



De-anonymization Attacks on Neuroimaging Datasets

Vikram Ravindra
ravindrv@purdue.edu
Purdue University
West Lafayette, Indiana

Ananth Grama
ayg@cs.purdue.edu
Purdue University
West Lafayette, Indiana

ABSTRACT

Advances in imaging technologies, combined with inexpensive storage, have led to an explosion in the volume of publicly available neuroimaging datasets. Effective analyses of these images hold the potential for uncovering mechanisms that govern functioning of the human brain, and understanding various neurological diseases and disorders. The potential significance of these studies notwithstanding, a growing concern relates to the protection of privacy and confidentiality of subjects who participate in these studies. In this paper, we present a de-anonymization attack rooted in the innate uniqueness of the structure and function of the human brain. We show that the attack reveals not only the *identity of an individual*, but also the *efficacy with which they performing cognitive tasks*. Our attack relies on novel matrix analyses techniques that are used to extract discriminating features in neuroimages. These features correspond to individual-specific signatures that can be matched across datasets for highly accurate identification. We present data preprocessing, signature extraction, and matching techniques that are computationally inexpensive, and can scale to large datasets. We characterize the efficacy of our de-anonymization attacks on publicly available databases. Finally, we discuss implications of the attack and challenges associated with defending against such attacks.

CCS CONCEPTS

• **Social and professional topics** → **Patient privacy**; • **Security and privacy** → *Pseudonymity, anonymity and untraceability; Data anonymization and sanitization*; • **Applied computing** → *Imaging*.

KEYWORDS

de-anonymization attack; privacy concerns in medical images

ACM Reference Format:

Vikram Ravindra and Ananth Grama. 2021. De-anonymization Attacks on Neuroimaging Datasets. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*, June 20–25, 2021, Virtual Event, China. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3448016.3457234>

1 INTRODUCTION

Functional Magnetic Resonance Images (fMRIs) are low resolution 4-dimensional images (three spatial dimensions and one temporal dimension) of the brain. At a high level, an fMRI is a group of time-varying signals, each of which has an associated 3-D physical location

corresponding to a region of the brain. fMRI measures local changes in blood flow to quantify the activity of different regions of the brain.

In this paper, we focus on privacy aspects of fMRIs. We present a de-anonymization attack on neuroimages in the public domain. An illustrative attack is shown in Figure 1. Our attack model assumes that there is an entity (the attacker) with a set of personally identifiable neuroimages (potentially from the past) from third party imaging facilities. We also assume a public domain database of de-identified patient records that contain neuroimages, along with sensitive information such as the patient genome, and associated pathologies. This is indeed a typical scenario today, wherein imaging centers have access to identifiable imaging data but not the complete Electronic Health Record (EHR). Separately, repositories from research efforts such as the Human Connectome Project (HCP) release large amounts of de-identified neuroimaging data, metadata, and often, highly sensitive genomic and phenotypic (disease) data. We demonstrate that using our technique, imaging centers can reference imaging data across these datasets, and gain access to patient metadata and health records (including highly sensitive genomic variants and pathologies). We use matrix sampling to identify subjects with high accuracy. Our proposed methods provide theoretical guarantees, while performing extremely well in practice, in terms of de-identification accuracy, computational cost, and scalability to large datasets.

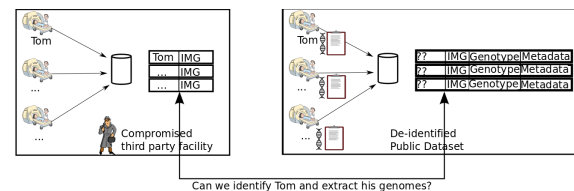


Figure 1: A schematic representation of a de-anonymization attack on neuroimages. Here, a patient (Tom) has a functional MRI in a compromised third party facility. Tom is also part of a de-identified public dataset, which has more information such as genotypic data and other metadata (age, ethnicity, medical record, etc.). In this paper, we show that we can identify Tom's de-identified record in the public dataset.

We note that while some preliminary work (see Section 4) has addressed the problem of identifying individuals from neuroimages, there have been *no* prior work that explores the privacy implications of identifying individuals. In this context, our results are entirely novel. Even for the problem of identifying individuals, our methods are much more powerful than existing techniques, in that they map directly back to data (image) features and regions. This is critical both for identifying signature regions in the brain, as well as for constructing effective defenses against de-anonymization attacks.



