A deep learning approach for automated diagnosis of pulmonary embolism on computed tomographic pulmonary angiography

Pranav Ajmera
Dr D.Y. Patil Medical College
Amit Kharat
Dr D.Y. Patil Medical College
Jitesh Seth
DeepTek Medical Imaging Pvt. Ltd
Snehal Rathi
Mahatma Gandhi Mission Medical College and Hospital
Richa Pant (richa.pant@deeptek.ai)
DeepTek Medical Imaging Pvt. Ltd
Manish Gawali
DeepTek Medical Imaging Pvt. Ltd
Viraj Kulkarni
DeepTek Medical Imaging Pvt. Ltd
Ragamayi Maramraju
Dr D.Y. Patil Medical College
Isha Kedia
Dr D.Y. Patil Medical College
Rajesh Botchu
Royal Orthopedic Hospital
Sanjay Khaladkar
Dr D.Y. Patil Medical College

Research Article

Keywords: Artificial Intelligence, Pulmonary Embolism, Computed tomographic pulmonary angiography, U-Net architecture

Posted Date: August 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1909034/v1

Abstract

Background

Computed tomographic pulmonary angiography (CTPA) is the diagnostic standard for confirming Pulmonary Embolism (PE). Since PE is a life-threatening condition, early diagnosis and treatment are critical to avoid PE-associated morbidity and mortality. However, the diagnosis of PE remains subject to misdiagnosis.

Methods

We retrospectively identified 251 CTPAs performed at a tertiary care hospital between January 2018 to January 2021. The scans were classified as positive (n = 55) and negative (n = 196) for PE based on the annotations made by board-certified radiologists. A fully anonymized CT slice served as input for detection of PE by the 2D segmentation model comprising U-Net architecture with Xception encoder. The diagnostic performance of the model was calculated at both the scan and the slice levels.

Results

The model correctly identified 44 out of 55 scans as positive for PE and 146 out of 196 scans as negative for PE with a sensitivity of 0.80 [95% CI: 0.68, 0.89], a specificity of 0.74 [95% CI: 0.68, 0.80], and an accuracy of 0.76 [95% CI: 0.70, 0.81]. On slice level, 4817 out of 5183 slices were marked as positive for the presence of emboli with a specificity of 0.89 [95% CI: 0.88, 0.89], a sensitivity of 0.93 [95% CI: 0.92, 0.94], and an accuracy of 0.89 [95% CI: 0.887, 0.890]. The model also achieved an AUROC of 0.85 [0.78, 0.90] and 0.94 [0.936, 0.941] at scan level and slice level, respectively for the detection of PE.

Conclusion

The development of an AI model and its use for the identification of pulmonary embolism will support healthcare workers by reducing the rate of missed findings and minimizing the time required to screen the scans.

Introduction

Pulmonary embolism (PE) is an emergency condition associated with a high mortality and morbidity rate. The annual incidence of embolism is approximately 60–70 cases per 1,00,000 people, resulting in up to 1,00,000–3,00,000 deaths annually [1]. PE is the third most common cause of cardiovascular disease events, with only myocardial infarction and stroke having a higher prevalence [2]. The most common cause of death from PE is a failure to diagnose [3]. The diagnostic approach to PE usually involves a series of investigations, including echocardiography and D-dimer, establishing the necessity for further

confirmatory examinations like Computed tomographic pulmonary angiography (CTPA) and ventilation/perfusion scans [4]. Because of its convenience of operation, CTPA enables a definitive diagnosis of PE and has practically eliminated the use of ventilation/perfusion scans [5]. According to multiple published studies, CTPA has the sensitivity and specificity of approximately 90% for detecting thrombus in the main pulmonary and segmental arteries [6][7]. CTPA has no contraindications and a low (0.3–1.8%) associated morbidity [6]. Because of COVID-19 pneumonia, the frequency of cases diagnosed with PE on CTPA has increased. Planquette et al. reported a 5% prevalence of PE in the total COVID-19 affected population and a 20% prevalence in the clinically suspected population with acute PE [8].

