetV2	2018	71.3%	3.5M
EfficientNetB7	2019	84.3%	66.7M
EfficientNetV2L	2021	85.7%	119.0M
Vision Transformer L/16	2021	85.2%	307M
ConvNeXtXLarge	2022	86.3%	350.1M

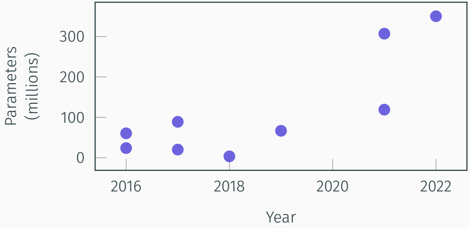
Deep learning on MRI data

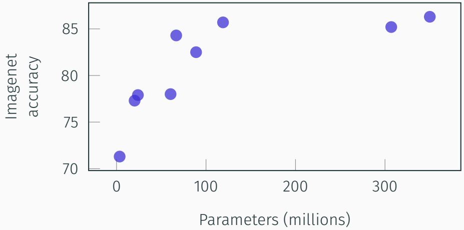
Model	Year	Imagenet top-1 accuracy	Parameters
InceptionV3	2016	77.9%	23.9M
ResNet152V2	2016	78.0%	60.4M
DenseNet201	2017	77.3%	20.2M
NASNetLarge	2017	82.5%	88.9M
MobileNetV2	2018	71.3%	3.5M
EfficientNetB7	2019	84.3%	66.7M
EfficientNetV2L	2021	85.7%	119.0M
Vision Transformer L/16	2021	85.2%	307M
ConvNeXtXLarge	2022	86.3%	350.1M



Deep learning on MRI data

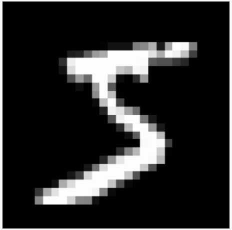
Model	Year	Imagenet top-1 accuracy	Parameters
InceptionV3	2016	77.9%	23.9M
ResNet152V2	2016	78.0%	60.4M
DenseNet201	2017	77.3%	20.2M
NASNetLarge	2017	82.5%	88.9M
MobileNetV2	2018	71.3%	3.5M
EfficientNetB7	2019	84.3%	66.7M
EfficientNetV2L	2021	85.7%	119.0M
Vision Transformer L/16	2021	85.2%	307M
ConvNeXtXLarge	2022	86.3%	350.1M





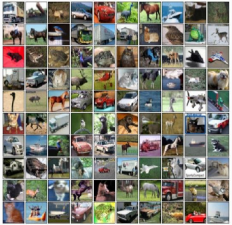
Deep learning on MRI data

Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels



Deep learning on MRI data

Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels
CIFAR	∼60K	32x32 pixels



Deep learning on MRI data

Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels
CIFAR	∼60K	32x32 pixels
COCO	∼330K	640x640 pixels
ImageNet (21k)	∼14M	224x224 pixels

Deep learning on MRI data

Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels
CIFAR	∼60K	32x32 pixels
COCO	∼330K	640x640 pixels
ImageNet (21k)	∼14M	224x224 pixels
CLIP	∼400M	224x224 pixels
JFT-3B	∼3B	256x256 pixels
LAION-5B	∼2.3B	336x336 pixels

Deep learning on MRI data

Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels
CIFAR	∼60K	32x32 pixels
COCO	∼330K	640x640 pixels
ImageNet (21k)	∼14M	224x224 pixels
CLIP	∼400M	224x224 pixels
JFT-3B	∼3B	256x256 pixels
LAION-5B	∼2.3B	336x336 pixels

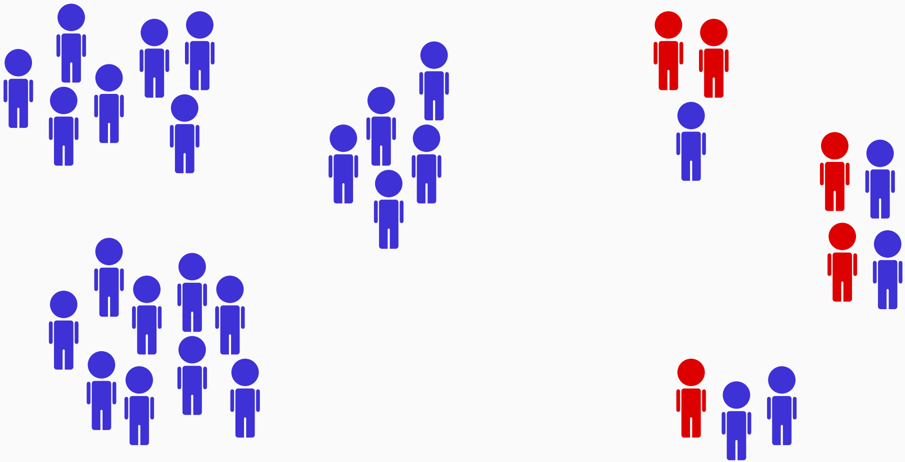
UK Biobank	45,891	256x256x256 voxels
ABCD	11,660	256x256x256 voxels
ADNI	2,729	256x256x256 voxels
TOP	2,222	256x256x256 voxels
HBN	1,365	256x256x256 voxels
QTIM	1,201	256x256x256 voxels
PING	1,176	256x256x256 voxels
CoRR	1,156	256x256x256 voxels
HCP	1,113	256x256x256 voxels
OASIS3	1,062	256x256x256 voxels

Deep learning on MRI data

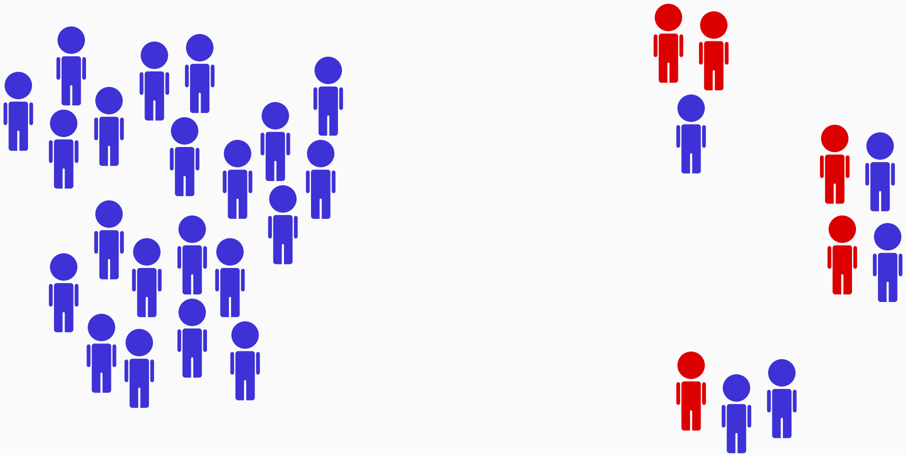
Dataset	Size	Dimensionality
MNIST	∼70K	28x28 pixels
CIFAR	∼60K	32x32 pixels
COCO	∼330K	640x640 pixels
ImageNet (21k)	∼14M	224x224 pixels
CLIP	∼400M	224x224 pixels
JFT-3B	∼3B	256x256 pixels
LAION-5B	∼2.3B	336x336 pixels

UK Biobank	45,891	256x256x256 voxels
ABCD	11,660	256x256x256 voxels
ADNI	2,729	256x256x256 voxels
TOP	2,222	256x256x256 voxels
HBN	1,365	256x256x256 voxels
QTIM	1,201	256x256x256 voxels
PING	1,176	256x256x256 voxels
CoRR	1,156	256x256x256 voxels
HCP	1,113	256x256x256 voxels
OASIS3	1,062	256x256x256 voxels

