# Towards an evolution in the characterization of the risk of re-identification of medical images

Antoine Boutet, Carole Frindel, Mohamed Maouche

# Towards an evolution in the characterization of the risk of re-identification of medical images

Antoine Boutet
*Univ. Lyon, INSA Lyon, Inria, CITI*
*Lyon, France*
antoine.boutet@insa-lyon.fr

Carole Frindel
*Univ. Lyon, INSA Lyon, CREATIS*
*Lyon, France*
carole.frindel@creatis.insa-lyon.fr

Mohamed Maouche
*Univ. Lyon, Inria, INSA Lyon, CITI*
*Lyon, France*
mohamed.maouche@inria.fr

***Abstract*—As facial recognition technology proliferates, concerns emerge regarding its application to medical imaging, specifically Magnetic Resonance Imaging (MRI). This paper investigates privacy risks associated with MRI data, including re-identification through social network photographs and sensitive attribute inference. The exponential growth in MRI quality coincides with the increasing sophistication of facial recognition tools, raising the potential for re-identification using medical images. Our attack involves reconstructing faces and applying facial recognition techniques to extract identifying features that can be compared to photographs. Legal frameworks like GDPR mandate the assessment and protection of personal data, necessitating continuous risk evaluation. Beyond re-identification, we explore the inference of individual attributes from MRI images, such as age, gender, and ethnic group. This research assesses the privacy risks associated with MRI data by taking into account the evolution of facial recognition and reconstruction tools that have become increasingly accessible. We also show that facial hair removal technique on photographs increases the risk of re-identification. Overall, our results highlight vulnerabilities in sharing MRI data, emphasizing the need for enhanced privacy safeguards.**

***Index Terms*—Privacy, Risk Assessment, Re-Identification, Medical Images**

## I. Introduction

Face recognition systems are increasingly deployed for the authentication process as well as mass surveillance programs [1]. These systems are typically built by scraping publicly available images from social media. Smart cameras equipped with facial recognition are becoming a new threat to privacy. However, this risk does not only concern images and facial recognition tools can be used also on medical imaging.

In recent years, medical professionals have increasingly relied on various imaging technologies for diagnosing patients. Brain imaging, in particular, has seen remarkable advancements, with Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Computed Tomography (CT) being among the key imaging modalities. Moreover, MRI, which continues to evolve in terms of both modalities and contrast techniques, provide diverse capabilities for visualizing brain tissues, ranging from highlighting fat (T1-weighted), fluids (T2-weighted), and lesions (FLAIR, DWI) to mapping functional activity (fMRI) and vascular structures (SWI, MRA) and assessing macromolecular content (MTI).

However, this rapid enhancement in brain imaging quality, along with the widespread sharing of tagged personal images on social media, raises concerns regarding the potential for re-identification using facial recognition software. Instances of re-identification have been reported for anatomical MRI images [2, 3, 4], PET images [5, 5], CT images [6], as well as functional MRI (fMRI, dMRI, and ASF) images [7]. To mitigate this risk, de-facing software has been developed to obscure or replaces part of the human faces [8, 9]. Nevertheless, these mitigation measures often compromise image utility and do not provide complete protection against re-identification [10, 11].

In the pursuit of automating face recognition from medical images, methods for face reconstruction typically involve creating an isosurface and applying skin rendering [2, 3]. Recognition is then achieved through geometric structural properties [12] or dedicated neural networks [13]. Commercial face recognition software, such as Amazon Rekognition[1], Deep Vision AI[2], Face++[3] or Microsoft Azure Cognitive Services Face API[4] is increasingly integrated into face databases.

The rapid evolution of imaging technologies and the accessibility of certain tools gives rise to significant privacy questions. Legal frameworks like the European General Data Protection Regulation (GDPR), the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA), and the California's Consumer Privacy Act (CCPA) mandate the quantification of privacy risks and the effective protection and anonymization of personal data, particularly sensitive health-related information. These laws take into account evolving practices, available tools, and adversaries's capacities for identification. For example, under GDPR Article 29 [14], there is a strong emphasis on considering contextual factors and all potential identification methods, especially in light of recent technological advancements and increased computing power. Therefore, the evolving landscape of face reconstruction and recognition tools calls a reassessment of re-identification risk and the implementation of appropriate protection measures.

