Preserving privacy in medical images while still enabling AI-driven research: A comprehensive review

Laçi Hafsa
Department of Statistics and Applied Informatics, Faculty of Economy, University of Tirana,
Tirana, Albania
hafsa.laci@unitir.edu.al

Sevrani Kozeta
Department of Statistics and Applied Informatics, Faculty of Economy, University of Tirana,
Tirana, Albania
kozeta.sevrani@unitir.edu.al

2024 13th Mediterranean Conference on Embedded Computing (MECO) | 979-8-3503-8756-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/MECO62516.2024.10577795

Abstract— **The application of AI algorithms in medical image processing requires access to a large amount of data containing protected health information (PHI). Preserving individuals' privacy is not only an ethical issue, but it is also dictated by personal privacy laws such as HIPAA or GDPR.**

Health authorities and hospitals should be aware that is not possible to fully anonymize medical images without losing their research utility, meaning that some level of risk for potentially reidentifying patient information will be present in any case. On the other hand, researchers and software developers should be informed about de-identification or anonymization approaches and should consider them as part of any solution.

Review papers published in this area are mainly focused on preserving privacy in structured medical data or exploring defense mechanisms and approaches against adversarial attacks. This paper takes into consideration unstructured medical data and provides an overview of the: techniques and tools adopted for medical image de-identification or anonymization; faced limitations in ensuring patient privacy; and researchers' future directions.

Keywords: Medical images, De-identification, Anonymization, Privacy

I. INTRODUCTION

The use of AI has an increasing impact in the field of medicine and in terms of scientific research, medical images play an important role [1]. Such data is collected by health institutions on a daily basis and stored into patient's health records. The latter, are not exposed for public use due to privacy issues related to personal health information disclosure. On the other hand, a vast amount of image processing AI models can be built, trained and tested only with the help of large medical image datasets [2].

The large amount of data needed, the complex nature of AI models, and private health information preservation are some of the main factors that complicate the process of medical image data sharing [3]. Medical AI-driven research is striving for better solutions while keeping patients' privacy intact [4]. To ease things up, negotiation with respective data owners is required, but this is not an easy task, especially when performed under privacy regulation laws [5,6].

However, seen from another point of view, through the presence of data regulation laws (HIPAA, GDPR, CCPA, etc.) the process of accessing data for research purposes, is facilitated. These laws suggest the use of technical solutions to de-identify or anonymize PHI before using the data that contains it and incentivize the development of new techniques to protect privacy [7].

For a better understanding of the de-identification or anonymization techniques applied by researchers to unstructured medical data, particularly to medical images, a review is conducted. Our focus will be only in the two forementioned tasks employed to preserve privacy in this type of data. Additionally, due to the complexity of these operations, authors have to deal with numerous challenges and limitations.

Our review aims to answer the following questions:

- **RQ1:** Does the state-of-art suggest any standard de-identification/anonymization approach for preserving privacy in medical images?
- **RQ2:** What are the limitations and challenges faced by the authors during medical image de-identification/anonymization?

The subsequent sections of this study are organized as follows: Section 2 presents a brief state-of-art summary. Section 3 describes the methodology used to conduct the research. Section 4 interprets the results obtained from the selected relevant papers. Section 5 draws the conclusions obtained and gives suggestions about steps that can be considered in the future.

II. LITERATURE REVIEW

Medical images are an important tool for medical diagnosis [1] and are vital to increase the performance accuracy of AI tasks, such as image classification, image segmentation, transfer learning, etc. To be able to use these images as an input to AI algorithms, the personal data that they contain must be protected. However, before trying to prevent personal health information disclosure, one should first understand what type of data is considered to be personal. A set of direct identifiers that can be found in images and should be de-identified are specified by HIPAA privacy rules [8].

Besides de-identification and anonymization, other privacy protection approaches can be undertaken. Examples of such methods are federated learning, differential privacy and homomorphic encryption [6]. In the literature, these tasks are performed in cases where privacy breaches can occur and there is a need to protect diagnostic models published to third parties from adversarial attacks [9,10].

