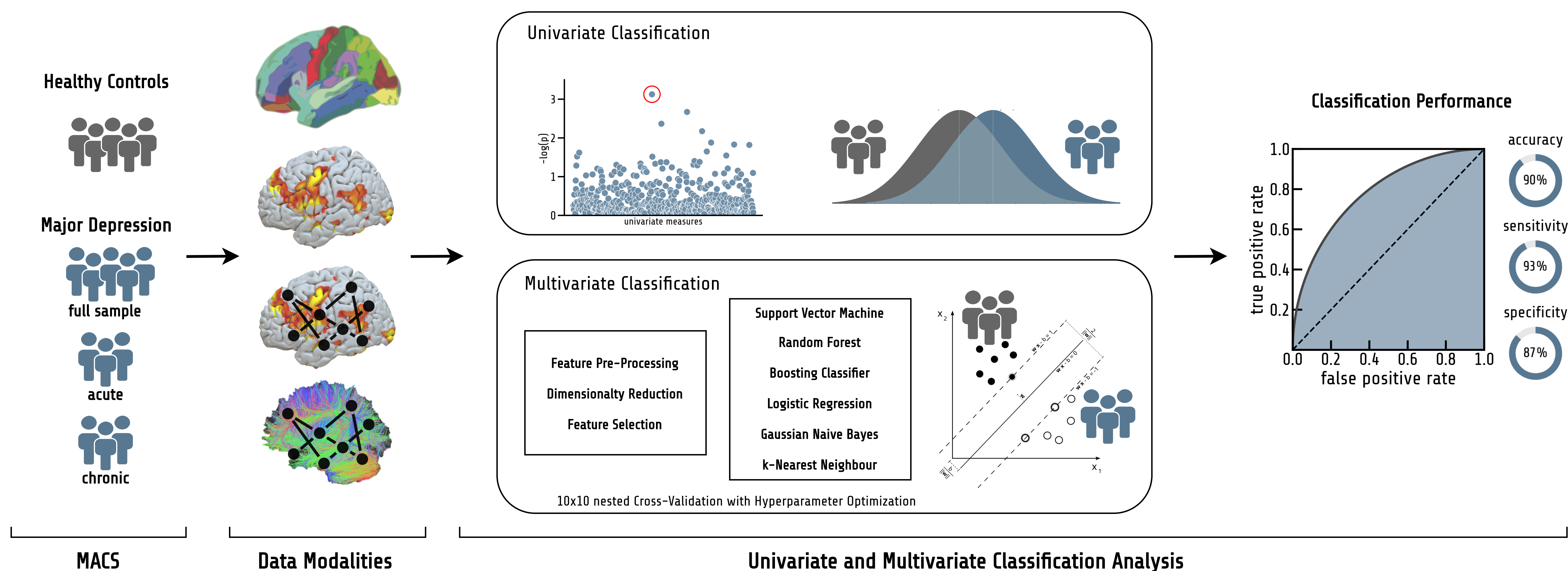
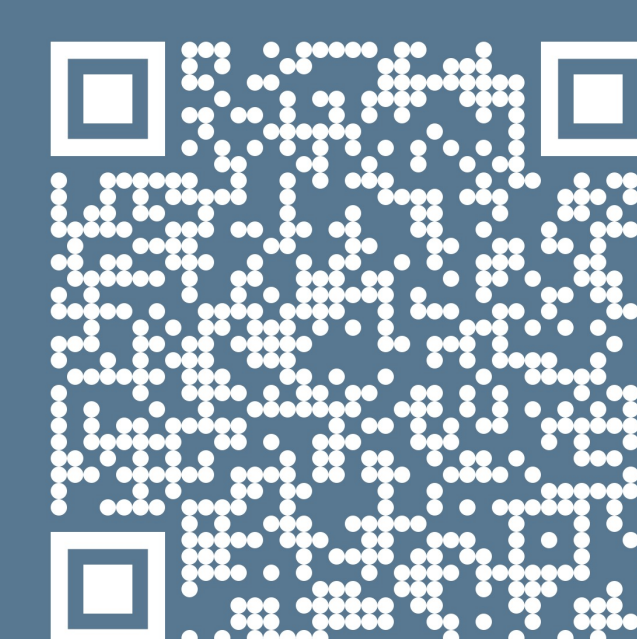




Healthy and depressive individuals are more alike than different.

Single-subject classification seems not possible based on neuroimaging.

MT375



Rethinking Univariate and Multivariate Biomarker for Major Depression A Multimodal Evaluation

Nils R. Winter¹, Julian Blanke¹, Ramona Leenings^{1,2}, Jan Ernsting^{1,2}, Kelvin Sarink¹, Daniel Emden¹, Lukas Fisch¹, Carlotta Barkhau¹, Frederike Stein³, Katharina Brosch³, Tilo Kircher³, Udo Dannlowski¹, and Tim Hahn¹

¹University of Münster, Institute for Translational Psychiatry, Germany, ²University of Münster, Faculty of Mathematics and Computer Science, Germany

³Department of Psychiatry and Psychotherapy, Philipps University Marburg, Germany

INTRODUCTION

With the emergence of **precision medicine**, there is an increasing awareness for the necessity of individualized prediction, fuelling hopes for a **personalized psychiatry** to finally **improve the classification and treatment of psychiatric disorders** such as major depression [Ozomaro et al., 2013].