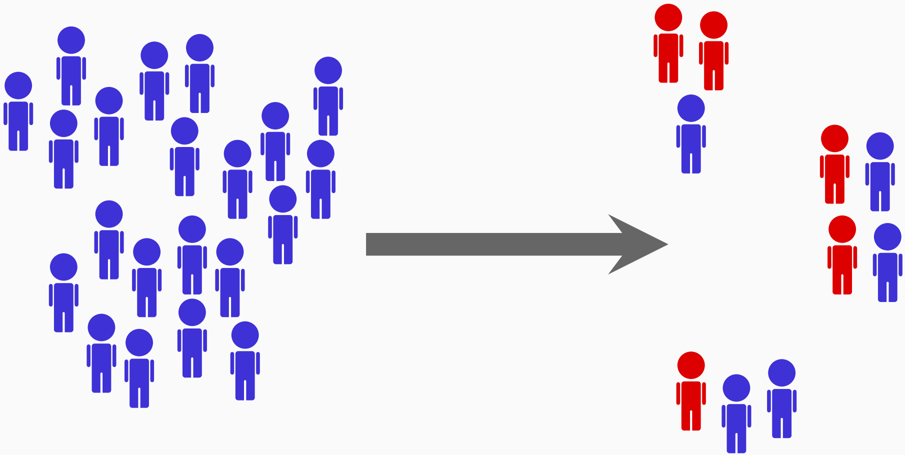
Deep learning on MRI data



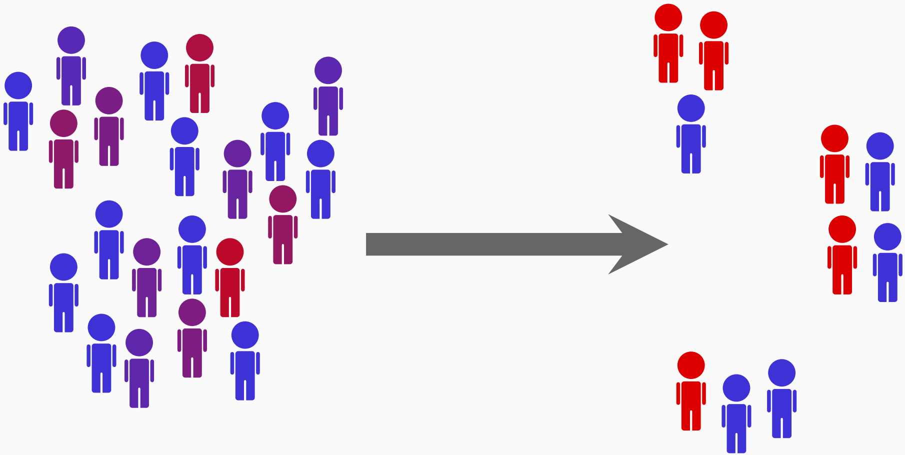
Deep learning on MRI data



Deep learning on MRI data

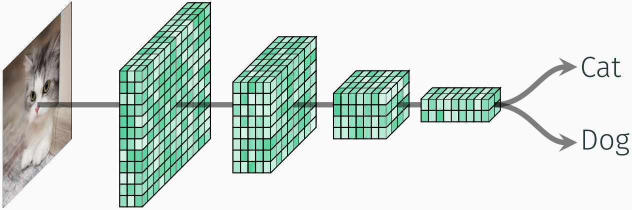


Deep learning on MRI data

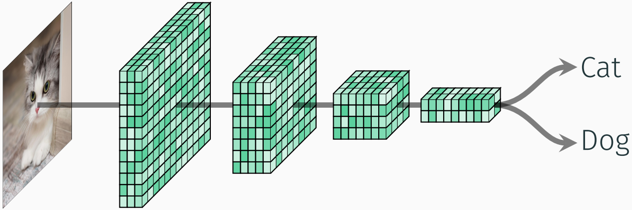


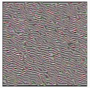
Transfer learning

How does CNNs work

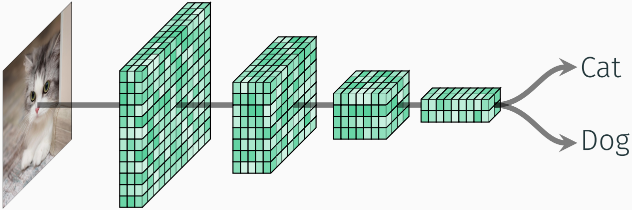


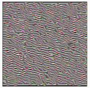
How does CNNs work



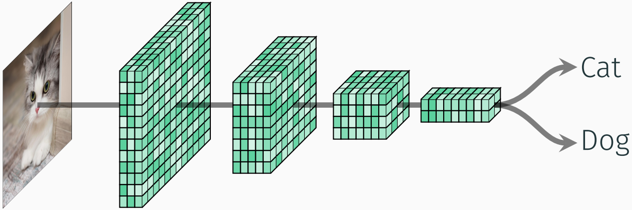


How does CNNs work

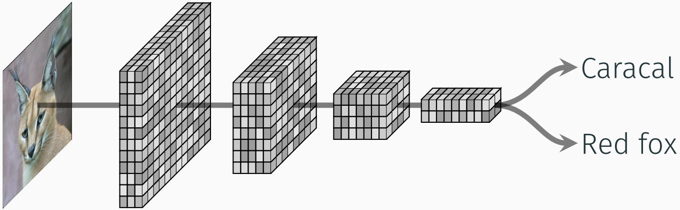




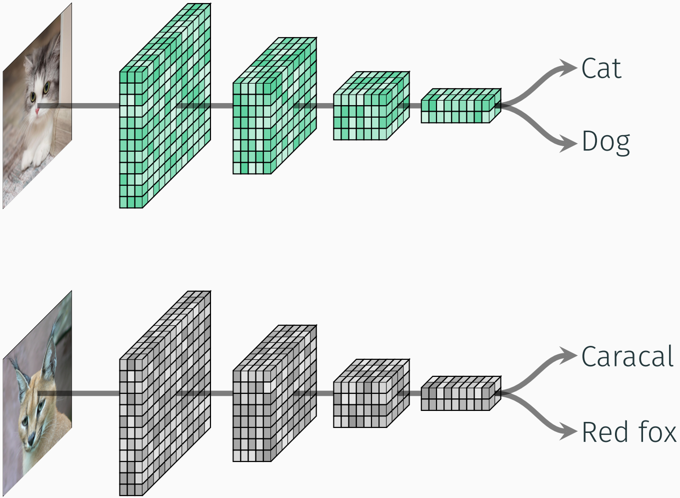
How does CNNs work



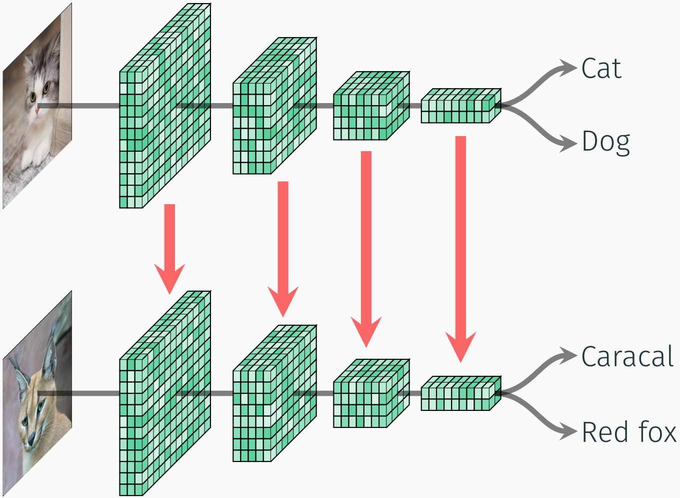
Transfer learning



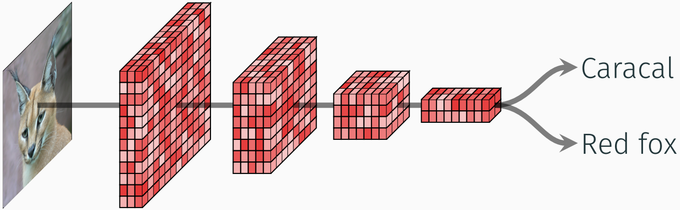
Transfer learning



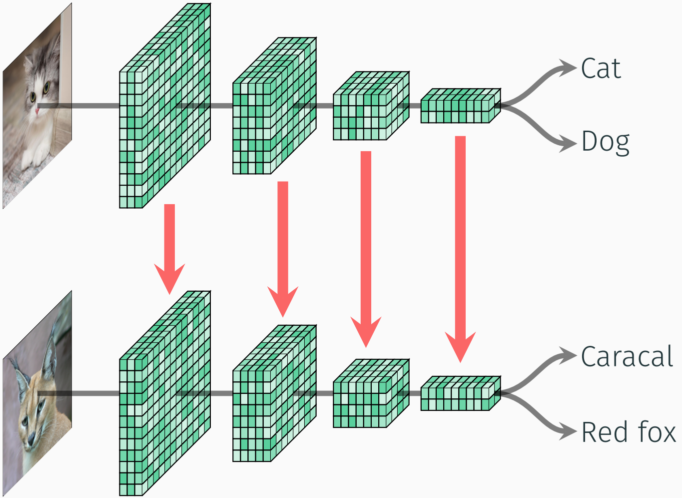
Transfer learning



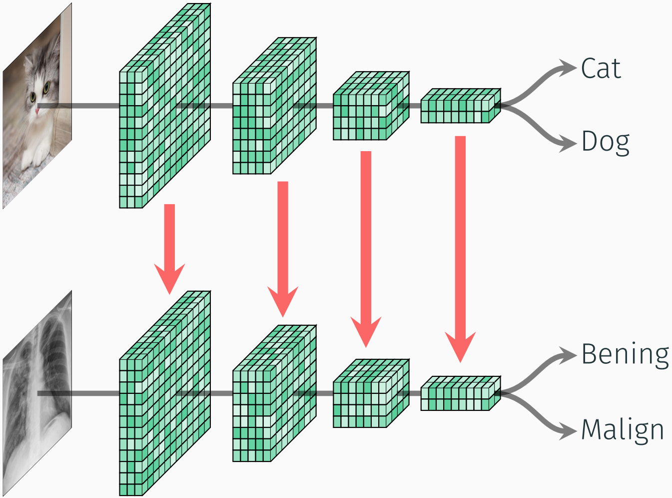
Transfer learning



Transfer learning

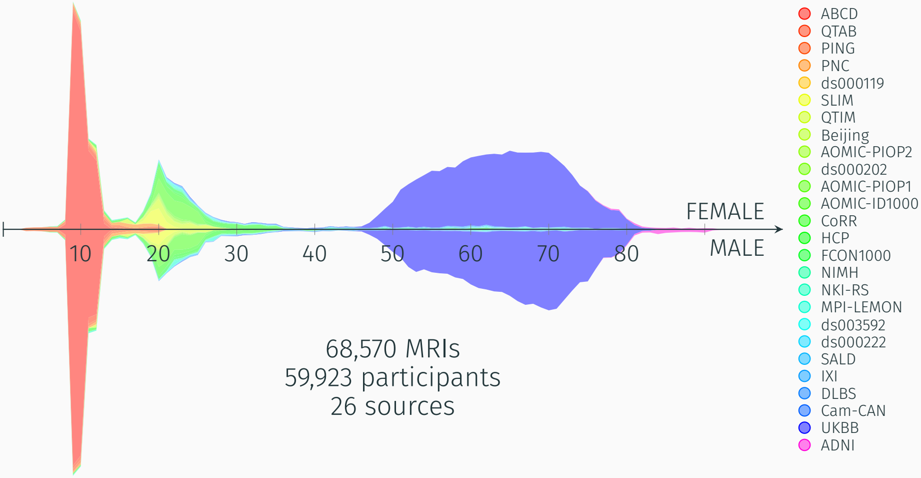


Transfer learning



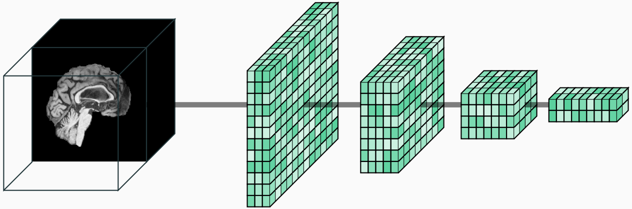
Our work

Methods: Training dataset

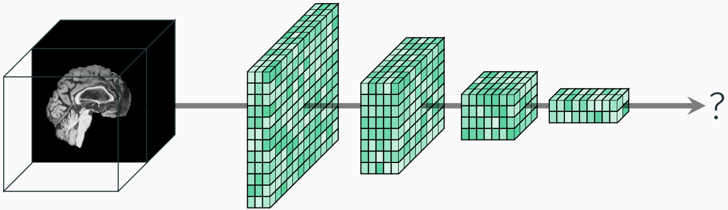


68,570 MRIs
59,923 participants
26 sources

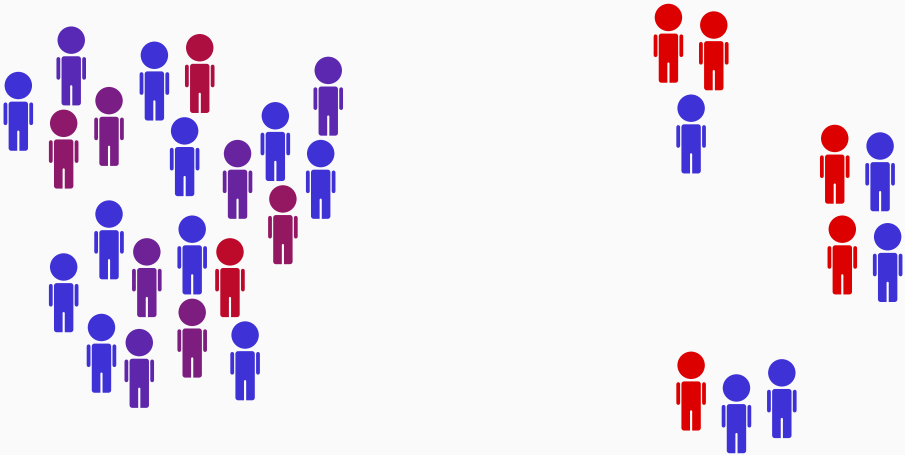
Methods: Modelling



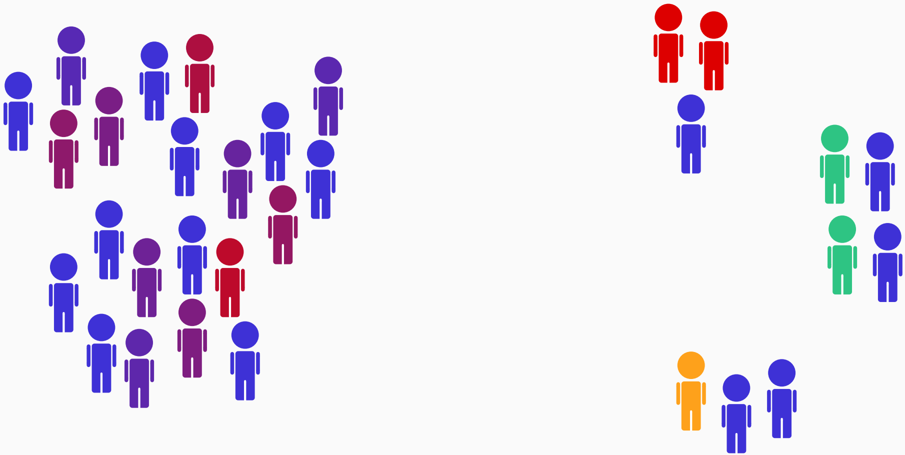
Methods: Modelling



Methods: Modelling



Methods: Modelling

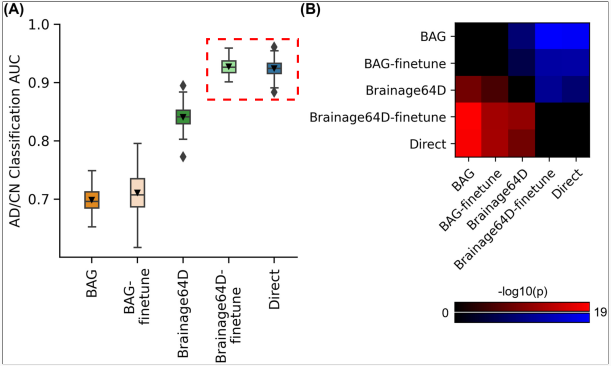


Methods: Modelling

