International Journal of Public Health Science (IJPHS)
Vol. 13, No. 1, March 2024, pp. 330~337
ISSN: 2252-8806, DOI: 10.11591/ijphs.v13i1.23032

Software development to anonymize patient information from clinical computed tomography images

Choirul Anam[1], Ariij Naufal[1], Jhon Hadearon Saragih[2], Mohd Hanafi Ali[3]
[1]Department of Physics, Faculty of Sciences and Mathematics, Diponegoro University, Semarang, Indonesia
[2]Department of Radiation Oncology, Mayapada Hospital Tangerang, Banten, Indonesia
[3]Department of Medical Imaging, Faculty of Health Sciences, Qaiwan International University, KRG, Iraq

Article Info

Article history:

Received Feb 1, 2023
Revised Apr 6, 2023
Accepted Apr 15, 2023

Keywords:

Anonymization
Computed tomography
DICOM
IndoQCT
Patient data

ABSTRACT

A free software for digital imaging and communications in medicine (DICOM) image anonymization is needed to protect patient information from third parties. This study aimed to develop a software for the anonymization of patient information in computed tomography (CT) image. There was a total of 17 informations to be anonymized, such as Patient's Name, Other Patient Names, Patient's ID, Patient's Birth Date, Patient's Sex, Patient's Age, Study ID, Study Description, Series Description, Institution Name, Institution Adress, Referring Phisician's Name, Consulting Physician's Name, Performing Physician's Name, Name of Physician(s) Reading Study, Operator's Name, and Protocol Name. In every information, its initial value was replaced with a dummy value with the string value of "N/A". For testing, patient CT images from four different hospitals with different scanners were collected. It is found that each scanner had different information stored in DICOM information. However, the anonymization process in the four hospitals works well with accuracy of 100%. The developed software can anonymize DICOM images flexibly and successfully. This software can be used for anonymization of patient information in order to protect patient information.

Corresponding Author:

Choirul Anam
Department of Physics, Faculty of Sciences and Mathematics, Diponegoro University
Prof. Soedarto SH Street, Tembalang, Semarang 50275, Central Java, Indonesia
Email: anam@fisika.fsm.undip.ac.id

1. INTRODUCTION

Computed tomography (CT) is a main medical imaging modality used in health care system [1]. CT produces image displaying map of coefficient attenuation of the tissue inside the patient [2]. CT images are reconstructed from sinogram data into axial images [3], [4]. The following axial images can be distributed in the picture archiving and communication system (PACS) for diagnosis, communication and storage purposes [5], making them easy to share with various medical devices in the network within healthcare facilities [6].

CT images are stored in certain standardized formats, namely the digital imaging and communications in medicine (DICOM) [7]–[11]. On a single examination, CT reconstructs multiple DICOM images at once. Number of images depends on the scan length, slice thickness, slice interval, and pitch. In addition to containing image details, DICOM also contains information about examinations, patient information, institutional information, and others called as DICOM tags [9]. The standardization of DICOM makes it widely used, not only for clinical purposes, but also for education and research purposes [12], [13]. However, sharing or exchanging confidential information about patients requires data protection to ensure privacy and security [14]. Leakage of important information can result in privacy violations, as well as intrusion into the system [15], [16]. Therefore, tools are needed to protect sensitive information, one of which is the anonymization of DICOM [17]–[20].

While several tools are available for de-identifying DICOM data, including those integrated into DICOM Viewer software [21]–[24], many of these tools are commercial programs and may not be accessible to everyone. Additionally, these tools may not always work effectively in anonymizing specific data [7], highlighting the need for a freely available and user-friendly software for DICOM anonymization. Therefore, this study aims to develop a free software with a simple graphical user interface (GUI) for anonymization of patient information stored in CT image. The developed software will be a valuable too for medical staff seeking to protect patient privacy.

2. METHOD

2.1. DICOM standard

DICOM is a standard protocol for the medical image communication system. This includes information and transmission management to facilitate communication and diagnostics that are widely used in many healthcare facilities [25]. Because of its digital nature, this can be a substitute for film media for image storage and reading purposes. Initially, DICOM is introduced by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology (ACR). DICOM is a trademark that is recognized and used as a standard for broad medical imaging communication and various types of modalities.

