


A comparative study between state-of-the-art MRI deidentification and AnonyMI, a new method combining re-identification risk reduction and geometrical preservation

Ezequiel Mikulan¹  | Simone Russo¹ | Flavia Maria Zauli¹ | Piergiorgio d'Orio² | Sara Parmigiani¹ | Jacopo Favaro³ | William Knight⁴ | Silvia Squarza⁵ | Pierluigi Perri⁶ | Francesco Cardinale² | Pietro Avanzini⁷ | Andrea Pigorini¹

¹Dipartimento di Scienze Biomediche e Cliniche "L.Sacco", Università degli Studi di Milano, Milano, Italy

²"Claudio Munari" Epilepsy and Parkinson Surgery Center, ASST Grande Ospedale Metropolitano Niguarda, Milano, Italy

³Department of Women's and Children's Health, Pediatric Neurology and Neurophysiology Unit, University of Padua, Padua, Italy

⁴Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

⁵Department of Neuroradiology, "Claudio Munari" Epilepsy and Parkinson Surgery Center, ASST Grande Ospedale Metropolitano Niguarda, Milano, Italy

⁶Dipartimento di Scienze Giuridiche Cesare Beccaria, Università degli Studi di Milano, Milano, Italy

⁷Istituto di Neuroscienze, Consiglio Nazionale delle Ricerche, Parma, Italy

Correspondence

Andrea Pigorini, Dipartimento di Scienze Biomediche e Cliniche "L.Sacco", Università degli Studi di Milano, Milano, Italy.
Email: andrea.pigorini@unimi.it

Funding information

European Commission, Grant/Award Number: Horizon 2020 Framework Programme for Research; Horizon 2020 Framework Programme, Grant/Award Numbers: 785907, 945539

Abstract

Deidentifying MRIs constitutes an imperative challenge, as it aims at precluding the possibility of re-identification of a research subject or patient, but at the same time it should preserve as much geometrical information as possible, in order to maximize data reusability and to facilitate interoperability. Although several deidentification methods exist, no comprehensive and comparative evaluation of deidentification performance has been carried out across them. Moreover, the possible ways these methods can compromise subsequent analysis has not been exhaustively tested. To tackle these issues, we developed AnonyMI, a novel MRI deidentification method, implemented as a user-friendly 3D Slicer plugin-in, which aims at providing a balance between identity protection and geometrical preservation. To test these features, we performed two series of analyses on which we compared AnonyMI to other two state-of-the-art methods, to evaluate, at the same time, how efficient they are at deidentifying MRIs and how much they affect subsequent analyses, with particular emphasis on source localization procedures. Our results show that all three methods significantly reduce the re-identification risk but AnonyMI provides the best geometrical conservation. Notably, it also offers several technical advantages such as a user-friendly interface, multiple input-output capabilities, the possibility of being tailored to specific needs, batch processing and efficient visualization for quality assurance.

KEYWORDS

data sharing, geometrical preservation, MRI deidentification, privacy

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

1 | INTRODUCTION

A growing amount of evidence points to open science as the most efficient strategy for understanding the structural and functional organization of the human brain (Milham et al., 2018). This notion has driven several recent unprecedented large-scale collaborative endeavors in neuroscience (Amunts et al., 2019; Mott, Gordon, & Koroshetz, 2018). As a natural consequence, in the last years, data sharing has promptly become a fundamental pillar of contemporary neuroscience. Indeed, it allows otherwise impossible large-scale and multi-centric analyses and also promotes transparency and reproducibility (Ascoli, Maraver, Nanda, Polavaram, & Armañanzas, 2017). Data sharing also maximizes the scientific profit that can be obtained from the data (Brakewood & Poldrack, 2013), reduces the economic cost of new studies, and provides valuable resources for researchers without access to expensive data acquisition equipment (Poline et al., 2012).

However, human data sharing also entails some risks, mostly related to privacy and confidentiality of personal information and metadata. Indeed, protecting the privacy and confidentiality of research subjects is not always granted, even though considerable efforts have been directed to establish data deidentification protocols and tools, and data protection legislation (Kalkman, Mostert, Gerlinger, van Delden, & van Thiel, 2019). Researchers must ensure that they adhere to the applicable data protection legislation. Within Europe, the General Data Protection Regulation (GDPR; EU 679/2016) establishes the law on data protection issues. Brought into force in May 2018, the GDPR covers only those data where there is the possibility of identifying the data subject from its contents—and it applies a number of restrictions, such as how data can be shared and stored, who can access them, and how it might be ultimately used and reused. If it is not possible to re-identify subjects from the data, then that data falls outside of the provisions of GDPR—and so such restrictions do not apply. Thus, there is a pressing need, then, to classify which data entails (or not) the risk of identification of the participant.

The problem of possible re-identification of participants is fundamental for data types containing explicit physical features, as in the case of Magnetic Resonance Imaging (MRI). Typical MRI acquisitions in neuroscience, aside from the brain, contain detailed representations of the facial features of the participants (Figure 1) that can be used to create accurate three-dimensional representations of the subject's faces (Prior et al., 2009). To date, several solutions exist, which range from extracting the brain and sharing only this part of the volumes (Kalavathi & Prasath, 2016), to facial masking and facial removal (Bischoff-Grethe et al., 2007), but each of these methods has its downsides. For example, different masking and defacing procedures can affect subsequent steps of analysis and even lead to failure (de Sitter et al., 2020). Moreover, when MRIs are shared as part of EEG studies, brain extraction may impede accurate source modeling analyses that take advantage of head geometries (Hallez et al., 2007), and current facial masking or removal techniques may induce geometrical distortions.

Few studies have quantitatively tested the re-identification risk and the geometrical preservation offered by these methods (Budin, Zeng, Ghosh, & Bullitt, 2008; de Sitter et al., 2020; Mazura et al., 2012; Prior et al., 2009; Schwarz et al., 2021), and no comprehensive comparison among them has been performed. Treating these two aspects simultaneously is of paramount importance not only for investigators and research participants but also for properly informed legislation regarding data protection. Indeed, without such information, it is impossible to know whether currently accepted methods comply with the GDPR requirements and, at the same time, to what extent the geometrical properties of the subjects can be preserved without risking re-identification.