> Furthermore, we see this result as evidence that deep learning models trained to predict age in large multisite datasets constitute excellent starting points for transfer learning, which can subsequently be fine-tuned to a variety of tasks.

Leonardsen, E. H., Peng, H., Kaufmann, T., Agartz, I., Andreassen, O. A., Celius, E. G., ... & Wang, Y. (2022). Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage*, *256*, 119210.

Methods: Modelling



Tan, T. W. K., Nguyen, K. N., Zhang, C., Kong, R., Cheng, S. F., Ji, F., ... & Australian Imaging Biomarkers and Lifestyle Study of Aging. (2025). Mind the gap: Does brain age improve Alzheimer's disease prediction?. *Human Brain Mapping*, *46*(12), e70276.

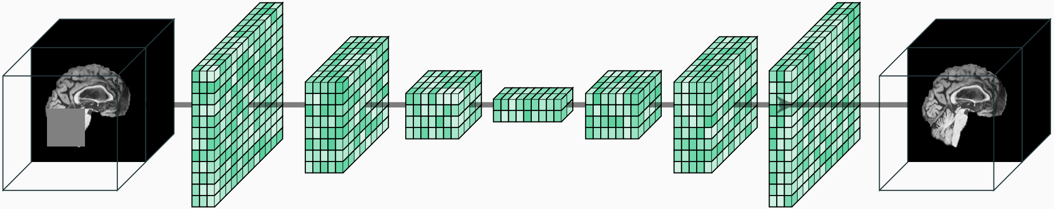
Methods: Modelling

Brain age

Brain age

Self-supervised approaches

Methods: Modelling

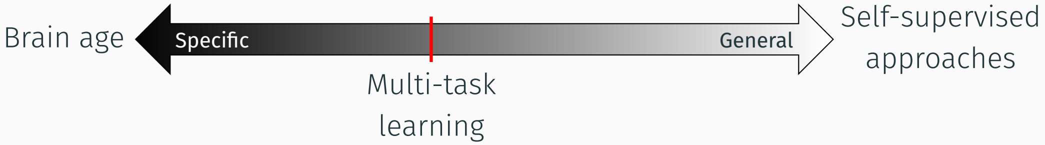


Methods: Modelling

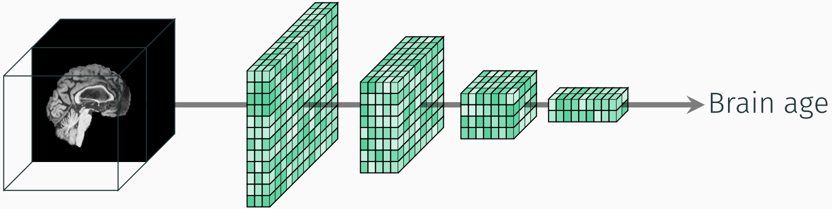
Brain age

Self-supervised approaches

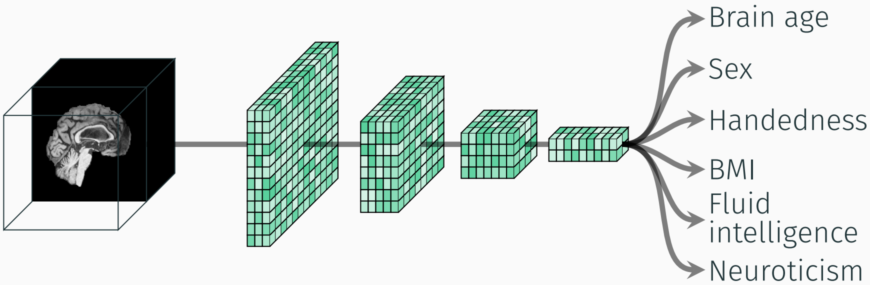
Methods: Modelling



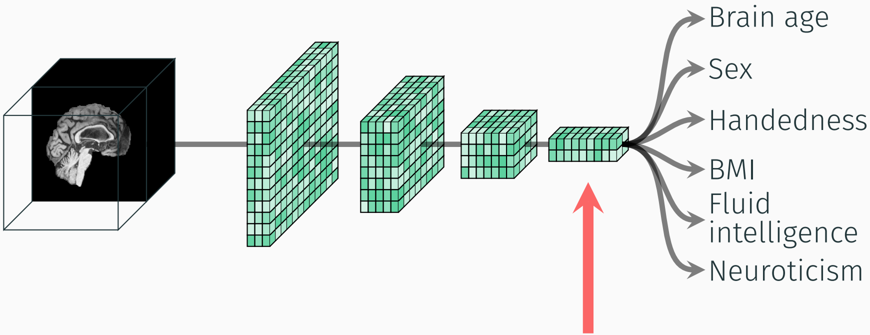
Methods: Multi-task learning



Methods: Multi-task learning



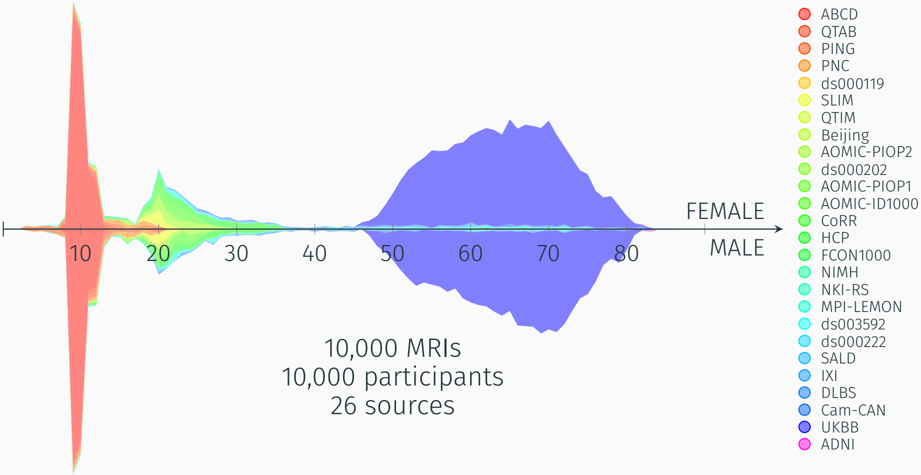
Methods: Multi-task learning



Methods: Multi-task learning

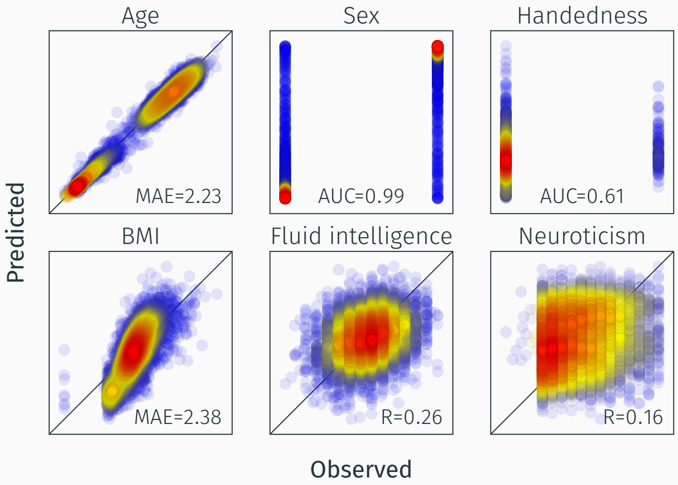
	Age	Sex	Handedness	BMI	Fluid intelligence	Neuroticism
Age	1	0	0.02	0.52	-0.1	-0.04
Sex	0	1	0.03	0.04	-0.15	0.04
Handedness	-0.02	0.03	1	0	0.03	-0.03
BMI	0.52	0.04	0	1	-0.03	-0.05
Fluid intelligence	-0.1	-0.15	0.03	-0.03	1	-0.04
Neuroticism	-0.04	0.04	-0.03	-0.05	-0.04	1