A test lacking a high degree of sensitivity results in a delay in the timely initiation of anticoagulant therapy for embolus-positive patients, consequently leading to a higher mortality rate. Early anticoagulation therapy has been documented to reduce the mortality rate from 30–8% [9]. However, the increase in the utilization of medical imaging has made a timely and reliable diagnosis of CTPE extremely difficult for the healthcare system and radiology practitioners. The usage of CTPA in emergency departments has increased over 27-fold in the last two decades [10, 11]. This challenge can be addressed by employing automated detection algorithms for diagnosing PE in clinical settings. A well-trained neural network can help healthcare practitioners by highlighting the exams that are positive for PE in the worklist, thereby accelerating the diagnosis and communication workflow [12]. Previous computer-aided detection (CAD) algorithms for the automatic identification of PE on CTPAs had low sensitivities [13–16]. Recent studies have focused on the use of CTPA imaging for PE diagnosis using deep learning and convolutional neural networks.

The objective of our study is to externally validate the performance of a 2D segmentation model with U-Net architecture for automated detection of PE in CTPAs using clinically relevant data from the CTPAs performed at a tertiary care hospital. In this study, we validate an externally developed deep learning model “DxPE AI Screen” for early detection of emboli on CTPA to provide radiologists and clinicians with a tested high-performance tool to help triage patients with PE. This tool has the potential to enable rapid and reliable radiology reporting while minimizing the substantial disparities between general radiologists and subspecialty-trained radiologists in interpreting scans. The U-Net architecture is generally used to address segmentation problems, but we effectively utilized it to solve a classification task by deriving probabilistic output from segmentation output. Our approach to converting slice level segmentation predictions to scan level classification predictions resulted in a high area under the curve and a short processing time.

Materials And Methods

Patient Selection (external validation dataset)

The study was reviewed and approved by the Institutional Review Board of Dr. D.Y. Patil Medical College, Hospital, and Research Centre, Pune, India. The study was performed in a Health Information Portability and Accountability Act (HIPAA) compliant manner. Due to the retrospective nature of data collection, the

need for written informed consent from the individual patients was waived. 251 CTPAs performed at the institution between January 2018- January 2021 were retrospectively collected. Extraction of datasets was performed utilizing the hospital Picture and Archiving Communication System (PACS). The diagnosis of PE included both clinical information (D-dimer test and echocardiography) and CTPA findings. CTPAs with poor-quality images or motion artifacts were excluded from the study. CTPA studies were performed on a 128-slice multidetector Philips Ingenuity CT scanner using the standard protocol. It was performed in the craniocaudal direction with the slice thickness of 0.9 mm for 248 scans, 1 mm for 2 scans, and 2 mm for 1 scan.

Establishing ground truth:

A total of 942 CT scans from two open-source datasets, Kaggle FUMPE [17] and RSNA STR Pulmonary Embolism Detection [18], were used for training and internal validation of the model. The external test dataset, consisting of 251 scans, was obtained from Dr. D.Y. Patil Medical College, Hospital, and Research Centre, Pune, India. To establish the ground truth, three board-certified expert radiologists with 23, 15, and 9 years of experience categorized scans in multiple batches of 25 and highlighted emboli in each slice, if present, using the ITK-Snap software (version 3.8.0). This data was used to calculate the performance of the model at both slice and scan levels.

Case definition:

The radiologists used several indications to identify the presence of PE. For annotating and categorizing the scan as positive or negative, we used the established definitions of PE. Acute PE is defined as (a) the blockage of an artery with the absence of enhancement of the arterial lumen due to a filling defect, which completely obstructs the lumen. Consequently, the artery may enlarge in comparison to neighboring patent vessels. (b) Partial blockage of the artery by a filling defect that partially obstructs the lumen, often producing the “polo mint” sign on sections acquired perpendicular to the long axis of the vessel, and the “railway-track” sign on longitudinal visualization of the vessel. (c) Peripherally located intraluminal defect which forms an acute angle with the arterial wall [19].

Model Details:

A 2D segmentation model consisting of U-Net architecture with a Xception encoder was used to detect pulmonary embolism from CTPAs. The input comprised a slice of a CT scan that was preprocessed as mentioned below. The DICOM images were converted to Hounsfield Units, and a window of (40, 400) (Window Level, Window Width) was applied. Subsequently, the images and masks were resized to 512x512 voxels for uniformity. The model was trained using Adam Weight Decay as an optimizer and a combination of binary cross-entropy loss (at voxel level) and Dice Loss (at slice level) as loss functions. The output comprised a 512x512 image containing the predicted mask. Each voxel of the output image had a value between 0 and 1, indicating confidence in the presence of embolism in that region. Because the model only predicts the masks at the slice level, a method was developed to determine the presence of embolism at the scan level. All voxels having a value greater than 0.5 were assigned the value 1 (corresponding to a ‘positive voxel’), while those with a value less than or equal to 0.5 were assigned the

