Pitfalls of Confound Regression in Machine Learning

Sami Hamdan[1,2], Bradley C. Love[3], Georg G. von Polier[2], Susanne Weis[1,2], Holger Schwender[4], Simon B. Eickhoff[1,2] & Kaustubh R. Patil[1,2]

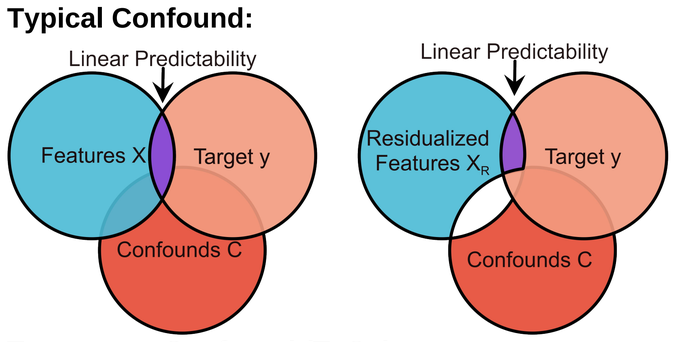
[1]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[2]Institute of Neuroscience and Medicine (INM-7: Brain and Behaviour), Research Centre Jülich, Jülich, Germany,
[3]University College London, Department of Experimental Psychology, London, UK
[4]Heinrich-Heine University Düsseldorf, Institute of Mathematics, Germany

Poster Number: 236

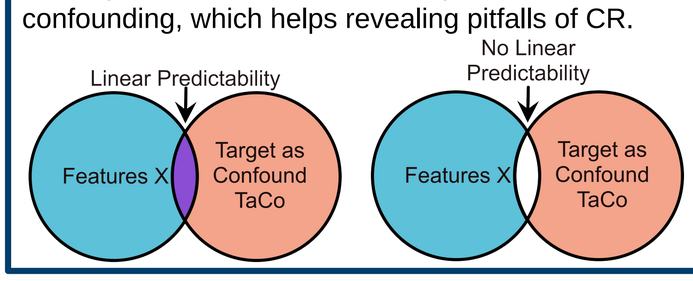
Introduction

Not dealing with confounding can threaten the interpretability & meaningfulness of machine learning (ML) models.
This can lead to: untrustworthy predictions & questionable insights.[1]
Therefore, researchers often remove them using Confound Regression (CR).[2]

Typical Confound:



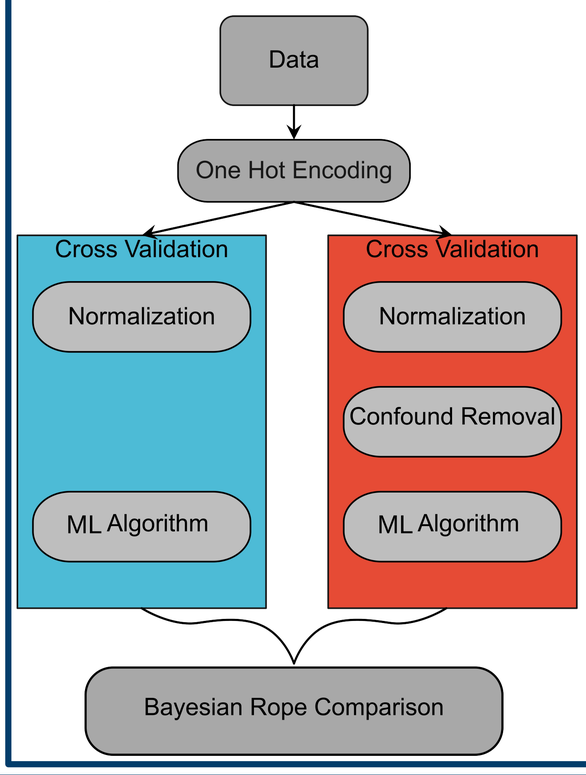
Target as a Confound (TaCo):
TaCo shares all variance any confound could share with features & the target. Therefore, the TaCo represents the most extreme case of confounding, which helps revealing pitfalls of CR.