Results: Validation

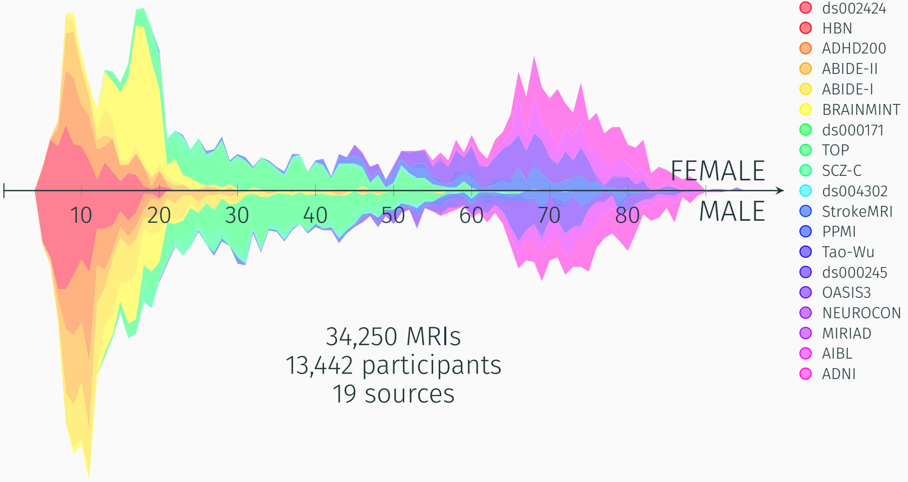


10,000 MRIs
10,000 participants
26 sources

Results: Validation

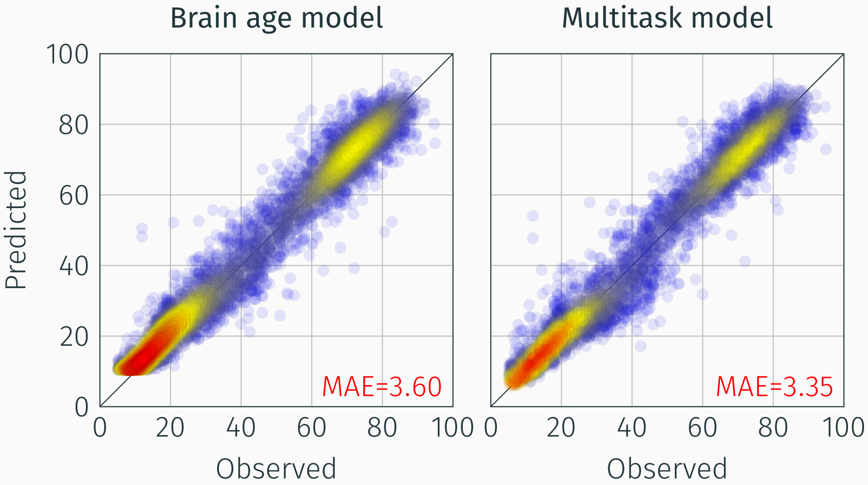


Results: Brain age

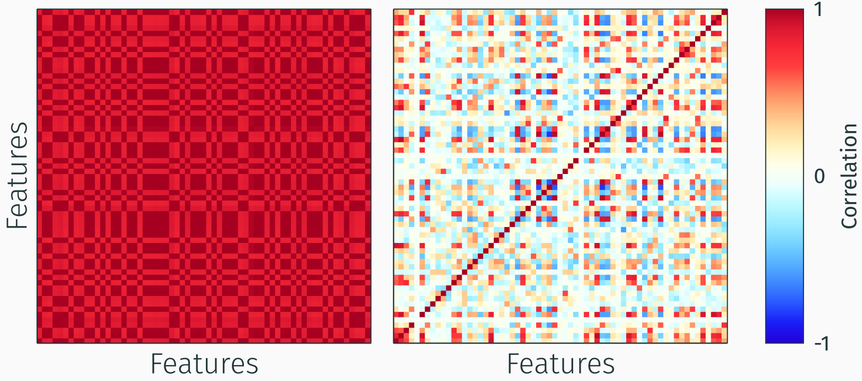


34,250 MRIs
13,442 participants
19 sources

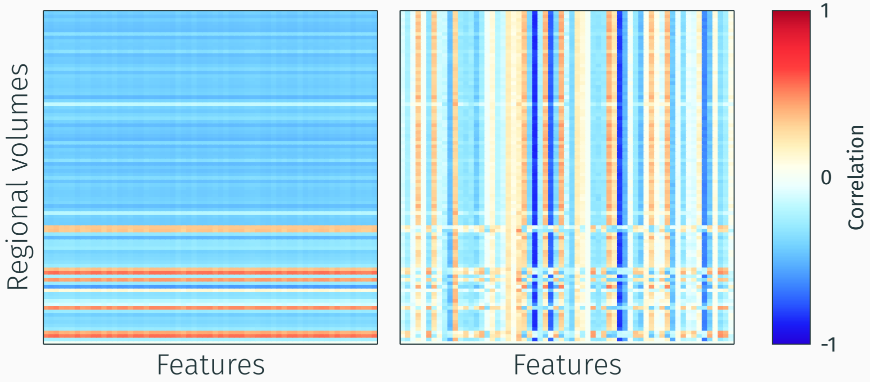
Results: Brain age



Results: Feature spaces



Results: Feature spaces



Transfer learning: Motivation

Will transfer learning from the pretrained multi-task model result in better predictive performance in downstream prediction tasks relying on small datasets for training?

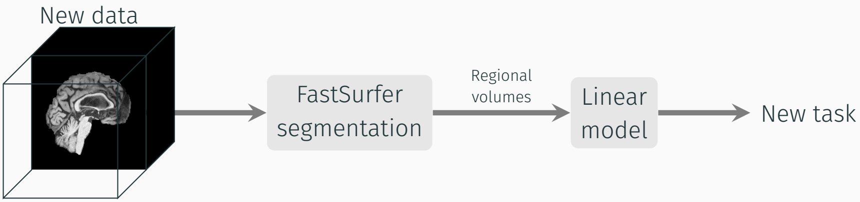
Transfer learning: Prediction tasks

Four prediction tasks:

1. Heterogeneous brain age prediction
   - 6,750 images from 20 sources
2. Homogeneous brain age prediction
   - 1,072 images from TOP
3. Classification of patients with Alzheimer's disease and healthy controls
   - 10,660 images from ADNI
4. Classification of patients with schizophrenia and bipolar disorder
   - 1,037 images from TOP

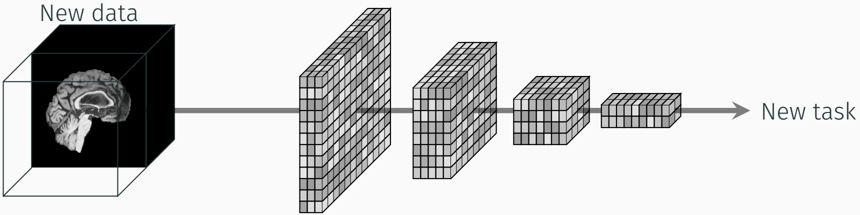
Transfer learning: Modelling approaches

Linear model | Baseline | Direct transfer | Finetuned multi-task model | Finetuned brain age model | Partially fine-tuned multi-task model

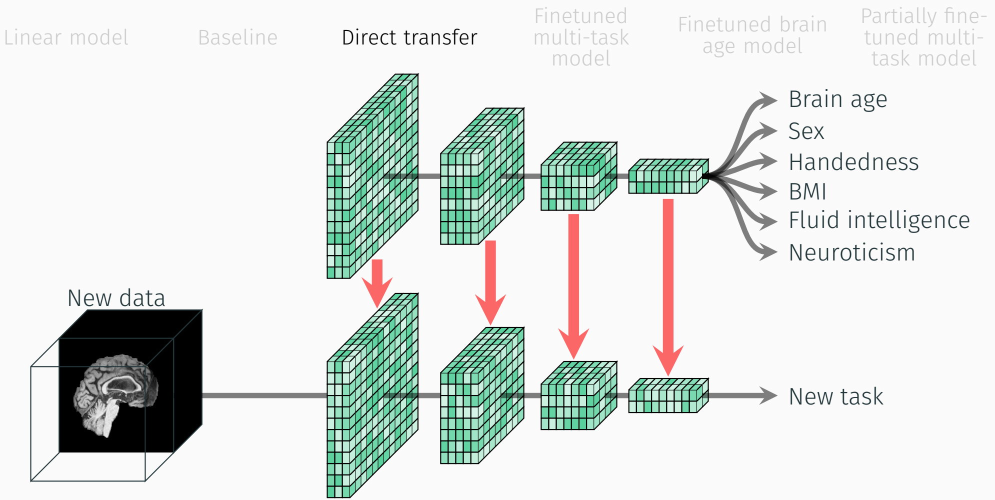


Transfer learning: Modelling approaches

Linear model | Baseline | Direct transfer | Finetuned multi-task model | Finetuned brain age model | Partially fine-tuned multi-task model