value 0 (corresponding to a 'negative voxel'). The average number of positive voxels per slice was computed by dividing the total number of positive voxels in a scan by the number of slices in the CT scan. If a CT scan with n slices contains x_1, x_2, x_3… x_n positive voxels, then the average number of positive voxels per slice (X) is calculated as:

$$X = \frac{\sum_{i=1}^{n} x_i}{n}$$

The model training was monitored by its performance on a separate validation data set. The model with the lowest validation loss was chosen as the final model. The PE mask prediction process is represented in Fig. 1. The distribution of the number of CT scans and slices in the training, validation, and external test set is given in Table 1.

Table 1
Distribution of the number of CT scans and slices in training, validation, and test sets.

Set	Training set		Validation set		Test set	
	Number of Slices	Number of Scans	Number of Slices	Number of Scans	Number of Slices	Number of Scans
Negative	167359	557 (65.4%)	13217	56 (63%)	140984	196 (78%)
Positive	14427	296 (34.6%)	1106	33 (37%)	5183	55 (22%)
Total	181786	853	14323	89	146167	251

Statistical analysis:

The comprehensive evaluation of the model performance on the test set included sensitivity, specificity, PPV, NPV, accuracy, F1 score, and ROC. To measure the variability in these values, we used 95% confidence intervals using the empirical bootstrapping method. To better understand the performance of the model in diagnosing PE, we also calculated confusion metrics on the entire test set.

Results

Patient Population:

A total of 251 CTPAs corresponding to 251 patients were used to evaluate the performance of the model on the test set. Among 251 patients, 145 were males and 106 were females, with an average age (standard deviation) of 49.7 (17.4) years. The age distribution of participants included in the study is represented in Fig. 2. The rate of exams positive for PE was 21.74% (55/251). A total of 3.55% slices (5183/146167) were positive for PE. There was no missing data.

Performance of U-Net model:

Table 2 illustrates the segmentation results of the U-Net model on the clinical dataset. The model had a sensitivity of 0.8 [0.68, 0.89] and 0.93 [0.92, 0.94] at scan level and slice level, respectively. Fig 3 represents the sensitivity and specificity values of the model when the average positive voxels per slice threshold were varied at the scan level. In this study, we set our operating point at a threshold that maximizes both sensitivity and specificity on the external test dataset. Although the threshold >15 resulted in high specificity values, the corresponding sensitivity values declined. However, in clinical settings, applications are usually tuned to maximize sensitivity in order to minimize false negative rates. Therefore, to calculate the threshold for our model, we calculated the geometric mean of sensitivity and specificity at different thresholds, varying the threshold on steps of 15 from 0 to 300. A threshold of 15 voxels per slice was chosen as the final scan level threshold because it maximized both sensitivity and specificity. This threshold allowed the model to achieve a sensitivity of 0.80 [0.68, 0.89] and specificity of 0.74 [0.68, 0.80].

Table 2

Performance metrics of 2D segmentation U-Net model.

Metrics	Performance at slice level	Performance at scan level
Sensitivity	0.93 [0.92, 0.94]	0.80 [0.68, 0.89]
Specificity	0.89 [0.88, 0.89]	0.74 [0.68, 0.80]
PPV	0.23 [0.23, 0.24]	0.47 [0.36, 0.57]
Accuracy	0.89 [0.88, 0.89]	0.76 [0.70, 0.81]
F-1 Score	0.37 [0.36, 0.38]	0.59 [0.49, 0.68]
ROC	0.94 [0.936, 0.941]	0.85 [0.78, 0.90]
NPV	0.997 [0.996, 0.997]	0.93 [0.89, 0.97]

The model achieved an AUROC of 0.85 [0.78, 0.90] and 0.94 [0.936, 0.941] at scan level and slice level, respectively (Fig. 4).

The low number of true positives in the dataset and our choice of operating threshold for high sensitivity resulted in lower PPV (47%) and higher NPV (93%). However, in the clinical setting, improving the sensitivity of the positive cases is more important than PPV (more false positives). The distribution of TPs, FPs, TNs, and FNs for both scan and slice levels are represented in the confusion metrics (Fig. 5).