According to Feng et al. [11] the current defense approaches have a bad impact to the functionality of AI models. Because of that they decided to create a Fake Gradient defense framework, which encrypts the model's output to protect it from attackers. It is focused on image classification models. Jarin & Eshete [12] introduced PRICURE system to prevent membership inference attacks by maintaining the accuracy of the models intact. It combines secure multi-party computation and differential privacy to enable private multiple model prediction. Additionally, Nikolaidis et al. have studied the ability of implicit learning from visually unrealistic stimuli to guarantee resilience against membership inference attacks [13].

Abuadbba et al. [14] have studied whether 1D CNN models can be trained while preserving privacy through split learning procedure. This is performed by splitting the model into client side and server side respectively. While the client processes the raw data, the server is not able to directly access it. Meanwhile, Santos & Rocha [15] mentioned the problems of content-based, network and DOS or DDOS attacks during data preparation stage.

Protecting demographics and diagnosis codes is important considering the possibility that they give to researchers to conduct medical studies. Existing studies are not able to guard identity disclosure, ensure useful anonymization and minimize the loss of information at the same time. For this reason, Poulis et al. [16] proposed a new approach by applying (k, k^m)-anonymity in datasets containing demographic and diagnoses codes.

To improve patient care, sharing health information data is crucial and challenging at the same time. De Kok et al. [17] explored accessible healthcare databases to dictate appropriate sharing. A seven-recommendation approach was formulated by the authors to guide future open healthcare data sharing initiatives.

Abouelmehdi et al. [18] have discussed about privacy and security concerns present in big health data. Despite the fact that they have reviewed encryption and anonymization methods applied to general healthcare data, their study states that techniques like hiding a needle in a haystack, attribute-based encryption access control, homomorphic encryption and storage path encryption are also present.

Majeed [19] proposed an anonymization scheme that aims to guarantee data privacy, but mainly to ensure protection of identity from adversarial attacks. His proposed solution was focused on general electronic health records in the clouds.

III. METHODOLOGY

A comprehensive protocol-driven review was conducted in three large databases (ScienceDirect, ACM, PubMed). Fig. 1 shows in details the reproducible systematic procedure followed by the authors.