Transfer learning: Modelling approaches



Transfer learning: Modelling approaches

Linear model

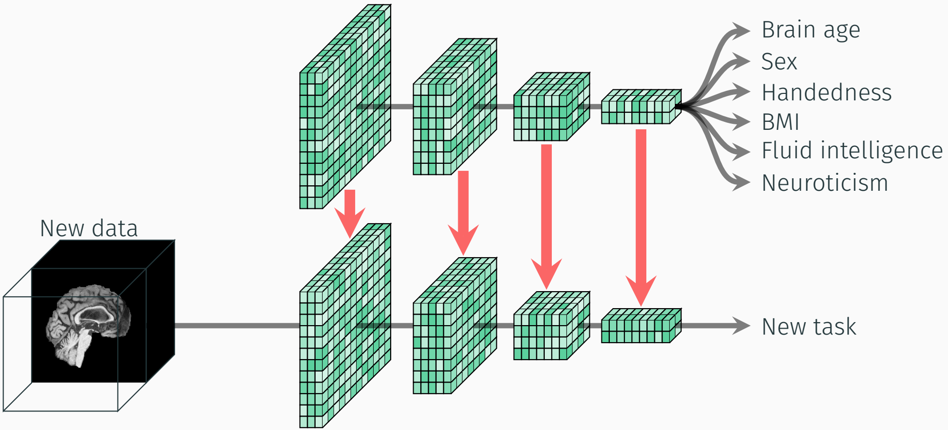
Baseline

Direct transfer

Finetuned multi-task model

Finetuned brain age model

Partially fine-tuned multi-task model



Transfer learning: Modelling approaches

Linear model

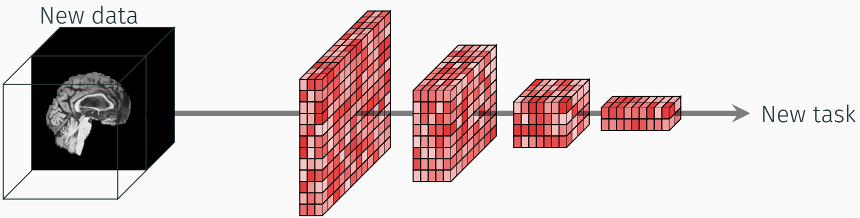
Baseline

Direct transfer

Finetuned multi-task model

Finetuned brain age model

Partially fine-tuned multi-task model



Transfer learning: Modelling approaches

Linear model

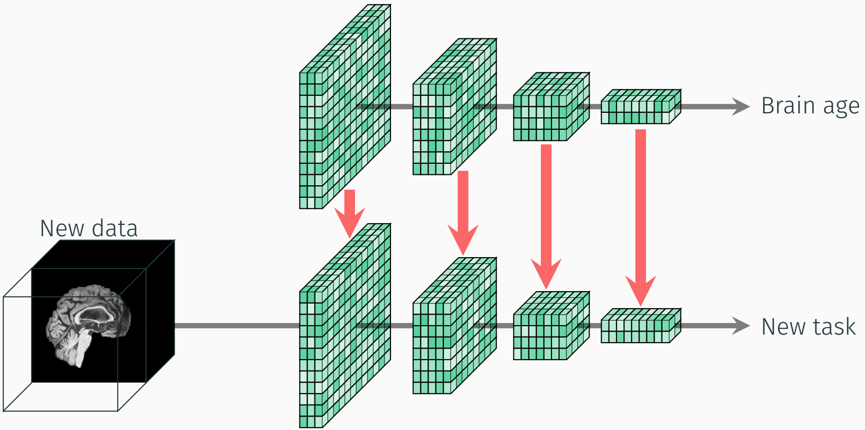
Baseline

Direct transfer

Finetuned multi-task model

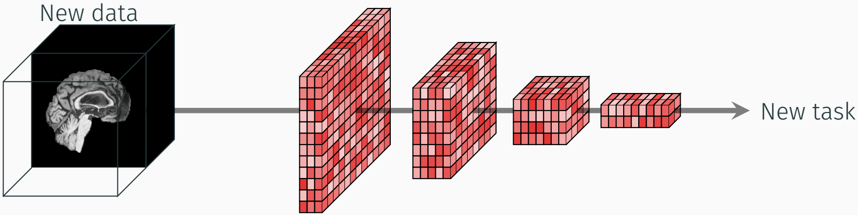
Finetuned brain age model

Partially fine-tuned multi-task model

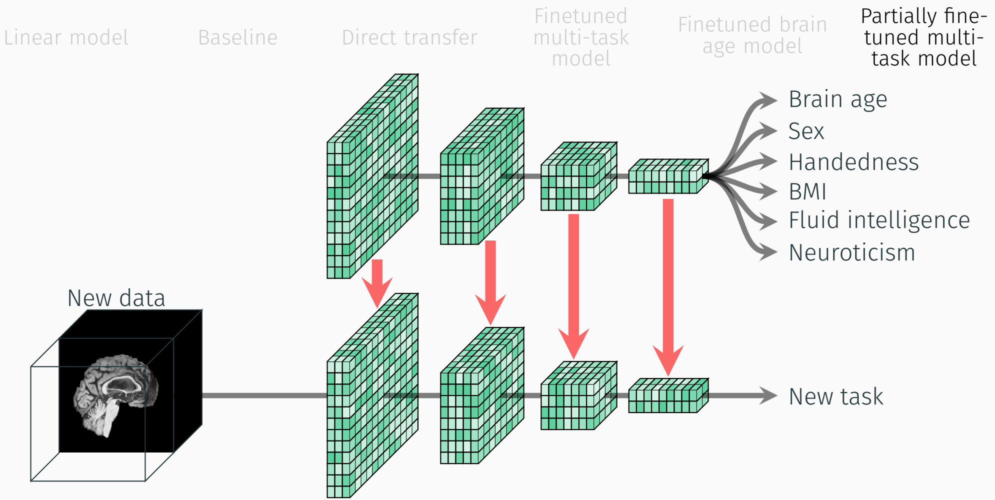


Transfer learning: Modelling approaches

Linear model | Baseline | Direct transfer | Finetuned multi-task model | Finetuned brain age model | Partially fine-tuned multi-task model

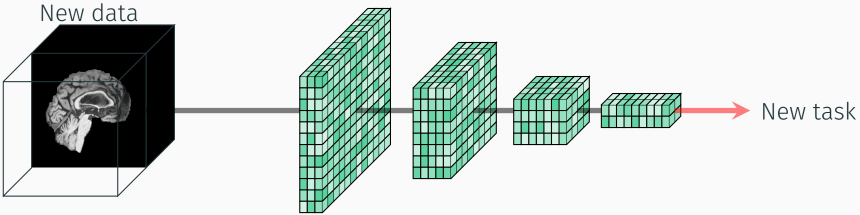


Transfer learning: Modelling approaches

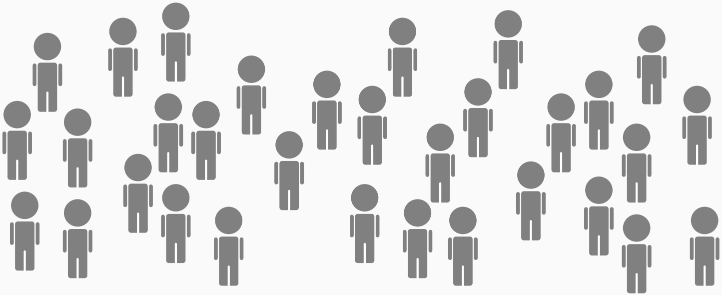


Transfer learning: Modelling approaches

Linear model | Baseline | Direct transfer | Finetuned multi-task model | Finetuned brain age model | Partially fine-tuned multi-task model

