scientific data

OPEN

DATA DESCRIPTOR

Pixel-level annotated dataset of computed tomography angiography images of acute pulmonary embolism

João Mario Clementin de Andrade[1]✉, Gabriel Olescki[2], Dante Luiz Escuissato[1], Lucas Ferrari Oliveira[2], Ana Carolina Nicolleti Basso[1] & Gabriel Lucca Salvador[1]

Pulmonary embolism has a high incidence and mortality, especially if undiagnosed. The examination of choice for diagnosing the disease is computed tomography pulmonary angiography. As many factors can lead to misinterpretations and diagnostic errors, different groups are utilizing deep learning methods to help improve this process. The diagnostic accuracy of these methods tends to increase by augmenting the training dataset. Deep learning methods can potentially benefit from the use of images acquired with devices from different vendors. To the best of our knowledge, we have developed the first public dataset annotated at the pixel and image levels and the first pixel-level annotated dataset to contain examinations performed with equipment from Toshiba and GE. This dataset includes 40 examinations, half performed with each piece of equipment, representing samples from two medical services. We also included measurements related to the cardiac and circulatory consequences of pulmonary embolism. We encourage the use of this dataset to develop, evaluate and compare the performance of new AI algorithms designed to diagnose PE.

Background & Summary

Pulmonary embolism (PE) has a high incidence and mortality. It occurs when a blood clot, most commonly from the deep venous system, moves into the pulmonary arterial circulation[1]. Up to 300,000 deaths per year are estimated to occur in the United States due to PE[2]. Less than 10% of deaths occur among diagnosed and treated patients, indicating a potential reduction in mortality by improving the diagnostic accuracy for the disease[3].

Computed tomography pulmonary angiography (CTPA) is the examination of choice for evaluating patients with PE[4,5]. After intravenous infusion of iodinated contrast medium, CT is performed when there is optimal opacification of the pulmonary arterial circulation, and the thrombus is identified as an intraluminal filling defect.

CTPA image interpretation is a complex task: radiologists must carefully search for contrast filling defects in the entire pulmonary arterial vasculature across a large number of images. Technical problems, patient-related factors, anatomical issues and the presence of other pathologies[6] can lead to misdiagnosis.

Computer-aided diagnosis (CAD) programs aimed at reducing these errors can reduce mortality. Several approaches have already been proposed[7–48]; however, a definitive solution has not yet been reached. More recently, there has been increased interest in the creation of artificial intelligence (AI) techniques, especially using artificial neural networks (ANNs), for addressing this problem.

The diagnostic performance of these techniques is highly dependent on the dataset used for training them, as it must contain examinations as diverse as those in real applications. The diagnostic accuracy of these methods tends to increase by augmenting the training dataset[49].

Obtaining reliable datasets is a considerable obstacle encountered by researchers, as it is time-consuming, requires radiologists with experience to recognize PE and depends on medical center cooperation. Furthermore, to be suitable for supervised learning applications, the dataset must be annotated. Two different annotation approaches have been used in the three public datasets available: a pixel-level annotation, in which all pixels of

[1]Department of Radiology and Image Diagnosis, Hospital de Clínicas, Federal University of Paraná, Curitiba, Brazil. [2]Department of Informatics, Federal University of Paraná, Curitiba, Brazil. ✉e-mail: joaoclementin@gmail.com