Methods

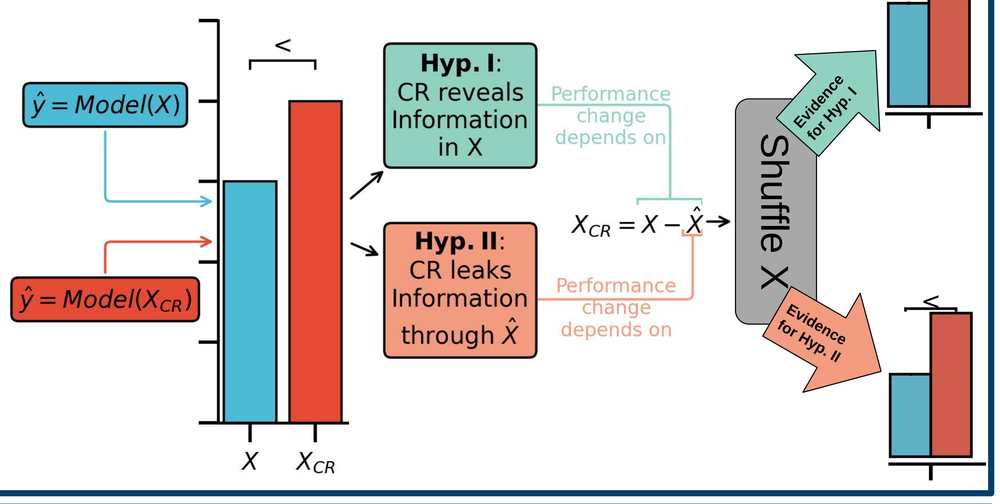
Pipeline to measure Performance

Analyzing performance of 10 benchmark data sets.
Confounds were either TaCo or simulated.
All comparisons were based on this structure:



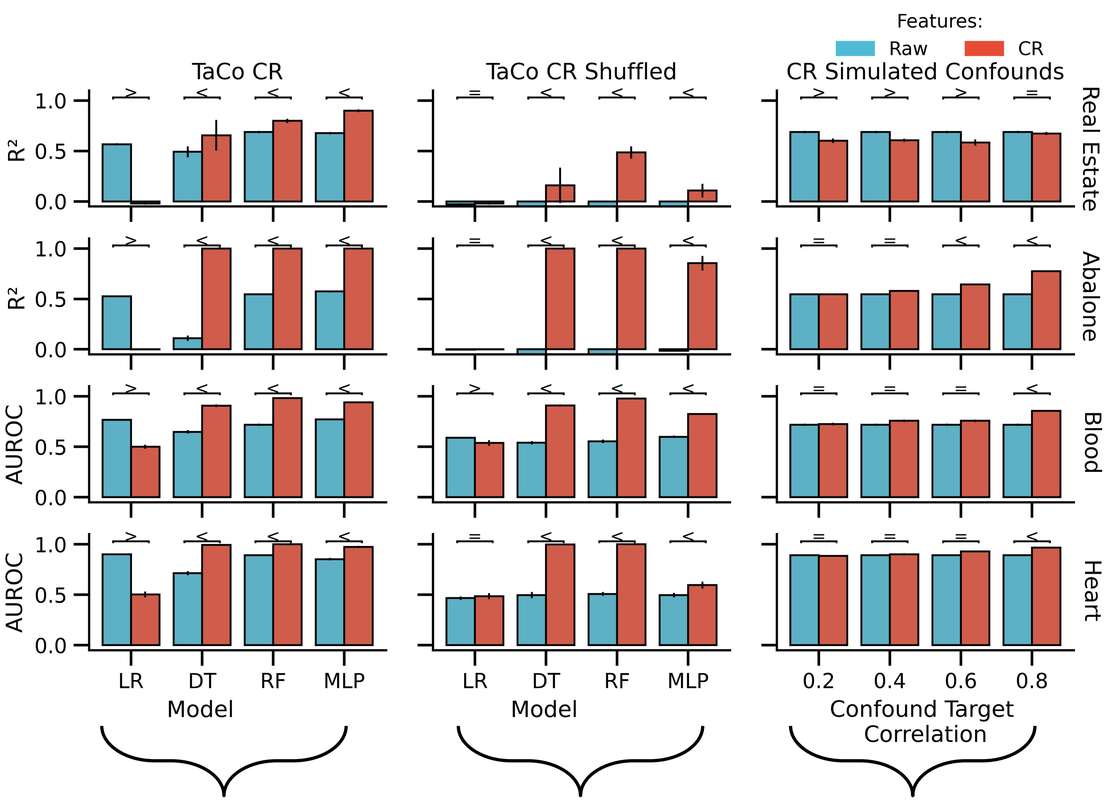
ML Algorithms used:
LR: Linear/Logistic Regression
DT: Decision Tree
RF: Random Forest
MLP: Multilayered Perceptron