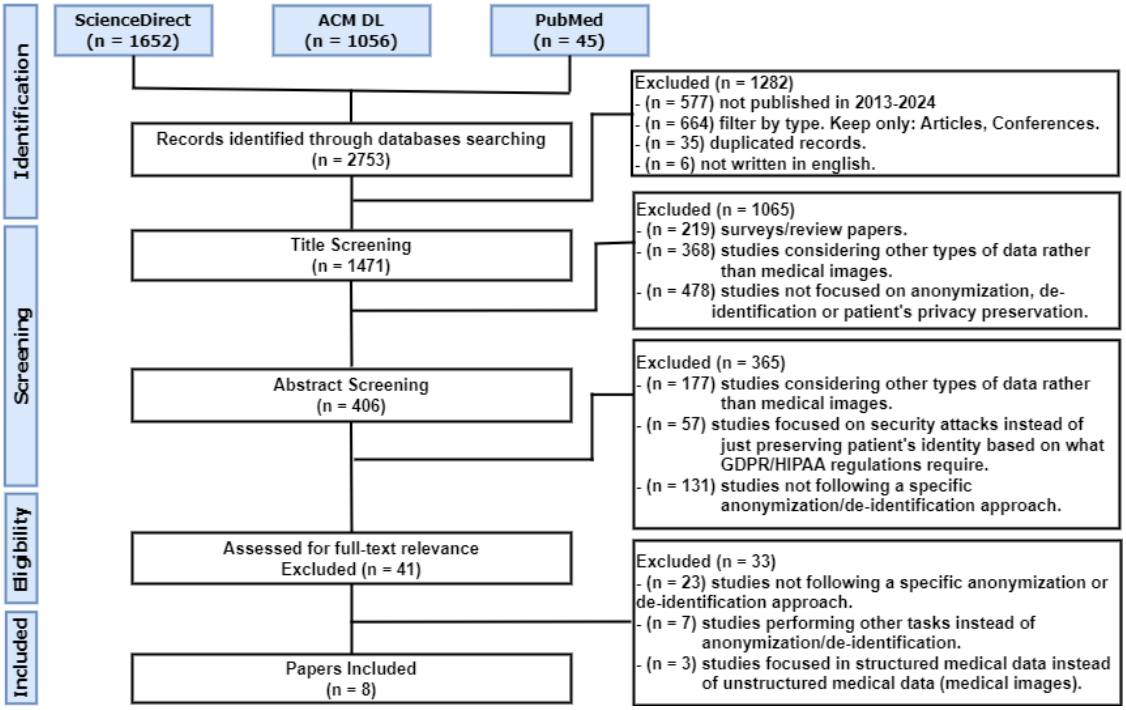

Figure 1. Study selection procedure
Source: Authors, February 2024

A. Inclusion Criteria

The search query performed in the selected databases was: medical image AND (anonymization OR anonymisation OR deidentification OR de-identification) AND (privacy OR confidentiality OR data protection). We considered only papers fulfilling the following inclusion criteria:

- Articles published from 2013 to 2024;
- Journal articles and conferences;
- Articles published in English;
- Articles following a specific de-identification or anonymization approach;

B. Exclusion Criteria

From the final analysis were excluded duplicates and:

- Articles performing de-identification/anonymization process in structured medical data instead of medical images (unstructured medical data);
- Non-experimental articles that do not give details about the de-identification/anonymization process;
- Articles focusing on security attacks and defense mechanisms instead of patient's privacy preservation based on what data privacy regulations (HIPAA, GDPR, etc.) require;

IV. Discussions

From the screening procedure, 41 papers were assessed for full-text relevance and only 8 papers were found eligible for our review. With the aim of answering the research questions forementioned in the introduction section, we have analyzed the relevant studies based on: the de-identification or anonymization process performed; the specific data modality used; the challenges they underwent and the suggested future directions.

A. De-identification or Anonymization Approach

Results shown in Table 1 give an outline of the approaches adopted by the relevant papers to preserve patient's privacy in medical image data, while still enabling research. The first thing to note is that all the studies have performed de-identification instead of anonymization. Besides that, they have all adopted an open-source solution and the URLs provided under the "Solution URL" column indicate the available source code.

It is important to highlight the fact that these two terms were often used interchangeably in the selected papers. Through anonymization we make sure that personal information cannot be used to re-identify a patient, meaning that this process once performed, is irreversible. By contrast, de-identification makes re-identification impossible unless individuals are authorized to do so [20].

Regarding the de-identification solution, half of the relevant studies have employed a RSNA CTP pipeline [21,22,23,8], which is a solution that can be personalized according to particular requirements. Moreover, it is utilized across diverse image modalities, but mostly in DICOM format data. Another preferred solution is distorting all [24] or partial [25] facial features using an image de-facer program. The latter is applied on head or brain MRI image modalities only. Moreover, generating synthetic MRIs using an auxiliary classifier Generative Adversarial Network [26] instead of using real data or simply de-identifying the DICOM header and checking further for pixel burned-in PHI [27] are considered as possible solutions as well.

B. Limitations and Challenges

Clinical picture archives and communication systems (PACS), use DICOM as a standard image format for different types of medical images. It is common for manufacturers or imaging system vendors, who are responsible for the documentation of PHI, to exclude some private attributes from the DICOM conformance statements. Sometimes, they encode acquisition parameters into private attributes or they do not use attributes fields for their intended purposes. PHI might be nested in different DICOM levels and as a consequence, it can be ignored if only the first DICOM level gets scanned. In addition, it is often difficult to find the conformance statement itself because vendor's model and SW version information are removed by mistake during image submission.

