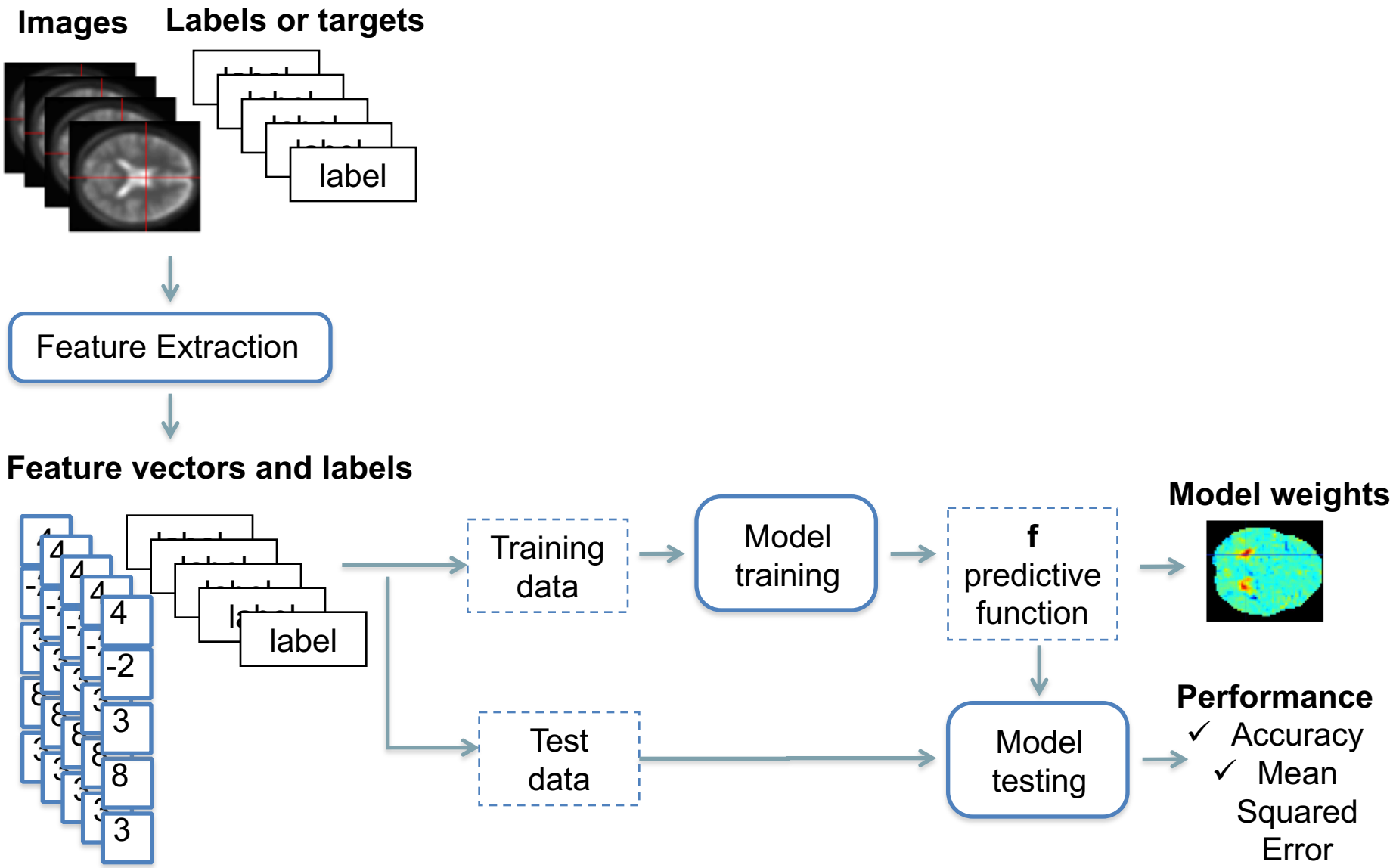
Confounding variables, how do we account for them?

Janaina Mourao-Miranda
Machine Learning and Neuroimaging Lab
Centre for Medical Image Computing (CMIC)
University College London, UK

Outline

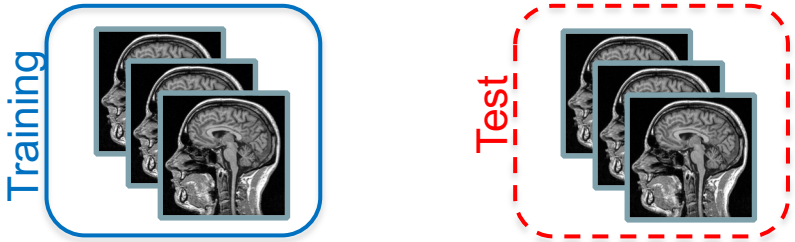
- ✓ Neuroimaging-based predictive models
- ✓ What are confounds in neuroimaging-based predictive models?
- ✓ Common strategies to account for confounds
- ✓ Comparative results
- ✓ A machine learning perspective