Higher Score After CR — Possible Explanations



Results

Performance after CR increases before applying non-linear ML



CR can increase performance when using TaCo?
Is this leakage of information?
Or revealing real information?

Shuffling
=> Non-informative features
Still performance increase
Here: Only leakage possible

Performance increase possible with weaker confounds (non TaCo)

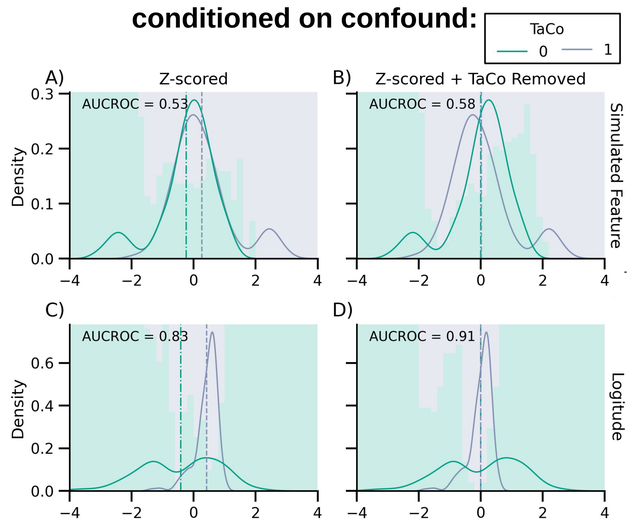
Confound removal using CR can leak information into your features!
=> confound-leakage

Confound-leakage:

CR increases information usable by (here non-linear) ML models inside of the features.
=> Danger for interpretability & meaningfulness of all CR-ML workflows

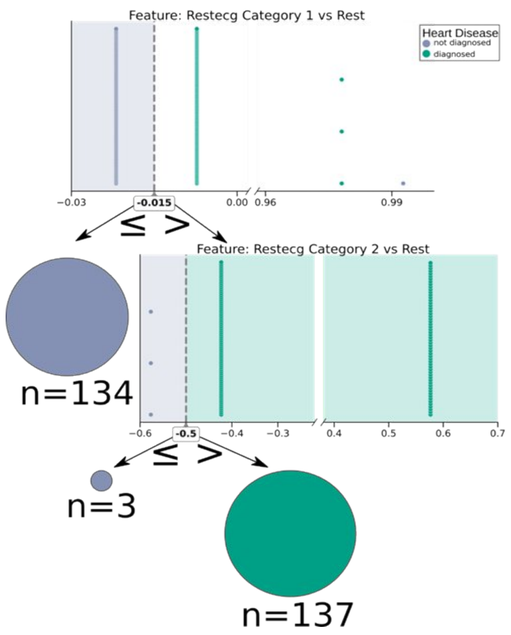
Possible mechanisms for confound-leakage