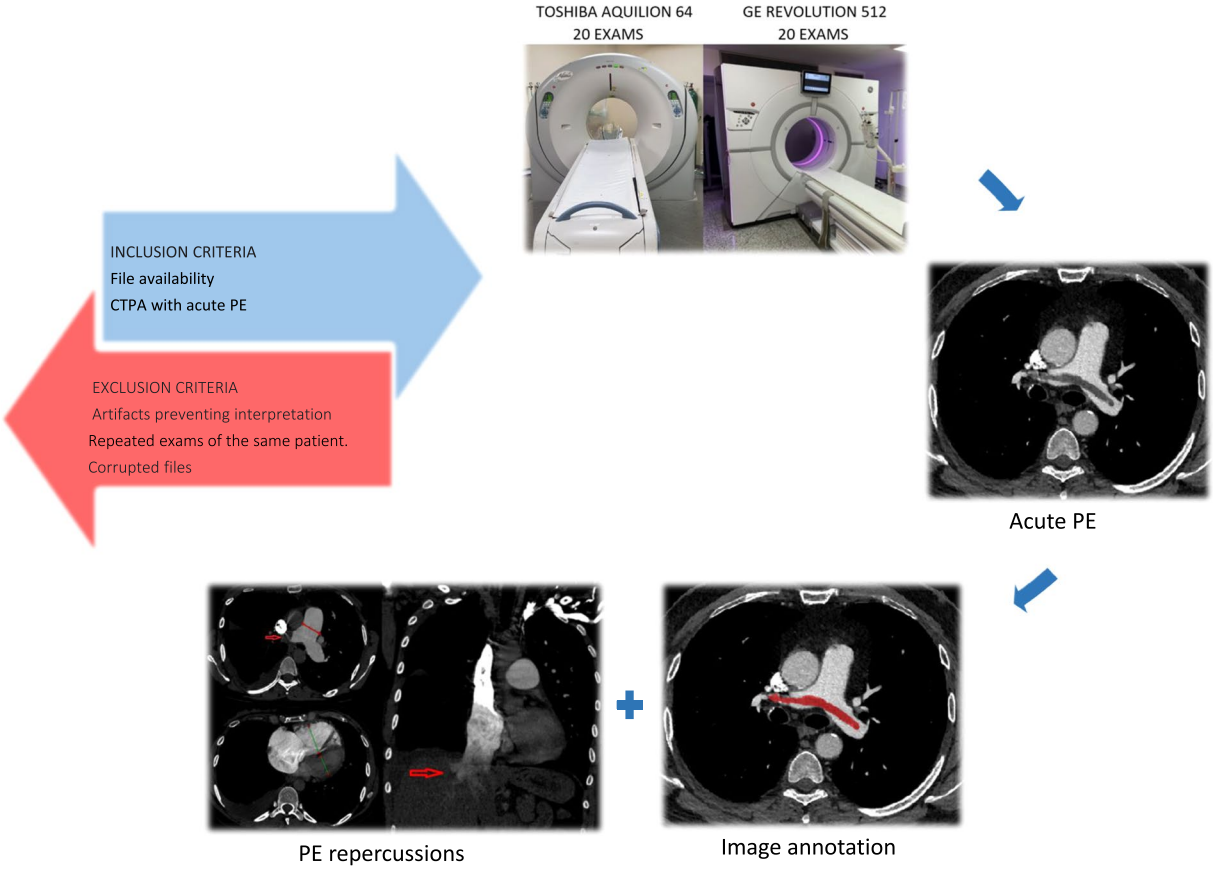

Fig. 1 Steps for producing the dataset. First step: definition of inclusion and exclusion criteria. Second and third steps: selection of 20 examinations performed with each machine on patients with acute PE. Fourth step: image annotation at the pixel and image levels. Fifth step: evaluation of features related to right heart strain and pulmonary artery hypertension.

the thrombus are demarcated as a ground truth, and annotations at the image and study levels[50], in which images with a visible PE receive a label, but the thrombus itself is not demarcated. The first approach is more versatile and can be converted into an annotation at the image level, but the opposite is not true.

The pixel-level annotation may help the algorithm to predict the exact region in which the embolus is located and to verify if this was the region used by the algorithm to generate the diagnostic output. It is especially useful for evaluating CADs designed to diagnose PE, as a high false-positive rate is a notable challenge for many of these algorithms.

To date, there are only two public datasets containing CTPA examinations with pulmonary emboli annotated at the pixel level. The first one contains 91 examinations of patients with PE obtained using SIEMENS CT scanners[51], and the second contains 35 examinations, 33 of which were conducted in patients with PE, obtained with CT scanners from PHILIPS and Neusoft Medical System Co[52]. The information published from these datasets does not disclose the examination selection process or the inclusion and exclusion criteria.

A recently published guide for research on AI[53] highlights the importance of using datasets containing images from devices from multiple vendors due to the variability inherent in such images.

There is a shortage of public datasets with a representative sample of examinations annotated at the pixel level that the AI algorithm can process in a real clinical setting. In medical practice, there is great variability among CTPA images, which occurs due to patient-related factors, such as different biotypes and comorbidities that can obscure PE, and technical factors, such as delays in image acquisition following infusion of the contrast media or with inadequate infusion flow, that lead to suboptimal image quality.

We developed a dataset with a sample of cases of acute PE annotated at the pixel and image levels, making it suitable for algorithms developed using both approaches (Fig. 1). Our dataset contains 40 examinations performed with multidetector scanners, half from a Toshiba CT and the other half from a GE CT[54].

The dataset was primarily used in conjunction with the two public datasets of CTPA examinations with pulmonary emboli annotated at the pixel level[51,52] to develop a program for the diagnosis of PE[48]. A method capable of finding and segmenting PEs in CT images using deep learning was subsequently developed.

We encourage the use of this dataset to develop, evaluate and compare the performance of AI algorithms designed to diagnose PE.