Model output:

The output of the model contained voxel-wise probability for the input images. The average number of positive voxels per slice was computed by adding all the positive voxels and dividing them by the number of axial slices of the CT scan. An illustration of the same is represented in Fig. 6. The localization of emboli is represented as masks in the CT slice. The model was able to mark the presence of pulmonary embolism in most positive cases. The average segmentation DICE coefficient score per scan for the test set was 0.743 ± 0.155, indicating the high degree of similarity between the reference and the AI-predicted embolus masks.

Reading time per scan:

The standalone model required a mean time of 30.15 seconds (std. dev: 11.03 seconds) to process the thin angiography sequences, make inferences, and categorize the scan as positive or negative for emboli.

Discussion

Suspected pulmonary embolism is a prevalent, and potentially life-threatening condition in emergency patients [20]. Therefore, an accurate and timely diagnosis of PE is critical for improving prognosis. Previous CAD solutions for the automatic detection of PE on CTPAs were limited by the low sensitivity of the model [13–16, 21]. Additionally, many of the evaluations were performed on a small dataset with less than 50 cases in the testing set [13, 21–23]. We developed a novel U-Net-based classification approach for PE detection and tested it on the clinical dataset containing 251 CTPAs. The model achieved high sensitivity, accuracy, and ROC in detecting pulmonary embolism at both slice level and scan level. We also observed that the slice level predictions were slightly better than scan level predictions for AUC, sensitivity, and specificity values. Our model had an AUC of 0.85 at scan level, which is comparable to the previously reported deep learning models. Rajan et al. developed and tested their Convolutional neural network (CNN) model on open-source datasets, reporting an AUC of 0.85 for severe PE and 0.70 for all other PE cases. However, their model was not tested on an independent external dataset [24]. Shi et al. tested a ResNet-based model on a multi-institutional dataset and obtained an AUC of 0.812 [25]. However, the source of their data was not well elucidated. Tajbakhsh et al. evaluated the performance of a CNN model on an internal dataset containing 121 CTPAs with 326 emboli and achieved a sensitivity of 83%. However, their model had a sensitivity of 34.6% on a relatively small test dataset of 20 CTPAs with 133 emboli from the PE challenge [26]. Our model had a slice level sensitivity of 93% and a scan level sensitivity of 80% when tested on an external dataset with 251 CTPAs (5183 emboli). Our model matched the performance of the PENet model developed by Huang et al. that achieved an AUC of 0.85 on the external dataset. While they had the advantage of testing their model on two independent external datasets, the maximum sensitivity of their model was 75% [27] at a given threshold. Similarly, Yang et al. reported a sensitivity of 75% for their CNN model, but their test set had only 129 scans (269 emboli) [28].

The use of automated algorithms with a human-in-the-loop approach to triage pulmonary embolism within seconds of a CTPA scan has the potential to be used in clinical practices. It can shorten the time between diagnosis and therapy by highlighting all suspected positive scans for immediate reporting by a radiologist. In places where there is insufficient night-time coverage by an experienced radiologist, the

algorithm can provide a preliminary diagnosis to the emergency physician, assuring a reliable diagnostic method. Studies have indicated that even in locations with a 24x7 radiologist, the quality of reporting is significantly affected, with a discrepancy of nearly 13% between daytime only and night-time faculties in detecting embolism in CTPA studies [29] [30] [31]. In addition to this, the number of CTPAs recommended by the emergency department in the previous two decades has shown a multi-fold increase [11]. The use of AI models for the detection of PE can serve as a reliable second opinion for resident radiologists. Additionally, the algorithm can run on a centralized picture archiving and communication system (PACS) or on any edge device in the absence of PACS. The use of edge devices on X-ray, CT, and MRI for diagnostic purposes has already proven to be a huge success, allowing the algorithm to be integrated into the basic system and multiplying its reach [32].

We reported an average DICE coefficient score per scan of 0.743 ± 0.155, which was better than the score reported by Cano-Espinosa et al. on 2D, 2.5D, and 3D networks [33]. Our DL model provides an explainable solution to detect PE, which has the potential to minimize the time required to interpret positive scans and ensure that the peripherally located emboli are not overlooked in cases where a more centrally located embolus is also present.