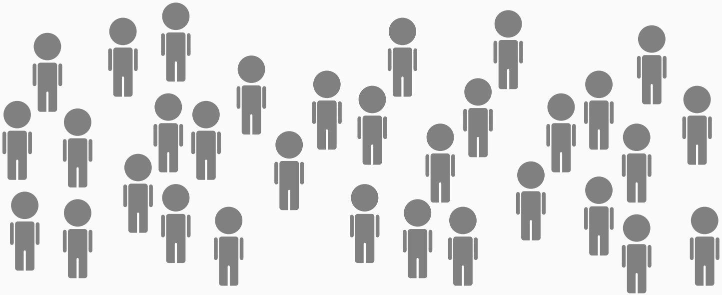


Transfer learning: Sampling

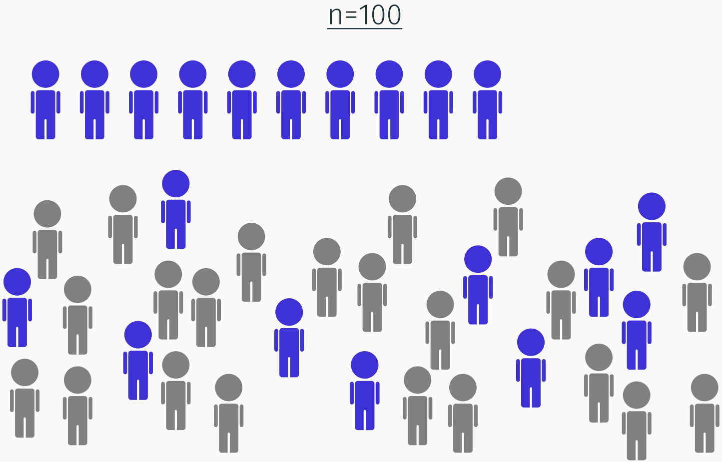


Transfer learning: Sampling

n=100

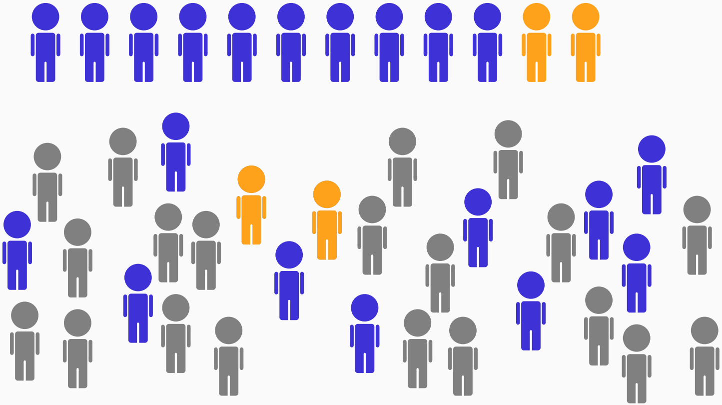


Transfer learning: Sampling



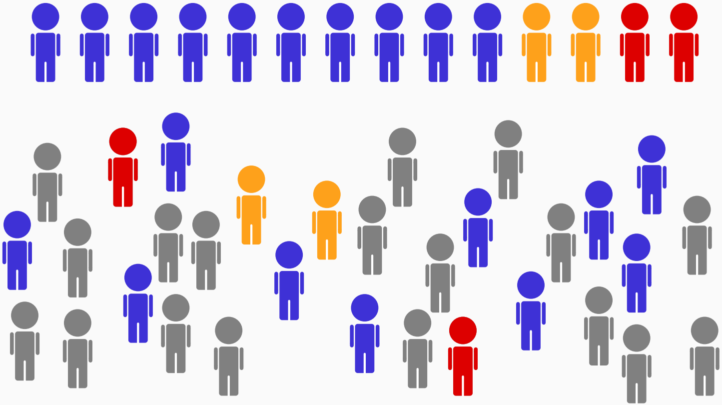
Transfer learning: Sampling

n=100

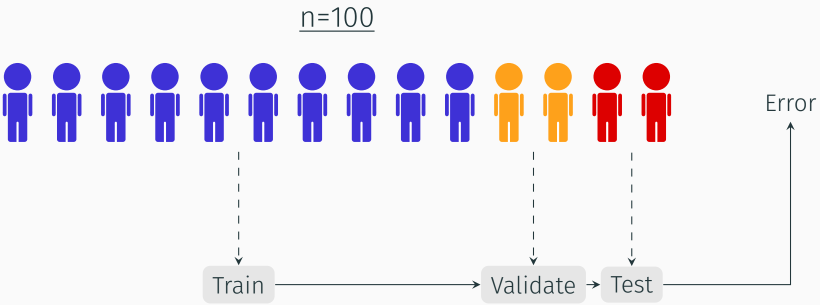


Transfer learning: Sampling

n=100

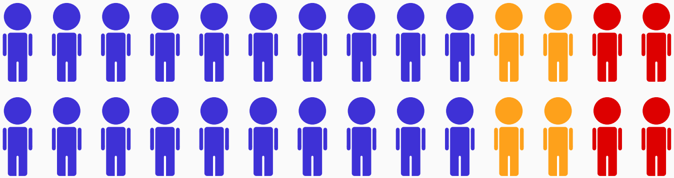


Transfer learning: Sampling



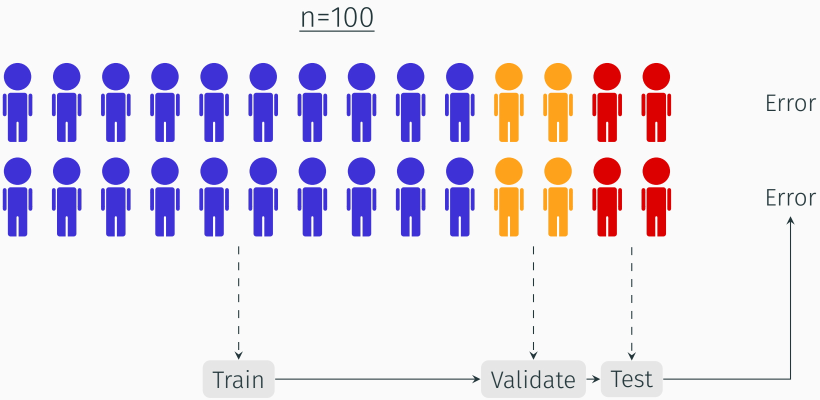
Transfer learning: Sampling

n=100

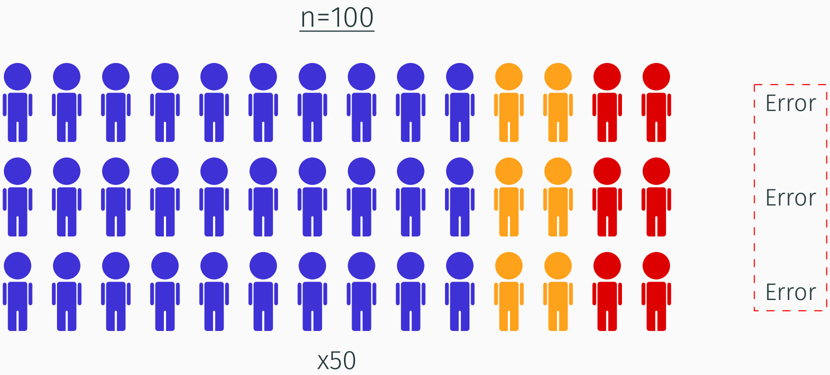


Error

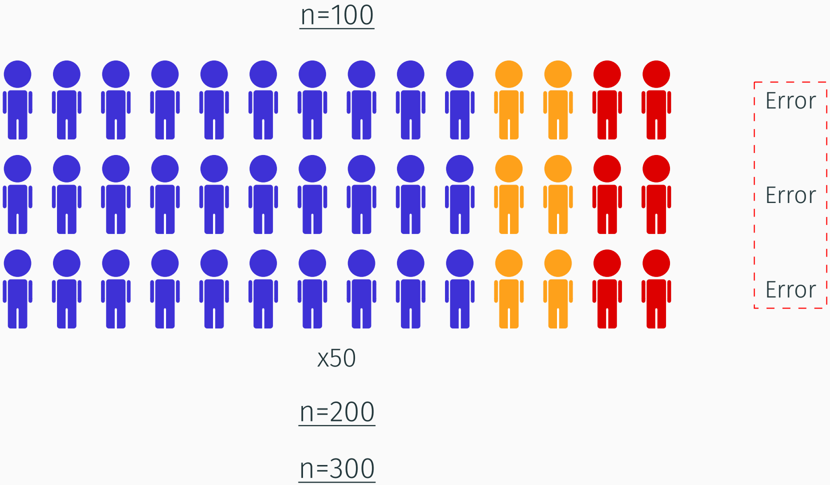
Transfer learning: Sampling



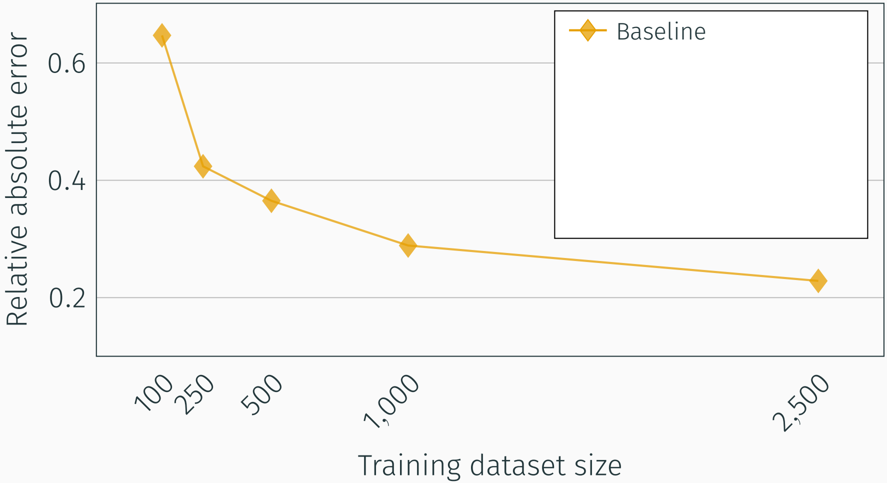
Transfer learning: Sampling



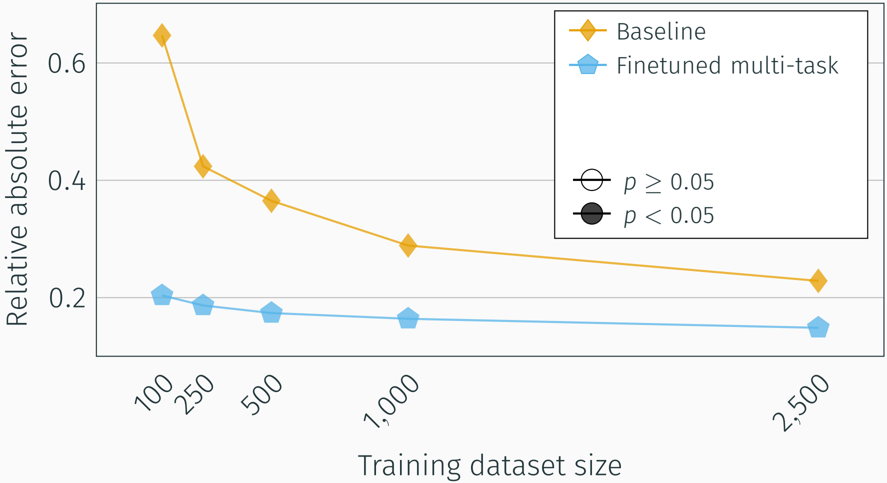
Transfer learning: Sampling



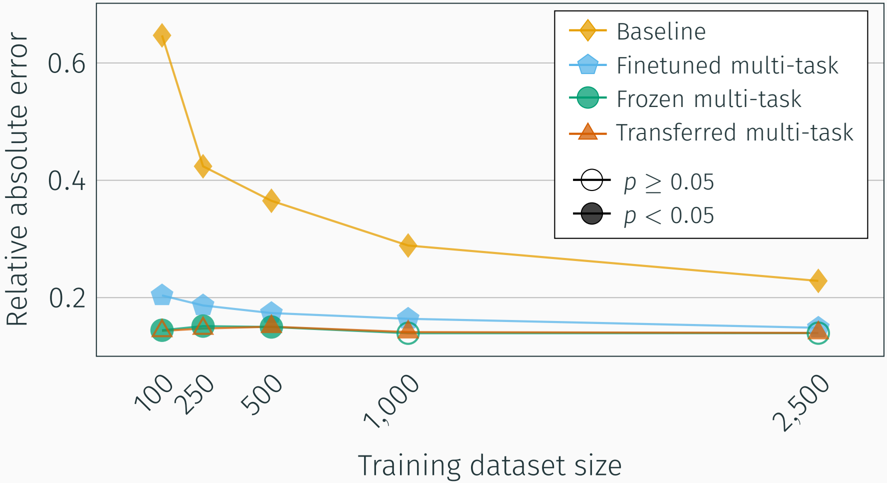
Results: Heterogeneous brain age predictions