While re-identification is a central concern, privacy violations encompass more than just this risk. GDPR and similar regulations emphasize safeguarding against three primary risks: singling out, linkability, and inference. These risks

[1]https://aws.amazon.com/fr/rekognition/
[2]http://deepvisionai.in/
[3]https://www.faceplusplus.com/
[4]https://azure.microsoft.com/en-gb/products/cognitive-services/face

extend beyond mere identification, allowing data to be linked, and potentially disclosing highly sensitive information. In this paper, we also explore a less-studied aspect: the possibility of inferring individual attributes from MRI images with a high degree of confidence.

Imaging data, although not publicly available, resides in health data warehouses. Yet, this data is increasingly shared for research purposes or with private laboratories. Hospitals, in particular, face cybersecurity attacks, leading to data breaches in recent years [5]. Therefore, safegaurding this medical data at its source is crucial to prevent re-identification risks.

This paper aims to assess privacy violations related to brain imaging, considering the evolving landscape of available tools and data. Indeed, to establish the privacy risk assessment, we take into account both the severity of the impacts (which depends on the significance of the consequences and how difficult it would be for subjects to overcome them) and likelihood (which depends on the feasibility of the threat and its motivation). With the evolution of facial recognition and reconstruction tools, as well as their easier accessibility, we discuss the need to consider a higher likelihood. Additionally, we present a comprehensive evaluation campaign to illustrate these risks. Specifically, the contributions of the papers can be summarized as follows:

- We utilize a more robust setting to evaluate the privacy risk compared to the state of the art. Taking inspiration from research in other data modalities (for instance VoicePrivacy Challenge [15]).
- We design two different attacks of MRI face recognition and we compare them within the same dataset.
- We propose a hair-removing technique on the photographs to increase the effectiveness of the attack.
- We evaluate attribute inference on Age, Gender and Ethnic group, and compare the effectiveness between the attack on photographs or MRI.

Our results show that there are leakages in sharing MRI images (both in terms of re-identification risk and sensitive attribute inference). Also, our removing facial hair technique is important when using deep-learning based face recognition techniques.

## II. Related Work

In the literature, works have addressed the privacy concerns associated with the sharing of brain MRI data in various manners. However, a prevailing consensus has emerged, indicating that the safeguarding of brain MRI images predominantly relies on the application of defacing techniques. A comprehensive examination of state-of-the-art techniques is provided in the study of Schwarz et al. [9]. Starting with this principle some works have tackled the reversing of such defacing method. For instance, Abramian et al. [10] have used a CycleGAN [16] in order to reverse the defacing methods. Most specifically, they showed that the method that removes the face (*mri_deface* [17] for the FreeSurfer package [18]) is harder to reverse compared to a blurring method *MaskFace* [19]. In addition some works have tried to compute rapidly deface masks that are considered effective. Such as the work of Khazane et al. [20], where a 3D U-net is used to construct a faster version of pydeface [8] with high accuracy (up-to 10 times faster).

As for the attack that illustrates the privacy risk of sharing MRI images. Some have focuses on the inter-MRI linkability. For instance, Ravindra et al. [21] aim at linking fMRIs from two different datasets without face reconstruction. To do this, they use matrix decomposition to reduce the dimension of each data sample to a smaller vector (100 dimension only) from the functional connectomes matrix. Authors assume that patients are present in both datasets. Thus, the attacker can conduct a full bipartite graph matching between two pseudonymized datasets (a single permutation is found). They noticed that the accuracy is high ($> 94\%$) when both datasets correspond to resting state fMRI compared to the accuracy for language processing and relational processing tasks which is $> 90\%$, and for social processing task which is $> 80\%$.

Similarly to our setup, some works aim at finding the identity of the MRI owner from photographs. For instance, Mikulan et al. [2] evaluate their defacing method with both a Human-recognition and a Machine-recognition experiment. The machine recognition is constructed using the Surf-Ice Sofwtare [22] to generate 2D renders. The dlib package [23] is then used with the pre-trained detector from [24]. The main drawback of this work compared to ours is that their evaluation of machine-recognition is flawed because the average number of true identifications is defined as *"proportion of subjects in which the correct subject was among the ones indicated by the algorithm"*. A subject would be indicated by their algorithm only if its score was above a fixed threshold. This is a typical error in Machine Learning (ML) model evaluation where we can always artificially increase the true positive rate by choosing a low threshold but we should also look at the impact on the false positive rate. In other words, a detection system could accept the correct subject by accepting way more false subjects on the way (e.g., 1 correct subject accepted with 99 other false ones). To avoid such mistakes, in our work we use a more robust setting with identity verification and AUC in case of classification evaluation (see Section III). As for their proposed method *AnonMI*, they use the watershed algorithm from FreeSurfer [18] to reconstruct the skin and skull of the subject. Then, they use a template constructed from the IXI dataset to determine the location of ears and face to fill in the voxels with new values. All the values outside of this skin model are set to zero.