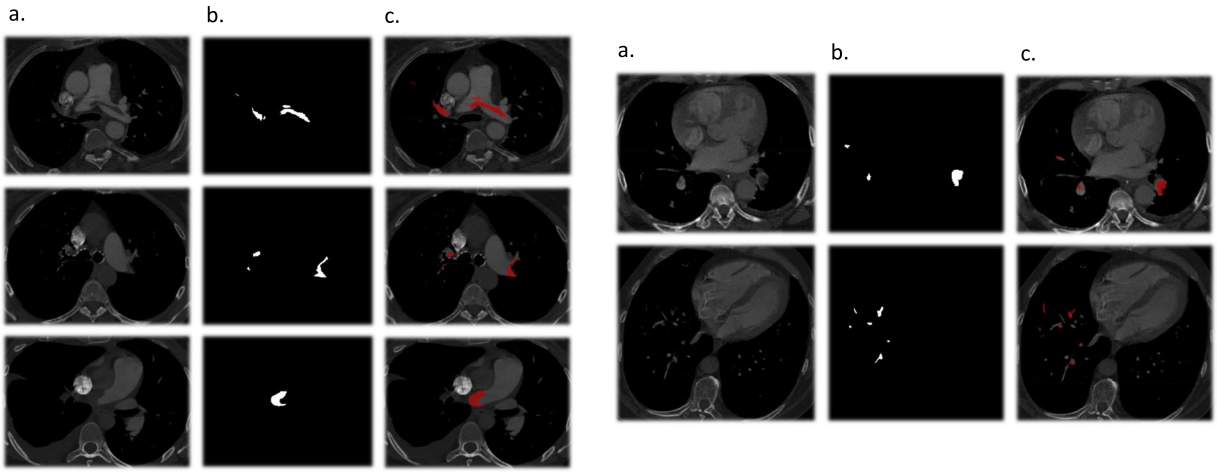

Fig. 2 Pixel level annotation. Column (**a**) shows CTPA images of pulmonary emboli in different anatomical locations. In column (**b**), pixel-level annotations show all pixels of the embolus of the corresponding CTPA image in column (**a**) in white. In column (**c**), the corresponding images of columns (**a**) and (**b**) are superimposed, and the thrombus is shown in red.

Methods

The study was approved by the Research Ethics Committee of Hospital de Clínicas-Federal University of Paraná. Given the retrospective nature of the study, only retrospective access to anonymized scan files was necessary, and so the need for informed consent was waived by the ethics committee. Aiming to preserve the identity of the participants, the examinations were deidentified by deleting patients' personal information, such as names, date of birth and identification numbers from the CT scans. Fields corresponding to time or numbers were replaced by "000000.00". Fields corresponding to dates were replaced by "00010101". Written Fields were replaced by "Anonymized" or removed.

Imaging. Twenty CTPA scans were performed in a public university hospital with a 64-channel Toshiba Aquilion CT scanner, with a tube voltage of 120 KVp, slice thickness of 1.0 mm, gantry rotation time of 0.5 seconds, beam pitch of 1.485 and dose modulation protocol.

The other twenty scans were performed in a private imaging practice with a GE Revolution 512 CT scanner, with a tube voltage of 120 KVp, slice thickness of 0.625 mm, slice interval slice thickness of 0.625 mm, gantry rotation time of 0.5 seconds, beam pitch of 0.9 and dose modulation protocol.

Image segmentation. All CTPA scans were used to make a diagnosis of acute PE by a staff radiologist. The diagnosis and location of the acute pulmonary embolism were confirmed by a thoracic radiologist with 32 years of experience. After this, a third-year resident generated the pixel-level ground-truth masks, which were revised by a certified radiologist.

The examinations were segmented using manual mode in ITK-SNAP[55], generating the ground-truth mask in which the thrombus pixels are demarcated (Fig. 2). Based on this mask, image-level segmentation was performed, labeling slices containing the thrombi.

CTPA features related to right heart strain and pulmonary artery hypertension. Obstruction of the pulmonary vasculature due to PE can increase vascular resistance, leading to an increase in pulmonary arterial pressure and right cardiac strain. Indirect signs such as pulmonary artery dilatation, right ventricular enlargement (increase in the right ventricle-to-left ventricle diameter ratio), inferior vena cava (IVC) contrast reflux, and abnormal positioning of the interventricular septum (flattening or even paradoxically bowing toward the left ventricle), can be observed on CTPA scans (Fig. 3).

In all CTPA scans, we evaluated the largest artery involved, inferior vena cava reflux, interventricular septum flattening or paradoxical bowing, pulmonary artery trunk diameter (PAD), transverse diameters of the right ventricle (RV) and left ventricle (LV)—measured between the endocardial surfaces in the largest place perpendicular to the longitudinal axis—and right ventricle-to-left ventricle diameter ratio (Tables 1, 2).