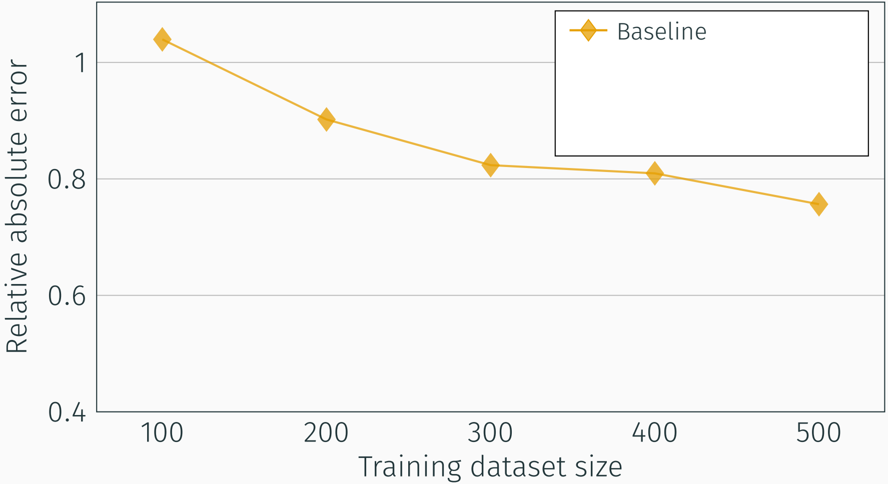
Results: Heterogeneous brain age predictions



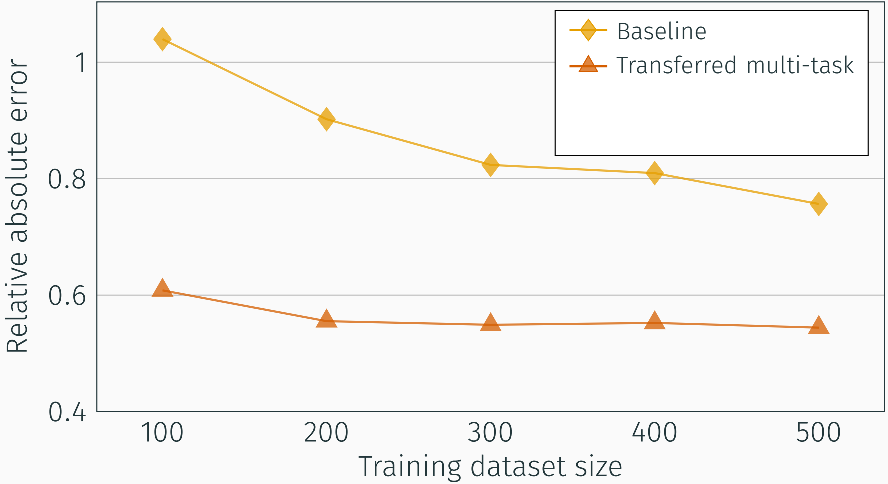
Results: Heterogeneous brain age predictions



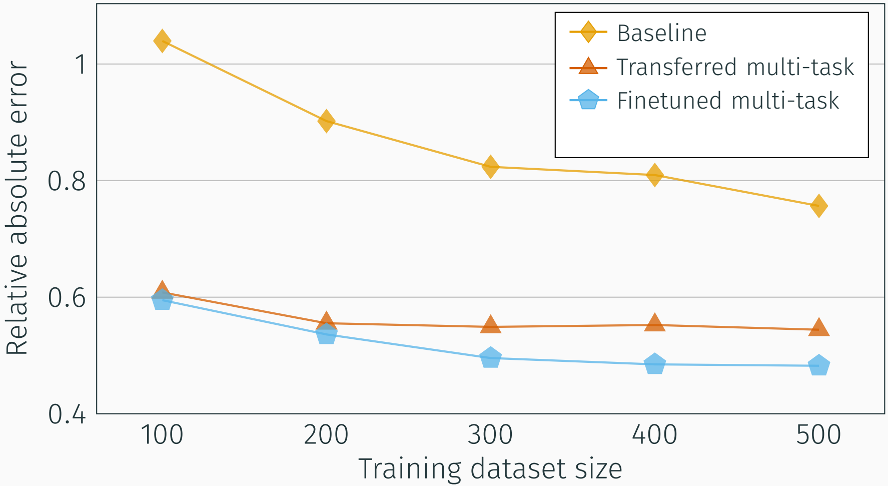
Results: Homogeneous brain age predictions



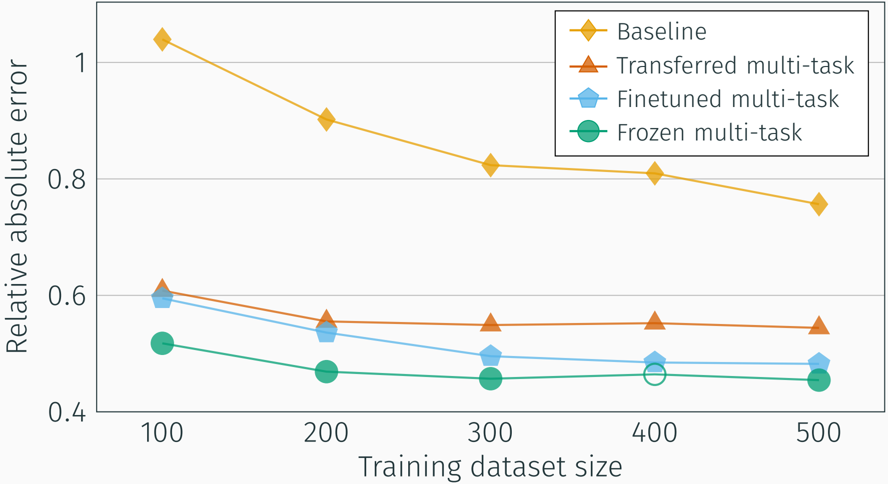
Results: Homogeneous brain age predictions



Results: Homogeneous brain age predictions



Results: Homogeneous brain age predictions

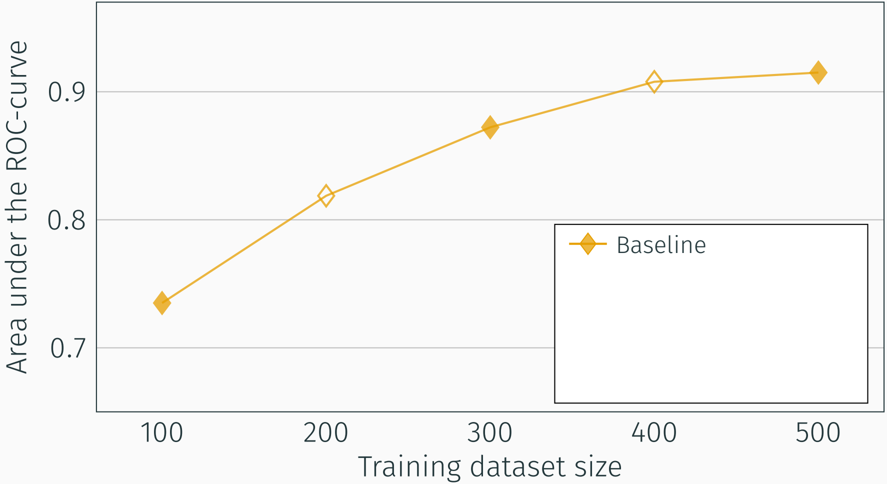


Results: Brain age

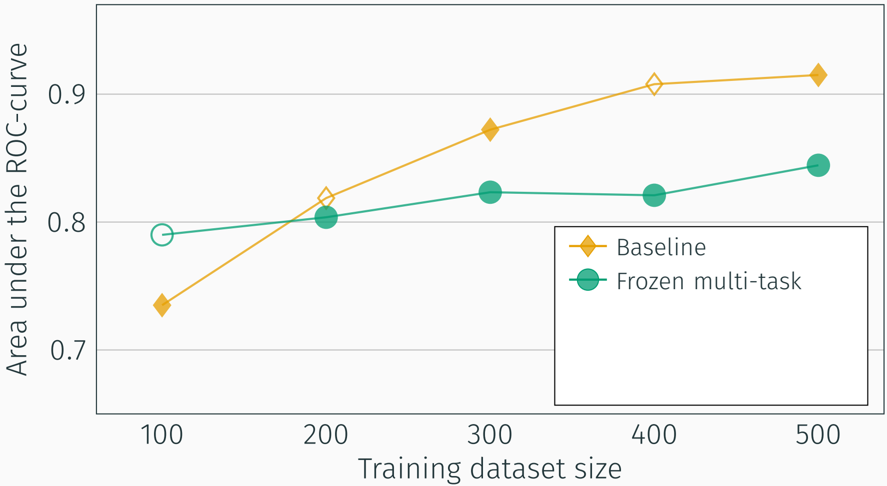
If you want to accurately predict brain age in new samples:

- You should *always* use a pretrained model
- In smaller, homogeneous, samples the model should ideally be finetuned, but using a pretrained model off-the-shelf works reasonably well