Furthermore, the de-identification system works separately from the original image center, image pixels might contain private data and DICOM attributes can vary based on image modalities [21,8]. More limitations come with the fact that different authorities (GDPR, HIPAA, etc.) specify different regulations, which complicates the process standardization. DICOM de-identification solutions need to be evaluated and checked for quality before being deployed, as well [27]. When it comes to using a de-facer program for de-identification, the scientific utility of the images can be compromised if all facial features get removed, but on the other hand, re-identification can be easily achieved from disclosed features [24,25]. Additionally, when generating a synthetic dataset is adopted as an alternative to de-identification, limited sample size and ad hoc computed privacy are challenges that need to be considered [26].

C. Future Directions

When dealing with the process of removing PHI from medical image data, we should be aware of the fact that various image modalities exist and different attributes might need different de-identification or anonymization strategies, such as randomizing, replacing, clearing or removing the attribute [27]. Also, real datasets used to train privacy preservation systems are limited in terms of the features that they consider [24], while synthetic generated datasets do not always greatly resemble the real ones [26]. These issues are part of the author's future plans.

Meanwhile, for the available de-identified datasets, authors plan to: improve their management; optimize their de-identification by adding machine learning OCR method to remove PHI from pixel data; provide researchers more metadata and increase the dataset utility. Authors state that in the future they will further improve the secure data storage and increase the variety of medical image data of these datasets [23].

TABLE I. MEDICAL IMAGE DE-IDENTIFICATION/ANONYMIZATION

Ref.	*Solution/Approach*	*Privacy Process*	*Open-source*	*Solution URL*	*Data/Dataset*
[26]	ac-GAN model to generate synthetic data	De-identification	✓	https://github.com/tcoroller/pGAN	MRI dataset
[27]	A two steps procedure: -De-identification at the image data source; -Pixel data de-identification;	De-identification	✓	https://csgitlab.ucd.ie/mldawn/dicom_de_identifier_public	DICOM CT-scan images (NIMIS)
[24]	Eyes, noses and ears were distorted from the images by using a Defacer SW; Header PHI removal through a function;	De-identification	✓	https://github.com/yeonuk-Jeong/Defacer	240 NIFTI+DICOM MRI images from ADNI dataset and 100 from OASIS
[21]	RSNA CTP+KNOWLEDGE BASE Customization of the de-identification scripts.	De-identification	✓	https://wiki.cancerimagingarchive.net/display/Public/Submission+and+De-identification+Overview	TCIA DICOM images https://wiki.cancerimagingarchive.net/display/Public/Wiki (CT, MR, PET)
[22]	Authors tested 10 DICOM de-identification tools. Tests were performed using: -Default settings; -Customized settings; Considering HIPAA requirements, customized RSNA CTP had the best performance.	De-identification	✓	https://mircwiki.rsna.org/index.php?title=MIRC_CTP	Tested with a dummy DICOM image. Specific modality not mentioned.
[23]	A RSNA CTP pipeline including these steps: -A PHI removal script; -Numeric attributes retaining using PyDicom; -Filter stage for pixel burned-in data; -Quality check; -Image report using TagSniffer.	De-identification	✓	https://ngdc.cncb.ac.cn/obia https://pypi.org/project/pydicom/ https://github.com/stlstevemoore/dicom-tag-sniffer RSNA MIRC Site http://mirc.rsna.org/query	OBIA (CT+MR+DX) https://ngdc.cncb.ac.cn/obia
[25]	-FreeSurfer to create a full de-faced version of the image (similar to skull-stripping); -Create a soft de-faced version of the image following three trimming steps;	De-identification	✓	https://github.com/rbruna/defacing https://github.com/rbruna/MRI-anonymization-for-MEG-coregistration	T1-weighted MRIs from 10 volunteers, members of the Cognition and Brain Sciences Unit (CBU) of the Medical Research Council, in Cambridge (UK).
[8]	RSNA CTP A two steps procedure: -De-identification of the header; -Pixel data de-identification;	De-identification	✓	https://mircwiki.rsna.org/index.php?title=MIRC_CTP	DICOM images for teaching and learning: https://mistr.usask.ca/odin/ TFS Server: https://mistrprodnew.usask.ca:8443/query (CT, DX)