In that vein, here we present a new method for deidentifying MRIs - called AnonyMI- and we compare it with two state-of-the-art deidentification methods by evaluating, for the first time, their performance with tailored behavioral tasks. With respect to other methods, AnonyMI aims at preserving most of the geometrical properties of the subject's head while effectively deidentifying the MRI. It takes advantage of the Boundary Element Method (BEM; Hallez et al., 2007), which is a standard procedure for creating head models for source

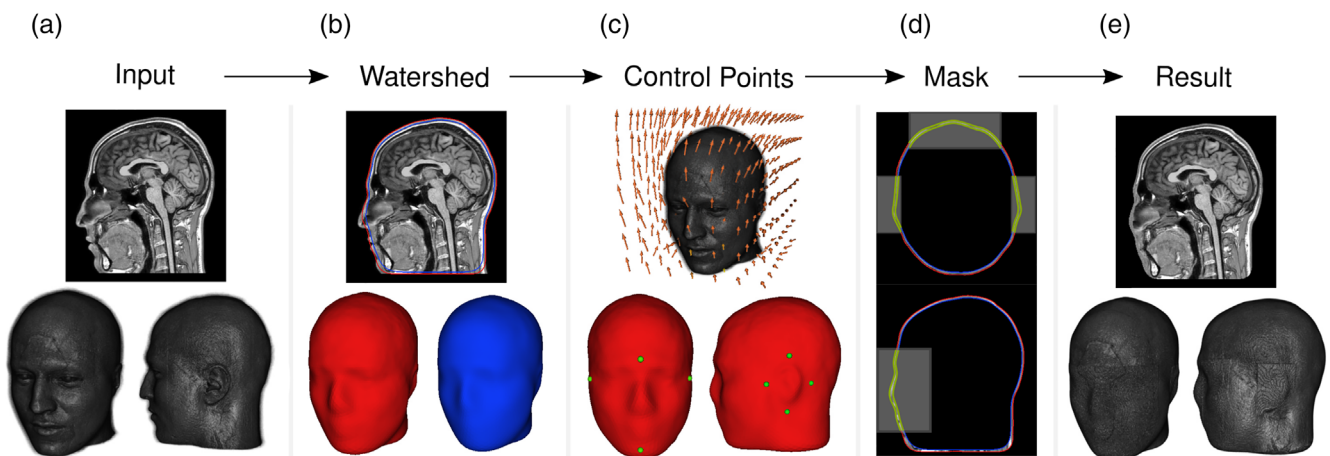


FIGURE 1 AnonyMI. Illustration of the deidentification procedure. The deidentification procedure is performed by first using a watershed algorithm to obtain 3D reconstructions of the skin and skull of the subject, then aligning the subject's MRI to a template that contains control points for the face and ears (or indicating them manually), and finally applying a mask to the intersection of these control points and the skin and skull surfaces. The MRI shown in this example is from a subject that provided informed consent for it to be shown

localization of scalp EEG activity. Specifically, AnonyMI employs the same three-dimensional surfaces that are used for source modeling (Akalin-Acar & Gençer, 2004; Hamalainen & Sarvas, 1987) to mask the subjects' facial features, which results in the same surfaces being obtained when re-computing the surfaces on the anonymized MRIs.

To test the performance of AnonyMI, we performed two sets of analyses, in both cases comparing it with other two state-of-the-art methods: PyDeface and Maskface (Gulban et al., 2019; Milchenko & Marcus, 2013). Specifically, first we performed a behavioral validation in which we tested and compared, on two experiments, how effectively each method deidentified the MRIs by asking participants to identify subjects from 3D reconstructions of their MRIs. In addition, we complemented these experiments by testing the identification performance on the same subjects using a machine learning face recognition algorithm. Second, we performed a geometrical evaluation in which we measured how much each method introduces spatial distortions and affects subsequent processing stages, and also compared their performance when being employed for source localization analysis using ground-truth data (Mikulan et al., 2020). Our results show that AnonyMI outperforms the other two methods in terms of geometrical conservation. Most importantly, we also show that AnonyMI, similarly to the other two tested deidentification methods, allows taking the re-anonymization risk close to chance level.

2 | METHODS

2.1 | AnonyMI

AnonyMI is an MRI deidentification tool that uses 3D surface modeling in order to deidentify MRIs while retaining as much geometrical information as possible. It can be run automatically or manually, which allows precise tailoring for specific needs. It is distributed as a plug-in of 3D Slicer, a widely used, open-source, stable, and reliable image processing software (Fedorov et al., 2012). It leverages the power of this platform for reading and saving images which makes it applicable on almost any MRI filetype, including all the most commonly used formats (e.g., DICOM, Nifti, Analyze, etc.). As it uses an algorithm from Freesurfer (Dale, Fischl, & Sereno, 1999) it currently runs on Linux and Mac platforms, but native Windows support will be possible in the near future and is indeed possible at the present with an elaborate setup. This algorithm is provided inside the toolbox and therefore does not require installing Freesurfer. Importantly, AnonyMI operates on the image data and does not apply changes to the file header, which might contain sensitive information. Several options exist for removing sensitive data from MRI file headers (Kushida et al., 2012). However, AnonyMI, by default, saves images in NIFTI format which is less prone to contain private information. Nevertheless, it is the responsibility of the user to ensure that the file header is free of sensitive data.

The procedure involves three main steps (Figure 1). First, the *watershed* algorithm, taken from Freesurfer (Dale et al., 1999; Ségonne et al., 2004), is applied in order to obtain 3D reconstructions of the skin and skull of the subject (Figure 1b). It is worth mentioning that these are the same surfaces that are often used in the creation of forward

solutions for source modeling (Hallez et al., 2007). Next, the location of the face and ears are determined by performing a nonlinear registration between the subject's MRI and a template, taken from the IXI dataset (Avants & Tustison, 2018), on which these control points have already been marked (Figure 1c). The location of these points can also be done manually. Indeed, since AnonyMI is integrated with 3D Slicer, manual procedures are eased by the extensive rendering and interactive capabilities of this image computing platform. This allows extending the areas to be anonymized at will, in order to cover specific parts that might increase the subject's identification risk (e.g., scars, malformations, etc.). In addition, creating a new template is fast and straightforward and can be done by simply importing the image to be used and marking the face and ears control points. This permits the creation of population-specific templates (e.g., age-specific) that can then be used to perform automatic anonymization. Finally, a subject-specific mask is created by taking the intersection of the control points and the skin and skull surfaces, which is filled with random numbers that follow the distribution of voxel intensities inside the space between the two surfaces (Figure 1d). In other words, in this last step the data of the space between the skull and the skin is filled with new values only in the places of intersection with the control points, and the values outside the skin model are set to zero. The resulting volume is an anonymized MRI that preserves most of the geometrical information of the subject (Figure 1e).

In contrast to other state-of-the-art methods, AnonyMI does not substantially modify the geometrical properties of the MRI, as in the case of PyDeface, which uses template matching to completely remove the face and cheeks area and leaves it empty. Additionally, it does not blur the face area, dissolving the boundaries between tissue types, as in the case of Maskface, which can preclude the possibility of doing source reconstructions. Another key difference with respect to previous solutions is that it provides an easy-to-use graphical interface that facilitates the procedure for nonprogrammers, even for batch processing of large datasets. Importantly, AnonyMI is an open-source software that can be freely downloaded from <https://github.com/ITCf/anonymi>.

2.2 | Validation

In order to assess the performance of our method and validate its results, we performed two series of analyses. First, we assessed the deidentification level achieved in two behavioral experiments. Second, we evaluated the similarity of the deidentified MRIs with respect to the original versions and tested how using deidentified MRIs affects EEG source localization results. In all cases, we compared the results obtained with our method to those obtained with other two state-of-the-art methods, PyDeface and Maskface. Data analyses were performed in Python and statistical analyses were performed in R (R Core Team, 2019).

