Demystifying the Risk of Reidentification in Neuroimaging Data
- A Technical and Regulatory Analysis - [1]

Anita S. Jwa,[1] Oluwasanmi Koyejo,[2] & Russell A. Poldrack[1]

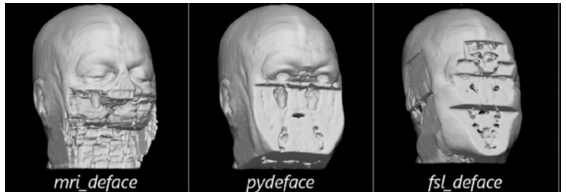
[1] Dept. of Psychology, Stanford University, United States. [2] Dept. of Computer Science, Stanford University, United States

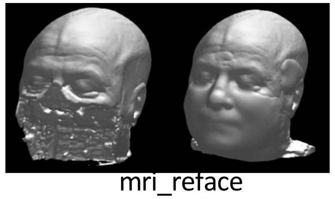
Introduction

- Emergence of novel software tools and algorithms, such as face recognition, has raised concerns about reidentification of defaced neuroimaging data.
- Despite the surge of privacy concerns,[2] the risk of reidentification has not yet been examined outside the limited settings for demonstration purposes.
- We will examine the likelihood of reidentification via face recognition in realistic settings and analyze the regulatory implications of this risk in neuroimaging data sharing.

Previous Study (Schwarz et al., 2021)[3]

- Matching accuracies of defaced images

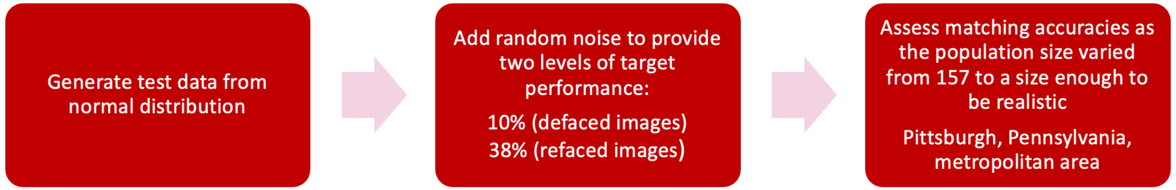




Defaced images	16/157(10%)	16/157(10%)	5/157(3%)
Refaced images	52/157(33%)	59/157(39%)	44/157(28%)

Methods

- Design a classification problem using simplified data to test the generalizability of the reported accuracies in real-world situations


Generate test data from normal distribution

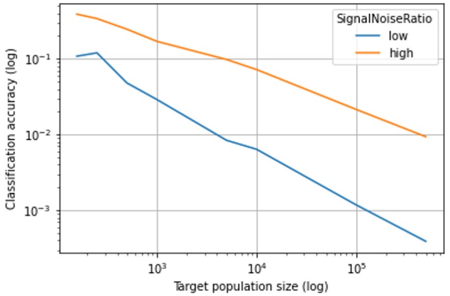
Add random noise to provide two levels of target performance:
10% (defaced images)
38% (refaced images)

Assess matching accuracies as the population size varied from 157 to a size enough to be realistic
Pittsburgh, Pennsylvania, metropolitan area

- Examine the impact of this novel risk of reidentification in achieving deidentification of neuroimaging data under the US regulatory regime.

Results

- Simulation Analysis



The relationship between accuracy and population size is roughly *linear in log-log space*, consistent with theoretical results.[4]

- Reidentification performance across different population sizes -

157 (Schwarz et al., 2021)	6,500 (narrow age range, gender & race)	70,000 (narrow age range & gender)	423,000 (broad age range & gender)	865,000 (broad age range only)
37.6%	8.6%	2.4%	0.9%	0.6%
9.6%	0.8%	0.2%	0.05%	0.03%

6,500 (a Black female, age 25–29), 70,000 (a female age, 25–29), 423,000 (a female age, 20–49), and 865,000 (an adult, age 20–49)

- Regulatory Analysis

It is *unlikely* that the risk of reidentification via face recognition would affect achieving deidentification under the current US regulatory standards.

	Deidentification standard	Would defaced data still meet the standard?
Common Rule	The identity of the subject is not readily ascertainable (OHRP Guidance, 2008)	Yes
HIPAA	Expert determination (45 CFR §164.514(b)(1))	Yes
	Safe harbor (45 CFR §164.514(b)(2))	Yes

Discussion

- The results of our study suggest a need for
  - A more rigorous empirical analysis of the risk of reidentification
  - Developing best practices for sharing human neuroimaging data to better inform researchers of the standards and due diligence beyond the regulatory requirements.
  - Implementing desirable measures and mechanisms in data repositories for preservation and sharing of human neuroimaging data.
  - Developing technical countermeasures to the novel privacy attack

[1] Jwa, Koyejo, and Poldrack, 2024., *Imaging Neuroscience;* [2] Eke et al., 2021. *Neuroimage Reports*, [3] Schwarz et al., 2021. *Neuroimage*, [4] Zheng et al., 2018. *J. Mach. Learn. Res.*