The DICOM image contains not only data pixels, but also almost all examination information. This includes information on the modality used, scan parameters, patient information, and institutional information. This information is stored in a single file as metadata that can be accessed using tags [20]. Most tags have keywords to get specific information and its value. NEMA as the initiator of the DICOM system applies a dictionary structure to sort the DICOM tags. For example, Table 1 shows some of the tags used in DICOM. The tags function is to sort information by category of information. Each tag has a standard name, and a keyword to access its value. The value can also be accessed directly using tags.

Table 1. Examples of some tags used by DICOM

Tag	Name	Keyword	Value
(0010,0010)	Patient's Name	PatientName	String
(0018,1151)	X-Ray Tube Current	XRayTubeCurrent	Integer
(0008,0022)	Acquisition Date	AcquisitionDate	String

In this study, the DICOM anonymization software was written using the Python programming language and has been integrated into the IndoQCT [26]. DICOM files were accessed using the Pydicom module package [27], [28]. Using this module, DICOM files were read and stored in memory as a dataset. The dataset is the main object of an image that is processed for various operations, both related to image processing and dosimetry measurements [29]. The dataset contains data elements that store DICOM image information. This includes those discussed in the previous section, namely tags, names, and values. To access data elements, specific tags or keywords can be used in the dataset as attributes. For example, (1) displays a command to call the data element of the patient's name. While (2) gives the command to call the value of the data element of the patient's name.

ds[‘PatientName’] (1)

ds.PatientName (2)

where ds is the dataset of the DICOM images that have been read. To change the value of the specific data element, someone can assign a replacement value on the value attribute of the specific data elements as in (3).

ds.[‘PatientName’].value = ‘new name’ (3)

Anonymization performs value replacement on some sensitive data elements in a dataset using a new anonymous value. By default, we set it to use a dummy value of “not applicable” or N/A. The assignment of dummy values was repeated according to the list of keywords that can be seen in Table 2.

Figure 1 displays the GUI for anonymizing patients’ information. The form will be enabled if the data element was detected with the keywords shown in Table 3. This adjustment was added because not every DICOM file contains the intended data element, even though it has been standardized. The data element contained in the file was vendor-specific. After the image was anonymized, it can be saved as a new file in the desired directory and file name. In the developed software, an option for anonymizing all slices was provided.

Table 2. List of anonymized data elements

Tag	Name	Keyword
(0010,0010)	Patient's Name	PatientName
(0010,1001)	Other Patient Names	OtherPatientNames
(0010,0020)	Patient ID	PatientID
(0010,0030)	Patient's Birth Date	PatientBirthDate
(0010,0040)	Patient's Sex	PatientSex
(0010,1010)	Patient's Age	PatientAge
(0020,0010)	Study ID	StudyID
(0008,1030)	Study Description	StudyDescription
(0008,103E)	Series Description	SeriesDescription
(0008,0080)	Institution Name	InstitutionName
(0008,0081)	Institution Address	InstitutionAddress
(0008,0090)	Referring Physician's Name	ReferringPhysicianName
(0008,009C)	Consulting Physician's Name	ConsultingPhysicianName
(0008,1050)	Performing Physician's Name	PerformingPhysicianName
(0008,1060)	Name of Physician(s) Reading Study	NameOfPhysiciansReadingStudy
(0008,1070)	Operator's Name	OperatorsName
(0018,1030)	Protocol Name	ProtocolName