This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGMOD '21, June 20–25, 2021, Virtual Event, China.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8343-1/21/06.

<https://doi.org/10.1145/3448016.3457234>

Our demonstrated attack does not rely on any meta-data such as age, sex, ethnicity, education, intelligence, etc. Instead, it relies on patterns in the functional image, which we show are unique to individuals. Learning from our attack, we also highlight challenges associated with simultaneously defending against the attack, while retaining integrity of data in a way that supports downstream data analyses.

2 METHODOLOGY

We assume an attacker has two functional MRI datasets (taken at different instances – potentially, over years) in their possession and that one of them is de-anonymized. The attacker wishes to de-anonymize the other dataset. However, such an attacker would first need to convert raw neuroimages into a format suitable for analysis, and then extract individual-specific brain signatures from these processed images. We first provide a brief-overview of the image pre-processing to motivate our problem. Then, we discuss our attack, which is based on matrix row-sampling that relies on leverage-scores.

2.1 Common de-identification procedure

The standard process of de-identification of publicly available neuroimages consists of scrubbing identifying voxels and de-identifying meta-data associated with images. Facial features (such as cheeks, jaw, teeth, and nose) can be used to reconstruct the face and match them to photographs. To overcome this threat to privacy, facial voxels are removed in a process called defacing [4, 5, 15, 19]. Further, many public repositories also remove voxels corresponding to the skull in a standard procedure called skull stripping. In our work, we use the Brain Extraction Tool (BET) by Smith [22], which is part of the open-source FMRI Standard Library (FSL). Meta-data such as name, SSN, contact details, and other directly identifiable data are also removed [14]. Images and corresponding meta-data (such as age, gender, presence of disease, etc.) are consolidated under subject IDs (an alpha-numeric identifier unique to each subject). We note that these steps are necessary to make datasets consistent with HIPAA's Standards for Privacy of Individually Identifiable Health Information [16]. However, GDPR (in EU) is much broader in scope. Hence, data that is "de-identified" according to HIPAA is not necessarily "anonymized" according to EU [26].

2.2 Standard Image Pre-processing

Conventionally, an fMRI is expressed as a matrix $\mathbf{M}$, where each row is the time-varying signal (time-series) corresponding to a small volume of the brain, and each column is a snapshot of the brain at a point in time. In other words, $\mathbf{M} \in \mathbb{R}^{n \times t}$, where n is the number of voxels and t is the number of time-steps. We use the Minimal Preprocessing Pipeline [13]¹ to deal with spatial artifacts, as well as to normalize images to a common coordinate system. We then average voxels belonging to a common anatomical region using the Glasser Atlas [12]. We note that the aforementioned steps are commonly used in the neuroimaging community, and implemented in open-source packages such as FSL² and FreeSurfer³.

This preprocessing results in a time-series matrix $\mathbf{T} \in \mathbb{R}^{r \times t}$, where r is the number of regions. Finally, we compute the similarity matrix $\mathbf{S} \in \mathbb{R}^{r \times r}$ by computing correlations (dot products over time series

vectors) across regions. These matrices are called Functional Connectomes [24], since each element in the matrix quantifies the extent of co-firing across regions of the brain in a specific functional context. This procedure is repeated for all images of all subjects in the dataset.

2.3 Our Setup

Each element in the similarity matrix $\mathbf{S}$ can be viewed as a feature of the corresponding connectome, and the goal is to find a compact set of features that distinguish subjects from each other. To accomplish this, we vectorize the similarity matrix by stacking its columns one on top of the other. In this manner, each 4D image is reduced to a single (column) feature vector with r^2 elements (recall that r is the number of regions in the atlas). We take all of the feature vectors corresponding to images in our de-anonymized set and organize them into a matrix $\mathbf{G}_1$. Each column of this matrix corresponds to a subject, and each row corresponds to a feature from the similarity matrix. We compute a second matrix $\mathbf{G}_2$ in a similar manner from all of the images in the anonymous image dataset. The goal is to use the de-anonymized matrix $\mathbf{G}_1$ to identify patients in $\mathbf{G}_2$.