Data Records

All data records described in this paper are available on a Figshare repository[54]. This repository contains three folders. The first contains CTPA images in Digital Imaging and Communications in Medicine (DICOM) format. The second contains the ground-truth pixel-level segmentation of the location of the pulmonary embolus in Neuroimaging Informatics Technology Initiative (NiFTI) format. The segmentations consist of a three-dimensional matrix in which each element corresponds to a voxel of the CT scan: the elements corresponding to the embolus have a value of "1", and the others have a value of "0". The third is in comma separate values (CSV) format, in which each element corresponds to a slice of the CT scan. The first element represents

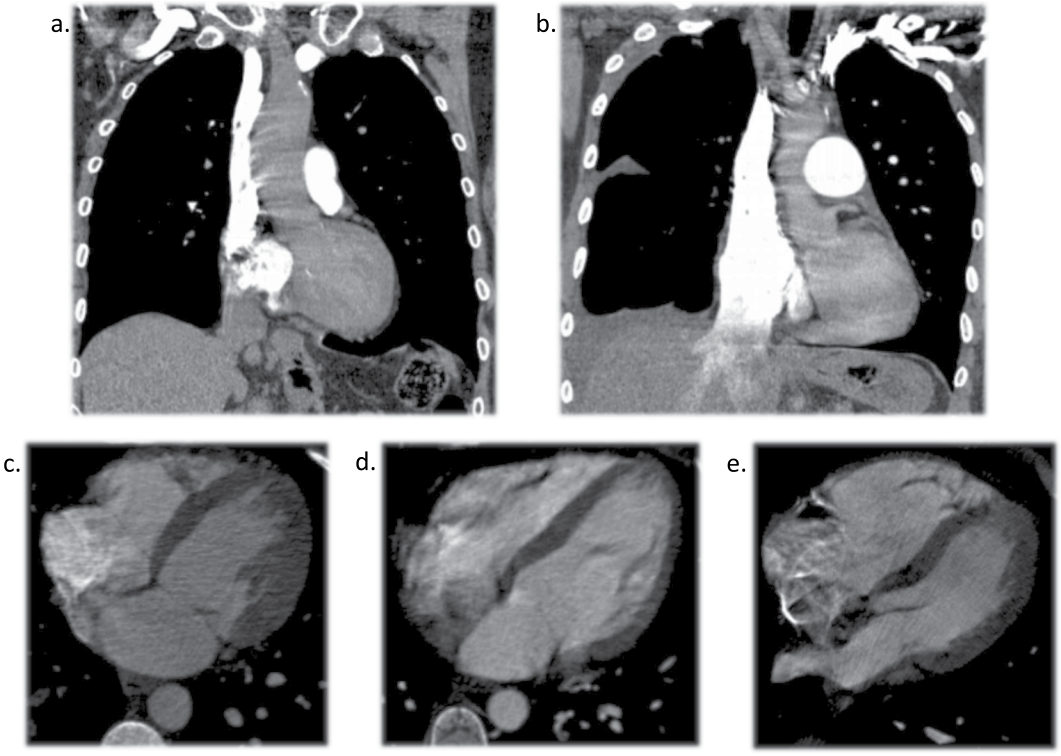

Fig. 3 CTPA features related to right heart strain and pulmonary artery hypertension. (**a**) Exam TS04 without IVC reflux. (**b**) Exam TS19 with IVC reflux. (**c**) Exam TS19 showing the interventricular septum in its normal position. (**d**) Exam TS02 showing a flattened interventricular septum. (**e**) Exam TS10 showing paradoxical interventricular septal bowing.

Case	Sex	Age	PAD (mm)	RV (mm)	LV (mm)	RV/LV	IVC Reflux	IV Septum	Largest affected vessel
1	F	63	24	43	57	0.75	Present	Normal	Lobar
2	F	11	20	37	40	0.92	Absent	Flattened	Trunk bifurcation
3	F	77	29	45	35	1.29	Absent	Flattened	Main pulmonary artery
4	F	38	34	52	29	1.79	Present	Paradoxically bowed	Main pulmonary artery
5	F	68	29	41	43	0.95	Absent	Flattened	Trunk bifurcation
6	F	54	28	39	54	0.72	Present	Flattened	Main pulmonary artery
7	F	58	35	33	35	0.94	Present	Normal	Lobar
8	F	64	31	37	48	0.77	Absent	Flattened	Segmental
9	M	29	28	23	31	0.74	Absent	Flattened	Segmental
10	M	68	26	52	35	1.49	Present	Paradoxically bowed	Main pulmonary artery
11	F	81	27	29	42	0.69	Present	Normal	Lobar
12	F	84	31	52	41	1.27	Present	Normal	Trunk bifurcation
13	F	41	24	32	38	0.84	Absent	Normal	Segmental
14	F	48	34	43	51	0.84	Present	Normal	Segmental
15	F	45	20	40	45	0.89	Absent	Normal	Main pulmonary artery
16	F	75	31	33	46	0.72	Present	Normal	Subsegmental
17	F	59	31	42	36	1.17	Present	Flattened	Trunk bifurcation
18	F	42	24	42	45	0.93	Absent	Normal	Segmental
19	F	26	25	35	50	0.70	Absent	Normal	Lobar
20	F	50	25	34	52	0.66	Absent	Normal	Main pulmonary artery

Table 1. Data from exams obtained with the Toshiba Aquilion 64. F: female. M: male. PAD: pulmonary artery diameter. RV: right ventricle diameter. LV: left ventricle diameter. RV/LV: right ventricle-to-left ventricle diameter ratio. IVC Reflux: inferior vena cava reflux of contrast media. IV Septum: interventricular septum position.