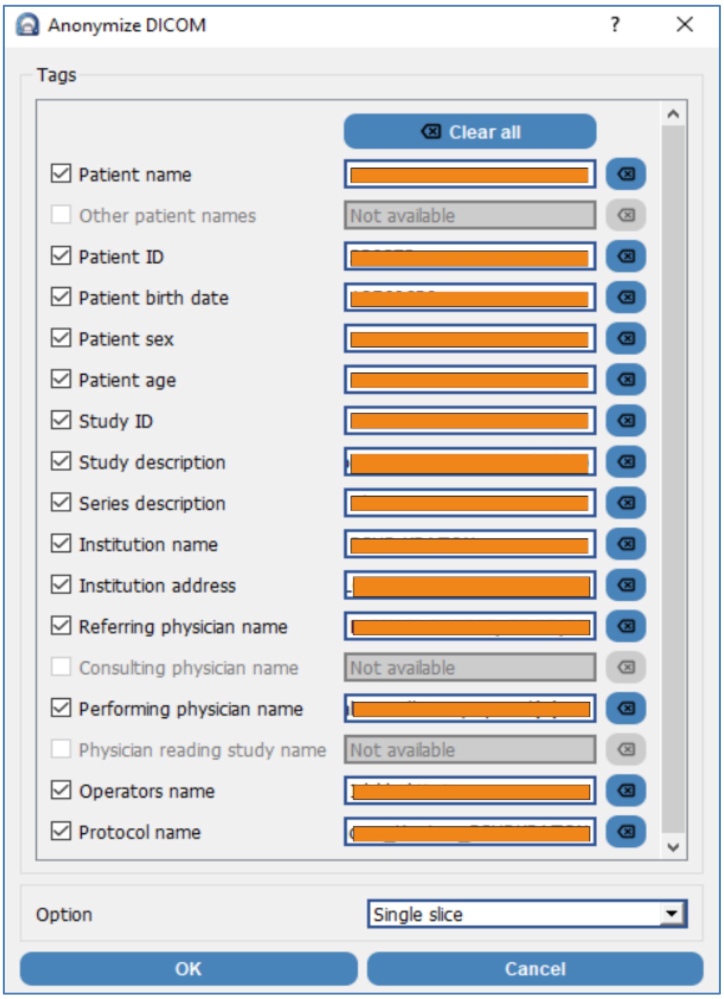

Figure 1. Graphical user interface (GUI) for anonymization

Table 3. Types of CT scanners used for testing the developed software

Hospital	Scanner	Protocol
Hospital A	Siemens Emotion 6	Abdomen
Hospital B	Siemens Somatom Definition AS	Thorax with contrast
Hospital C	Toshiba Aquillion	Chest
Hospital D	Hitachi ECLOS	Abdomen

2.2. Dataset

Four datasets of clinical images from four different hospitals with different types of CT scanners were used for testing the developed software. The type of scanner can be seen in Table 3. Both files (before and after anonymization) were then opened using the Microdicom Viewer to see the comparison.

3. RESULTS AND DISCUSSION

3.1. Comparison using microdicom viewer

Figures 2 to 5 show the original images with the patient and institutional information (blurred) and the anonymized images of hospital A as presented in Figure 2, hospital B as shown in Figure 3, hospital C as presented in Figure 4, and hospital D as shown in Figure 5, respectively. It can be seen that the original image contains patient and hospital information which is displayed on the viewer screen by default. However, after the anonymization process was carried out, the information was replaced with a new value, namely "N/A". The results can be seen directly in the anonymized figure so that patient and hospital information does not appear. From four hospitals, the anonymization process works well.