2.4 Principal Features Subspace Method

Let $\mathbf{G}_1$ be the matrix of connectomes, and $\mathbf{U}$ be the orthonormal matrix that spans the column space of $\mathbf{G}_1$. An example of such a matrix $\mathbf{U}$ is constructed using left singular vectors from a Singular Value Decomposition (SVD) of $\mathbf{A}$. Let t be the number of features that need to be retained. We can then compute the leverage(l) scores of $\mathbf{G}_1$ as $l_i = \|\mathbf{U}_i\|_2^2$, $\forall i \in \{1 \dots m\}$

We sort the leverage scores in descending order, and select the features corresponding to the top t leverage scores. Then, we create $\tilde{\mathbf{G}}_1 \in \mathbb{R}^{t \times n}$ and $\tilde{\mathbf{G}}_2 \in \mathbb{R}^{t \times n}$ by sampling the rows corresponding to the top t leverage scores of $\mathbf{G}_1$. Finally, we compute the Pearson Correlations $\mathbf{C}$ between all pairs of columns (patients) of $\tilde{\mathbf{G}}_1$ and $\tilde{\mathbf{G}}_2$. For each row of $\mathbf{C}$, we find the maximum entry. Let $C_{i,j}$ be the maximum entry of row i in $\mathbf{C}$. We predict that patient indexed i in $\mathbf{G}_1$ is the same as patient indexed j in $\mathbf{G}_2$. We note that this approach is adapted from Ravindra et al. [20], where it was shown that $t = 100$ provides high prediction accuracy, which is what we use in our results as well.

In contrast to prior randomized approaches [7, 9], we select features in a deterministic manner; Cohen et al. [6] provide theoretical bounds for this selection process. In our application, if the features corresponding to top leverage scores are common across different sub-populations, it suggests that these features are sufficient to launch a de-anonymization attack on neuroimaging datasets. Furthermore, for a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, with $m \gg n$ (as is our setting), Drineas et al. [8] show that computational cost of leverage scores is $O(mn^2)$. Note that the squared term is the smaller dimension, hence computing leverage scores is inexpensive.

3 EXPERIMENTAL EVALUATION

3.1 Datasets

The Washington University-University of Minnesota (WU-Minn) project, which is part of Human Connectome Project consortium conducted a Healthy Young Adult study [10] resulting in 3T structural and functional MRI for 1113 adults, 7T resting and task magnetoencephalography (MEG) from 184 subjects, and 3T and 7T diffusion data. In this paper, we use resting state functional MRI and task functional MRI from the 100 "unrelated subjects" subset of the Young

¹<https://github.com/Washington-University/HCPpipelines>

²<https://fsl.fmrib.ox.ac.uk/>

³<https://surfer.nmr.mgh.harvard.edu/>

Adult database. We refer interested readers to Smith et al. [23] and Barch et al. [1] for details about the acquisition protocols for rest and task datasets respectively. The ADHD-200 dataset was collected by the International Neuroimaging Datasharing Initiative (INDI), as part of a collaboration of eight imaging sites around the world. It consists of multiple scans from 362 children and adolescents diagnosed with ADHD, along with 585 controls. The publicly available dataset consists of both structural MRI, resting state functional MRI, and associated phenotypic information [2].

3.2 Results

3.2.1 De-anonymization of subjects across tasks. We first demonstrate an efficient de-anonymization attack in task functional MRI. We assume that one data-set is de-anonymized, and we are interested in characterizing its impact on the anonymity on other datasets. The tasks performed by subjects in the two datasets are potentially different. For instance, the REST1 may be de-anonymized and we are interested in finding the correspondences in the LANGUAGE task. The accuracy in de-anonymizing a dataset of resting-state functional MRIs in the HCP was found to be in excess of 94%. However, task fMRIs are typically more complex, since task driven brain activities are more complex than spontaneous firings recorded in resting state sessions.

