

Susanne Weis^{1,2}, Lisa Wiersch^{1,2}, Sami Hamdan^{1,2}, Felix Hoffstaedter^{1,2}, Mikhail Votinov^{3,4}, Ute Habel^{3,4}, Benjamin Clemens^{3,4}, Birgit Derntl^{5,6}, Simon B. Eickhoff^{1,2}, Kaustubh R. Patil^{1,2}

WTh051

¹Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Germany; ²Institute of Neuroscience and Medicine (INM-7), Research Centre Jülich, Germany; ³Department of Psychiatry, Psychotherapy and Psychosomatics, Faculty of Medicine, RWTH Aachen University, Aachen, Germany; ⁴Institute of Neuroscience and Medicine (INM-10), Research Centre Jülich, Jülich, Germany; ⁵Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany; ⁶LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany. s.weis@fz-juelich.de

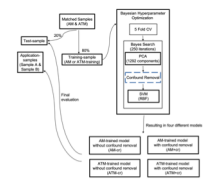
Introduction

- Total intracranial volume (TIV) differs prominently between males and females
- This difference can cause bias in machine learning (ML) based investigation of sex differences in brain structure
- High sex classification accuracies might be driven by TIV, rather than qualitative structural sex differences
- Such TIV bias might be especially relevant for populations for which modulations of brain size have been suggested ([1])
- Reduced sex classification accuracies in transgender individuals have been taken to indicate an interactive effect of biological sex and gender identity on brain organization ([1,2])
- Here, TIV bias in sex classification models applied to cisgender and transgender individuals was investigated by controlling for brain size either through feature-wise confound removal or by matching training samples for TIV, i.e. debiasing the model on the data level.