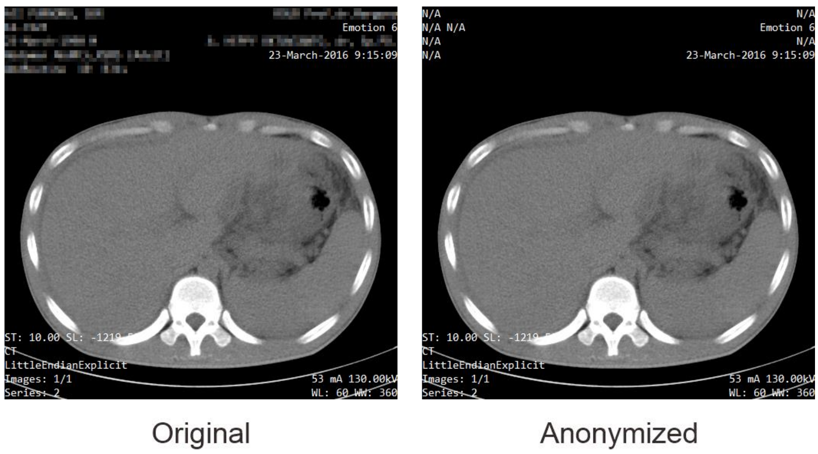

Figure 2. Anonymized DICOM image at hospital A

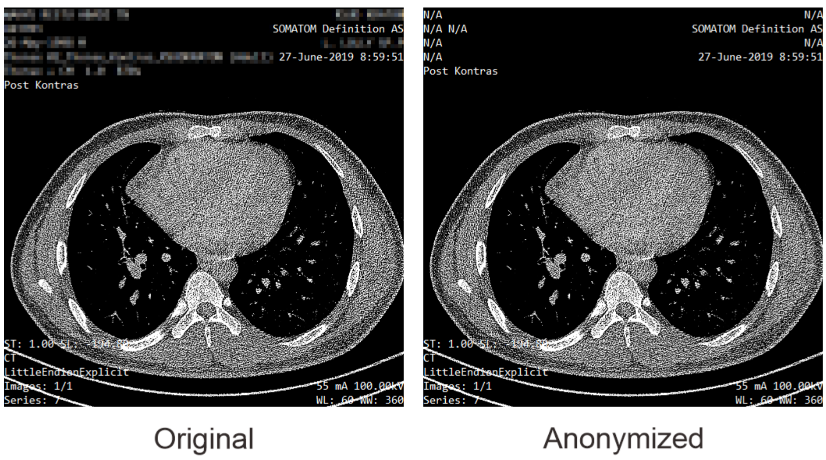

Figure 3. Anonymized DICOM image at hospital B

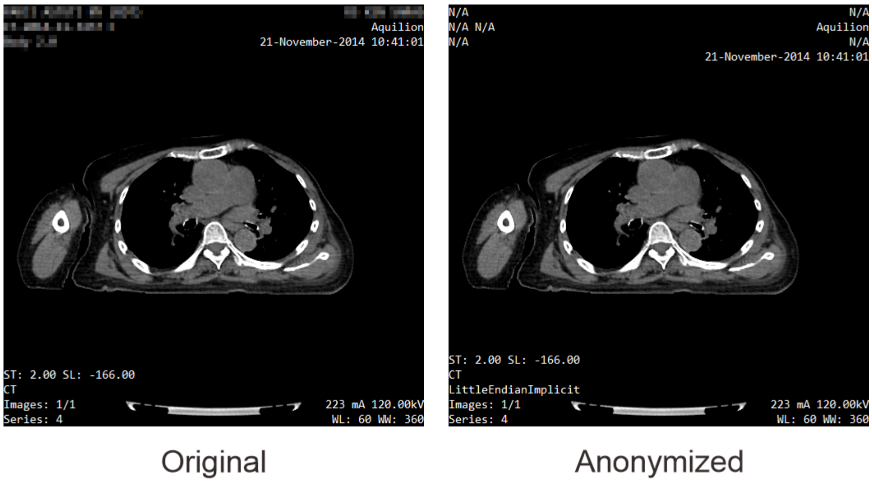

Figure 4. Anonymized DICOM image at hospital C

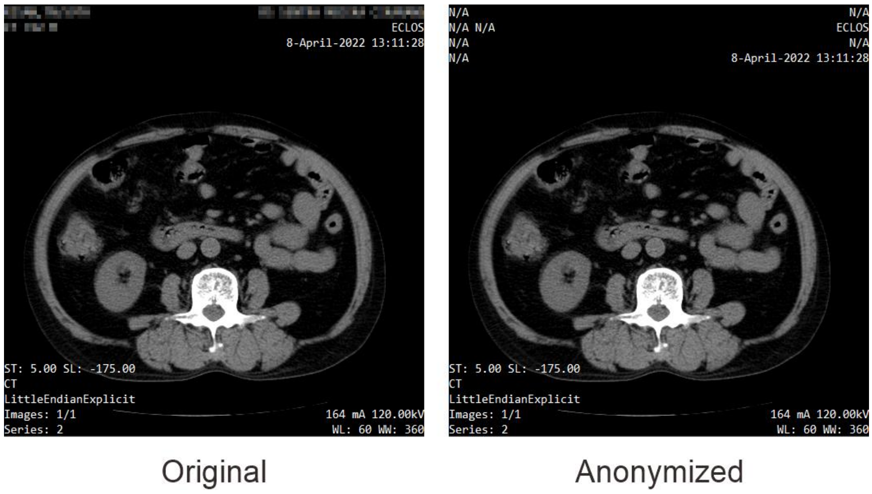

Figure 5. Anonymized DICOM image at hospital D

3.2. Data elements anonymization

Table 4 shows the anonymization results per data element. It can be seen that confidential information can be properly anonymized if it is available. Of the four hospitals, the information contained in DICOM is different. This depends on the type of scanner and settings. Some hospitals provide a tag whose value is empty.

The DICOM anonymization program is important to secure sensitive information, both information about patients and about institutions [30]. The aim of this study is to develop a DICOM anonymization program that can be run on various types of CT scanners in various hospitals. The success rate of the program is tested by running anonymization on many predefined tags.

Table 4 indicates that all the information contained in the DICOM file can be accessed and its value can be changed with the dummy value "N/A" properly. The availability of data elements varies depending on the type of scanner. Not all CT scanners provide the same specific information as other scanners (e.g. SpiralPitchFactor) even though they are standardized. The PerformingPhysicianName tag in Table 4 indicates the different availability. However, the developed software has only been tested using only CT images because the parent program (i.e. IndoQCT) is a program whose main purpose is to evaluate the quality of CT images [26]. Images from other modalities still need further investigation.