Neuroimaging-based predictive models

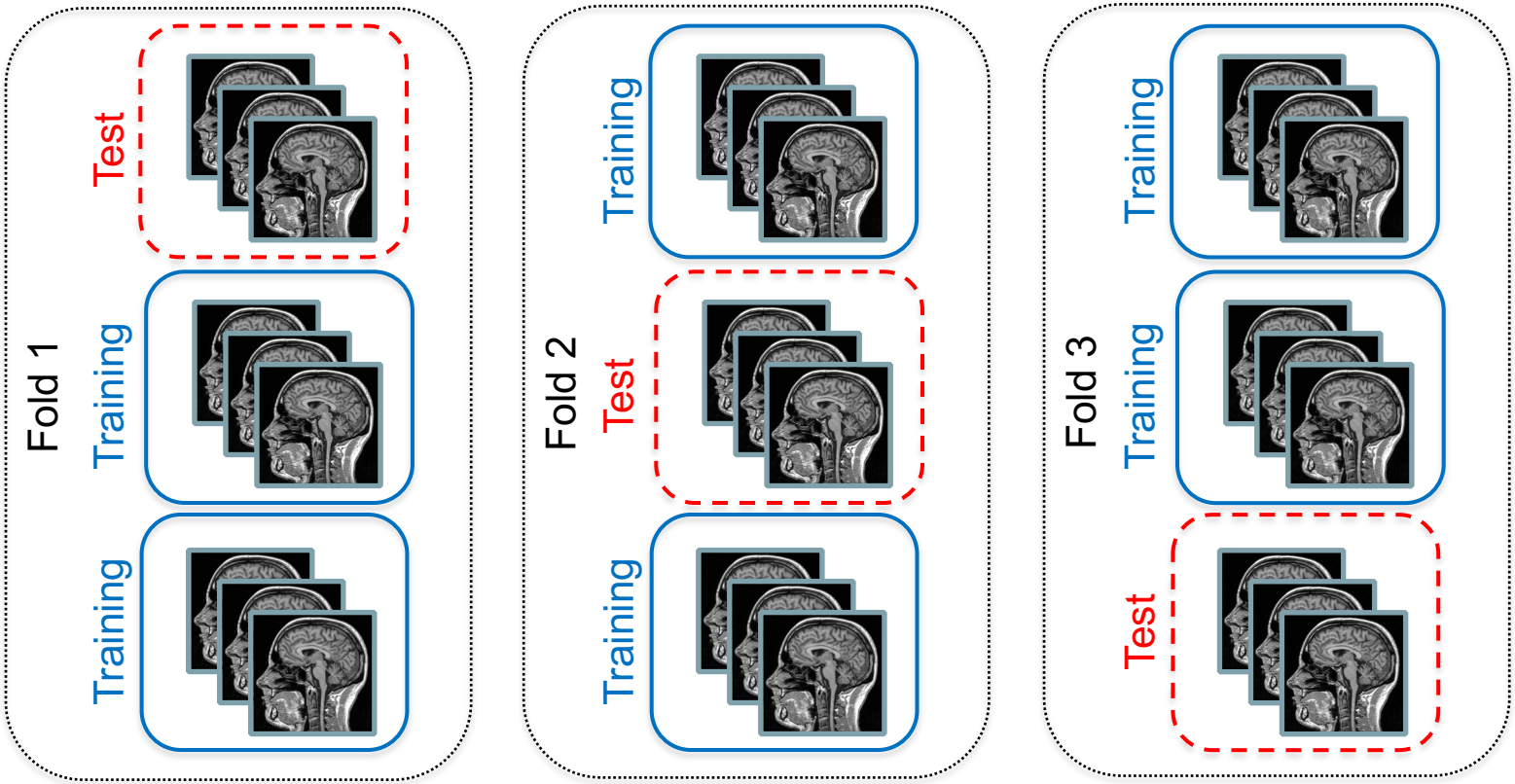


How do we split the data into training/test?

- Half split



- K-fold cross-validation strategy (e.g. k=3, 5, 10)



What are confounds in neuroimaging-based predictive models?

Definition 1: "***variable*** *that affects the* ***neuroimaging data*** *and/or the* ***targets*** *but is not of interest from a neuroscience or clinical perspective.*"

Definition 2: "***variable*** *that is not of primary interest, correlates with the* ***to-be-decoded variable*** *(the target), and is encoded in the* ***neuroimaging data***" (Snoek et al. 2019).

Definition 3: "***variable*** *which affects the* ***imaging data*** *and has an association with the* ***target*** *variable in the sample that differs from that in the* ***population-of-interest***" (Rao et al., 2017).

Example of confounds in neuroimaging-based predictive models

Example 1: Predicting ***age*** from structural MRI data considering ***site/scanner*** as confound.

Example 2: Predicting ***cognitive decline*** in demented patients from structural MRI data considering ***age*** and **sex** as confounds.

Example 3: Predicting ***object category*** from fMRI activity patterns considering ***luminance*** and ***spatial frequency*** as confounds.

What are the reasons to account for confounds?

- ✓ **Improve the generalizability of the models:**
  - Account for scanner related noise in the images to improve the performance of the model.

- ✓ **Improve the interpretability of the models:**
  - Account for total brain volume in order to identify predictive brain features that are not influenced by overall brain size.

Common strategies to account for confounds

1. **Matching / Counter-balancing**
2. **Image adjustment by confounds or regressing-out**
3. Incorporating confounds as predictors
4. Instance-based weighting

Notation

X: features

y: labels

C: confounds

Subject 1	x_{11}	…	x_{1j}	…	x_{1J}	y_1	c_{11}	…	c_{iK}
⋮	…	…	…		…				
Subject *i*	x_{i1}		x_{ij}		x_{iJ}	y_i	c_{i1}	…	c_{iK}
⋮	…	…	…		…				
Subject N	x_{N1}		x_{Nj}		x_{NJ}	y_N	c_{N1}	…	c_{NK}

Feature vector

$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots x_{iJ})$

Vector for feature *j*

$\mathbf{x}_j = (x_{1j}, x_{2j}, \dots x_{Nj})^{\mathsf{T}}$

Label or target for $\mathbf{x}_i$

y_i

Confounds for $\mathbf{x}_i$

$\mathbf{c}_i = (c_{i1}, c_{i2}, \dots, c_{iK})$

1. Matching / Counterbalancing

- The sample is chosen so that there is no significant correlation between the confound variables ($\mathbf{c}$) and the target ($\mathbf{y}$).