2.2.1 | Behavioral validation

We conducted two behavioral experiments to evaluate how much different deidentification methods reduce the probability of identification of a subject from a 3D reconstruction of his/her MRI. On each trial of

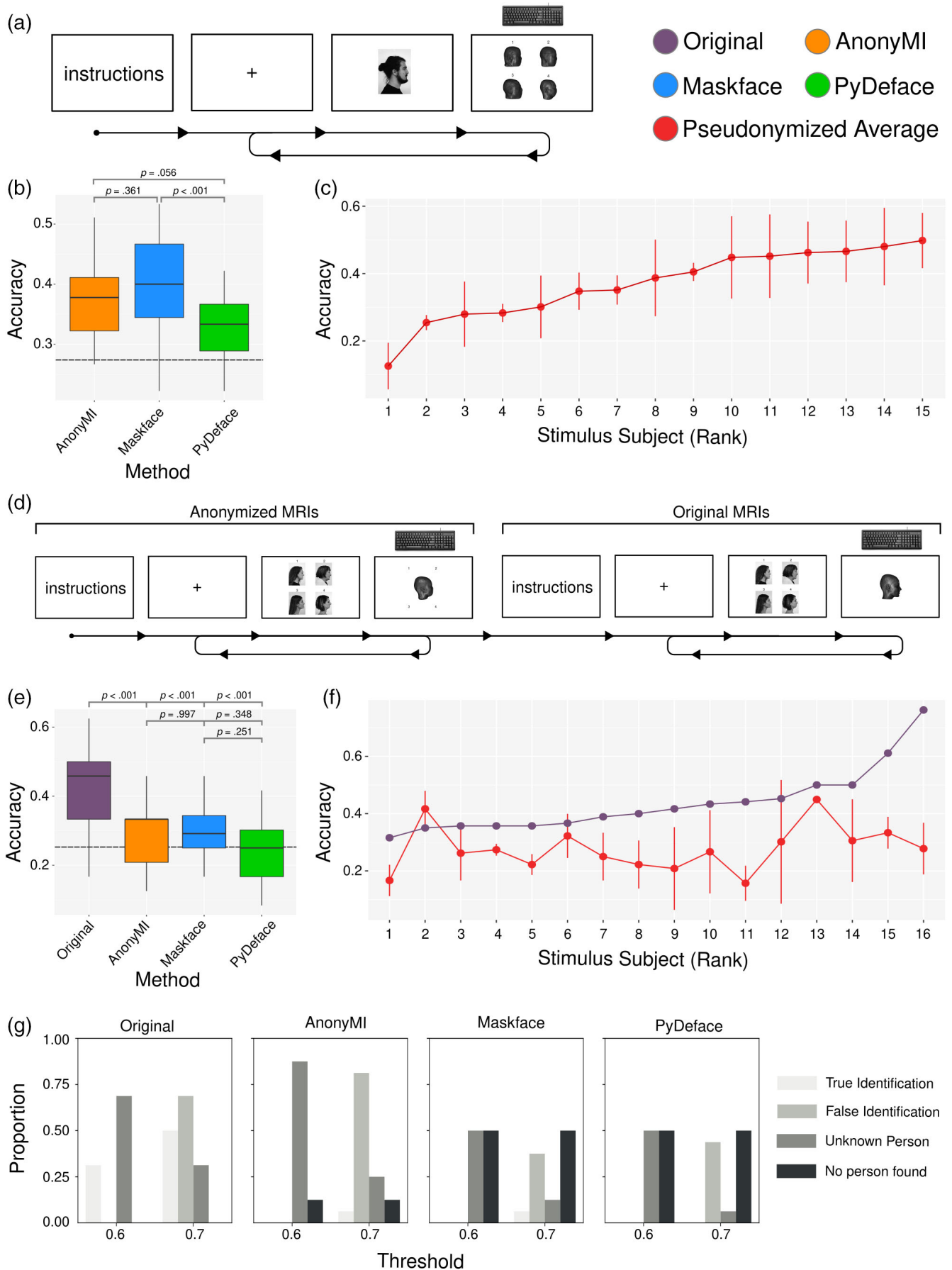


FIGURE 2 Legend on next page.

the first experiment, participants were presented with one photograph and then four deidentified MRIs, and they had to indicate to which of them the photograph belonged. In the second experiment, the procedure was inverted, that is, participants were presented with four photographs and then one deidentified MRI, and had to indicate to which photograph the MRI belonged. The two versions of the experiment aimed to provide ecologically valid estimates of the deidentification performance of each method as they correspond to trying to identify a person from a series of available MRIs, and vice versa, trying to identify an MRI from a series of available candidates.

Experiment 1

Stimuli. Stimuli consisted in 3D reconstructions of MRIs from 15 subjects ($F = 7$, \bar{X} age = 33.8, sd age = 7.28) that were deidentified with three different methods (PyDeface, Maskface and AnonyMI) and their corresponding photographs (front, left and right views). Reconstructions were obtained from 3D Slicer's Volume Rendering module and photographs were acquired with a light gray background. All images were postprocessed using the GNU Image Manipulation Program (GIMP), where they were converted to grayscale, centered, aligned (horizontally across the eyeline), cropped to the same size, and finally normalized using the histogram equalization technique (see Figure S1a). In the case of MRI reconstructions that presented noise around the edges of the head, they were manually cleaned by background cloning (see Figure S1b).

Participants and procedure. A total of 31 subjects participated in the experiment ($F = 19$; \bar{X} age = 20.16, sd age = 0.58). The experiment was performed in three consecutive groups in the same room with one desktop computer per participant. Participants were instructed that on each trial they would first see the photograph of a person, followed by four 3D reconstructions of deidentified MRIs, and that their task was to indicate, using the numbers from 1 to 4 on the keyboard, which was the MRI that corresponded to the photograph they had seen. Each participant performed a total of 135 trials in ~25 min. The experiment was programmed using PsychoPy3 (Peirce et al., 2019). At the beginning of the experiment, the software created a random sequence of trials for each participant by choosing for each trial a photograph, its corresponding MRI and three random sex-matched distractor MRIs. Each trial started with a fixation cross (1 s), followed by the presentation of the photograph at the center of the screen (3 s), followed by four MRI reconstructions displayed on a 2 by 2 grid, each one with a number from 1 to 4 on top (Figure 2a). Participants could take all the time they needed on each trial to respond.

Analysis. In order to quantify the differences in the deidentification performance of these methods, we performed a mixed-effects logistic regression analysis with participant as random factor, as it is one of the recommended approaches for accuracy data (Dixon, 2008; Jaeger, 2008). This analysis was performed using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), which provides p values via Wald tests. We defined *deface* as the reference level (as it is arguably the most performant method due to the complete removal of facial attributes) and assessed how much the probability of correct responses changed by using the other two methods. Model fits were judged by visual inspection of simulated residuals plot (i.e., diagonal patterns in Q-Q plots) using the DHARMA package with 10^3 replications. Post-hoc comparisons were performed using estimated marginal means and Tukey p value adjustment for multiple comparisons.

Experiment 2

Stimuli. The same stimuli of Experiment 1 were used, with the addition of one more MRI (and its corresponding photographs), adding up to a total of 16 subjects ($F = 7$, \bar{X} age = 34.1, sd age = 7.11), and 40 photographs from new subjects that were used as distractors ($F = 14$, \bar{X} age = 38.6, sd age = 6.81). All images were post-processed as described in Experiment 1. Test and control stimuli matching was performed separately for each participant as each one was presented with only a subset of the stimuli pool (see below).