Case	Sex	Age	PAD (mm)	RV (mm)	LV (mm)	RV/LV	IVC Reflux	IV Septum	Largest affected vessel
1	M	59	27	52	57	0.91	Absent	Normal	Main Artery (unilateral)
2	F	72	32	60	40	1.50	Absent	Flattened	Main Artery (unilateral)
3	M	71	36	38	32	1.19	Present	Normal	Lobar
4	M	62	28	33	34	0.97	Absent	Flattened	Trunk bifurcation
5	M	73	30	32	42	0.76	Absent	Normal	Segmental
6	M	82	40	47	30	1.57	Present	Flattened	Trunk bifurcation
7	F	21	32	39	47	0.83	Absent	Normal	Segmental
8	F	82	29	32	42	0.76	Absent	Normal	Segmental
9	M	50	28	22	40	0.55	Present	Normal	Lobar
10	F	57	26	55	38	1.45	Absent	Paradoxically bowed	Subsegmental
11	M	31	27	25	51	0.49	Absent	Normal	Segmental
12	F	36	22	30	35	0.86	Absent	Normal	Lobar
13	F	86	26	35	36	0.97	Present	Normal	Subsegmental
14	F	57	38	50	48	1.04	Present	Flattened	Main Artery (unilateral)
15	F	32	27	44	42	1.05	Absent	Flattened	Segmental
16	M	38	30	46	50	0.92	Absent	Normal	Lobar
17	F	81	31	50	50	1.00	Present	Normal	Lobar
18	M	54	28	44	50	0.88	Absent	Normal	Lobar
19	M	41	29	40	49	0.82	Absent	Normal	Trunk bifurcation
20	M	54	29	38	64	0.59	Present	Normal	Segmental

Table 2. Data from exams obtained with the GE Revolution 512. F: female. M: male. PAD: pulmonary artery diameter. RV: right ventricle diameter. LV: left ventricle diameter. RV/LV: right ventricle-to-left ventricle diameter ratio. IVC Reflux: inferior vena cava reflux of contrast media. IV Septum: interventricular septum position.

the first slice, and the last element represents the lowest slice. Slices in which the embolus can be visualized are represented by the number "1", and the others by the number "0".

The files corresponding to patients scanned with the GE equipment are named 01GE to 20GE, and the files corresponding to patients scanned with the Toshiba equipment are named 01TS to 20TS, according to the number of each patient in Tables 1, 2, which are also available on the repository in Excel binary file (XLS) format.

Technical Validation

Exam selection. The CTPA scans were selected retrospectively by a search of the digital files from a public university hospital and a private imaging practice.

An arbitrary starting date was defined for each medical facility. From the defined starting date, all CTPA scans were sequentially reviewed until we reached 20 examinations with PE from each device that fit the inclusion and exclusion criteria.

For the Toshiba device, the examinations were performed from November 5, 2018 to February 5, 2019. For the GE device, the examinations were performed from November 9, 2018 to September 20, 2019.

Inclusion criteria. Chest CT scans performed using the PE protocol (CTPA) for diagnosing acute PE. File availability in the Picture Archive and Communication System (PACS) from each medical center.

Exclusion criteria. Artifacts that prevented the radiologist from visually interpreting the images. Examinations with fully or partially corrupted files. Follow-up examinations (only the first CTPA was used).

There was no restriction on age, patient status (inpatients or outpatients) or any other inclusion or exclusion criteria different from those mentioned.

Code availability

All code for loading and normalization of the dataset is available in GitHub (https://github.com/glescki/dicom_image_parser).

For parsing the data, the PyDicom library is recommended, and the loading of the labels can be performed with a parser available in GitHub. Each DICOM file should be loaded separately and then joined within a data structure.

For normalization, it is recommended that the spacing in the z-axis of all slices be changed to 1.