As described in Section 2.3, we construct the de-identified matrix G_1 and anonymized matrix G_2 . In this experiment, G_1 has the first rest session and the first half of the sessions for the seven tasks in HCP. In G_2 , we have the second rest session, along with the the second part of the same seven tasks. We stack the vectorized network representations of the images into two matrices – one for each group. The feature space of both groups are restricted to the features with top 100 leverage scores (i.e., $t = 100$, see Section 2.4) in the first group, as explained in Section 2.3. The results for this experiment are presented in Figure 2(a). It shows that the accuracy of identification is stronger along the diagonal, i.e., when both the datasets correspond the same task. The accuracy is particularly high ($> 94\%$) when they both are resting state functional images. Similarly, accuracies for language processing and relational processing tasks are $> 90\%$, and for social processing task is $> 80\%$. Surprisingly, MOTOR and WM tasks are ineffective in predicting the correspondence, even for the same task. We also can see that the matrix is clearly asymmetric, implying that de-anonymization of different tasks has varying impact on the anonymity of other datasets. On the whole, we observe that if a resting-state dataset is de-anonymized, it can be used to de-anonymize other tasks with high accuracy. This is particularly worrisome, since most studies produce resting state functional images. On the other hand, de-anonymizing tasks such as working memory and motor tasks reveals relatively less about the identities of individuals in other datasets.

3.2.2 Prediction of performance during task. In this experiment, we predict how well they perform at the task. HCP provides task-performance metrics, which is a measure of effectiveness during the language processing, emotion processing, relational processing, and working memory tasks. This measure is expressed as a percentage accuracy of correct replies for the respective tasks. We show that our signatures can be used to predict the performance metrics of new subjects.

Task	Train nRMSE (in %)	Test nRMSE (in %)
Language	0.33 ± 0.11	1.52 ± 0.20
Emotion	0.28 ± 0.07	0.60 ± 0.37
Relational	0.44 ± 0.04	2.74 ± 0.34
Working Memory	0.57 ± 0.12	1.93 ± 0.41

Table 1: Task-wise prediction error expressed as normalized root-mean-squared error.

We divide the subjects into two sets, chosen at random – train and test, of sizes 80 and 20 subjects, respectively. For each task, we compute the $G_{1,train}$ and $G_{1,test}$ from the first half of each task session and $G_{2,train}$ and $G_{2,test}$ from the second half. All the matrices consist of vectorized correlation matrices of the subjects for a particular task. The target vector consists of the performance metric for the same task. Since we have the ground truth, i.e., the performance metrics for the test set, we can measure the goodness-of-fit for our model.

First, we compute leverage scores for the $G_{1,train}$. We restrict the feature-space to the top leverage scores. We use Support Vector Machine (SVM) for regression, with the performance metric as the target. The coefficients learned from the train-set are used to predict the scores in the test set. To avoid biases due to choices of train and test set, we repeat the experiment 1000 times.

The results are shown in Table 1 for both train and test. All the errors are within 4% of the true value. We note that in our setup, we do not use timing information. This information encodes the exact time frames for the different blocks of stimuli within an experiment. Most studies also provide performance metrics within each time-block. The use of this additional data further improves prediction, and provides deeper insights that predict the neuronal response of individuals to particular sub-types of stimuli, such as math and story inputs, which is part of the language task.

3.2.3 Prediction of Identity of subjects with ADHD. In this section, we extend the concept of brain signatures to subjects with neurological disorders such as Attention Deficit Hyperactive Disorder (ADHD). We will use the dataset from the ADHD-200 dataset. As before, we stack the vectorized connectomes of subjects into two matrices G_1 and G_2 , one per session. Since we use the AAL2 atlas for the ADHD-200 dataset, each data-point now has 6670 features.

The stark differences between intra-subject similarity and inter-subject similarity for ADHD Subtypes 1 and 3 are shown in Figures 2(b) and 2(c). We divide the subjects into train and test sets, and pick features that correspond to the highest leverage scores of the train matrix, as described in Section 2.3. When we restrict the feature space in the test subjects to the same features, we find the test accuracy to be $97.2 \pm 0.9\%$. This result is of significant interest for the following reasons. First, the acquisition protocol of ADHD-200 is different from HCP; also, the atlas used here is drawn differently from the one used in the previous experiments, and has different number of regions. Second, the subjects in this dataset are children with ADHD, whereas the HCP dataset is comprised of healthy young adults. Despite all of these differences, we find a consistent signature that can correctly identify subjects with high accuracy. Finally, functional MRIs of patients exist in patient records outside of neuroimaging studies. Hence, identifiable data of such patients is often not stored with the same strict protocols, thereby increasing vulnerability to de-anonymization attacks. Finally, even if the dataset is a combination of cases and controls, which is often the case, we can correctly match scans across datasets with prediction accuracy of $94.12 \pm 3.4\%$.