Participants & procedure. A total of 24 subjects participated in the experiment ($F = 11$; \bar{X} age = 30.5, sd age = 6.57). The experiment was programmed in PsychoPy and it was performed online by participants through the Pavlovia platform (<https://pavlovia.org>). They were instructed that on each trial they would first see four photographs from different persons followed by a 3D reconstruction of an MRI, and that their task was to indicate, using the numbers from 1 to 4 on the keyboard, to which of the subjects they had seen it belonged. The numbers that indicated each stimulus remained on the screen while the MRI image was being presented, in order to ease the memory load of the task. Participants completed a total of 96 trials in ~20 min. One hundred random combinations of stimuli were created offline and each participant was assigned one of them using a random number generator at the beginning of the experiment. These combinations of stimuli were created by selecting six males and six females with MRIs, and six males and six females from the distractors pool. For each deidentification method, each of these stimuli subjects was presented the same number of times (once from the front view and once from one of the lateral views) in random combinations. After all trials using

FIGURE 2 Behavioral results. (a) Illustration of the experimental procedure of Experiment 1. (b) Boxplot of mean accuracy by participant for each deidentification method of Experiment 1. The horizontal dashed line represents chance level. (c) Average pseudonymized accuracy across methods by stimulus subject of Experiment 1. (d) Illustration of the experimental procedure of Experiment 2. (e) Boxplot of mean accuracy by participant for each deidentification method of Experiment 2. The horizontal dashed line represents chance level. (f) Average pseudonymized accuracy across methods by stimulus subject of Experiment 2. (g) Barplots of the face recognition machine learning algorithm performance on the original a pseudonymized MRIS with two thresholds (stringent: 0.6; permissive: 0.7). Statistical analysis of panel b and e were carried out employing binomial generalized linear mixed effects models with random intercept per participant. Post-hoc comparisons were performed using estimated marginal means and Tukey p value adjustment for multiple comparisons

deidentified MRIs were completed, a control block was presented in which stimuli were created in the same way but using the 3D reconstructions of the original MRIs (i.e., non deidentified) in order to obtain a reference score for the performance of each participant. The control block was presented at the end of the experiment, instead of aside to the deidentification methods, in order to avoid that subjects recognized distinctive features on the nonanonymized stimuli that could have later served as cues in anonymized trials. On each trial subjects were presented with a fixation cross (1 s), followed by the four photographs (7 s) on a 2 by 2 grid, followed by the MRI (maximum of 10 s to respond; Figure 2d).

Analysis. The same analyses as in Experiment 1 were used (see above), with the only difference of using the performance with the original MRIs as reference level. Namely, we assessed how much was the probability of correct responses reduced by using deidentified MRIs with respect to the performance with nondeidentified MRIs.

Machine learning face recognition

Recent advances in artificial intelligence have shown that it is possible to identify a subject from his/her MRI without the need of human intervention (Prior et al., 2009; Schwarz et al., 2019; Schwarz et al., 2021). These methods can process databases several orders of magnitude bigger than humans and therefore pose a threat to the protection of the privacy of research participants. We therefore performed a comparative analysis to test how well a state-of-the-art machine learning face recognition algorithm could identify the same subjects of the behavioral experiment.

Following a recently published article, which showed high recognition accuracy on nonpseudonymized MRIs, we created 3D models of the subjects faces using the *Surfice Ice* software (Schwarz et al., 2019; <https://www.nitrc.org/projects/surfice/>) and saved 2D renders of them as images. We then used the *dlib* package (King, 2009), with the pre-trained detector as implemented in the *face_recognition* library (https://github.com/ageitgey/face_recognition). Finally, we obtained the recognition results by employing two thresholds, a more stringent (0.6) and a more permissive one (0.7). We reported, for each method, the average number of “true identifications” (proportion of subjects in which the correct subject was among the ones indicated by the algorithm), “false identifications” (proportion of wrongly indicated subjects), “unknown persons” (proportion of subjects on which the algorithm indicated that the subject was unknown) and “no persons found” (proportion of subjects on which the algorithm indicated that it was unable to identify a face on the image). Of note, the “false identifications” are not univocal as more than one wrong answer can be produced by the algorithm for each subject, whereas for the other categories the maximum per subject is one.

2.2.2 | MRI geometrical similarity

In order to assess the geometrical similarity between the original and deidentified MRIs, we performed three analyses on a publicly available dataset (Souza et al., 2018), from which we randomly selected 75 MRIs

that included data from three different scanner manufacturers (25 Siemens; 25 Phillips; 25 General Electric). First, we computed the Jaccard Similarity, which is computed as the ratio between the intersection and the union of two sets (Jaccard, 1912; Milchenko & Marcus, 2013). To this end, performed skull-stripping, using the ANTs toolbox (Avants et al., 2011), on the original and deidentified MRIs and calculated the similarity between each deidentified MRI and its corresponding original MRI. The skull-stripping procedure failed (i.e., removed brain tissue) on eight of the original MRIs and therefore these were removed from the analysis. The template used for skull-stripping for AnonyMI and Maskface was the IXI template while the one used for PyDeface was the OASIS template from ANTs, as using a nondefaced template for the defaced images would have unfairly reduced its performance (Tustison et al., 2014). The comparison among methods was performed using pairwise Wilcoxon Rank Sum tests with Holm–Bonferroni correction for multiple comparisons.

Next, we employed the watershed algorithm from Freesurfer to create 3D reconstructions of the brain, inner skull, outer skull and outer skin, which are the same that are used for EEG/MEG source modeling. We assessed how many anonymized MRIs yielded usable reconstructions, that is, did not present deformations as appraised by visual inspection (Figure 3). We tested if the different deidentification methods produced significantly different numbers of usable reconstructions using Cochran's Q test followed by post-hoc pairwise McNemar's Chi-squared tests with Holm–Bonferroni correction for multiple comparisons.

Finally, we computed the Hausdorff Distance (Huttenlocher, Klanderman, & Rucklidge, 1993), a measure of similarity of 3D models, between each of the 3D reconstructions of deidentified MRIs that did not present deformations and their corresponding reconstructions obtained from the original MRIs. We tested if these distances were different across deidentification methods for each of the four surfaces (brain, inner skull, outer skull, and outer skin) using pairwise Wilcoxon Rank Sum tests within surfaces with Holm–Bonferroni multiple comparisons correction.

2.2.3 | Source localization

In order to assess how much the geometrical distortions induced by the deidentification procedures affect source localization results we used the *LocalizeMI* dataset (Mikulan et al., 2020), which contains ground-truth data for assessing source localization performance, and tested how much the localization accuracy varied as a function of the deidentification method with respect to the localization obtained using the original MRI.