V. CONCLUSIONS

Large image datasets are essential for training AI models, performing complex machine learning tasks, or simply conducting AI-driven research. This type of data contains sensitive information belonging to patients that need to be protected at all costs. Different techniques and approaches to preserve the privacy of individuals are currently available, but the majority of authors in the state-of-art, have adopted them for two main reasons: to preserve privacy in structured medical data; and to protect against adversarial attacks.

In this review paper authors have taken into consideration unstructured medical data, to be more precise, medical images. Because of their format, private information can be present not only in the visible parts but also hidden in the image pixel data. Given that, privacy protection, in this case, becomes an arduous task.

This study explores de-identification and anonymization techniques applied to medical images, as a means to preserve privacy while still being able to conduct accurate AI-driven research. From the results of our screening process, we wanted to understand if a standard de-identification or anonymization procedure is available in the existing literature. The relevant selected studies suggested different solutions, but the most adopted was the RSNA CTP pipeline. Two main reasons made the latter more preferred: the flexibility, since it could be easily customized; and the variety of imaging modalities that it could consider for de-identification.

However, there are plenty of limitations regarding image de-identification that arise because of the image nature itself, but the most important is input data quality. Since the DICOM format is a standard way to store and share medical images, a common process beginning from the first step of image submission, should be followed by different vendors or data owners. PHI must be correctly documented and conformance statements must be clearly defined.

REFERENCES

[1] J. M. Silva, E. Pinho, E. Monteiro, J. F. Silva, C. Costa, "Controlled searching in reversibly de-identified medical imaging archives", vol. 77, January 2018.

[2] R. Venugopal, N. Shafqat, I. Venugopal, B. M. J. Tillbury, H. D. Stafford, A. Bourazeri, "Privacy preserving Generative Adversarial Networks to model Electronic Health Records", vol. 153, September 2022.

[3] G. Li, R. Togo, T. Ogawa, M. Haseyama, "Compressed gastric image generation based on soft-label dataset distillation for medical data sharing", vol. 227, December 2022.

[4] D. Eke, I. E. J. Aasebø, S. Akintoye, W. knight, A. Karakasidis, E. Mikulan, et al., "Pseudonymisation of neuroimages and data protection: increasing access to data while retaining scientific utility", Neuroimage: Reports, vol. 25, September 2021.

[5] O. Lytvyn, G. Nguyen, "Secure Multi-Party Computation for Magnetic Resonance Imaging Classification", vol. 220, pp. 24-31, 2023.

[6] A. Gopalakrishnan, N. P. Kulkarni, Ch. B. Raghavendra, R. Manjappa, P. Honnavalli, S. Eswaran, "PriMed: Private federated training and encrypted inference on medical images in healthcare", March 2023.

[7] European Society of Radiology, "The new EU General Data Protection Regulation: what the radiologist should know", vol. 8, pp. 295-299, April 2017.

[8] B. Burbridge, "Dicom image anonymization and transfer to create a diagnostic radiology teaching file", International Journal of Radiology and Imaging Technology, vol. 6, September 2020.

[9] S. T. Arasteh, A. Ziller, Ch. Kuhl, M. Makowski, S. Nebelung, R. Braren, D. Rueckert, et al., "Private, fair and accurate: Training large-scale, privacy-preserving AI models in medical imaging", March 2023.

[10] A. Ziller, D. Usynin, N. W. Remerscheid, M. Knolle, M. Makowski, R. Braren, et al., "Differentially private federated deep learning for multi-site medical image segmentation", April 2022.