A mainstay of biological psychiatry has been the identification of **group-level differences between healthy (HC) and depressed (MDD) individuals**. Yet, recent meta-analyses question the reliability of these case-control differences [Gray et al., 2020].

Machine learning-based, **multivariate** approaches have entered the field of neuroimaging emphasizing **single-subject prediction** rather than explanation [Janssen et al., 2018]. Although initial results seem promising, small sample sizes might have **inflated reported classification accuracy** [Flint et al., 2021].

METHODS

Here, we systematically investigate the **univariate and multivariate classification accuracy** in a large, **single-study cohort of N=1,809** participants (n=861 MDD) across a broad **range of neuroimaging modalities** [MACS, Vogelbacher et al., 2018].

Univariate classification: Accuracy of single variables showing the largest difference between depressive and healthy individuals.

Multivariate classification: Accuracy of extensive machine learning pipeline with nested cross-validation using PHOTONAI [Leenings et al., 2020].

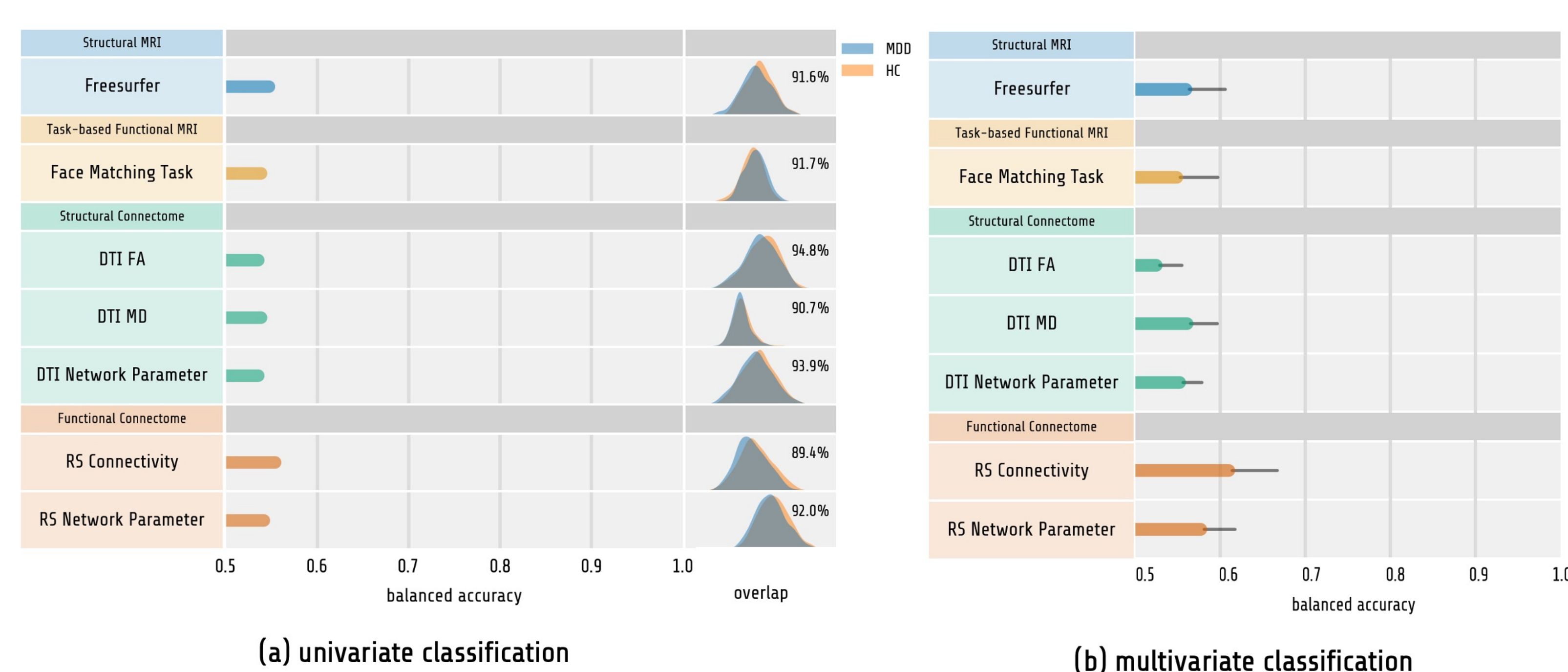
Structural MRI

Task-based fMRI

Structural Connectome

Functional Connectome

RESULTS



Univariate analyses yield accuracies between 52.5% and 55.4%. **Overlap** between HC and MDD in these variables is between 89.4% and 94.7%. **Multivariate machine learning** analyses yield accuracies between 52.4% and 60.9%. This remains **unchanged** for **acutely** or **chronically** depressed MDD.

DISCUSSION

Based on evidence from a large dataset, we conclude that **neither univariate nor multivariate methods** reliably **differentiate between healthy and depressive subjects** across neuroimaging modalities.

Strikingly, even the **best performing predictive model did not exceed 60% accuracy** despite an in-depth machine learning optimization. These results call into question the clinical utility of neuroimaging-based classification models of depression.

We argue that a stronger focus on symptoms rather than syndromes, a transparent communication of predictive utility as well as data-driven or clinically derived patient stratification is key to bridging this substantial gap between neurobiological data and clinical phenotype.