Received: 20 June 2022; Accepted: 11 July 2023;
Published online: 04 August 2023

References

1. Beckman, M. G., Hoopeer, W. C., Critchley, S. E. & Ortel, T. L. Venous thromboembolism: a public health concern. *Am J Prev Med.* **4**, 495–501, https://doi.org/10.1016/j.amepre.2009.12.017 (2010).
2. Konstantinides, S. V. *et al*. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): the task force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). *Eur. Heart Journal.* **41**, 543–603, https://doi.org/10.1093/eurheartj/ehz405 (2020).
3. Dalen, J. E. Pulmonary embolism: what have we learned since Virchow? *Chest.* **122**, 1440–1456, https://doi.org/10.1378/chest.122.4.1440 (2002).
4. Konstantinides, S. V. *et al*. Task force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur. Heart J.* **35**, 3033–3069, https://doi.org/10.1093/eurheartj/ehu283 (2014).
5. Remy-Jardin, M. *et al*. Management of suspected acute pulmonary embolism in the era of CT angiography: a statement from the Fleischner society. *Radiology.* **245**, 315–329, https://doi.org/10.1148/radiol.2452070397 (2007).
6. Wittram, C. *et al*. CT angiography of pulmonary embolism: diagnostic criteria and causes of misdiagnosis. *RadioGraphics.* **24**, 219–1238, https://doi.org/10.1148/rg.245045008 (2004).
7. Masutani, Y., Macmahon, H. & Doi, K. Computerized detection of pulmonary embolism in spiral CT angiography based on volumetric image analysis. *IEEE Trans. Med. Imaging.* **21**, 1517–1523, https://doi.org/10.1109/TMI.2002.806586 (2002).
8. Pichon, E., Novak, C. L., Kiraly, A. P. & Naidich, D. P. A novel method for pulmonary emboli visualization from high-resolution CT images. *Proceedings of the SPIE.* **5367**, 161–170, https://doi.org/10.1117/12.532892 (2004).
9. Zhou, C. *et al*. Preliminary investigation of computer-aided detection of pulmonary embolism in three-dimensional computed tomography pulmonary angiography images. *Acad. Radiol.* **12**, 782–792, https://doi.org/10.1016/j.acra.2005.01.014 (2005).
10. Maizlin, Z. V., Vos, P. M., Godoy, M. C. & Cooperberg, P. L. Computer-aided detection of pulmonary embolism on CT angiography: initial experience. *J. Thorac. Imaging.* **22**, 324–329, https://doi.org/10.1097/RTI.0b013e31815b89ca (2007).
11. Buhmann, S. *et al*. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. *Acad. Radiol.* **14**, 651–658, https://doi.org/10.1016/j.acra.2007.02.007 (2007).
12. Schoepf, U. J. *et al*. Pulmonary embolism: computer-aided detection at multidetector row spiral computed tomography. *J. Thorac. Imaging.* **22**, 319–323, https://doi.org/10.1097/RTI.0b013e31815842a9 (2007).
13. Liang, J. & Bi, J. Computer aided detection of pulmonary embolism with tobogganing and multiple instance classification in CT pulmonary angiography. *Inf. Process. Med. Imaging.* **20**, 630–641, https://doi.org/10.1007/978-3-540-73273-0_52 (2007).
14. Sebbe, R. *Computer-aided Diagnosis of Pulmonary Embolism in Opacified CT Images*, dissertation, Faculte Polytechnique de Mons. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.3563&rep=rep1&type=pdf (2006).
15. Engelke, C., Schmidt, S., Bakai, A., Auer, F. & Marten, K. Computer-assisted detection of pulmonary embolism: performance evaluation in consensus with experienced and inexperienced chest radiologists. *Eur. Radiol.* **18**, 298–307, https://doi.org/10.1007/s00330-007-0770-3 (2008).
16. Bouma, H. *Vessel-Diameter Quantification and Embolus Detection in CTA Images*, Phd Thesis, Technische Universiteit Eindhoven. https://research.tue.nl/nl/publications/vessel-diameter-quantification-and-embolus-detection-in-cta-image (2008).
17. Das, M. *et al*. Computer-aided detection of pulmonary embolism: influence on radiologists' detection performance with respect to vessel segments. *Eur. Radiol.* **18**, 1350–1355, https://doi.org/10.1007/s00330-008-0889-x (2008).
18. Walsham, A. C. *et al*. The use of computer-aided detection for the assessment of pulmonary arterial filling defects at computed tomographic angiography. *J. Comput. Assist. Tomogr.* **32**, 913–918, https://doi.org/10.1097/RCT.0b013e31815b3ed0 (2008).