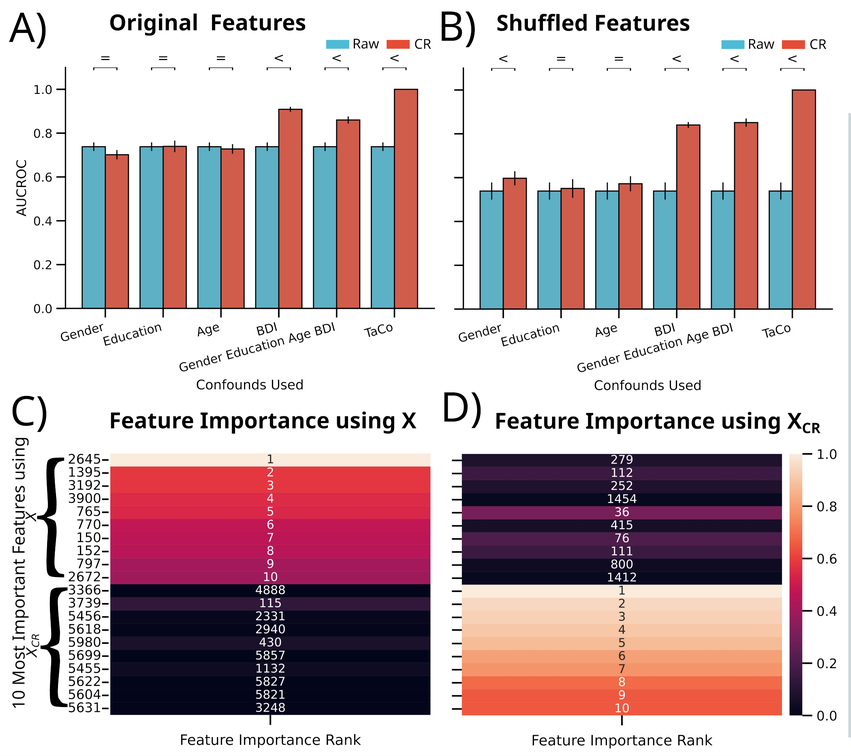
Imbalance in distributions conditioned on confound:



CR + binary TaCo
=> Mean adjusting features separated by TaCo
This can expose other parts of the distributions!

CR leaks information when using low precision features:
e.g. categorical features, counts & integers





Confound-leakage: confirmed in medical data

A) Predicting ADHD using speech features and different confounds. BDI leads to highest performance after CR.

B) Highest performance after CR same even with shuffled features

C-D) Different features are important in a RF using either original features (X) or features after CR (X_{CR})

Summary CR changes:
=> performance ↑
=> feature importance

Discussion

Summary

We show confound-leakage in
- Benchmark data
- Simulations
- Medical data

We found possible mechanisms
- imbalance in distributions conditioned on confound
- low precision features

Recommendations CR-ML Workflows

1) Check performance without CR
If CR-ML higher proceed.
2) *Assess* confounding strength
Stronger confounds to target relationship pose greater danger of confound-leakage.
3) *Gain evidence* for/against confound-leakage
TaCo + shuffling of features
4) Carefully *choose* alternatives
other procedures may also entail unknown pitfalls.

Limitations

No full-proof method for detection & elimination of confound leakage.

Several complex combinations of:
- Features
- Target
- Confounds

could lead to confound-leakage

References:
1) Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. A survey on bias and fairness in machine learning. ACM Comput. Surv. 54, 1–35 (2021).
2) Snoek, L., Miletić, S. & Scholte, H. S. How to control for confounds in decoding analyses of neuroimaging data. Neuroimage 184, 741–760 (2019).

Acknowledgments: This work was partly supported by the Helmholtz-AI project DeGen, the Helmholtz Portfolio Theme 'Supercomputing and Modeling for the Human Brain' and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)