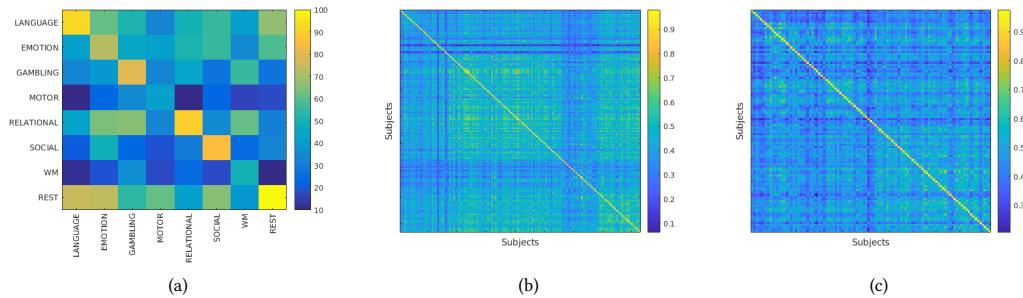


Figure 2: Identifiability of subjects under different settings. (a) Subjects are performing different tasks. The rows consist of de-anonymized datasets and columns consist of datasets in which subjects are anonymous. This figure shows the effect of de-anonymizing different datasets. (b) Correlation matrix of subjects with ADHD subtype 1 and (c) subtype 3 in the ADHD-200 dataset. The strong values on the diagonal suggests strong inter-session similarity between scans of the same subject. The low off-diagonal values suggest a strong dissimilarity between scans of different subjects.

4 RELATED LITERATURE

The United States Department of Health and Human Services (HHS) has specified standards for the privacy of identifiable health information. The legal requirements of these standards include removal of identifiable information, such as name, social security number (SSN), contact details, home address, etc. These standards are broadly drafted for all studies involving human subjects. The applicability of these standards to specific neuroimaging studies for medical doctors or non-medical scientists is complicated, and is described in detail by Tovino [25]. Similar standards are established by organizations such as the National Institute of Health (NIH) and Organization for Economic Cooperation and Development (OECD) [18]. To address these standards, Plis et al. [17] propose a de-centralized approach to simplify data-sharing whilst fostering collaborative research. The anonymization of subjects, as mandated by the HHS Privacy Rules, protects against naive attackers, but does not account for attacks by a competent adversary. Raw 3D images of the brain, regardless of modality – structural MRI, CT scans, MEGs, etc. also include facial features. The problem of de-identification was first studied by Chen et al. [5]. Budin et al. [4] showed that the chances of matching a 3D reconstructed face with the photograph of the subject was greater than what was expected by chance. With increasing resolution of structural images, various techniques have been developed to address threats to privacy. The process of skull stripping – removing non-brain voxels, was developed well before this threat was first studied. Skull stripping classifies voxels as brain and non-brain, and masks the latter. This procedure is effective in removing the skull, when provided with a faceless image, but ineffective in presence of facial features, since the algorithms confuse facial voxels and brain voxels [3].

The security and privacy implications of functional MRI data are relatively unexplored in the security and database communities, as they relate to existence, identification, and use of brain signatures. Brain signatures or brain fingerprints are patterns in the functional brain image that are unique to individuals. Finn et al. [11] show that two functional connectomes belonging to the same subject are more similar than functional connectomes drawn from two different subjects. It has also been shown that these distinct profiles can be used to predict fluid intelligence. The accuracy in identification was shown to be around 90%, and when restricting the analysis to the parieto-frontal region, the accuracy of identification was in excess

of 95% in resting-state images. Further, it has also been shown that image registration is not required to identify individuals [21].

A leverage-score based sampling method was first presented by Ravindra et al. [20]. Leverage-score sampling identified a small set of discriminative features that strongly expressed the signature. These features are used to identify small regions, called parcels on the cerebral cortex, which are important to the task of identification. While these methods and the subsequent findings are important to science, they are yet to be studied in an adversarial setting. Stated otherwise, while many methods have been developed to identify distinguishing patterns in functional images, the goal of our paper is to use a combine known and new methods to demonstrate how an adversary could launch a de-anonymization attack on a neuroimaging dataset.