We used data from simultaneous HD-EEG (High-Density Electroencephalography) recordings and intracranial stimulation in order to assess how much results varied by employing deidentified MRIs in the creation of forward models. The description of the experimental procedure, preprocessing and source localization procedures can be found in the open-access publication that accompanies the dataset (Mikulan et al., 2020) and the corresponding scripts are also publicly available. Here, we repeated the analysis with all the parameter's

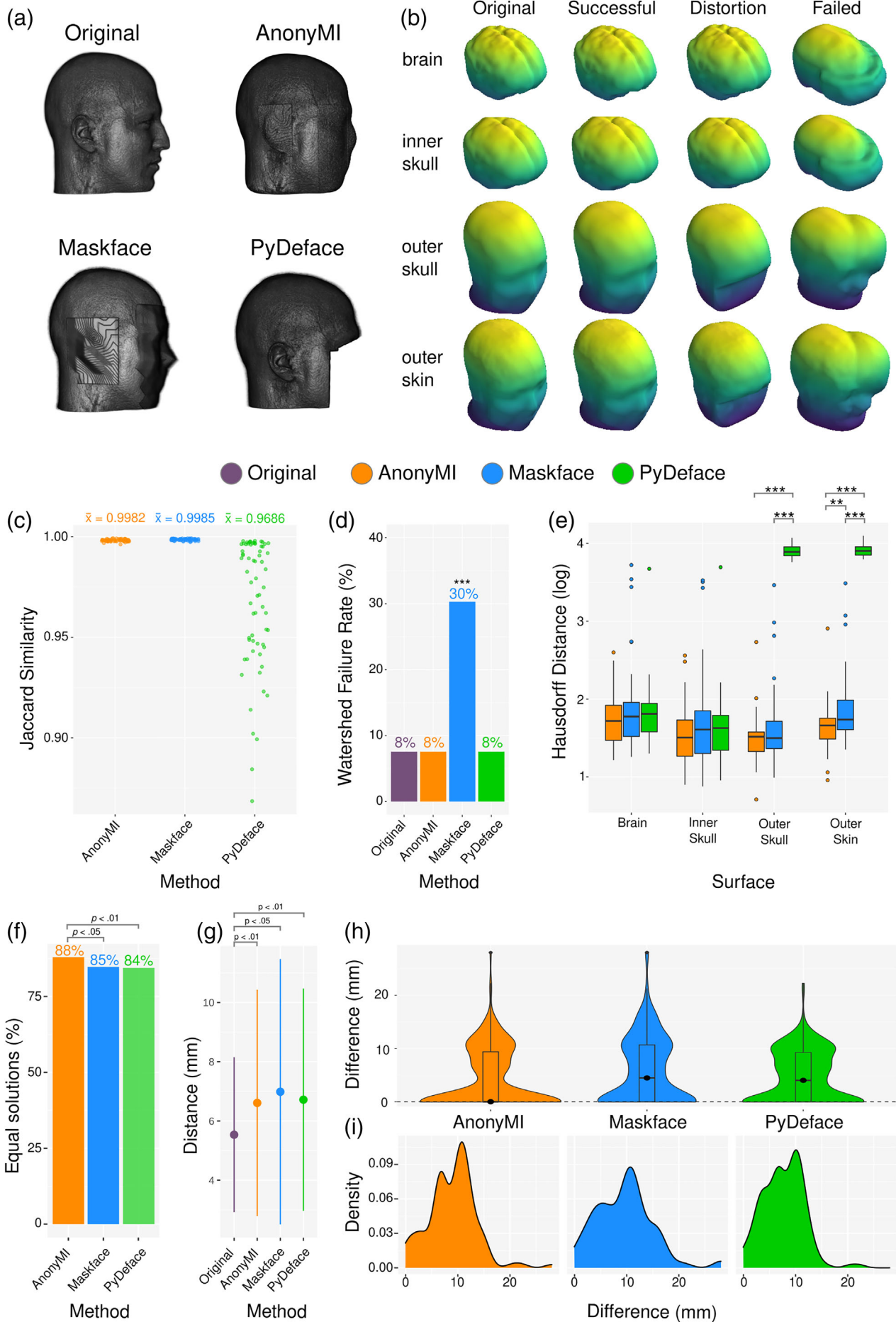


FIGURE 3 Legend on next page.

configurations that yielded optimal solutions for each of the 61 stimulation sessions, but this time using forward models created with deidentified MRIs. We then computed the percentage of solutions that resulted in the same localization error as the one obtained with the original MRIs. We tested if the different deidentification methods produced significantly different percentages of equal solutions using Cochran's Q test followed by post-hoc pairwise McNemar's Chi-squared tests with Holm-Bonferroni correction for multiple comparisons. Next, on all the configurations that did not result in the same localization error as in the original case for at least one method, we calculated the distance from the real source and tested how different it was from the distance obtained with the original MRIs by means of a mixed effects model. Specifically we used the log-transformed distance to the real source as dependent variable, each MRI deidentification method as predictor with the original MRIs as reference level, and subject as random factor (intercept and slope).

3 | RESULTS

3.1 | Behavioral validation

3.1.1 | Experiment 1

The overall performance (i.e., percentage of correct trials) was 33% for PyDeface, 37% for AnonyMI and 40% for Maskface. The mixed effects logistic regression analysis showed that AnonyMI increased the probability of correct responses ($\beta = 0.18, p < .05$) with respect to PyDeface, but this increase was smaller than in the case of Maskface ($\beta = 0.28, p < .001$). See Table S1 for a complete report of this analysis. Interestingly the overall performance across stimulus subjects ranged from 12% to 49% ($\bar{X} = 36\%, sd = 10\%$), showing that there were subjects that were almost unidentifiable whereas the performance on others was above chance level.

3.1.2 | Experiment 2

The overall performance was 46% for the original MRIs, 25% for PyDeface, 29% for AnonyMI and 30% for Maskface. All three

deidentification methods reduced the probability of correct responses in a statistically significant manner. PyDeface provided the highest reduction in performance with respect to the original MRIs ($\beta = -0.92, p < .001$), followed by AnonyMI ($\beta = -0.70, p < .001$) and Maskface ($\beta = -0.67, p < .001$). See Table S2 for a complete report of this analysis. The overall performance across stimulus subjects varied from 15% to 46% ($\bar{X} = 28\%, sd = 8\%$), showing again that there were subjects that were almost unidentifiable whereas others were above chance level.

3.1.3 | Machine learning face recognition

This analysis showed that ~30% and 50% of the subjects were correctly identified when using the original MRIs with the stringent and permissive thresholds, respectively (Figure 2g). In contrast, with the pseudonymized MRIs no subjects were identified with any of the tested methods when using a stringent threshold, and only one subject was identified with both AnonyMI and Maskface when using a more permissive threshold. Interestingly, PyDeface and Maskface produced a higher number of "no person found" results than AnonyMI in which the algorithm detected that there was a person but was unable to identify it, in line with the improved geometrical preservation of the latter.

3.2 | MRI geometrical similarity and application to source localization

3.2.1 | Geometrical similarity

The Jaccard Distance on skull-stripped images analysis showed that AnonyMI resulted in a mean similarity of 0.9982 and for Maskface of 0.9985 with respect to the original MRIs, whereas PyDeface resulted in 0.9686 (with some volumes going below 0.9). The differences between methods were all statistically significant ($p < .001$), however, the difference of .0001 between AnonyMI and Maskface holds little practical significance, especially if compared to the difference of ~0.03 between both of these methods and PyDeface.