19. Bouma, H., Sonnemans, J. J., Vilanova, A. & Gerritsen, F. A. Automatic detection of pulmonary embolism in CTA images. *IEEE Trans. Med.Imaging.* **28**, 1223–1230, https://doi.org/10.1109/TMI.2009.2013618 (2009).
20. Dinesh, M. S., *et al* Adaptive contrast-based computer aided detection for pulmonary embolism. *Proc. SPIE. 7260, Medical Imaging 2009: Computer-Aided Diagnosis*. https://doi.org/10.1117/12.812223 (2009)
21. Zhou, C. *et al*. Computer-aided detection of pulmonary embolism in computed tomographic pulmonary angiography (CTPA): performance evaluation with independent data sets. *Med. Phys.* **36**, 3385–3396, https://doi.org/10.1118/1.3157102 (2009).
22. Wittenberg, R., Peters, J. F., Sonnemans, J. J., Prokop, M. & Schaefer-Prokop, C. M. Computer-assisted detection of pulmonary embolism: evaluation of pulmonary CT angiograms performed in an on-call setting. *Eur. Radiol.* **20**, 801–806, https://doi.org/10.1007/s00330-009-1628-7 (2010).
23. Dewailly, M. *et al*. Computer-aided detection of acute pulmonary embolism with 64-slice multidetector row computed tomography: impact of the scanning conditions and overall image quality in the detection of peripheral clots. *J. Comput. Assist. Tomogr.* **34**, 23–30, https://doi.org/10.1097/RCT.0b013e3181b2e383 (2010).
24. Lee, C. W. *et al*. Evaluation of computer aided detection and dual energy software in detection of peripheral pulmonary embolism on dual-energy pulmonary CT angiography. *Eur. Radiol.* **21**, 54–62, https://doi.org/10.1007/s00330-010-1903-7 (2011).
25. Park, S. C., Chapman, B. E. & Bin, Z. B. A multistage approach to improve performance of computer-aided detection of pulmonary embolisms depicted on CT images: preliminary investigation. *IEEE Trans Biomed Eng.* **58**, 1519–1527, https://doi.org/10.1109/TBME.2010.2063702 (2011).
26. Blackmon, K. N. *et al*. Computer-aided detection of pulmonary embolism at CT pulmonary angiography: can it improve performance of inexperienced readers? *Eur Radiol.* **21**, 1214–1223, https://doi.org/10.1007/s00330-010-2050-x (2011).
27. Wittenberg, R. *et al*. Impact of image quality on the performance of computer-aided detection of pulmonary embolism. *Am. J. Roentgenol.* **196**, 95–101, https://doi.org/10.2214/AJR.09.4165 (2011).
28. Wittenberg, R. *et al*. Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis—an observer study. *Radiology.* **262**, 305–313, https://doi.org/10.1148/radiol.11110372 (2012).
29. Kligerman, S. J., Lahiji, K., Galvin, J. R., Stokum, C. & White, C. S. Missed pulmonary emboli on CT angiography: assessment with pulmonary embolism–computer-aided detection. *Am. J. Roentgenol.* **202**, 65–73, https://doi.org/10.2214/AJR.13.11049 (2014).
30. Özkan, H., Osman, O., Şahin, S. & Boz, A. F. A novel method for pulmonary embolism detection in CTA images. *Comput. Methods Programs Biomed.* **113**, 757–766, https://doi.org/10.1016/j.cmpb.2013.12.014 (2014).
31. Lahiji, K., Kligerman, S., Jeudy, J. & White, C. Improved accuracy of pulmonary embolism computer-aided detection using iterative reconstruction compared with filtered back projection. *Am. J. Roentgenol.* **203**, 763–771, https://doi.org/10.2214/AJR.13.11838 (2014).
32. Tajbakhsh, N., Gotway, M. B. & Liang, J. Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science.* **9350**, 62–69, https://doi.org/10.1007/978-3-319-24571-3_8 (2015).
33. Ozkan, H., Tulum, G., Osman, O. & Sahin, S. Automatic detection of pulmonary embolism in cta images using machine learning. *Elektron. Elektrotech.* **23**, 63–67, https://doi.org/10.5755/j01.eie.23.1.17585 (2017).
34. Tajbakhsh, N., Shin, J. Y., Gotway, M. B. & Liang, J. Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation. *Med. Image Anal.* **58**, 101541, https://doi.org/10.1016/j.media.2019.101541 (2019).
35. Rajan, D., Beymer, D., Abedin, S. & Dehghan, E. Pi-PE: A pipeline for pulmonary embolism detection using sparsely annotated 3D CT images. *Proceedings of the Machine Learning for Health NeurIPS Workshop*, **116**, 220–232, http://proceedings.mlr.press/v116/rajan20a/rajan20a.pdf (2020).