The attack described by Schwarz et al. [3] was used in various state-of-art studies [5, 9] in order to compare various defacing (or refacing) techniques. This attack was also used to study the privacy risk of different type of imaging techniques beyond classical MRI (dMRI, fmRI, ASL compared to classical T1-W, T2...) [7]. The base of the attack is to first apply thresholding and remove artifacts disconnected from the head, then create an isosurface using Matlab's built-in function, and

[5]https://www.upguard.com/blog/biggest-data-breaches-in-healthcare

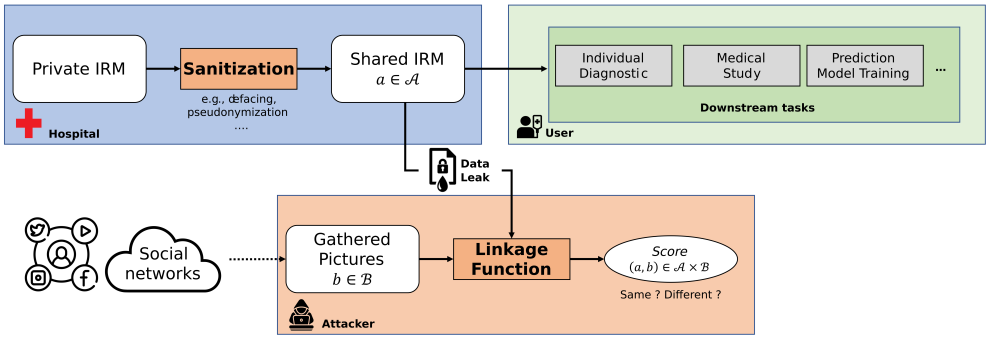


Fig. 1. Threat model: an adversary having access to MRI data tries to identify the associated individual from photographs found on social networks.

render it to static 2D images using Surf-Ice [22]. Then, if needed, the attacker conducts a refacing method using an average template. As for the face recognition itself, for each image, the attacker generates a single 3D surface that they transform into 10 renders (2D, png files) under a range of simulated viewing and lighting angles. These renders are used as training for the Microsoft Azure Face API PersonGroup classifier which is tested using 5 real photographs of each participant. The classifiers used Microsoft's pre-trained models for face detection (detection_03) and recognition (recognition_02). Unfortunately, these models are commercial Microsoft products, and their underlying algorithms have not been published. This why in our work we make use of deepface [25] to extract face recognition features with open source programs. Another differentiating aspect of our work is that to the best of our knowledge, we are the only one to inspect the effect of removing hair on photographs for the face recognition.

## III. Threat Model and Privacy Metrics

We opted for privacy evaluation protocols used in state-of-the-art competitions. Notably, the VoicePrivacy Challenge [15]. First, lets define the threat model depicted in Fig. 1. We consider a setting where MRI data become available online (honestly or dishonestly) after going through an anonymization method. This data is shared to participate in a given downstream taks. We call it the *utility task*, which can be as simple as a diagnostic by a physician or as advanced as the training of large deep learning models.

We consider that an attacker obtains this data $a \in \mathcal{A}$ (e.g., through a data leak in a hospital) and compares it to known data she has accumulated $b \in \mathcal{B}$ using a linkage function that outputs a score $s = LF(a, b)$. In our setting the known data $\mathcal{B}$ is a list of social network photos that could include the real photograph of the subject that produced $a$. For the metrics, we consider an oracle evaluator that analyzes the list of scores $s \in \mathcal{S}$ outputted by the attacker for given sets of $\mathcal{A}$ and $\mathcal{B}$ in order to assess how much information the attacker could have exploited to better link any given $a$ with the correct element $b$. The setting is often used to evaluate verification systems (e.g., for authentication) where we evaluate the atomic capacity of an attack to differentiate if a pair of data $(a, b)$ come from the same identity (called a *mated* pair). Unlike, the metrics that are based on finding a fixed correct match, our setting has the advantage to generalize to an arbitrary number of identities in both $\mathcal{A}$ and $\mathcal{B}$. For instance, compared to the re-identification rate which is highly sensitive to the number and the nature of elements in $\mathcal{B}$.