FIGURE 3 MRI similarity and source localization analyses. (a) Example of 3D reconstructions of an original MRI and the three deidentification methods under evaluation. (b) Example of original, successful, distorted, and failed surface reconstructions obtained employing the watershed algorithm, (c) Results of Jaccard Similarity between skull-stripped MRIs between the original image and each deidentification method. (d) Percentage of failed watershed reconstructions (as depicted in rightmost column of panel b). Statistical analyses were performed using Cochran's Q test and pairwise McNemar tests with Holm-Bonferroni correction for multiple comparisons. (e) Results of the Hausdorff Distance analysis between the surfaces obtained with the original MRIs and those obtained with deidentified MRIs. Only images whose watershed reconstructions did not fail in any of the three deidentification methods were employed in this analysis. Statistical analyses were performed using pairwise Wilcoxon Rank Sum tests within surfaces with Holm-Bonferroni multiple comparisons correction. (f) Percentage of solutions on which the reconstructed source was equal to the source obtained using the original MRI for each deidentification method. Statistical analyses were performed using Cochran's Q test and pairwise McNemar tests with Holm-Bonferroni correction for multiple comparisons. (g) Mean and standard deviation of distances to real source (i.e., location of the stimulating intracranial contact) when using the original and deidentified MRIs. (h) Violin plot and boxplot of difference between the source obtained with the original MRI and with each deidentification method when considering the solutions on which at least one method was different from the original solution. (i) Density plot of difference with respect to the source obtained with the original MRI when considering only the different solutions for each method

The analysis on the 3D reconstructions of the brain, skull, and skin surfaces revealed that the original MRIs, AnonyMI and PyDeface yielded 92% of usable models (all three failing on the exact same MRIs), whereas Maskface produced 69%. This difference was statistically significant as assessed by McNemar's Chi-squared test ($\chi^2 = 13.06$; $p < .001$).

The pairwise analysis of the Hausdorff Distance by surface and method showed significant differences between PyDeface ($\bar{X} = 49.3$) versus AnonyMI ($\bar{X} = 4.69$; $p < .001$) and PyDeface versus Maskface ($\bar{X} = 6.00$; $p < .001$) for the outer skull surface; and between PyDeface ($\bar{X} = 50.20$) versus AnonyMI ($\bar{X} = 5.42$; $p < .001$), PyDeface versus Maskface ($\bar{X} = 7.32$; $p < .001$) and AnonyMI versus Maskface ($p < .05$) for the outer skin surface. See Table S3 for a complete report of this analysis. Overall, these results showed that AnonyMI provided the highest level of geometrical preservation.

3.2.2 | Source localization

The source localization analysis showed that AnonyMI provided 88% of equal solutions with respect to the results obtained with the original MRIs, while Maskface provided 85% and PyDeface 84% (Figure 3f). The difference between methods was statistically significant as assessed by McNemar's Chi-squared tests. The difference in the overall distance between the real sources and the reconstructed ones was not statistically significant between methods, due to the relatively high number of equal solutions yielded by each method (>84%). However, all of them differed significantly from the results obtained with the original MRIs (Figure 3g). The overall distance between the sources reconstructed with the original MRI and the sources reconstructed with each method is shown in Figure 3h for each solution in which at least one method showed a different result (same number of observations across methods), and on Figure 3i for the solutions on which each of the methods showed a different result (different number of observations across methods).

4 | DISCUSSION

Deidentifying MRIs constitutes a challenge as it has to balance two opposite goals. It should prevent the possibility of re-identification of a participant while keeping as much geometrical information as possible. To tackle this issue, we developed a novel MRI deidentification method that aims at providing a balance between these objectives. In order to assess the performance on these two fronts, and to compare it to other state-of-the-art methods we performed two series of analysis: a behavioral and a geometrical validation.

To the best of our knowledge, this is the first study that jointly evaluated and compared the deidentification and geometrical preservation performance of state-of-the-art MRI deidentification methods. In two behavioral experiments we tested how well could participants identify subjects from 3D reconstructions of their MRIs. On Experiment 1, where subjects were first presented with a single photograph

which was followed by four deidentified MRIs, the performance was 33, 37, and 40% for PyDeface, AnonyMI and Maskface, respectively. The only statistically significant difference was found between PyDeface and Maskface. Even though the average results were above chance level (25%), the general performance of all three methods can be arguably considered satisfactory, as the experimental design corresponded more to a discrimination task than to an identification task, which would have been more difficult. In other words, if subjects were able to discard one of the four testing images (i.e., instead of identifying the correct subject they recognized that one of them was not the correct answer) and guess randomly, chance-level would rise to ~33%, which is close to the observed performance levels. To complement these results, we performed a second experiment, with a more ecological design, in which participants were presented with four photographs and then with a single reconstructed MRI and had to decide to which of the four photographs it corresponded. This task included a memory component as subjects had to remember the faces and reduced the possibility of using a discrimination strategy, which made it more similar to a real-life situation. In this case, the performance was 25, 29, and 30%, for PyDeface, AnonyMI and Maskface, respectively. In this task we also included a control block on which the MRIs were not deidentified, which allowed us to estimate a baseline level of identification performance, which was 46%. It is worth noting that the performance with the original MRIs might be inflated with respect to the anonymized ones as this block was always carried out at the end of the task, when subjects had had more training. This experimental design was chosen in order to avoid that subjects found distinctive features on the nondeidentified stimuli that could be used on deidentified trials and bias the results of the main aim of the experiment. The differences among methods were not statistically significant, but the identification performance of all three methods was significantly lower than in the nondeidentified condition. Interestingly, in both behavioral experiments the analysis on individual stimulus subjects showed that different methods worked better with different subjects. Several articles have consistently shown that the eyes, mouth, nose, eyebrows and also the head outline are the most important features for facial recognition (Davies, Ellis, & Shepherd, 1977; Haig, 1986; Kamps, Morris, & Dilks, 2019; Karczmarek, Pedrycz, Kiersztyn, & Rutka, 2017; Sadr, Jarudi, & Sinha, 2003). Future studies employing eye-tracking techniques will be required to assess which of these features are playing a role in the recognition of pseudonymized MRIs, in order to make these pseudonymization techniques more robust.

The overall pattern of results indicates that even though PyDeface was the most performant deidentification method in terms of reducing the re-identification risk, followed by AnonyMI and finally by Maskface, all three methods efficiently reduced the risk of identification almost at chance level. This notion is of paramount importance in light of the new legislation for data privacy. Indeed, if it is assumed that all currently used methods are sufficient to appropriately deidentify MRI data, that is, that it is not possible to re-identify a data subject from such a "deidentified" dataset, and this assumption is incorrect then it leaves open the possibility of a major data breach.

Data breaches can seriously infringe the privacy of patients and research subjects, and also, under GDPR, they can have significant financial implications for the offending scientist and their institution. Administrative fines are detailed in Article 83 of the GDPR and stipulate that, depending upon the severity of the breach, fines could be up to €20 million or 4% of the institutions' global annual turnover of the previous financial year, whichever is higher (European Parliament and Council of European Union, 2016). This is not to mention the potential reputational damage and distress to data subjects such a breach might cause.

Our study focused on the identification risk by a human operator, however, a growing body of literature shows that machine learning algorithms could potentially be used to individuate research participants from shared data and even to reconstruct the subject's face from deidentified MRIs (Abramian & Eklund, 2019; Schwarz et al., 2019; Schwarz et al., 2021). An in-depth analysis of this risk is of utmost importance for future studies in order to provide a comprehensive picture of the risks associated with making deidentified MRIs public. To this end, we employed a face recognition machine learning algorithm to test how well it could identify the same subjects used in the behavioral task. This analysis showed that with the original MRIs, depending on the threshold, ~30 to ~50% of the subjects could be correctly recognized. When using the pseudonymized MRIs, only one out of 16 were identified both with AnonyMI and Maskface and none with PyDeface.