Our approach to testing the 2D segmentation model on the external dataset has certain limitations. Firstly, validation of our model was performed on a single external test dataset. Therefore, future studies will include datasets from multiple clinical institutions to validate the model performance in a robust manner. Secondly, our model had a low PPV on the external test dataset at both slice and scan levels. This could have resulted because of the low prevalence of the positive CTPAs i.e., 55 scans (though our scans were feature-rich in terms of having a high number of emboli, i.e., 5183) in our dataset and due to our choice of particularly high sensitivity operating point, which resulted in a lower PPV and higher NPV of 0.47 and 0.93, respectively at scan level. While this approach could falsely label negative scans as positive and increase the urgency of diagnosis, it is important to interpret this in terms of the role this DL algorithm will play in the clinical setting. In a clinical setting, it is acceptable to spend time reassigning a false positive scan into the negative category rather than the disease cost of missing a true positive finding. The model is primarily built for detecting PE and all the scans should be interpreted by a clinician and/ or a radiologist to provide the final verdict. This human-in-the-loop approach is favored as it reduces the likelihood of missing or delay in reporting a positive scan.

In conclusion, our DL algorithm is on par with the current state-of-the-art DL algorithm and exceeds the performance of some previous DL algorithms. Our approach allows the algorithm to categorize the scan as positive or negative for an embolus in minimal time (mean time of 30.15 s). The mask highlights the region affected by the embolus, thus providing the clinician with a rapid, reliable, and explainable solution.

Abbreviations

AI

Artificial Intelligence

AUROC

Area Under the Receiver Operating Characteristics

CTPA

Computed Tomographic Pulmonary Angiography

HIPAA

Health Information Portability and Accountability Act

PACS

Picture Archival and Communication System

DICOM

Digital Imaging and Communications in Medicine

CT

Computed Tomography

DL

Deep Learning

NPV

Negative Predictive Value

PPV

Positive Predictive Value

TP

True Positive

FP

False Positive

TN

True Negative

FN

False Negative

CAD

Computer-Aided Detection

Declarations

Ethics approval and consent to participate: The study was approved by the institutional review board of D.Y. Patil hospital and research center as a part of its ethical committee meeting (DYPV/EC/646/2021). The Institutional Review Board (IRB) waived the need for separate informed consent from individual patients in view of the retrospective and observational nature of the study. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication: NA

Availability of data and materials: All data generated or analyzed during this study are included in this published article

Competing interests: Amit Kharat is a Professor at Dr. DY Patil Medical College, Pune, and is also a co-founder of DeepTek Medical Imaging Pvt Ltd whose expertise was employed in externally validating the model on the dataset from the hospital. Rest of the authors do not have any conflict of interest.

Funding: This study did not receive any funding

Author's contributions:

PA was involved in conceptualization, methodology, validation, formal analysis, investigation, data curation, writing and visualization.

AK was involved in conceptualization, methodology, validation, project administration, resources, data curation, and reviewing and validating the manuscript from the perspective of clinical and radiology workflows.

JS was involved in developing the deep learning model, technical writing, analyzing the results, and preparation of figures.

SR was involved in writing and figures.

RP was involved in writing and reviewing the manuscript, formal analysis, result analysis, and interpretation.

MG was involved in reviewing the manuscript.

VK was involved in validating the manuscript from the perspective of technical workflow.

RM was involved in data curation.

IK was involved in data curation.

RB was involved in validation and visualization.

SK was involved in project supervision.

All the authors made significant contributions to this project and the study was approved by all of them.