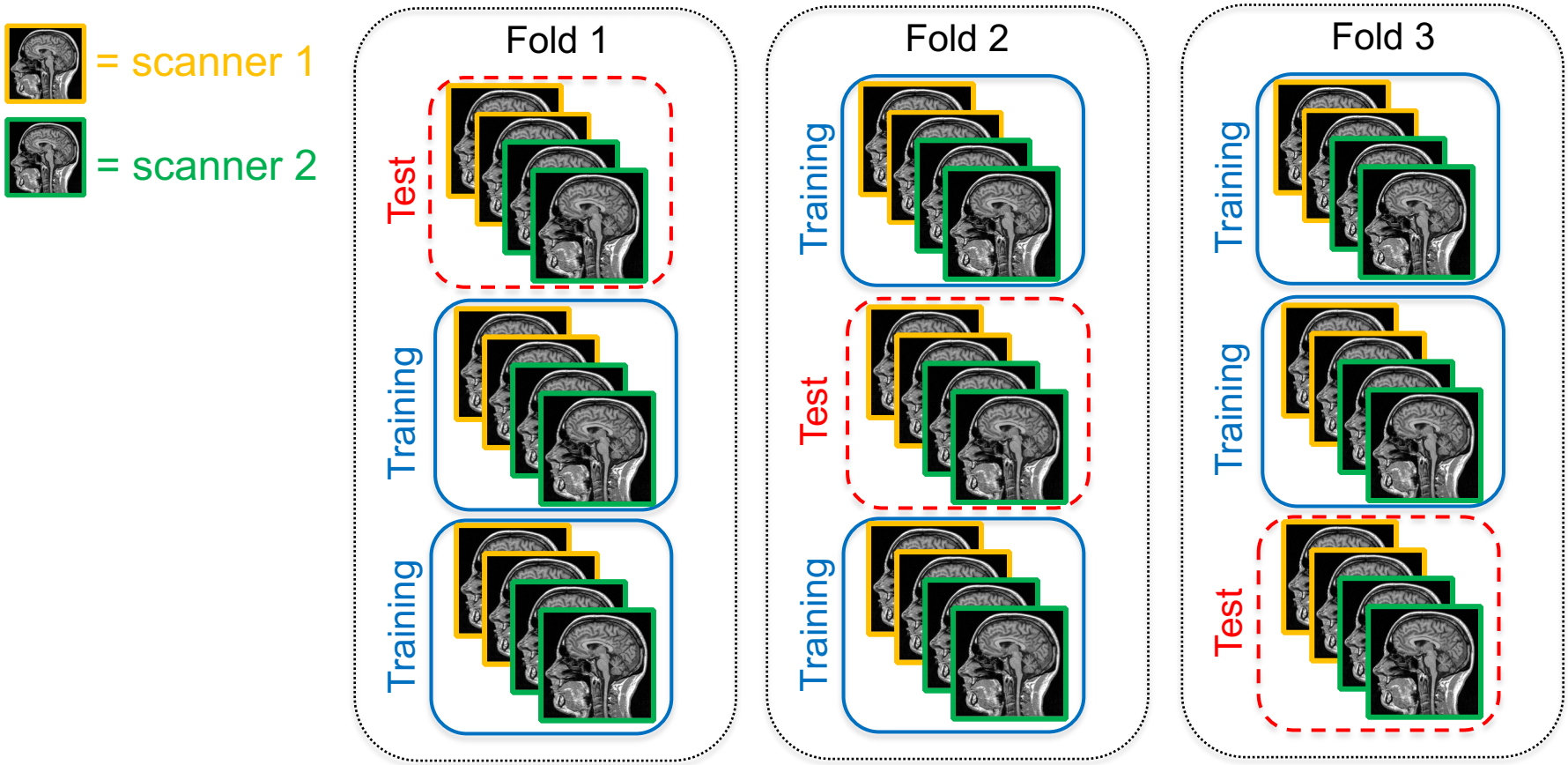
- Matching can be done a priori in the experimental design or post hoc by selecting a subset of a previously acquired sample.

- **Example:** match patients and controls by age and sex in the training and test data.

- **<u>Disadvantages:</u>**
  - **A priori:** the matched sample might not be representative of the population of interest.
  - **Post hoc:** it can reduce the sample size significantly.

1. Matching / Counterbalancing

- Matching should be done across classes (in classification problems) and across folds (during cross-validation).



2. Image adjustment or regressing-out

- Fit a linear regression model for each image feature ($\mathbf{x}_j$) using the confounds ($\mathbf{C}$) as predictors.

- The residuals of the linear regression model (i.e. adjusted images) are used as the input features in the predictive model.

- **Example:** regress-out age and sex from structural MRI and use the adjusted images as input to a predictive model.

- **<u>Disadvantages:</u>**
  - Assumes a linear relationship between the confound and the image features.
  - If the confound variable is correlated with the targets it removes predictive information from the image features.

2. Image adjustment or regressing-out

- For adjusting all image features simultaneously the following matrix equations can be used to estimate **the model parameters** and the **adjusted features**:

$$\boldsymbol{\beta} = \left(\hat{\mathrm{C}}^{T}\hat{\mathrm{C}}\right)^{-1}\hat{\mathrm{C}}^{T}\mathrm{X}$$

$$\mathrm{X}^{A} = \mathrm{X} - \hat{\mathrm{C}}\boldsymbol{\beta}$$

where $\hat{\mathrm{C}}$ is C augmented with a column of ones.

- However, this approach is often applied before splitting the data into training and test which can lead to bias (Snoek et al, 2019).

2. Image adjustment accounting for training/test separation

- Image adjustment should be done within the cross-validation framework to avoid potential bias.

- In this case the confound-imaging relationship is estimated based on the training data only.