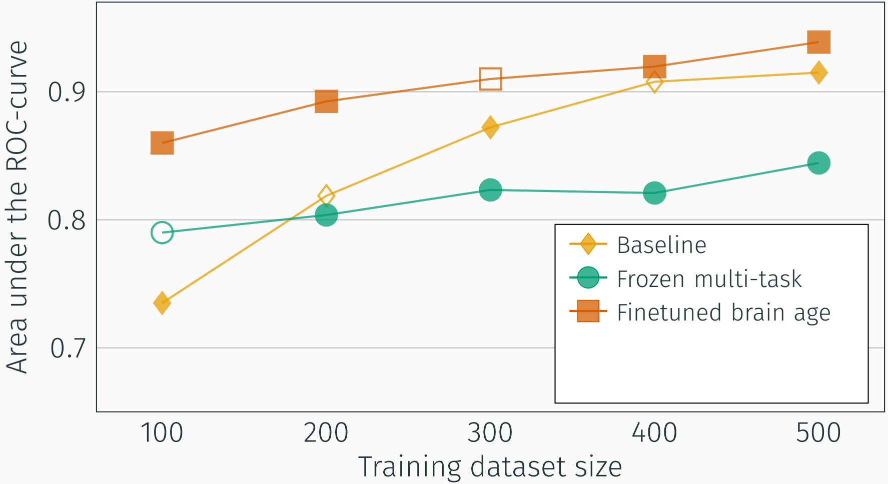
Results: AD vs HC classification



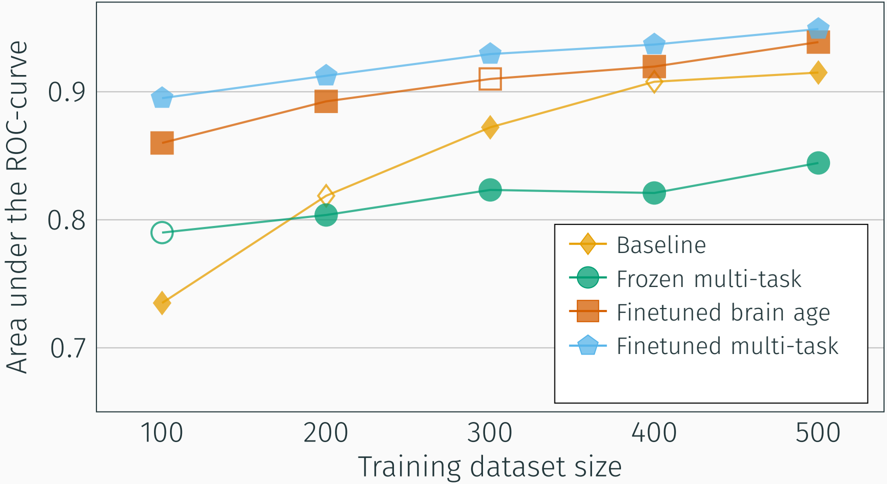
Results: AD vs HC classification



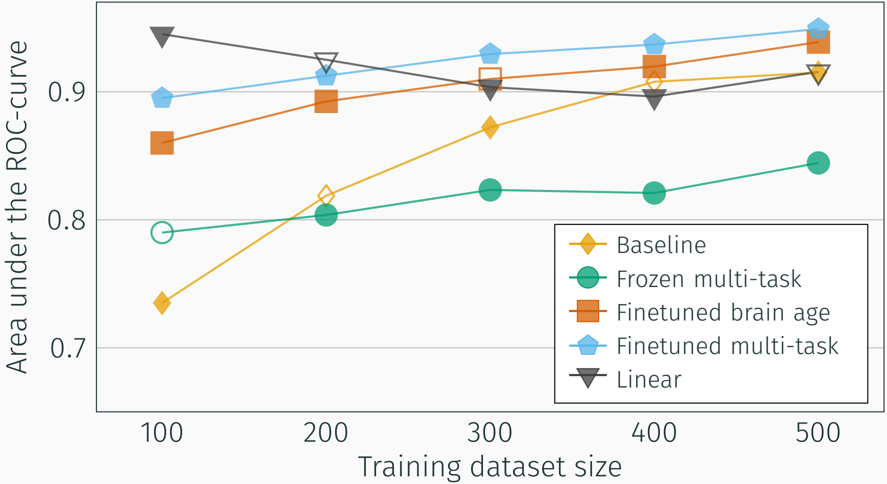
Results: AD vs HC classification



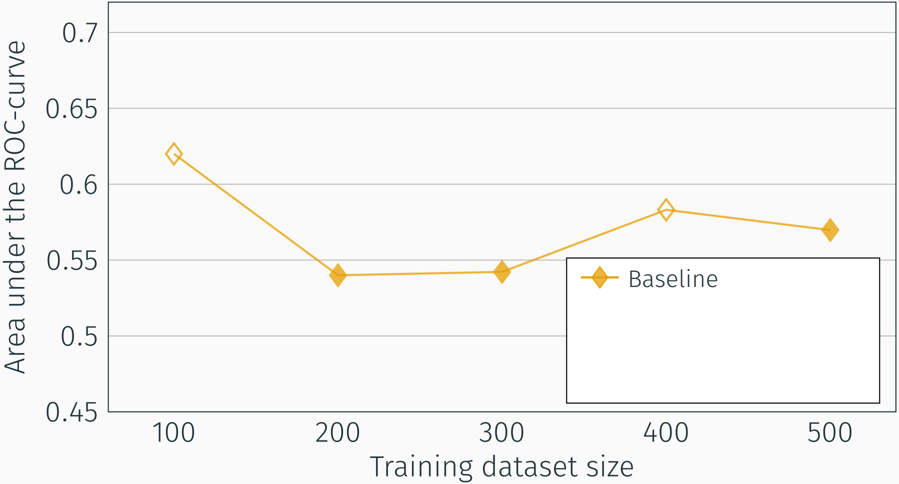
Results: AD vs HC classification



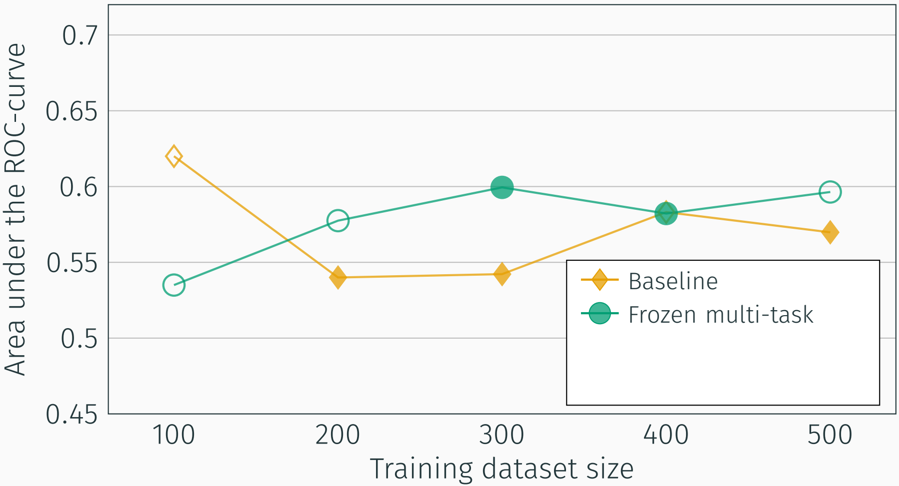
Results: AD vs HC classification



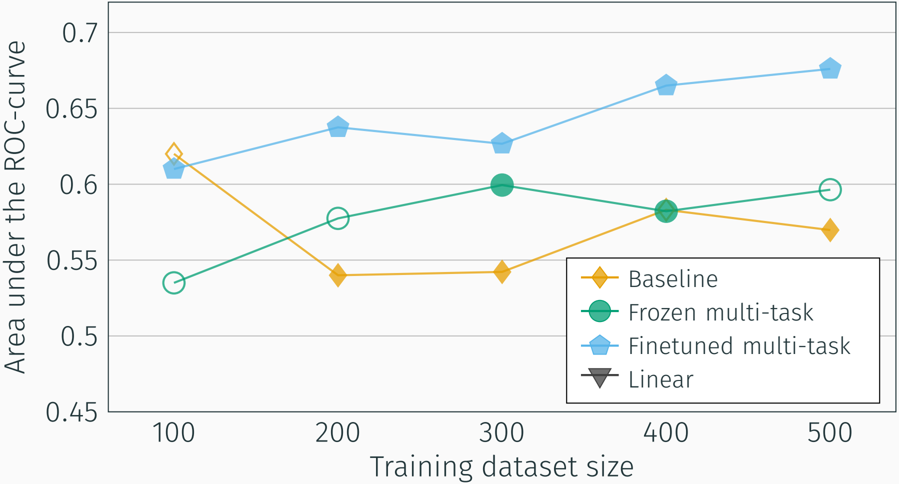
Results: SCZ vs BD classification



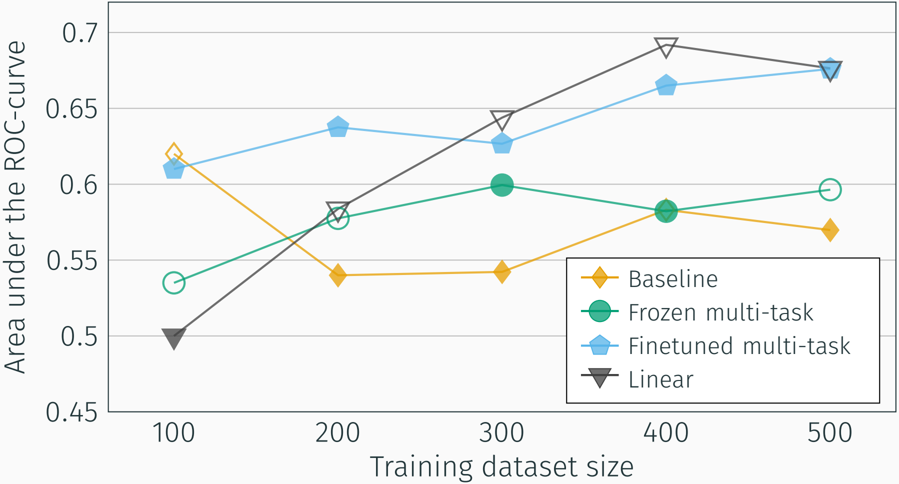
Results: SCZ vs BD classification



Results: SCZ vs BD classification



Results: SCZ vs BD classification



Results: Brain age

If you want to accurately predict anything[1] in new samples:

- If you are using a CNN, you should *always*[2] use a pretrained model, and a multi-task model outperforms a single-task model even when the pretraining task is closely related to the downstream task
- However, apparently[3], you can get quite far with simple linear models

1. At least clinical binary classification problems
2. Unless you are training a classifier for SCZ vs BD with 100 samples
3. It seems quite complicated to determine when the linear model is best

Thank you for your attention!

estenhl@uio.no