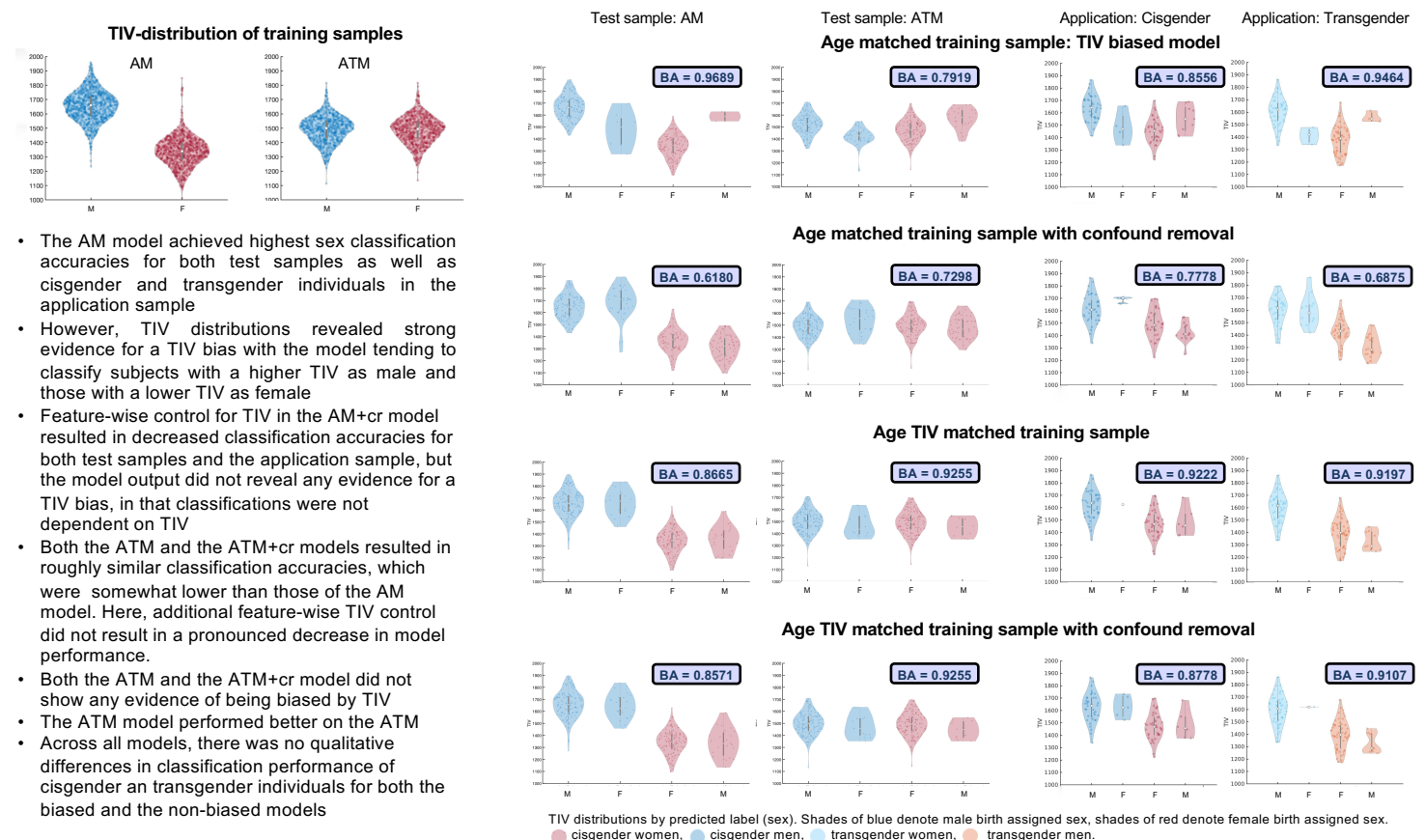
Methods

- Two mutually exclusive cisgender training samples (N = 1292, 646 f):
 - AM training set: Matched for age
 - ATM training set: Matched for age and TIV
- CAT 12.5 preprocessing, whole brain voxel-wise GMV extraction, PCA: n=1292 components [3]
- Support Vector Machine (SVM) classifiers, rbf kernel, Bayesian Hyper-Parameter Optimisation, nested 5-fold CV inner loop, stratified 10-fold CV outer loop) trained to classify sex [4]
- Models tested on:
 - held out AM and ATM test samples (n=322, 161f) and
 - Application sample (n = 202) [5,6]:
 - 43 cisgender men (CM)
 - 47 cisgender women (CW)
 - 62 transgender men (TM)
 - 50 transgender women (TW)

- Four sex classification models:
- 1) AM training sample without feature-wise TIV confound removal (AM);
 - 2) AM training sample with feature-wise TIV confound removal (AM+cr);
 - 3) ATM training sample without feature-wise TIV confound removal (ATM)
 - 4) ATM training sample with feature-wise TIV confound removal (ATM+cr),



Results



Discussion

- The TIV-biased AM model achieved very high sex classification accuracy, but inspection of the TIV-distributions shows for a strong TIV bias of the model
- The TIV-biased AM model performed worse on the ATM as compared to the AM test sample, since in the ATM sample the model could rely less on sex differences in TIV
- Feature-wise TIV confound removal in the AM+cr model reduced the bias of the model at the cost of reduced classification accuracies
- While a lack of bias in a model is desirable, so is high accuracy, suggesting that feature-wise confound removal might not be the ideal approach to reduce bias in structural sex classification
- Both the ATM and the ATM+cr model achieved high classification accuracies for both test samples and the application sample, while not showing any indication of a TIV bias
- We propose that matching men and women for TIV in the training sample provides an appropriate approach for creating an unbiased model that achieves high classification accuracies for both TIV matched and nonmatched test samples
- All models performed at similar levels of accuracy for cisgender and transgender individuals in the application sample
- Present results suggest that non-TIV-biased models are able to classify the sex of both cisgender and transgender individuals with high accuracy, highlighting the relevance of appropriate TIV control in sex classification studies.
- Furthermore, present results contradict previous studies reporting reduced sex classification accuracies and higher uncertainty for transgender individuals