[11] X. Feng, Y. Xie, M. Ye, Zh. Tang, B. Yan, Sh. Wei, "Fake Gradient: A Security and Privacy Protection Framework for DNN-based Image Classification", pp. 5510-5518, October 2021.

[12] I. Jarin, B. Eshete, "PRICURE: Privacy-Preserving Collaborative Inference in a Multi-Party Setting", pp. 25-35, February 2021

[13] K. Nikolaidis, S. Kristiansen, Th. Plagemann, V. Goebel, K. Liestøl, M. Kankanhalli, et al., "Learning Realistic Patterns from Visually Unrealistic Stimuli: Generalization and Data Anonymization", December 2021.

[14] Sh. Abuadbba, K. Kim, M. Kim, Ch. Thapa, S. A. Camtepe, Y. Gao, et al., "Can We Use Split Learning on 1D CNN Models for Privacy Preserving Training?", pp. 305-318, October 2020.

[15] M. Santos, N. P. Rocha, "A Big Data Approach to Explore Medical Imaging Repositories Based on DICOM", vol. 219, pp. 1224-1231, 2023.

[16] G. Poulis, G. Loukides, S. Skiadopoulos, A. Gkoulalas-Divanis, "Anonymizing datasets with demographics and diagnosis codes in the presence of utility constraints", vol. 65, pp. 76-96, January 2017.

[17] J. W. T. M. deKok, M. A. Armengol de la Hoz, Y. de Jong, V. Brokke, P. W. G. Elbers, P. Thoral, et al., "A guide to sharing open healthcare data under the General Data Protection Regulation", vol. 10, 2023.

[18] K. Abouelmehdi, A. Beni-Hessane, H. Khalouf, "Big healthcare data: preserving security and privacy", vol. 5, 2018.

[19] A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data", vol. 31, pp. 426-435, 2019.

[20] R. Chevrier, V. Foufi, Ch. Gaudet-Blavignac, A. Robert, Ch. Lovis, "Use and Understanding of Anonymization and De-Identification in the Biomedical Literature: Scoping Review", vol. 21, 2019.

[21] S. M. Moore, D. R. Maffitt, K. E. Smith, J. S. Kirby, K. W. Clark, J. B. Freymann, et al., "De-identification of medical images with retention of scientific research value", Radiographics, vol. 35, pp. 727-735, 2015.

[22] K. Y. E. Aryanto, M. Oudkerk, P. M. A. van Oijen, "Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy", European Radiology, vol. 25, June 2015.

[23] E. Jin, D. Zhao, G. Wu, J. Zhu, Zh. Wang, Zh. Wei, "OBIA: an open biomedical imaging archive", Genomics, Proteomics & Bioinformatics, September 2023, in press.

[24] Y. U. Jeong, S. Yoo, Y. H. Kim, W. H. Shim, "De-identification of facial features in magnetic resonance images: software development using deep learning technology", Journal of Medical Internet Research, vol. 22, 2020.

[25] R. Bruña, D. Vaghari, A. Greve, E. Cooper, M. O. Mada, R. N. Henson, "Modified MRI anonymization (de-facing) for improved MEG coregistration", Bioengineering, vol. 9, October 2022.

[26] H. Sun, J. Plawinski, S. Subramaniam, A. Jamaludin, T. Kadir, A. Readie, et al., "A deep learning approach to private data sharing of medical images using conditional generative adversarial networks (GANs)", PLoS ONE, vol. 18, July 2023.

[27] A. Shahid, M. H. Bazargani, P. Banahan, B. M. Namee, T. Kechadi, C. Treacy, et al., "A two-stage de-identification process for privacy-preserving medical image analysis", Healthcare, vol. 10, April 2022.

[28] N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, J. Qadir, "Privacy-Preserving artificial intelligence in healthcare: Techniques and Applications", Computers in Biology and Medicine, vol. 158, May 2023.