### A. Equal-Error-Rate – EER

The EER is the most known metric in verification systems. It assumes a threshold-based decision on the score $s$ (i.e., comparing it to a fixed threshold $t$). In that case, two types of error can be occur: (1) False alarms with rate $P_{fa}(t)$ and (2) Misses $P_{miss}(t)$. The EER corresponds to the error rate given with the threshold $t_e$ where both the error rates are equal.

$$EER = P_{fa}(t_e) = P_{miss}(t_e) \tag{1}$$

### B. Area under the ROC curve – AUC

We could consider the attacker as a classifier that tries to distinguish between mated and non-mated pairs. Then, we use classical method to evaluate binary classifiers such as the ROC-AUC which evaluates the quality of a binary classification task with every threshold used.

### C. Linkability

Linkability was also proposed in [26] for biometric template protection systems. The goal of this metric is to estimate the non-overlapping regions between the scores of mated pairs and non-mated pairs (Equation 2). The intuition is to evaluate how much an attacker could deduce the nature of a pair (i.e., mated) depending on how non-confusing the scores are (Equation 3).

$$L(s) = max(0, p(H|s) - (\bar{H}|s)), \tag{2}$$

$$L(\mathcal{S}) = \int p(H|s) \cdot L(s) ds, \tag{3}$$

where $s$ represents a given score and $p(H|s)$ represent the probability that a pair is mated $H$ (same subject) given the score $s$ (respectively, non-mated with $\bar{H}$). Thus $L(s)$ measures the chances that a score describes a mated pair rather than a non-mated pair. Finally $L(\mathcal{S})$ is the global linkability computed across all scores.

Fig. 2. Landmarks detected by *GEO* on both photographs and reconstruction.

## IV. Designed Attacks

### A. Face Recognition

In this study, we employed a multi-faceted approach to reconstruct facial images from MRI data and subsequently compared these reconstructions with photographs of subjects' identification images (id images) as well as facial photographs sourced from the LFW (Labelled Faces in the Wild) database [27] (social network images). The reconstruction process was accomplished using the 3D Slicer software [28], which allowed us to generate 3D models of subjects' faces from the MRI data. To optimize the reconstruction process, we employed various parameterizations within 3D Slicer: *Crop Volume with Isotropic Spacing* for uniform voxel sizes, *Thresholding in Segment Editor* to isolate the facial region, and *Smoothing Effect* to refine the 3D model's surface. These parameterization steps were crucial in creating accurate and visually appealing 3D models of subjects' faces from the MRI data, facilitating subsequent analyses and comparisons with reference photographs. For face recognition, we employed two distinct approaches: one based on geometrical features (named *GEO*) and the other on deep learning features extracted using the VGG neural network architecture (named Deepface).

*1) GEO:* For face recognition using geometrical features, we utilized the face_recognition library [24], which relies on the powerful dlib package (a versatile Python library renowned for its facial recognition capabilities [23]). It works by first detecting facial landmarks on the input images, which are critical points such as eyes, nose, mouth, and chin. These landmarks are located using a deep learning-based facial landmark detection model, typically a shape predictor trained on a large dataset of annotated facial landmarks as depicted in Fig. 2. Once the facial landmarks were detected using dlib, we computed a set of geometrical features, such as the distances between these landmarks, angles formed by specific points, and ratios between different facial measurements.

*2) Deepface:* To extract deep learning features, we utilized the *deepface* library [25], which incorporates the VGG-Face model. The VGG-Face model is a deep convolutional neural network pretrained on a vast dataset of facial images. It is specifically designed for face recognition and feature extraction. The VGG-Face model operates by passing facial images through its layers, extracting high-level features at various abstraction levels, effectively encoding unique facial characteristics into compact feature vectors.

To quantify the similarity between the reconstructed images and the reference photograph or LFW database images, we computed the Euclidean L2 distance between the feature vectors. This distance metric allowed us to quantitatively measure the dissimilarity between features extracted from the reconstructed faces and those from the photographs. This distance is converted to a similarity metric to it use as a linkage function $LF$ as described in Section III.