- Considering $\mathbf{X}_{train}$ and $\mathbf{X}_{test}$ two partitions of the data and $\mathbf{C}_{train}$ and $\mathbf{C}_{test}$ being the respective confounds, the following matrix equations be be applied:

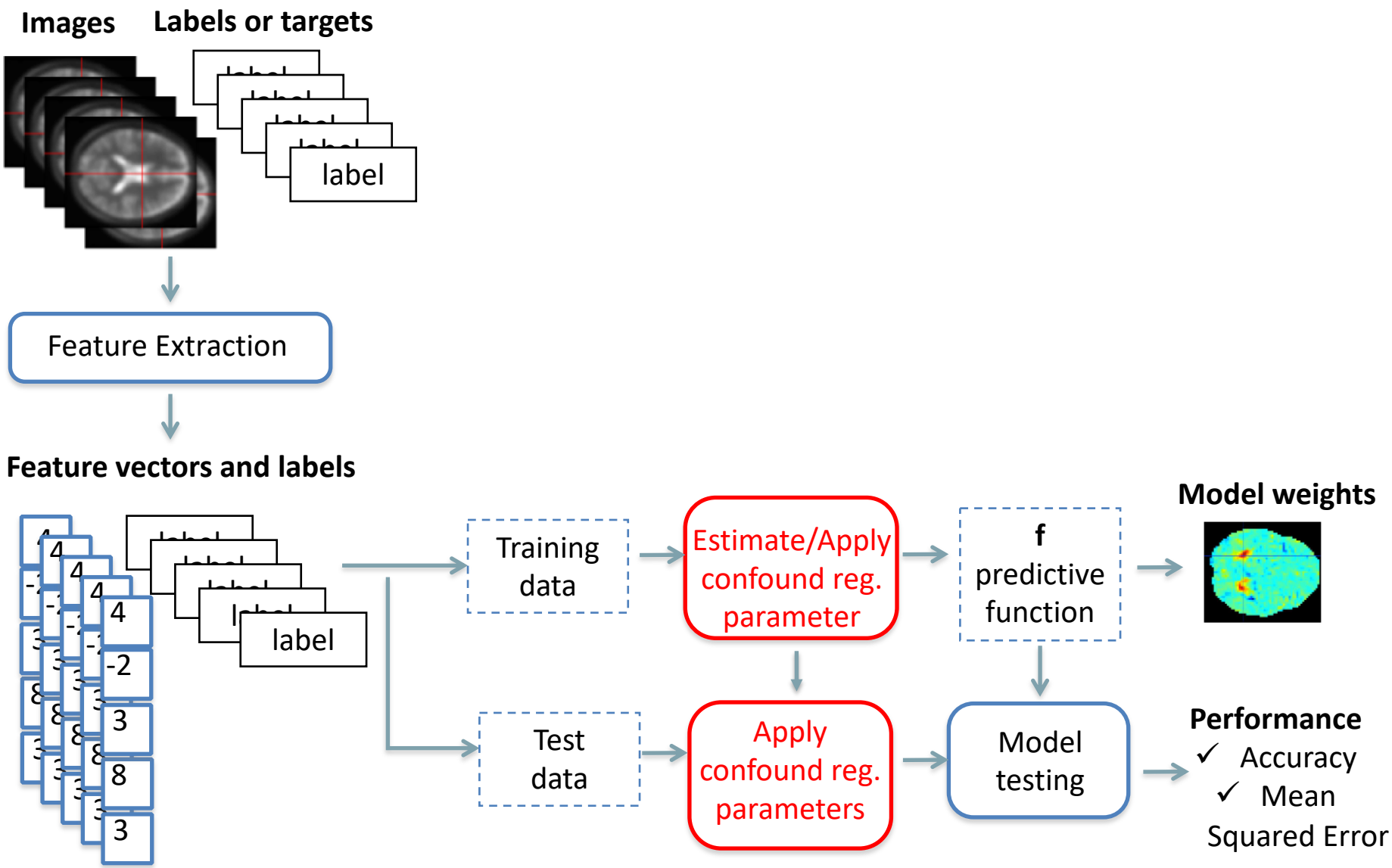
$$\boldsymbol{\beta}_{train} = \left(\hat{\mathbf{C}}_{train}{}^{T}\hat{\mathbf{C}}_{train}\right)^{-1}\hat{\mathbf{C}}_{train}{}^{T}\mathbf{X}_{train}$$

$$\mathbf{X}_{train}{}^{A} = \mathbf{X}_{train} - \hat{\mathbf{C}}_{train}\,\boldsymbol{\beta}_{train}$$

$$\mathbf{X}_{test}{}^{A} = \mathbf{X}_{test} - \hat{\mathbf{C}}_{test}\boldsymbol{\beta}_{train}$$

where $\hat{\mathbf{C}}_{train}$ and $\hat{\mathbf{C}}_{test}$ are $\mathbf{C}_{train}$ and $\mathbf{C}_{test}$ augmented with a column of ones