Several limitations exist in this study. The element data included for anonymization is still very limited and does not meet the standards of NEMA (Sup 142) [31]. The anonymization process is also still limited to replacing the value with a new value in accordance with the purpose of anonymization. In Sup 142, there is a recommended protocol for the anonymization of each data element. For example, PatientName is recommended to be cleaned completely, or replaced with a new value according to anonymization purposes. PatientBirthDate is recommended to be replaced with a zero-length value, and PatientAge can be used to replace PatientBirthDate if needed for educational purposes. Dummy values can be substitutes for data elements that point to institutions such as InstitutionName or InstitutionAddress. Private tags also need to be considered because they may contain patient information, as well as important exposure information that should be protected. Hence, the development of our program still needs further improvements.

Table 4. Anonymization results on various hospitals

Name	Hospital A	Hospital B	Hospital C	Hospital D
Patient's Name	S	S	S	S
Other Patient Names	S	NA	NA	NA
Patient ID	S	S	S	S
Patient's Birth Date	S	S	S	S
Patient's Sex	S	S	S	S
Patient's Age	S	S	S	S
Study ID	S	S	S	S
Study Description	S	S	NA	S
Series Description	S	S	S	S
Institution Name	S	S	S	S
Institution Address	S	S	NA	NA
Referring Physician's Name	S	S	S	S
Consulting Physician's Name	NA	NA	NA	S
Performing Physician's Name	NA	S	NA	S
Name of Physician(s) Reading Study	NA	NA	NA	S
Operators' Name	S	S	NA	S
Protocol Name	S	S	S	S

*Abbreviation: S = Success; NA = Not available

There are also other limitations regarding the burnt-in information in the image pixels [14], [32], [33]. This includes the dose report image, as well as the tomogram image. In this case, it is necessary to change the image on the actual pixels. If it is not, then anonymization at the metadata level will lose its purpose [21]. Sensitive information in dose reports and DICOM tags can be a way for third parties to infiltrate, both online and local systems [34], [35]. Therefore, it is important for the anonymization process to be enforced according to standards.