5 DISCUSSION

Imaging data comprises the largest component of clinical datasets, representing immense value in diagnostics and prognostics for diverse pathologies. A particularly important class of imaging datasets corresponds to the brain – with its associated use in understanding structure, function, and pathology in neurodegeneracy. Effective management and analyses of brain imaging data is therefore of profound and increasing importance, particularly in view of the rapidly increasing brain imaging data in public domain repositories. Our paper is the first to address important privacy concerns in fMRIs caused by individual-specific connectomic signatures. We note that these datasets are released not just for medical applications, but also to understand brain response to various stimuli, including watching media content, playing video games, and partaking in different activities. Such technologies will be increasingly adopted in applications ranging from VR/AR environments and video games, to semi-autonomous military platforms. This paper identifies major privacy issues in functional neuroimages that must be resolved before large databases are released for public access. Our observations motivate important new interdisciplinary efforts between neuroimaging experts, data management researchers, application scientists, and privacy experts to identify acceptable tradeoffs between utility, accuracy of analytics, and loss in subject privacy.

ACKNOWLEDGMENT

This work is supported by NSF grants IIS-1546488, OCI-1908691, CCF-0939370.

REFERENCES

- [1] Deanna M. Barch, Gregory C. Burgess, Michael P. Harms, Steven E. Petersen, Bradley L. Schlaggar, Maurizio Corbetta, Matthew F. Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, Dan Nolan, Edward Bryant, Tucker Hartley, Owen Footer, James M. Bjork, Russ Poldrack, Steve Smith, Heidi Johansen-Berg, Abraham Z. Snyder, and David C. Van Essen. 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* 80 (2013), 169–189. <https://doi.org/10.1016/j.neuroimage.2013.05.033> Mapping the Connectome.
- [2] Pierre Bellec, Carlton Chu, François Chouinard-Decorte, Yassine Benhajali, Daniel S. Margulies, and R. Cameron Craddock. 2017. The Neuro Bureau ADHD-200 Preprocessed repository. *NeuroImage* 144, Pt B (2017), 275–286.
- [3] Amanda Bischoff-Grethe, I Burak Ozyurt, Evelina Busa, Brian T Quinn, Christine Fennema-Notestine, Camellia P Clark, Shauna Morris, Mark W Bondi, Terry L Jernigan, Anders M Dale, Gregory G Brown, and Bruce Fischl. 2007. A technique for the deidentification of structural brain MR images. *Human brain mapping* 28, 9 (2007), 892–903.
- [4] François Budin, Donglin Zeng, Arpita Ghosh, and Elizabeth Bullitt. 2008. Preventing facial recognition when rendering MR images of the head in three dimensions. *Medical Image Analysis* 12, 3 (2008), 229–239.
- [5] Joseph J Chen, Khan M Siddiqui, Leslie Fort, Ryan Moffitt, Krishna Juluru, Woojin Kim, Nabile Safdar, and Eliot L Siegel. 2007. Observer success rates for identification of 3D surface reconstructed facial images and implications for patient privacy and security. In *Medical Imaging 2007: PACS and Imaging Informatics*, Vol. 6516. International Society for Optics and Photonics, San Diego, USA, 65161B.
- [6] Michael B. Cohen, Sam Elder, Cameron Musco, Christopher Musco, and Madalina Persu. 2015. Dimensionality Reduction for k-Means Clustering and Low Rank Approximation. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing (STOC '15)*. ACM, New York, NY, USA, 163–172. <https://doi.org/10.1145/2746539.2746569>
- [7] P. Drineas, R. Kannan, and M. W. Mahoney. 2006. Fast Monte Carlo Algorithms for Matrices I: Approximating Matrix Multiplication. *SIAM J. Comput.* 36, 1 (2006), 132–157.
- [8] Petros Drineas, Malik Magdon-Ismael, Michael W Mahoney, and David P Woodruff. 2012. Fast approximation of matrix coherence and statistical leverage. *The Journal of Machine Learning Research* 13, 1 (2012), 3475–3506.