2. Image adjustment accounting for train/test



3. Incorporating confounds as predictors

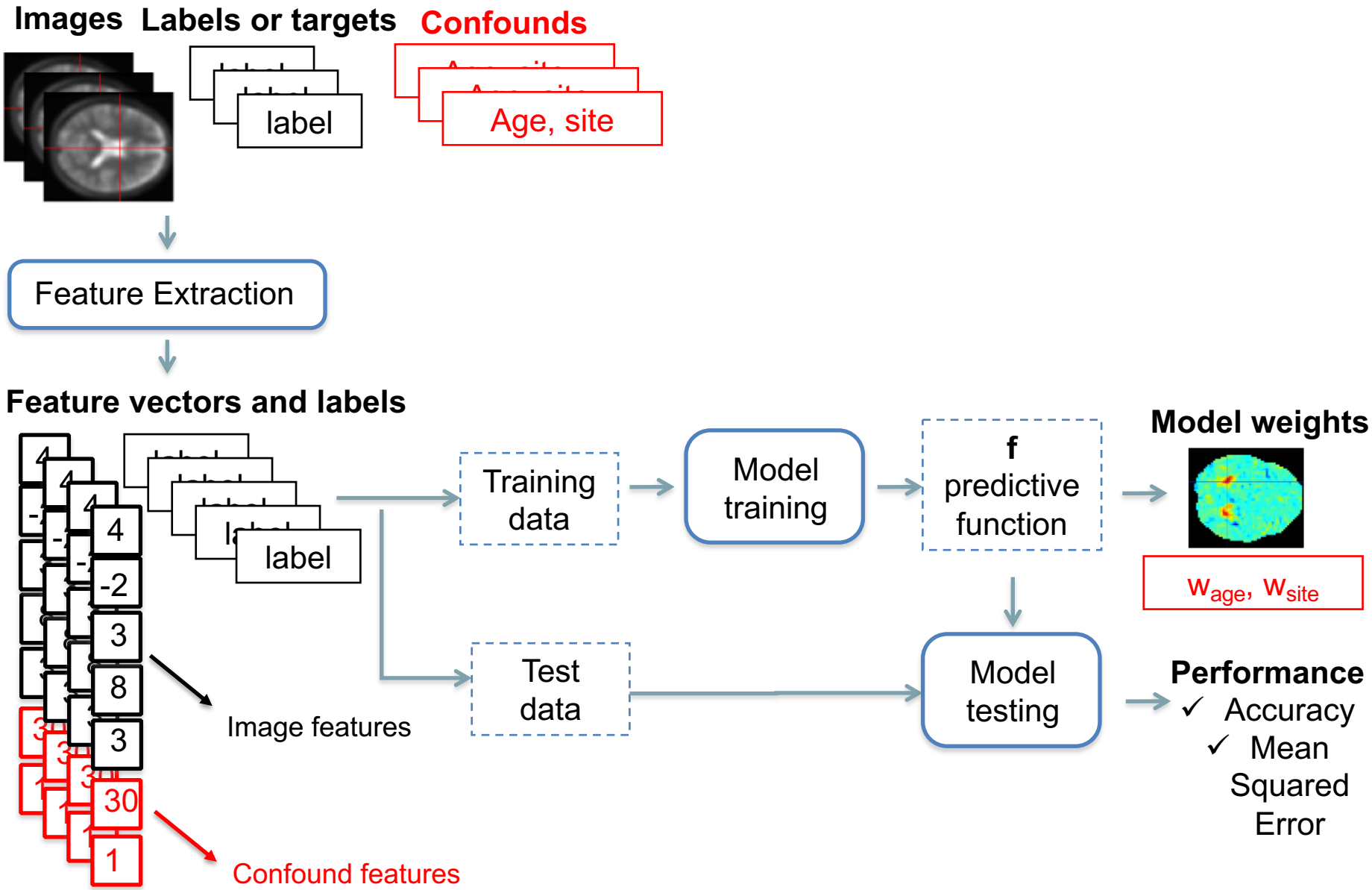
- The confounds are included as predictors (additional features) in the model.

- The goal is to learn the relative contribution of the confounds for the predictive model.

- **Example:** adding age and sex as an additional features to the feature vector containing the image features.

- **<u>Disadvantages:</u>**
  - It is not easy to separate the contribution of individual features in predictive models (model's weights are usually difficult to interpret).

3. Incorporating confounds as predictors



4. Instance-based weighting

- Weights the samples to produce a pseudo-sample in which the association between the target ($\mathbf{y}$) and the confounds ($\mathbf{C}$) has been removed (Linn et al., 2016 and Rao et al., 2017).

- Instance-based weighting proceeds in two stages:
  - In the first stage it models the relationship between the confound and the target.
  - In the second stage this relationship is used to weight the training examples in the predictive modelling.

- **<u>Disadvantages:</u>**
  - In case of high dimensional data it does not demonstrate a clear advantage over the other methods in terms of overall performance (Rao et al, 2017).

4. Instance-based weighting

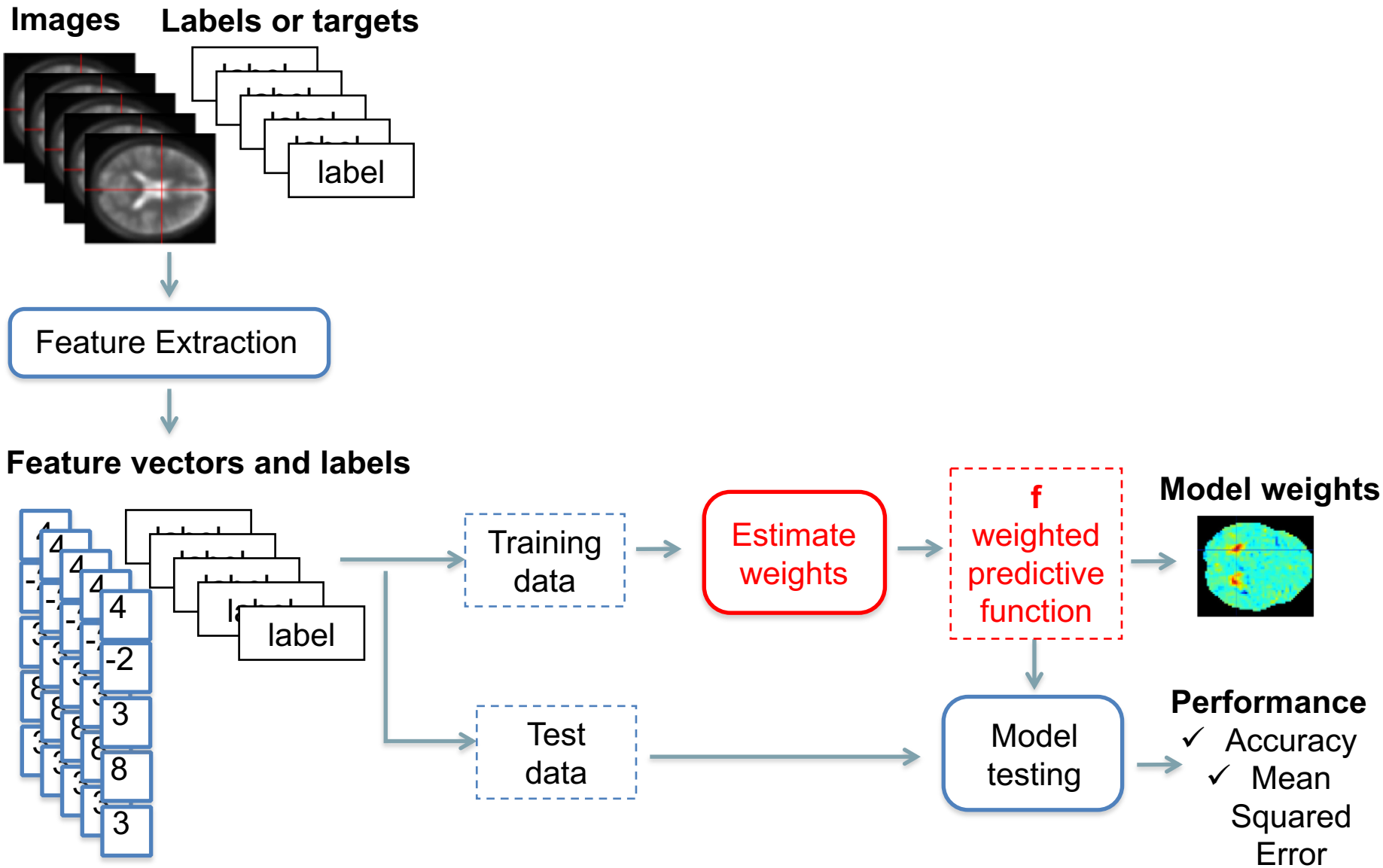
- In order to perform instance weighting, we first need to **estimate the weights** w_i for each training example:

$$w_i = \frac{P(y_i)}{P(y_i/\mathbf{c}_i)}$$

- This involves estimating the ratio between the marginal and conditional distributions $P(y)$ and $P(y/\mathbf{c})$ from the training data and evaluating w_i at each training point i.

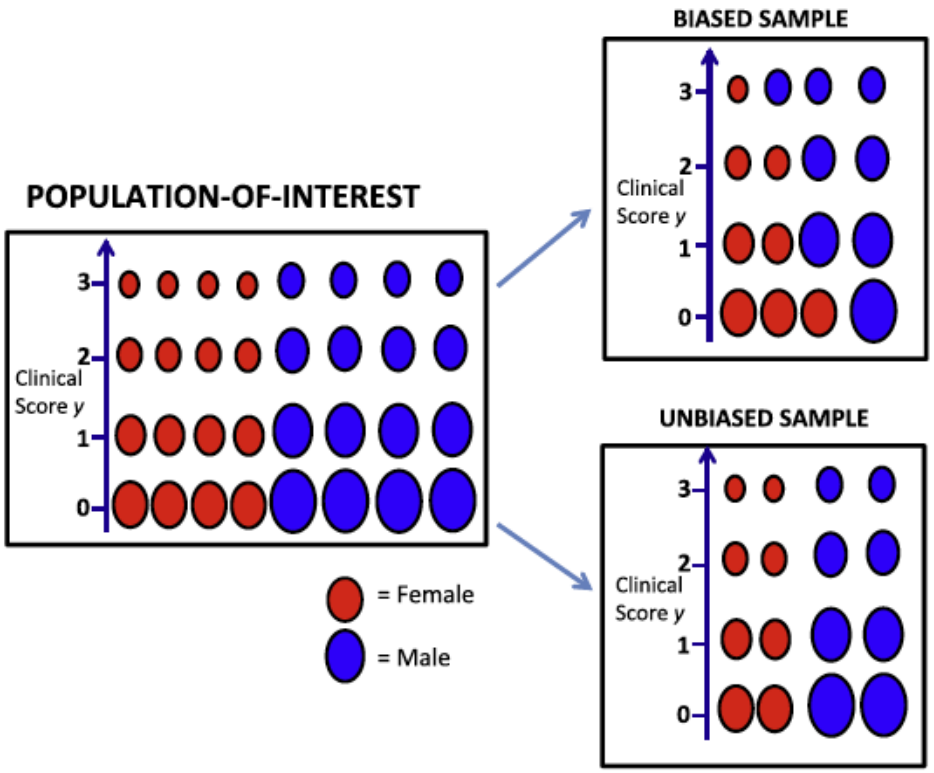
- The next step is to learn a **weighted predictive function**: e.g. a ***weighted* Kernel Ridge Regression** in which the loss associated with the training point i is weighted by w_i.

4. Instance-based weighting



Comparative Results

Example illustrating biased and unbiased samples from a population of interest



- The **population of interest** contain no association between the **clinical score** and **sex**.
- The **biased sample** contains a correlation between **sex** and the **clinical score**.
- The **unbiased sample** does not contain a correlation between **sex** and the **clinical score**.
- **Sex** is a **confound** in the biased sample.

Source: Rao et al. 2017

Example 1: Predicting Mini-Mental State Examination (MMSE) from sMRI

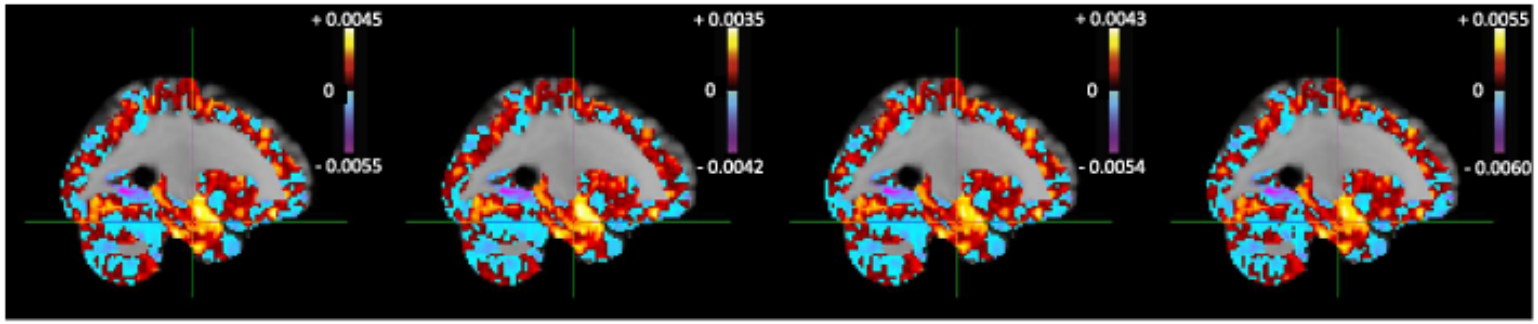
- ✓ A **biased** and an **unbiased** samples were selected from the **Alzheimer's Disease Neuroimaging Initiative** (**ADNI**) dataset.
- ✓ The **biased sampled** contained a correlation between **sex** and **MMSE scores**.
- ✓ The **unbiased sample** did not contain a correlation between **sex** and **MMSE scores**.