Anonymization programs perform their tasks locally or offline. This means all processes are executed after the DICOM image is exported to the user's device. To ensure the protection of patient information, medical staff must monitor the anonymization of the user's device from third parties before handing over anonymized DICOM to them. In other cases, medical staff may access DICOM directly from the PACS server to the user's device. It's certainly a concern to organize the original and anonymized files separately. Another solution can be run, namely by anonymizing DICOM on the direct export feature of PACS [16]. This will be carried out in our future work.

4. CONCLUSION

Software to anonymize patient's information stored CT images has been successfully developed. A total of 17 information can be anonymized by a dummy value. Although each information stored in each vendor is different, the developed software can handle these differences well. With this developed software, patient data can be anonymized. This is very useful for various purposes, for example for research purposes in order to maintain the confidentiality of patient.

ACKNOWLEDGEMENTS

This work was funded by the *Riset Publikasi International Bereputasi Tinggi (RPIBT)*, Diponegoro University (No. 569-187/UN7.D2/PP/IV/2023).

REFERENCES

[1] D. L. Miglioretti *et al.*, "The Use of Computed Tomography in Pediatrics and the Associated Radiation Exposure and Estimated Cancer Risk," *JAMA Pediatrics*, vol. 167, no. 8, p. 700, Aug. 2013, doi: 10.1001/jamapediatrics.2013.311.

[2] S. Mauf *et al.*, "Flat chest projection in the detection and visualization of rib fractures: A cross-sectional study comparing curved and multiplanar reformation of computed tomography images in different reader groups," *Forensic Science International*, vol. 303, p. 109942, Oct. 2019, doi: 10.1016/j.forsciint.2019.109942.

[3] D. Karimi, P. Deman, R. Ward, and N. Ford, "A sinogram denoising algorithm for low-dose computed tomography," *BMC Medical Imaging*, vol. 16, no. 1, p. 11, Dec. 2016, doi: 10.1186/s12880-016-0112-5.

[4] Y. Li, K. Li, C. Zhang, J. Montoya, and G.-H. Chen, "Learning to Reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10,

pp. 2469–2481, Oct. 2019, doi: 10.1109/TMI.2019.2910760.

[5] D. Haak, C.-E. Page, S. Reinartz, T. Krüger, and T. M. Deserno, "DICOM for clinical research: PACS-integrated electronic data capture in multi-center trials," *Journal of Digital Imaging*, vol. 28, no. 5, pp. 558–566, Oct. 2015, doi: 10.1007/s10278-015-9802-8.

[6] Z. M. Alalawi, M. M. Eid, and A. I. Albarrak, "Assessment of picture archiving and communication system (PACS) at three of ministry of health hospitals in Riyadh region – Content analysis," *Journal of Infection and Public Health*, vol. 9, no. 6, pp. 713–724, Nov. 2016, doi: 10.1016/j.jiph.2016.09.004.

[7] D. Haak, C.-E. Page, and T. M. Deserno, "A survey of DICOM viewer software to integrate clinical research and medical imaging," *Journal of Digital Imaging*, vol. 29, no. 2, pp. 206–215, Apr. 2016, doi: 10.1007/s10278-015-9833-1.

[8] M. Larobina and L. Murino, "Medical image file formats," *Journal of Digital Imaging*, vol. 27, no. 2, pp. 200–206, Apr. 2014, doi: 10.1007/s10278-013-9657-9.

[9] B. Chen, S. Leng, L. Yu, D. Holmes, J. Fletcher, and C. McCollough, "An open library of CT patient projection data," in *Proceedings SPIE 9783, Medical Imaging 2016: Physics of Medical Imaging*, D. Kontos, T. G. Flohr, and J. Y. Lo, Eds., Mar. 2016, p. 97831B. doi: 10.1117/12.2216823.

[10] S. G. Langer, "DICOM Data Warehouse: Part 2," *Journal of Digital Imaging*, vol. 29, no. 3, pp. 309–313, Jun. 2016, doi: 10.1007/s10278-015-9830-4.