While, in the case of data sharing, reducing the re-identification risk of the MRI is obviously the main aim of any MRI deidentification method, the associate distortions of the geometrical feature of the images should be maximally reduced in order to maximize reuse and interoperability (Wilkinson et al., 2016). For this reason, we performed a spatial analysis in which we tested how well each method preserved the geometrical characteristics of the MRIs. We first tested the Jaccard Similarity of skull stripped brain volumes done with the deidentified MRIs with respect those obtained with the original MRIs. In this analysis Maskface and AnonyMI showed ~99% convergence, whereas PyDeface showed 96%, with some images going below 90% with the latter. It is worth mentioning that the skull stripping procedure was carried out with a defaced template in the case of the PyDeface method while it was done with a full-face template for Maskface and AnonyMI, because using the latter for PyDeface would have unfairly underestimated its performance. These results indicate that both Maskface and AnonyMI provide deidentified volumes that give more similar results to those obtained with the original MRIs when applying a subsequent processing algorithm as skull stripping. This is important to ensure replicability of results when using deidentified images.

Next, we extracted 3D surfaces of the brain, skull and skin using the watershed algorithm, a standard procedure for performing M/EEG source localization. We found that Maskface provided 30% of unusable surfaces due to severe geometrical distortions, whereas with PyDeface and AnonyMI it was 8%, the same percentage obtained with the original MRIs. Then, employing the successfully

reconstructed surfaces we computed how similar they were to the surfaces obtained with the original MRIs. This analysis showed that AnonyMI significantly outperformed PyDeface for the outer skull, and also outperformed both PyDeface and Maskface for the skin. This pattern of results indicates that AnonyMI was the method that induced less geometrical distortions.

To assess how much using deidentified MRIs affects source localization results, we reconstructed electrical sources using a dataset with ground-truth and compared the results obtained with each deidentification method to those obtained with the original MRIs. This analysis showed that the percentage of equal solution with respect to the original MRIs was 88, 85 and 84% for AnonyMI, Maskface and PyDeface, respectively. In other words, AnonyMI provided an increase of ~4% of equal solution across multiple parameter configurations. The overall pattern of geometrical results shows that AnonyMI provides the most reliable MRIs for performing structural and source localization analysis. The increase in the number of equal solutions might be more pronounced if using FEM instead of BEM models (Hallez et al., 2007), as the former creates a more detailed geometrical reconstruction. We chose to use the BEM method, as it is arguably the most widely used method for distributed source modeling due to its balance of low computational costs with geometrical correctness (Hallez et al., 2007).

Aside from the aforementioned benefits, AnonyMI presents many other advantages. It leverages the power of 3D Slicer for importing and saving in any filetype, avoiding unnecessary format changes that are mandatory with other methods and that can lead to errors. It also takes advantage of 3D Slicer for fast and practical visualization of results, which eases the quality assurance of the procedure. It can also create population specific templates in minutes, simplifying the analysis of particular datasets (i.e., age specific). It provides a visual interface to mask specific areas of a subject that might have severely increased the re-identification risk. It can easily run multiple subjects automatically, making it easy to process large datasets and has both a command-line and a graphical interface, providing an essential solution for nonprogrammers.

In sum, AnonyMI is a novel method that offers several technical advantages that make MRI deidentification easier, and that provides the optimal balance between deidentification performance and geometrical preservation.

5 | LIMITATIONS

The present study is the first to comparatively analyze the performance of different MRI deidentification methods, however, further studies will be required to obtain a more comprehensive characterization. We focused on two particular behavioral tasks to assess the identification risk and to quantify the improvements provided by each method, aiming at tasks that could provide ecologically valid estimates. Nevertheless, different tasks might provide different results, and therefore, more experiments with different experimental designs will be required to thoroughly describe each method's performance.

Similarly, the stimuli employed, and the participants of the behavioral tasks corresponded to relatively homogeneous samples of white caucasian subjects, therefore the generalization of our results to other ethnicities is not granted and should be addressed in future studies (Dotson & Duarte, 2020). With regards to the geometrical analysis, we focused on three measures (Jaccard similarity, Hausdorff distance and BEM outcome). A comparison of more measures and analysis pipelines would be of great importance to appraise how much each method allows reusability under different circumstances. Finally, our source localization analysis focused on one forward model type (i.e., BEM), three inverse solution methods (dSPM, eLORETA, and MNE), and used nonphysiological ground truth data. We chose these methods as they are among the most used methods in contemporary M/EEG source localization analysis. However, an investigation of other forward models and inverse solution methods would provide a more nuanced comparison.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 785907 and No. 945539 (Human Brain Project SGA2 and SGA3).

AUTHOR CONTRIBUTIONS

Ezequiel Mikulan: Conceptualization, Methodology, Software, Formal analysis, Writing—Original Draft. Simone Russo: Investigation, Validation. Flavia Maria Zauli: Investigation, Data Curation. d'Orio Piergiorgio: Investigation, Validation. Sara Parimigiani: Investigation, Validation. Jacopo Favaro: Investigation, Validation, Writing. William Knight: Writing—Original Draft. Silvia Squarza: Validation. Pierluigi Perri: Writing—Original Draft. Francesco Cardinale: Investigation, Validation, Writing—Review & Editing. Pietro Avanzini: Conceptualization, Methodology, Writing—Original Draft, Writing—Review & Editing. Andrea Pigorini: Conceptualization, Methodology, Writing—Original Draft, Writing—Review & Editing, Project administration, Supervision.

ETHICS STATEMENT

The study was approved by the local Ethics Committee (Approval Number: 114/20).

DATA AVAILABILITY STATEMENT

AnonymI is an open-source software that can be freely downloaded from <https://github.com/ITCf/anonymi>. The data and code of the experiments and analyses are available upon request, with the exception of the photographs of the stimulus subjects, which are not included in order to respect their privacy.