Acknowledgments: None

References

1. Agnelli G, Anderson F, Arcelus J, Bergqvist D, Brecht J, Greer I, et al. Venous thromboembolism (VTE) in Europe. Thromb Haemost. 2007;98:756–64.
2. Clemens S, Leeper KV. Newer Modalities for Detection of Pulmonary Emboli. The American Journal of Medicine. 2007;120:S2–12.
3. Dalen JE. Pulmonary embolism: what have we learned since Virchow? Natural history, pathophysiology, and diagnosis. Chest. 2002;122:1440–56.
4. Righini M, Robert-Ebadi H, Le Gal G. Diagnosis of acute pulmonary embolism. J Thromb Haemost. 2017;15:1251–61.
5. Shujaat A, Shapiro JM, Eden E. Utilization of CT Pulmonary Angiography in Suspected Pulmonary Embolism in a Major Urban Emergency Department. Pulmonary Medicine. 2013;2013:1–6.
6. Rossum ABD, Donkers-van Rossum AB. Diagnostic strategies for suspected pulmonary embolism. European Respiratory Journal. 2001;18:589–97.
7. The role of spiral volumetric computed tomography in the diagnosis of pulmonary embolism. https://paperpile.com/app/p/bce7c0b1-6e11-05a5-8b47-6d4c891248a7. Accessed 18 Dec 2021.
8. Planquette B, Le Berre A, Khider L, Yannoutsos A, Gendron N, de Torcy M, et al. Prevalence and characteristics of pulmonary embolism in 1042 COVID-19 patients with respiratory symptoms: A nested case-control study. Thromb Res. 2021;197:94–9.
9. Donato AA, Scheirer JJ, Atwell MS, Gramp J, Duszak R Jr. Clinical outcomes in patients with suspected acute pulmonary embolism and negative helical computed tomographic results in whom anticoagulation was withheld. Arch Intern Med. 2003;163:2033–8.
10. Prologo JD, Gilkeson RC, Diaz M, Asaad J. CT pulmonary angiography: a comparative analysis of the utilization patterns in emergency department and hospitalized patients between 1998 and 2003. AJR Am J Roentgenol. 2004;183:1093–6.
11. Chandra S, Sarkar PK, Chandra D, Ginsberg NE, Cohen RI. Finding an alternative diagnosis does not justify increased use of CT-pulmonary angiography. BMC Pulm Med. 2013;13:9.
12. Weikert T, Winkel DJ, Bremerich J, Stieltjes B, Parmar V, Sauter AW, et al. Automated detection of pulmonary embolism in CT pulmonary angiograms using an AI-powered algorithm. Eur Radiol. 2020;30:6545–53.
13. Lee CW, Seo JB, Song J-W, Kim M-Y, Lee HY, Park YS, 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. 2011;21:54–62.
14. Website. Missed Pulmonary Emboli on CT Angiography: Assessment With Pulmonary Embolism–Computer-Aided Detection Read More: https://www.ajronline.org/doi/full/10.2214/AJR.13.11049.
15. Bouma H, Sonnemans JJ, Vilanova A, Gerritsen FA. Automatic detection of pulmonary embolism in CTA images. IEEE Trans Med Imaging. 2009;28:1223–30.
16. Maizlin ZV, Vos PM, Godoy MB, Cooperberg PL. Computer-aided Detection of Pulmonary Embolism on CT Angiography. Journal of Thoracic Imaging. 2007;22:324–9.

17. Larxel. Pulmonary Embolism in CT images.
18. RSNA STR Pulmonary Embolism Detection. https://kaggle.com/c/rsna-str-pulmonary-embolism-detection. Accessed 24 Feb 2022.
19. Wittram C, Maher MM, Yoo AJ, Kalra MK, Shepard J-AO, McLoud TC. CT angiography of pulmonary embolism: diagnostic criteria and causes of misdiagnosis. Radiographics. 2004;24:1219–38.
20. Buhmann S, Herzog P, Liang J, Wolf M, Salganicoff M, Kirchhoff C, et al. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. Acad Radiol. 2007;14:651–8.
21. A Multistage Approach to Improve Performance of Computer-Aided Detection of Pulmonary Embolisms Depicted on CT Images: Preliminary Investigation. https://ieeexplore.ieee.org/document/5540279. Accessed 18 Feb 2022.
22. Schoepf UJ, Schneider AC, Das M, Wood SA, Cheema JI, Costello P. Pulmonary embolism: computer-aided detection at multidetector row spiral computed tomography. J Thorac Imaging. 2007;22:319–23.
23. Özkan H, Osman O, Şahin S, Boz AF. A novel method for pulmonary embolism detection in CTA images. Comput Methods Programs Biomed. 2014;113:757–66.
24. Rajan D, Beymer D, Abedin S, Dehghan E. Pi-PE: A Pipeline for Pulmonary Embolism Detection using Sparsely Annotated 3D CT Images. 2019.
25. Shi L, Rajan D, Abedin S, Yellapragada MS, Beymer D, Dehghan E. Automatic Diagnosis of Pulmonary Embolism Using an Attention-guided Framework: A Large-scale Study. 2020.
26. Tajbakhsh N, Gotway MB, Liang J. Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag; 2015. p. 62–9.
27. Huang S-C, Kothari T, Banerjee I, Chute C, Ball RL, Borus N, et al. PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. npj Digital Medicine. 2020;3:1–9.
28. Yang X, Lin Y, Su J, Wang X, Li X, Lin J, et al. A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images. 2019.
29. Joshi R, Wu K, Kaicker J, Choudur H. Reliability of on-call radiology residents' interpretation of 64-slice CT pulmonary angiography for the detection of pulmonary embolism. Acta radiol. 2014;55:682–90.
30. Rufener SL, Patel S, Kazerooni EA, Schipper M, Kelly AM. Comparison of on-call radiology resident and faculty interpretation of 4- and 16-row multidetector CT pulmonary angiography with indirect CT venography. Acad Radiol. 2008;15:71–6.
31. Yavas US, Calisir C, Ozkan IR. The interobserver agreement between residents and experienced radiologists for detecting pulmonary embolism and DVT with using CT pulmonary angiography and indirect CT venography. Korean J Radiol. 2008;9:498–502.