[11] R. D. Ernst, B. R. Baumgartner, E. P. Tamm, and W. E. Torres, "Development of a teaching file by using a DICOM database," *RadioGraphics*, vol. 22, no. 1, pp. 217–221, Jan. 2002, doi: 10.1148/radiographics.22.1.g02ja10217.

[12] Y. Chen *et al.*, "Constructing an experiential education model in undergraduate radiology education by the utilization of the picture archiving and communication system (PACS)," *BMC Medical Education*, vol. 19, no. 1, p. 383, Dec. 2019, doi: 10.1186/s12909-019-1827-0.

[13] P. Kathiravelu *et al.*, "A DICOM framework for machine learning and processing pipelines against real-time radiology images," *Journal of Digital Imaging*, vol. 34, no. 4, pp. 1005–1013, Aug. 2021, doi: 10.1007/s10278-021-00491-w.

[14] P. Vcelak, M. Kryl, M. Kratochvil, and J. Kleckova, "Identification and classification of DICOM files with burned-in text content," *International Journal of Medical Informatics*, vol. 126, pp. 128–137, Jun. 2019, doi: 10.1016/j.ijmedinf.2019.02.011.

[15] Z. Xu, "An empirical study of patients' privacy concerns for health informatics as a service," *Technological Forecasting and Social Change*, vol. 143, pp. 297–306, Jun. 2019, doi: 10.1016/j.techfore.2019.01.018.

[16] M. Eichelberg, K. Kleber, and M. Kämmerer, "Cybersecurity challenges for PACS and medical imaging," *Academic Radiology*, vol. 27, no. 8, pp. 1126–1139, Aug. 2020, doi: 10.1016/j.acra.2020.03.026.

[17] W. Newhauser *et al.*, "Anonymization of DICOM electronic medical records for radiation therapy," *Computers in Biology and Medicine*, vol. 53, pp. 134–140, Oct. 2014, doi: 10.1016/j.compbiomed.2014.07.010.

[18] E. Monteiro, C. Costa, and J. L. Oliveira, "A de-identification pipeline for ultrasound medical images in DICOM format," *Journal of Medical Systems*, vol. 41, no. 5, p. 89, May 2017, doi: 10.1007/s10916-017-0736-1.

[19] J. D. Robinson, "Beyond the DICOM header: additional issues in deidentification," *American Journal of Roentgenology*, vol. 203, no. 6, pp. W658–W664, Dec. 2014, doi: 10.2214/AJR.13.11789.

[20] A. Shahid *et al.*, "A two-stage de-identification process for privacy-preserving medical image analysis," *Healthcare*, vol. 10, no. 5, p. 755, Apr. 2022, doi: 10.3390/healthcare10050755.

[21] K. Y. E. Aryanto, M. Oudkerk, and P. M. A. van Ooijen, "Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy," *European Radiology*, vol. 25, no. 12, pp. 3685–3695, Dec. 2015, doi: 10.1007/s00330-015-3794-0.

[22] J. S. Wadali, S. P. Sood, R. Kaushish, S. Syed-Abdul, P. K. Khosla, and M. Bhatia, "Evaluation of free, open-source, web-based dicom viewers for the indian national telemedicine service (eSanjeevani)," *Journal of Digital Imaging*, vol. 33, no. 6, pp. 1499–1513, Dec. 2020, doi: 10.1007/s10278-020-00368-4.

[23] M. E. Suleiman, P. C. Brennan, E. Ekpo, P. Kench, and M. F. McEntee, "Integrating mammographic breast density in glandular dose calculation," *The British Journal of Radiology*, vol. 91, no. 1085, p. 20180032, Feb. 2018, doi: 10.1259/bjr.20180032.

[24] A. Brühschwein *et al.*, "Free DICOM-Viewers for Veterinary Medicine," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 54–63, Feb. 2020, doi: 10.1007/s10278-019-00194-3.

[25] J. Bruthans, "The successful usage of the DICOM images exchange system (ePACS) in the czech republic," *Applied Clinical Informatics*, vol. 11, no. 01, pp. 104–111, Jan. 2020, doi: 10.1055/s-0040-1701252.