36. Yang, X. *et al*. A two-stage convolutional neural network for pulmonary embolism detection from CTPA images. *IEEE Access*. **7**, 84849–84857, https://doi.org/10.1109/ACCESS.2019.2925210 (2019).
37. Cano-espinosa, C., Cazorla, M. & González, G. Computer aided detection of pulmonary embolism using multi-slice multi-axial segmentation. *Appl. Sci.* **10**, 2945, https://doi.org/10.3390/app10082945 (2020).
38. Huang, S., Pareek, A., Zamanian, R., Banerjee, I. & Lungren, M. P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci. Rep.* **10**, 22147, https://doi.org/10.1038/s41598-020-78888-w (2020).
39. Yu, C. -Y., Cheng, Y. -C. & Kuo, C. Early pulmonary embolism detection from computed tomography pulmonary angiography using convolutional neural networks. 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306659 (2020).
40. Huang, S. *et al*. PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. *npj Digit. Med.* **3**, 61, https://doi.org/10.1038/s41746-020-0266-y (2020).
41. Liu, W. *et al*. Evaluation of acute pulmonary embolism and clot burden on CTPA with deep learning. *Eur. Radiol.* **30**, 3567–3575, https://doi.org/10.1007/s00330-020-06699-8 (2020).
42. Weikert, T. *et al*. Automated detection of pulmonary embolism in CT pulmonary angiograms using an AI-powered algorithm. *Eur. Radiol.* **30**, 6545–6553, https://doi.org/10.1007/s00330-020-06998-0 (2020).
43. Vainio, T., Mäkelä, T., Savolainen, S. & kangasniemi, M. Performance of a 3D convolutional neural network in the detection of hypoperfusion at CT pulmonary angiography in patients with chronic pulmonary embolism: a feasibility study. *Eur. Radiol. Exp.* **5**, 45, https://doi.org/10.1186/s41747-021-00235-z (2021).
44. Li, X., Wang, X., Yang, X., Lin, Y. & Huang, Z. Preliminary study on artificial intelligence diagnosis of pulmonary embolism based on computer in-depth study. *Ann. of transl. med.* **9**, 838, https://doi.org/10.21037/atm-21-975 (2021).
45. Long, K. *et al*. Probability-based mask R-CNN for pulmonary embolism detection. *Neurocomputing*. **422**, 345–353, https://doi.org/10.1016/j.neucom.2020.10.022 (2021).
46. Huhtanen, H. *et al*. Automated detection of pulmonary embolism from CT-angiograms using deep learning. *BMC Med. Imaging*. **22**, 43, https://doi.org/10.1186/s12880-022-00763-z (2022).
47. Yadlapalli, P. *et al*. *Segmentation of pulmonary embolism using deep learning. 2022 International Conference for Advancement in Technology (ICONAT)* https://doi.org/10.1109/ICONAT53423.2022.9726048 (2022).
48. Olescki, G., Andrade, J. M. C., Escuissato, D. L. & Oliveira, L. F. A two-step workflow for pulmonary embolism detection using deep learning and feature extraction. *Comput. Methods in Biomech. Biomed. Eng.: Imaging Vis.* https://doi.org/10.1080/21681163.2022.2060866 (2022).
49. Paiva, O. A. & Prevedello, L. M. The potential impact of artificial intelligence in radiology. *Radiol. Bras.* **50**, 5–6, https://doi.org/10.1590/0100-3984.2017.50.5e1 (2017).
50. Colak, E. *et al*. The RSNA pulmonary embolism CT dataset. *Radiology: Art. Int.* **3**, 2, https://doi.org/10.1148/ryai.2021200254 (2021).
51. González, G., et al. Computer aided detection for pulmonary embolism challenge (CAD-PE). Preprint at https://arxiv.org/abs/2003.13440 (2019).
52. Masoudi, M. *et al*. A new dataset of computed-tomography angiography images for computer-aided detection of pulmonary embolism. *Sci. Data*. **5**, 180180, https://doi.org/10.1038/sdata.2018.180 (2018).
53. Bluemke, D. A. *et al*. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the *Radiology* editorial board. *Radiology*. **294**, 487–489, https://doi.org/10.1148/radiol.2019192515 (2020).
54. Andrade, JMC. *et al*. Annotated dataset of acute pulmonary embolism computed-tomography images - READ PE CT, *figshare*, https://doi.org/10.6084/m9.figshare.c.6033134.v1 (2023).
55. Yushkevich, P. A. *et al*. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. **3**, 1116–1128, https://doi.org/10.1016/j.neuroimage.2006.01.015 (2006).

Acknowledgements

We are grateful to the Hospital de Clínicas of the Federal University of Paraná and to the *Clínica Diagnóstico Avançado por Imagem (DAPI)* for allowing the conduction of this project. We are also grateful to the postgraduate programs in internal medicine and health sciences and informatics from the Federal University of Paraná that made this project possible.

Author contributions

João Mario Clementin de Andrade. Project planning. Patient selection following the preestablished inclusion and exclusion criteria. Assistance in the definition of the segmentation method. Pixel-level segmentation. Review of all ground-truth masks. Writing of the manuscript. Gabriel Olescki. Project planning. Definition of the segmentation method. Image-level segmentation. Writing of the manuscript. Dante Luiz Escuissato. Project planning. Assistance in the selection of patients following the preestablished inclusion and exclusion criteria. Review/confirmation of the acute pulmonary embolus location of all patients. Orientation/revision of the manuscript. Lucas Ferrari Oliveira. Project planning. Definition of the segmentation method. Review of this manuscript. Ana Carolina Nicolleti Basso. Pixel level segmentation. Review of the manuscript. Gabriel Lucca Salvador. Pixel level segmentation. Review of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.M.C.d.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.