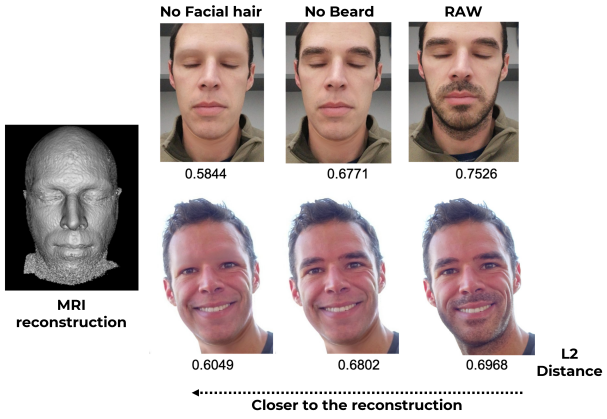


Fig. 3. Hair removal technique makes the reconstruction closer to the photograph (i.e., a smaller L2 distance).

*3) Baldness:* The rationale behind this approach lies in correcting the inherent inability of MRI reconstruction to faithfully reproduce hair details. Instead of fabricating artificial facial hair in these reconstructions, we opt to enhance the similarity of the photographs by eliminating facial hair from them. The method employed for hair removal in portrait images while maintaining facial structure and identity, as detailed in [29], consists of several steps. Initially, hair removal is achieved by creating paired latent codes for portraits, one with hair and one without hair. Subsequently, a HairMapper network is trained using these pairs, and the results are seamlessly blended with the original image through an image blending process. To enhance the quality of the blending, the hair mask is improved by dilation and blurring, ensuring a smoother transition. This pre-trained model was then applied to our dataset to generate "bald" versions of photographs.

### B. Sensitive Attribute Inference

To investigate the inclusion of gender, age, and ethnicity information within the feature vectors produced by Deepface, we employed the pre-trained model to extract feature vectors from each ID image and its corresponding MRI reconstruction. We omitted the final classification layer during this process. Next, we introduced a fully connected layer and conducted training specifically for attribute classification. Our evaluation involves a comprehensive assessment of the models' performance in attribute classification, based on ground truth data available for each subject in our dataset.

## V. Evaluation

### A. Dataset

This scientific paper reports on a study conducted at the University Hospital of Lyon between February and April

TABLE I
SUMMARY OF PRIVACY RESULTS

| Method | Facial hair | EER Max 50% | AUC Max 1 | Linkability Max 1 |
|---|---|---|---|---|
| Deepface | Raw | 41 | .54 | .07 |
| Deepface | Bald | **32** | **.71** | **.18** |
| Geo | Raw | 36 | .64 | .11 |
| Geo | Bald | 38 | .64 | .13 |

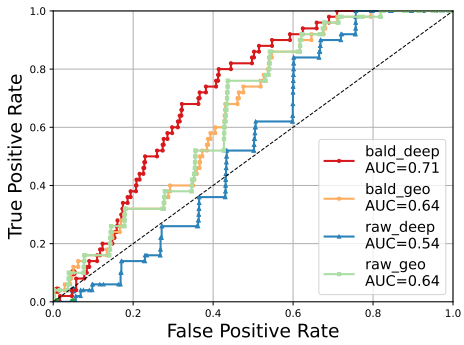


Fig. 4. ROC-Curve of the classification between mated and non-mated pairs: curves higher than the diagonal represent classifiers better than random (i.e., privacy leakage).

2022, involving 49 healthy volunteers. The study had stringent inclusion criteria, which required participants to be between 18 and 50 years of age, have no previous neurological issues or a history of neck surgery, and possess no contraindications for undergoing an MRI scan. Additionally, individuals with dental implants or braces were excluded from the study to reduce potential artifacts during image acquisition. Each volunteer provided their informed consent to participate in the study and to be part of this work, with a T2-weighted sagittal imaging using Turbo Spin Echo and their photograph collected for documentation purposes.

### B. Distinguishability of mated pairs

In Table I, we present the results of attacks conducted with the metrics defined in Section III. The results show that Deepface method is highly sensitive to the face hair of the subjects. But removing them using our method make the attack behave the best out of our 4 configurations. On the other hand, we notice that removing hair has little impact on *GEO* as expected since the face landmarks can be found even with hair present. Notably without removing the hair, *GEO* behaves better than deepface. This illustrates the sensitivity of deepface to hair. This is probably due to how the model of [24] was trained (similar hairstyle/facial hair in training).