	Biased training sample	Unbiased training sample
Model	**MSE (unbiased test sample)**	**MSE (unbiased test sample)**
Images only (baseline)	8.02*	7.58*
Adjusted Images	7.94*	7.64*
Images & Confounds	8.43*	7.57*
Instance Weighting	8.02*	7.61*

* Indicates better performance than chance, p-value <0.05
All models used a Gaussian Process Regression as predictive model

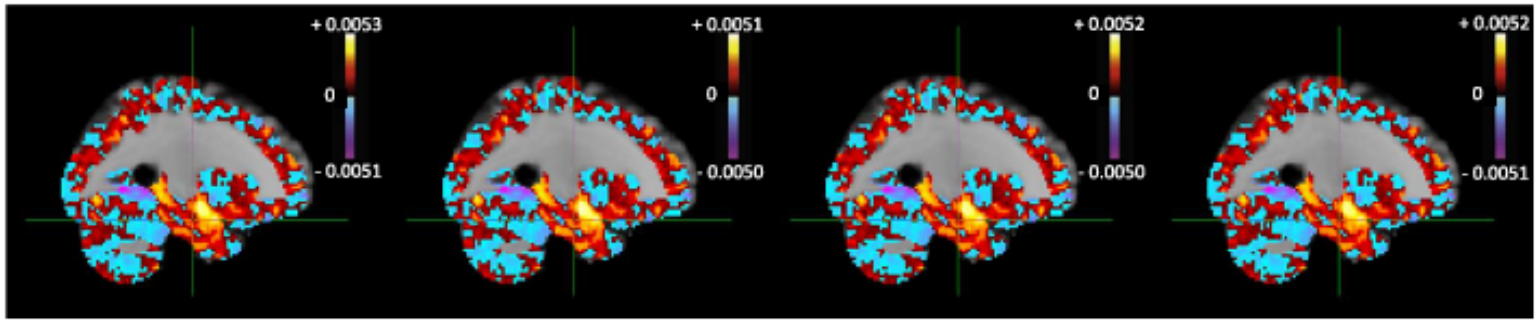
Source: Rao et al. 2017

Average weight map for each model trained to predict MMSE scores from structural MRI



Images Only | Adjusted Images | Images & Confounds | Instance Weighted

(a) Biased Training Samples



Images Only | Adjusted Images | Images & Confounds | Instance Weighted

(b) Unbiased Training Samples

Example 2: Predicting age from sMRI

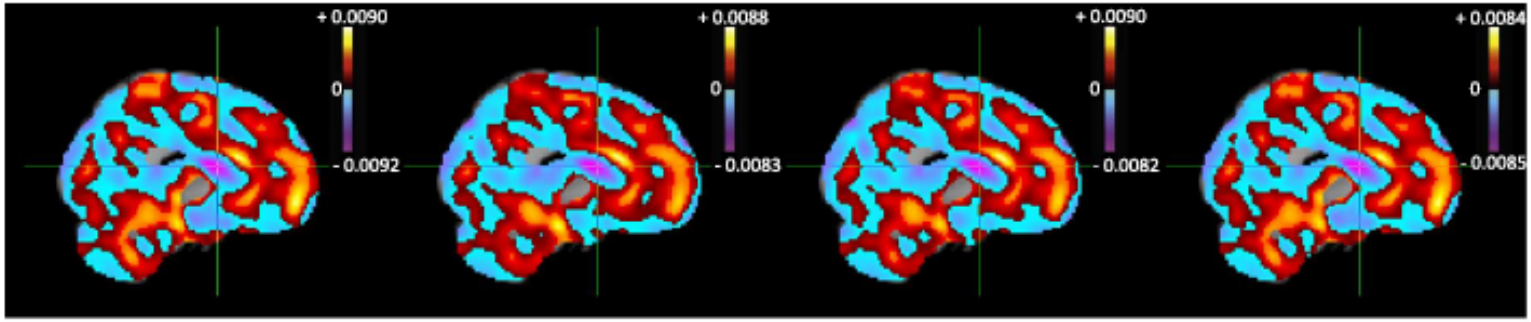
- ✓ A **biased** and an **unbiased** samples were selected from the IXI (http://brain-development.org/ixidataset/) dataset.
- ✓ The **biased sampled** contained a correlation between **age** and **site**.
- ✓ The **unbiased sample** did not contain a correlation between **age** and **site**.

	Biased training sample	Unbiased training sample
Model	**MSE (unbiased test sample)**	**MSE (unbiased test sample)**
Images only (baseline)	30.08*	25.21*
Adjusted Images	31.66*	25.34*
Images & Confounds	31.54*	25.84*
Instance Weighting	29.91*	25.21*

* Indicates better performance than chance, p-value <0.05
All models used a Gaussian Process Regression as predictive model