[26] C. Anam, A. Naufal, T. Fujibuchi, K. Matsubara, and G. Dougherty, "Automated development of the contrast–detail curve based on statistical low-contrast detectability in CT images," *Journal of Applied Clinical Medical Physics*, vol. 23, no. 9, Sep. 2022, doi: 10.1002/acm2.13719.

[27] M. D. Herrmann *et al.*, "Implementing the DICOM Standard for digital pathology," *Journal of Pathology Informatics*, vol. 9, no. 1, p. 37, Jan. 2018, doi: 10.4103/jpi.jpi_42_18.

[28] D. Cutright, M. Gopalakrishnan, A. Roy, A. Panchal, and B. B. Mittal, "DVH analytics: A DVH database for clinicians and researchers," *Journal of Applied Clinical Medical Physics*, vol. 19, no. 5, pp. 413–427, Sep. 2018, doi: 10.1002/acm2.12401.

[29] M. Porzio and C. Anam, "Real-time fully automated dosimetric computation for CT images in the clinical workflow: A feasibility study," *Frontiers in Oncology*, vol. 12, Aug. 2022, doi: 10.3389/fonc.2022.798460.

[30] T. White, E. Blok, and V. D. Calhoun, "Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed," *Human Brain Mapping*, vol. 43, no. 1, pp. 278–291, Jan. 2022, doi: 10.1002/hbm.25120.

[31] J. B. Freymann, J. S. Kirby, J. H. Perry, D. A. Clunie, and C. C. Jaffe, "Image data sharing for biomedical research—meeting HIPAA requirements for de-identification," *Journal of Digital Imaging*, vol. 25, no. 1, pp. 14–24, Feb. 2012, doi: 10.1007/s10278-011-9422-x.

[32] M. Aiello, G. Esposito, G. Pagliari, P. Borrelli, V. Brancato, and M. Salvatore, "How does DICOM support big data management? Investigating its use in medical imaging community," *Insights into Imaging*, vol. 12, no. 1, p. 164, Dec. 2021, doi: 10.1186/s13244-021-01081-8.

[33] A. Shah, P. S. Muddana, and S. Halabi, "A review of core concepts of imaging informatics," *Cureus*, vol. 14, no. 12, pp. 1–7, Dec. 2022, doi: 10.7759/cureus.32828.

[34] P. L. K. Mantos and I. Maglogiannis, "Sensitive Patient data hiding using a ROI reversible steganography scheme for DICOM Images," *Journal of Medical Systems*, vol. 40, no. 6, p. 156, Jun. 2016, doi: 10.1007/s10916-016-0514-5.

[35] A. Al-Haj, "Providing integrity, authenticity, and confidentiality for header and pixel data of DICOM images," *Journal of Digital Imaging*, vol. 28, no. 2, pp. 179–187, Apr. 2015, doi: 10.1007/s10278-014-9734-8.

BIOGRAPHIES OF AUTHORS

Choirul Anam completed his Ph.D in Physics Department, Bandung Institute of Technology (ITB). He received Master degree from University of Indonesia (UI) and the B.Sc degree from Diponegoro University (UNDIP). He is currently working as a Lecturer and Researcher at the UNDIP. His research interests are medical image processing and dosimetry for diagnostic radiology, particularly in CT. He is developer of two software, i.e. IndoseCT (for calculating and managing radiation dose of CT) and IndoQCT (for measuring CT image quality). He can be contacted at email: anam@fisika.fsm.undip.ac.id.

Ariij Naufal is a graduate student of the physics master's program at Diponegoro University. He is the developer of IndoQCT. He can be contacted at email: ariij.2019@fisika.fsm.undip.ac.id.

Jhon Hadearon Saragih is a medical physicist at Mayapada Hospital Tangerang. He is a medical staff in a radiotherapy and radiation protection facility. He can be contacted at email: jhonhadearonsaragih@gmail.com.

Mohd Hanafi Ali is a lecturer in medical imaging at Qaiwan International University, Iraq. His expertise is CT dosimetry and image quality. He can be contacted at email: han434@gmail.com.