- [9] Petros Drineas, Michael W. Mahoney, and S. Muthukrishnan. 2008. Relative-Error CUR Matrix Decompositions. *SIAM J. Matrix Anal. Appl.* 30, 2 (Sept. 2008), 844–881. <https://doi.org/10.1137/07070471X>
- [10] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil. 2013. The WU-Minn Human Connectome Project: An overview. *NeuroImage* 80 (2013), 62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- [11] Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience* 18, 11 (2015), 1664–1671.
- [12] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. Van Essen. 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536 (2016), 171–178.
- [13] Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R. Polimeni, David C. Van Essen, and Mark Jenkinson. 2013. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80 (2013), 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127> Mapping the Connectome.
- [14] Daniel S Marcus, Michael P Harms, Abraham Z Snyder, Mark Jenkinson, J Anthony Wilson, Matthew F Glasser, Deanna M Barch, Kevin A Archie, Gregory C Burgess, Mohana Ramaratnam, et al. 2013. Human Connectome Project informatics: quality control, database services, and data visualization. *NeuroImage* 80 (2013), 202–219.
- [15] Mikhail Milchenko and Daniel Marcus. 2013. Obscuring Surface Anatomy in Volumetric Imaging Data. *Neuroinformatics* 11, 1 (2013), 65–75.
- [16] HHS Office for Civil Rights. 2002. Standards for privacy of individually identifiable health information. Final rule. *Federal register* 67, 157 (2002), 53181–53273.
- [17] Sergey M. Plis, Anand D. Sarwate, Dylan Wood, Christopher Dieringer, Drew Landis, Cory Reed, Sandeep R. Panta, Jessica A. Turner, Jody M. Shoemaker, Kim W. Carter, Paul Thompson, Kent Hutchison, and Vince D. Calhoun. 2016. COINSTAC: A Privacy Enabled Model and Prototype for Leveraging and Processing Decentralized Brain Imaging Data. *Frontiers in Neuroscience* 10 (2016), 365. <https://doi.org/10.3389/fnins.2016.00365>
- [18] Jean-Baptiste Poline, Janis L. Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O. Halchenko, Michael Hanke, Christian Haselgrove, David B. Keator, Daniel S. Marcus, Russell A. Poldrack, Yannick Schwartz, John Ashburner, Karl G. Helmer, and David N. Kennedy. 2012. Data Sharing In Neuroimaging Research. *Frontiers in Neuroinformatics* 6, MARCH (2012), 9. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3319918/pdf/>
- [19] Fred W Prior, Barry Brunsden, Charles Hildebolt, Tracy S Nolan, Michael Pringle, S Neil Vaishnavi, and Linda J Larson-Prior. 2008. Facial recognition from volume-rendered magnetic resonance imaging data. *IEEE Transactions on Information Technology in Biomedicine* 13, 1 (2008), 5–9.
- [20] Vikram Ravindra, Petros Drineas, and Ananth Grama. 2018. Constructing Compact Brain Connectomes for Individual Fingerprinting. *CoRR* abs/1805.08649 (2018), 1–14. [arXiv:1805.08649](https://arxiv.org/abs/1805.08649) <http://arxiv.org/abs/1805.08649>
- [21] Vikram Ravindra, Huda Nassar, David F Gleich, and Ananth Grama. 2019. Rigid graph alignment. In *International Conference on Complex Networks and Their Applications*. Springer, Lisbon, Portugal, 621–632.
- [22] Stephen M Smith. 2002. Fast robust automated brain extraction. *Human brain mapping* 17, 3 (2002), 143–155.
- [23] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Ugurbil, D. C. Van Essen, and M. F. Glasser. 2013. Resting-state fMRI in the Human Connectome Project. *NeuroImage* 80 (2013), 144–168. <https://doi.org/10.1016/j.neuroimage.2013.05.039>
- [24] Olaf Sporns. 2010. *Networks of the Brain*. MIT press, USA.
- [25] Stacey A Tovino. 2005. The Confidentiality and Privacy Implications of Functional Magnetic Resonance Imaging. *The Journal of Law, Medicine & Ethics* 33, 4 (2005), 844–850.
- [26] Johns Hopkins University. 2018. GDPR Application in Research Settings. https://www.jhpsu.edu/offices-and-services/institutional-review-board/_pdfs-and-docs/GDPR_Application%20in%20Research%20Settings.pdf