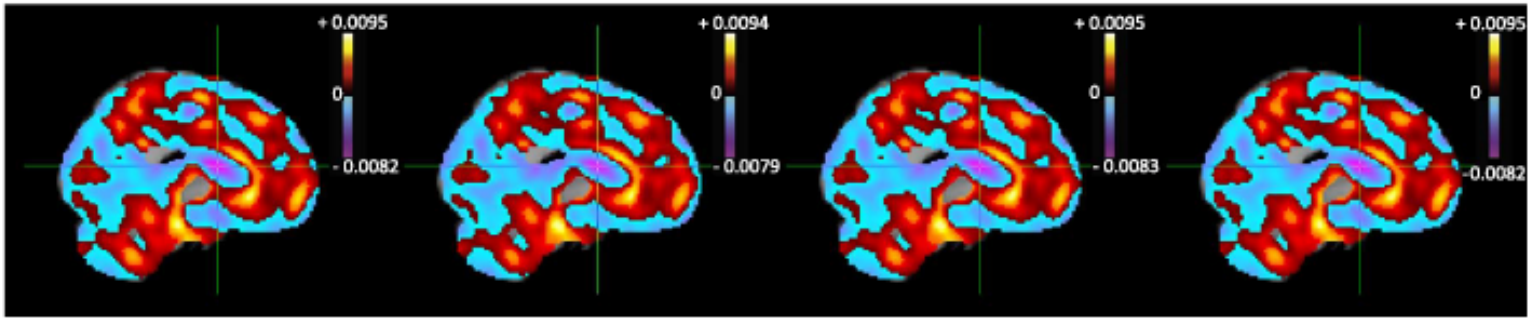
Source: Rao et al. 2017

Average weight map for each model trained to predict age from structural MRI



Images Only | Adjusted Images | Images & Confounds | Instance Weighted

(a) Biased Training Samples



Images Only | Adjusted Images | Images & Confounds | Instance Weighted

(b) Unbiased Training Samples

Results Summary

- **All models performed better than chance**, i.e. they were able to learn information that is useful for prediction of unbiased data despite the presence of confounding.

- **All models perform better when training with the unbiased samples** compared to when using the biased.

- The **performance of the different methods** for dealing with confounding **varied for the different datasets**.

- None of the methods gave more accurate predictions than the baseline ‘Images Only’ model for both datasets.

- Including the confound as a predictor gave less accurate predictions than the baseline model for both datasets.

Accounting for confounds during significance testing

- When permutation test is used one should use a **constrained permutation** (e.g. permuting the targets within site when considering site as confound).

- This procedure preserves the association between the targets and the confound, while breaking the relationship between the imaging data and the targets.

A machine learning perspective

- The problem of **confounds** is related to the **dataset shift** problem in machine learning.

- **Dataset shift** occurs when the distributions on the training and test set do not match.

- Examples of **dataset shift**:
  - **Covariate Shift:** the distribution of the features in the training and test sets are different.
  - **Prior Probability Shift:** the distribution of the targets in the training and test sets are different.
  - **Sample Selection Bias:** the distributions differ as a result of an unknown sample rejection process.

- Different approaches have been proposed in the machine learning field for dealing with **dataset shift** (e.g. Quionero-Candela et al. 2009).

Domain Adaptation and Transfer Learning

- **Domain Adaptation:** aims to learn a model from a **source data** distribution that performs well on a different (but related) **target data** distribution (e.g. Long et al. 2015, Bousmalis et al. 2016).

- **Example of domain adaptation:** adapting a predictive model trained on data collected from a specific scanner to make it compatible for testing data obtained from another scanner.

- **Transfer learning:** aims to transfer knowledge gained while solving one problem to a different but related problem (e.g. Shin et al. 2016).

- **Example of transfer learning:** training a classifier to discriminate Alzheimer patients from healthy controls and then fine-tune it to classify frontotemporal dementia patients from healthy controls.

Machine Learning & Neuroimaging (MLNL) Lab

Agoston Mihalik, PhD
Research Associate

Cemre Zor, PhD
Senior Research Associate

Rick Adams, PhD
MRC Skills Development Fellow & Consultant Psychiatrist

Isabel David, PhD
Associate Professor at FFU, Brazil
Sabbatical visitor

Taiane Ramos
Visiting PhD Student

Najiba Toron
PhD Student

Fabio Ferreira
PhD Student

James Chapman
PhD Student

Konstantinos Tsirlis
Research Assistant

Acknowledgements

Former MLNL members:

- Dr Anil Rao
- Dr Jessica Schrouff (Google Health)

- **Pattern Recognition for Neuroimaging Toolbox (PRoNTo) team** (http://www.mlnl.cs.ucl.ac.uk/pronto/)

Useful references

Chu, C., Mourao-Miranda, J., Chiu, Y.C., Kriegeskorte, N., Tan, G., Ashburner, J. (2011) Utilizing temporal information in fMRI decoding: Classifier using kernel regression methods. Neuroimage Sep 15;58(2):560-71. https://doi.org/10.1016/j.neuroimage.2011.06.053

Linn, K. A., Gaonkar, B., Doshi, J., Davatzikos, C., & Shinohara, R. T. (2015). Addressing Confounding in Predictive Models with an Application to Neuroimaging. The International Journal of Biostatistics, 12(1), 31–44. https://doi.org/10.1515/ijb-2015-0030.

Rao, A., Monteiro, J. M., Mourão-Miranda, J., & Alzheimer's Disease Initiative. (2017a). Predictive Modelling using Neuroimaging Data in the Presence of Confounds. NeuroImage. https://doi.org/10.1016/j.neuroimage.2017.01.066.

Rao, A., Mourao-Miranda, J. (2017b) Feature Adjustment in Kernel Space when using Cross-Validation. Research Note - UCL Department of Computer Science. https://pdfs.semanticscholar.org/37b3/c2e2be6c9579867c21381762fde01113f647.pdf

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, *184*, 741–760. https://doi.org/10.1016/j.neuroimage.2018.09.074

Useful references

Quionero-Candela, J., Sugiyama, M., Schwaighofer A., Lawrence, N.D. (2009) Dataset Shift in Machine Learning. The MIT Press.

Long, M., Cao, Y., Wang, J., Jordan, M. (2015) Learning transferable features with deep adaptation networks. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, 97–105.

Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D. (2016). Domain separation networks. In Proceedings of the 30th International Conference on Neural Information Processing System, 343–351.

Shin H-C., Roth, H. R., Gao M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R. M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. In IEEE Transactions on Medical Imaging, vol. 35, no. 5, 1285–1298, 2016.

Useful tools

PRoNTo (http://www.mlnl.cs.ucl.ac.uk/pronto/) includes operations for image adjustment accounting for training and test separation.

Confounds library
https://crossinvalidation.com/2020/03/04/conquering-confounds-and-covariates-in-machine-learning/

PALM toolbox **(https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM)** for generating restricted permutations.