ORCID

Ezequiel Mikulan  <https://orcid.org/0000-0001-7259-6120>

REFERENCES

- Abramian, D., Eklund, A. (2019). Refacing: Reconstructing Anonymized facial features using GANs, in: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Presented at the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1104–1108. <https://doi.org/10.1109/ISBI.2019.8759515>
- Akalin-Acar, Z., & Gençer, N. G. (2004). An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. *Physics in Medicine and Biology*, 49, 5011–5028. <https://doi.org/10.1088/0031-9155/49/21/012>
- Amunts, K., Knoll, A. C., Lippert, T., Pennartz, C. M. A., Ryvlin, P., Destexhe, A., ... Bjaalie, J. G. (2019). The human brain project: Synergy between neuroscience, computing, informatics, and brain-inspired technologies. *PLoS Biology*, 17, e3000344. <https://doi.org/10.1371/journal.pbio.3000344>
- Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S., & Armañanzas, R. (2017). Win-win data sharing in neuroscience. *Nature Methods*, 14, 112–116. <https://doi.org/10.1038/nmeth.4152>
- Avants, B., Tustison, N., 2018. ANTs/ANTsR brain templates. <https://doi.org/10.6084/m9.figshare.915436.v2>
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54, 2033–2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., ... Fischl, B. (2007). A technique for the deidentification of structural brain MR images. *Human Brain Mapping*, 28, 892–903. <https://doi.org/10.1002/hbm.20312>
- Brakewood, B., & Poldrack, R. A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage*, 82, 671–676. <https://doi.org/10.1016/j.neuroimage.2013.02.040>
- Budin, F., Zeng, D., Ghosh, A., & Bullitt, E. (2008). Preventing facial recognition when rendering MR images of the head in three dimensions. *Medical Image Analysis*, 12, 229–239. <https://doi.org/10.1016/j.media.2007.10.008>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9, 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the “Photofit” technique. *Perception*, 6, 263–269. <https://doi.org/10.1068/p060263>
- de Sitter, A., Visser, M., Brouwer, I., Cover, K. S., van Schijndel, R. A., Eijgelaar, R. S., ... MAGNIMS Study Group and Alzheimer's Disease Neuroimaging Initiative. (2020). Facing privacy in neuroimaging: Removing facial features degrades performance of image analysis methods. *European Radiology*, 30, 1062–1074. <https://doi.org/10.1007/s00330-019-06459-3>
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, 59, 447–456. <https://doi.org/10.1016/j.jml.2007.11.004>
- Dotson, V. M., & Duarte, A. (2020). The importance of diversity in cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464, 181–191. <https://doi.org/10.1111/nyas.14268>
- European Parliament, Council of European Union, 2016. Regulation (EU) 2016/679.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., ... Kikinis, R. (2012). 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30, 1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001>

- Gulban, O. F., Nielson, D., Poldrack, R., Lee, J., Gorgolewski, C., Vanessasaurus, Ghosh, S., 2019. poldracklab/pydeface: v2.0.0. Zenodo <https://doi.org/10.5281/zenodo.3524401>
- Haig, N. D. (1986). Exploring recognition with interchanged facial features. *Perception*, 15, 235–247. <https://doi.org/10.1068/p150235>
- Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., ... Lemahieu, I. (2007). Review on solving the forward problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 4, 46. <https://doi.org/10.1186/1743-0003-4-46>
- Hamalainen, M. S., & Sarvas, J. (1987). Feasibility of the homogeneous head model in the interpretation of neuromagnetic fields. *Physics in Medicine and Biology*, 32, 91–97. <https://doi.org/10.1088/0031-9155/32/1/014>
- Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, 850–863. <https://doi.org/10.1109/34.232073>
- Jaccard, P. (1912). The distribution of the Flora in the alpine Zone. 1. *The New Phytologist*, 11, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit mixed models. *Journal of Memory and Language*, 59, 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Kalavathi, P., & Prasath, V. B. S. (2016). Methods on skull stripping of MRI head scan images: A review. *Journal of Digital Imaging*, 29, 365–379. <https://doi.org/10.1007/s10278-015-9847-8>
- Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J. J. M., & van Thiel, G. J. M. W. (2019). Responsible data sharing in international health research: A systematic review of principles and norms. *BMC Medical Ethics*, 20, 21. <https://doi.org/10.1186/s12910-019-0359-9>
- Kamps, F. S., Morris, E. J., & Dilks, D. D. (2019). A face is more than just the eyes, nose, and mouth: fMRI evidence that face-selective cortex represents external features. *NeuroImage*, 184, 90–100. <https://doi.org/10.1016/j.neuroimage.2018.09.027>
- Karczmarek, P., Pedrycz, W., Kiersztyn, A., & Rutka, P. (2017). A study in facial features saliency in face recognition: An analytic hierarchy process approach. *Soft Computing*, 21, 7503–7517. <https://doi.org/10.1007/s00500-016-2305-9>
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 1755–1758.
- Kushida, C. A., Nichols, D. A., Jadrnicek, R., Miller, R., Walsh, J. K., & Griffin, K. (2012). Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. *Medical Care*, 50, S82–S101. <https://doi.org/10.1097/MLR.0b013e3182585355>
- Mazura, J. C., Juluru, K., Chen, J. J., Morgan, T. A., John, M., & Siegel, E. L. (2012). Facial recognition software success rates for the identification of 3D surface reconstructed facial images: Implications for patient privacy and security. *Journal of Digital Imaging*, 25, 347–351. <https://doi.org/10.1007/s10278-011-9429-3>
- Mikulan, E., Russo, S., Parmigiani, S., Sarasso, S., Zauli, F. M., Rubino, A., ... Pigorini, A. (2020). Simultaneous human intracerebral stimulation and HD-EEG, ground-truth for source localization methods. *Scientific Data*, 7, 1–8. <https://doi.org/10.1038/s41597-020-0467-x>
- Milchenko, M., & Marcus, D. (2013). Obscuring surface anatomy in volumetric imaging data. *Neuroinformatics*, 11, 65–75. <https://doi.org/10.1007/s12021-012-9160-3>
- Milham, M. P., Craddock, R. C., Son, J. J., Fleischmann, M., Clucas, J., Xu, H., ... Klein, A. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. *Nature Communications*, 9, 2818. <https://doi.org/10.1038/s41467-018-04976-1>
- Mott, M. C., Gordon, J. A., & Koroshetz, W. J. (2018). The NIH BRAIN initiative: Advancing neurotechnologies, integrating disciplines. *PLoS Biology*, 16, e3000066. <https://doi.org/10.1371/journal.pbio.3000066>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Poline, J.-B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., ... Kennedy, D. N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, 6, 9. <https://doi.org/10.3389/fninf.2012.00009>
- Prior, F. W., Brunsden, B., Hildebolt, C., Nolan, T. S., Pringle, M., Vaishnavi, S. N., & Larson-Prior, L. J. (2009). Facial recognition from volume-rendered magnetic resonance imaging data. *IEEE Transactions on Information Technology in Biomedicine*, 13, 5–9. <https://doi.org/10.1109/TITB.2008.2003335>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, 32, 285–293. <https://doi.org/10.1068/p5027>
- Schwarz, C. G., Kremers, W. K., Therneau, T. M., Sharp, R. R., Gunter, J. L., Vemuri, P., ... Jack, C. R. (2019). Identification of anonymous MRI research participants with face-recognition software. *The New England Journal of Medicine*, 381, 1684–1686. <https://doi.org/10.1056/NEJMc1908881>
- Schwarz, C. G., Kremers, W. K., Wiste, H. J., Gunter, J. L., Vemuri, P., Spychalla, A. J., ... Jack, C. R. (2021). Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives. *NeuroImage*, 231, 117845. <https://doi.org/10.1016/j.neuroimage.2021.117845>
- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22, 1060–1075. <https://doi.org/10.1016/j.neuroimage.2004.03.032>
- Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., ... Lotufo, R. (2018). An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170, 482–494. <https://doi.org/10.1016/j.neuroimage.2017.08.021>
- Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., ... Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, 99, 166–179. <https://doi.org/10.1016/j.neuroimage.2014.05.044>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mikulan, E., Russo, S., Zauli, F. M., d'Orio, P., Parmigiani, S., Favaro, J., Knight, W., Squarza, S., Perri, P., Cardinale, F., Avanzini, P., & Pigorini, A. (2021). A comparative study between state-of-the-art MRI deidentification and AnonymI, a new method combining re-identification risk reduction and geometrical preservation. *Human Brain Mapping*, 1–12. <https://doi.org/10.1002/hbm.25639>