The comparison is better illustrated in Fig. 4, where we compare the different classification between pairs using the Roc-curve. We notice that the Bald_deep combination behaves way better than random (diagonal line) showing that there is leakage of information and distinguishability between data pairs from same subjects compared to different subjects. To be noted, all of the metric order the attack in the same manner. EER also shows that there is still error in conducting the attack by setting up a threshold-based decision with an EER of 32%.

TABLE II
ACCURACY OF AGE, GENDER, AND ETHNICITY INFERENCE IN ID IMAGES AND MRI RECONSTRUCTIONS FOR THE SAME SUBJECTS

| Type | Age | Gender | Ethnicity |
|---|---|---|---|
| ID images | 0.4 | 1.0 | 0.4 |
| MRI reconstruction | 0.6 | 0.4 | 0.8 |

### C. Evaluation of sensitive attribute inference

In Table II, we present the results of our inference attacks on both the social media photograph and MRI. We notice for the Gender using the photograph was more effective compared to MRI. But, for both age and ethnic origin the MRI was more effective. Showing the high risk of sharing such type of data.

## VI. DISCUSSION AND CONCLUSION

In this paper, we contribute to evaluating privacy risks associated with MRI data. By designing attacks to do recognition through social network photographs and through sensitive attribute inference. Our results highlight vulnerabilities in sharing MRI data, emphasizing the need for enhanced privacy safeguards and continued vigilance. More precisely, we evaluate the leakage with both the *GEO* method and deepface, illustrating the impact of hair removal. We also discovered discrepancies in the inference of attributes when using an MRI or a photograph. However, there are some limitations that include the lack of testing of defacing methods and we did not evaluate the usage of commercial solutions for face recognition. Nonetheless, we introduced a more robust evaluation protocol in the context of MRI imaging that we advocate for. Also, to the best of our knowledge, we are the first work that compares different attacks and we also study the effect of facial hair removal.

In future work, we want to explore attack capacity beyond face recognition but rather find direct patterns on the brain tissues (e.g., with fingerprinting) in order to evaluate if defacing techniques are appropriate anonymization techniques.

## ACKNOWLEDGMENT

This work has been supported by the ANR 22-PECY-0002 IPOP (Interdisciplinary Project on Privacy) project of the Cybersecurity PEPR and by the FIL RIVIERA project.

## REFERENCES

[1] K. Hill, “Meet clearview ai, the secretive company that might end privacy as we know it,” 2023. [Online]. Available: https://www.chicagotribune.com/nation-world/ct-nw-nyt-clearview-facial-recognition-20200119-dkdqz7ypaveb3id42tpz7ymase-story.html

[2] E. Mikulan, S. Russo, F. M. Zauli, P. d’Orio, S. Parmigiani, J. Favaro, W. Knight, S. Squarza, P. Perri, F. Cardinale *et al.*, “A comparative study between state-of-the-art mri deidentification and anonyMI, a new method combining re-identification risk reduction and geometrical preservation,” Wiley Online Library, Tech. Rep., 2021.

[3] C. G. Schwarz, W. K. Kremers, T. M. Therneau, R. R. Sharp, J. L. Gunter, P. Vemuri, A. Arani, A. J. Spychalla, K. Kantarci, D. S. Knopman *et al.*, "Identification of anonymous MRI research participants with face-recognition software," *New England Journal of Medicine*, vol. 381, no. 17, pp. 1684–1686, 2019.

[4] C. G. Schwarz, R. C. Petersen, and C. R. Jack Jr, "Identification from MRI with face-recognition software. Reply." *The New England journal of medicine*, vol. 382, no. 5, pp. 490–490, 2020.

[5] C. G. Schwarz, W. K. Kremers, V. J. Lowe, M. Savvides, J. L. Gunter, M. L. Senjem, P. Vemuri, K. Kantarci, D. S. Knopman, R. C. Petersen *et al.*, "Potential for re-identifying brain PET research participants using face recognition," *Alzheimer's & Dementia*, vol. 18, p. e063651, 2022.

[6] C. L. Parks and K. L. Monson, "Automated facial recognition of computed tomography-derived facial images: patient privacy implications," *Journal of digital imaging*, vol. 30, no. 2, pp. 204–214, 2017.

[7] C. G. Schwarz, W. K. Kremers, A. Arani, M. Savvides, R. I. Reid, J. L. Gunter, M. L. Senjem, P. M. Cogswell, P. Vemuri, K. Kantarci, D. S. Knopman, R. C. Petersen, and C. R. Jack, "A face-off of MRI research sequences by their need for de-facing," *NeuroImage*, vol. 276, p. 120199, 2023.

[8] O. F. Gulban, D. Nielson, john lee, R. Poldrack, C. Gorgolewski, Vanessasaurus, and C. Markiewicz, "poldracklab/pydeface: Pydeface v2.0.2," Jul. 2022.

[9] C. G. Schwarz, W. K. Kremers, H. J. Wiste, J. L. Gunter, P. Vemuri, A. J. Spychalla, K. Kantarci, A. P. Schultz, R. A. Sperling, D. S. Knopman, R. C. Petersen, and C. R. J. Jr., "Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives," *NeuroImage*, vol. 231, p. 117845, 2021.

[10] D. Abramian and A. Eklund, "Refacing: Reconstructing anonymized facial features using GANs," in *ISBI*, 2019, pp. 1104–1108.

[11] C. G. Schwarz, W. K. Kremers, T. M. Therneau, R. R. Sharp, J. L. Gunter, P. Vemuri, A. Arani, A. J. Spychalla, K. Kantarci, D. S. Knopman *et al.*, "Popular MRI defacing software does not sufficiently protect participants from re-identification via face recognition: Neuroimaging/optimal neuroimaging measures for tracking disease progression," *Alzheimer's & Dementia*, vol. 16, p. e045157, 2020.

[12] Y.-A. Li, Y.-J. Shen, G.-D. Zhang, T. Yuan, X.-J. Xiao, and H.-L. Xu, "An efficient 3D face recognition method using geometric features," in *International Workshop on Intelligent Systems and Applications*, 2010, pp. 1–4.

[13] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in *international conference on computer vision workshops*, 2015, pp. 142–150.

[14] "Article 29 data protection working party," 2014. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

[15] N. Tomashenko, X. Wang, X. Miao, H. Nourtel, P. Champion, M. Todisco, E. Vincent, N. Evans, J. Yamagishi, and J.-F. Bonastre, "The VoicePrivacy 2022 Challenge evaluation plan," Technical Report, 2022.

[16] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *International Conference on Computer Vision*, 2017, pp. 2242–2251.

[17] A. Bischoff-Grethe, I. B. Ozyurt, E. Busa, B. T. Quinn, C. Fennema-Notestine, C. P. Clark, S. Morris, M. W. Bondi, T. L. Jernigan, A. M. Dale *et al.*, "A technique for the deidentification of structural brain mr images," *Human brain mapping*, vol. 28, no. 9, pp. 892–903, 2007.

[18] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[19] M. Milchenko and D. S. Marcus, "Obscuring surface anatomy in volumetric imaging data," *Neuroinformatics*, vol. 11, no. 1, pp. 65–75, 2013.

[20] A. Khazane, J. Hoachuck, K. J. Gorgolewski, and R. A. Poldrack, "Deepdefacer: Automatic removal of facial features via U-Net image segmentation," *CoRR*, vol. abs/2205.15536, 2022.

[21] V. Ravindra and A. Grama, "De-anonymization attacks on neuroimaging datasets," in *International Conference on Management of Data*, G. Li, Z. Li, S. Idreos, and D. Srivastava, Eds., 2021, pp. 2394–2398.

[22] C. Rorden, "Surf-Ice: A neuroimaging visualization and analysis tool," https://www.nitrc.org/projects/surfice/, 2021, accessed on Date 2023-09-26.

[23] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.

[24] A. Geitgey, "Face recognition," Jul. 2018. [Online]. Available: https://github.com/ageitgey/face_recognition

[25] S. I. Serengil and A. Ozpinar, "Lightface: A hybrid deep face recognition framework," in *Innovations in Intelligent Systems and Applications Conference*, 2020, pp. 23–27.

[26] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch, "General framework to evaluate unlinkability in biometric template protection systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 6, pp. 1406–1420, 2018.

[27] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[28] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[29] Y. Wu, Y.-L. Yang, and X. Jin, "Hairmapper: removing hair from portraits using gans," in *IEEE/CVF CVPR*, 2022, pp. 4227–4236.