32. Kharat A, Duddalwar V, Saoji K, Gaikwad A, Kulkarni V, Naik G, et al. Role of Edge Device and Cloud Machine Learning in Point-of-Care Solutions Using Imaging Diagnostics for Population Screening. 2020.
33. Cano-Espinosa C, Cazorla M, González G. Computer Aided Detection of Pulmonary Embolism Using Multi-Slice Multi-Axial Segmentation. Applied Sciences. 2020;10:2945.

Figures

Figure 1

A schematic representation of the PE mask prediction process. Each 2D slice of the CT scan is processed and passed through the model that produces the corresponding mask. The mask highlights emboli. The

results of all the slices are combined to form the scan level prediction.

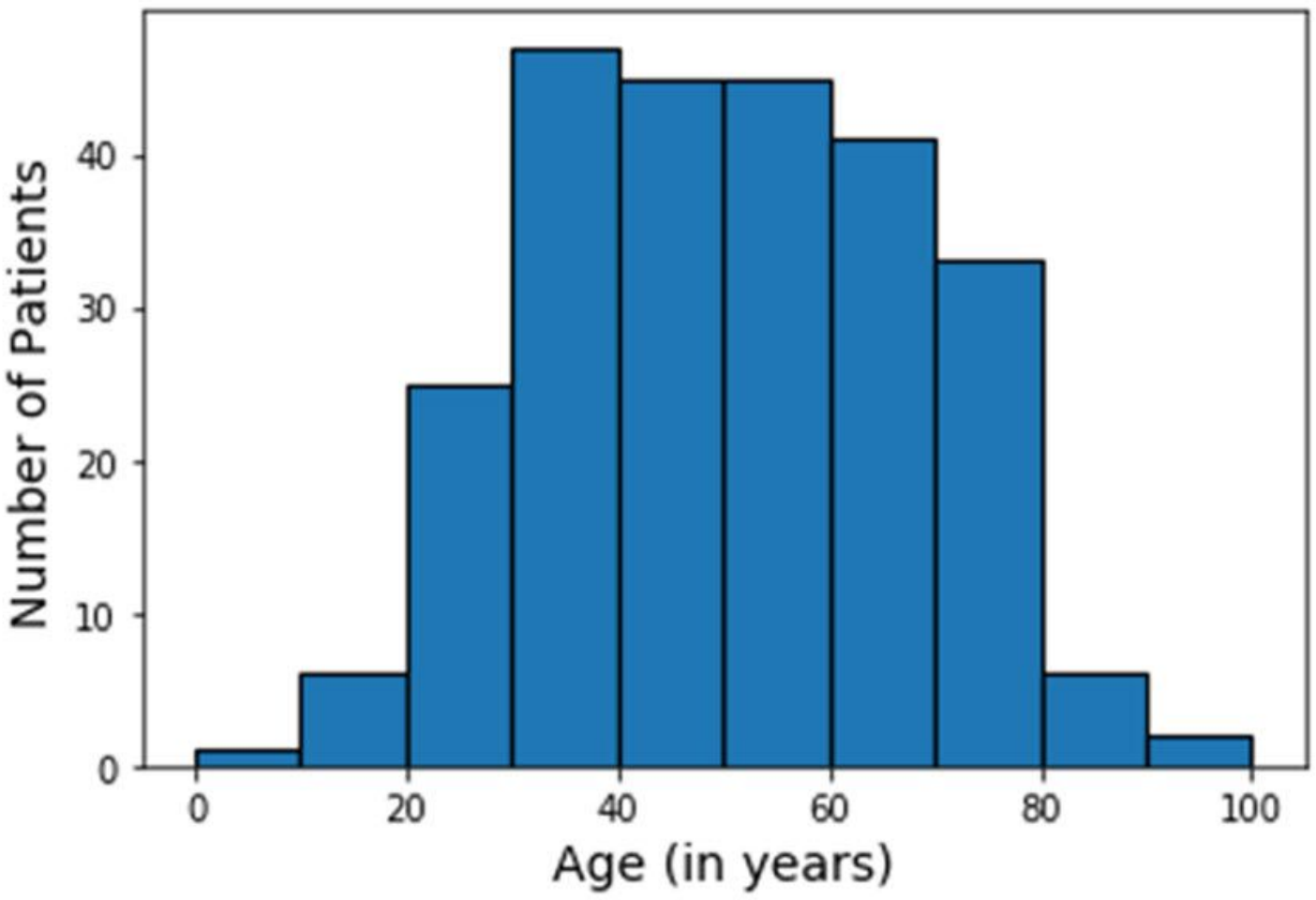

Figure 2

Histogram representing the age distribution of patients included in the study.

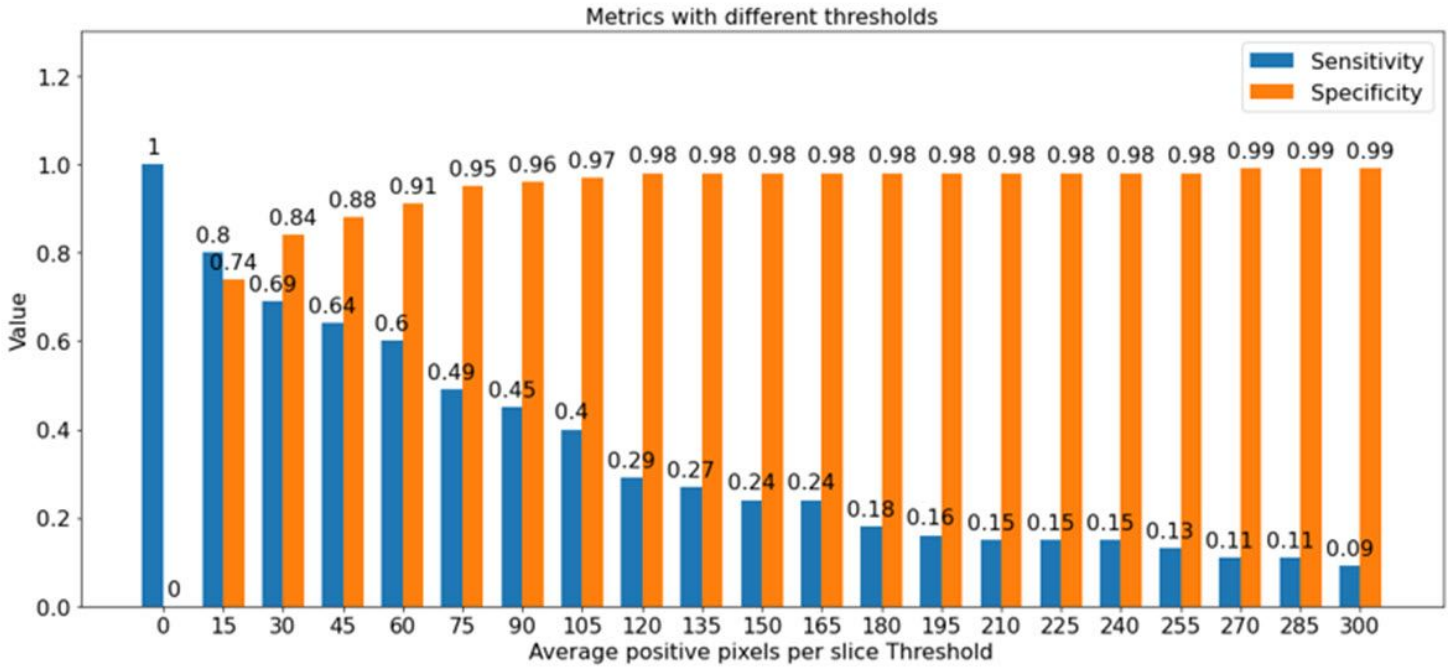

Figure 3

Sensitivity and specificity of the model across different operating points (average positive voxels per slice threshold) on the external test dataset.

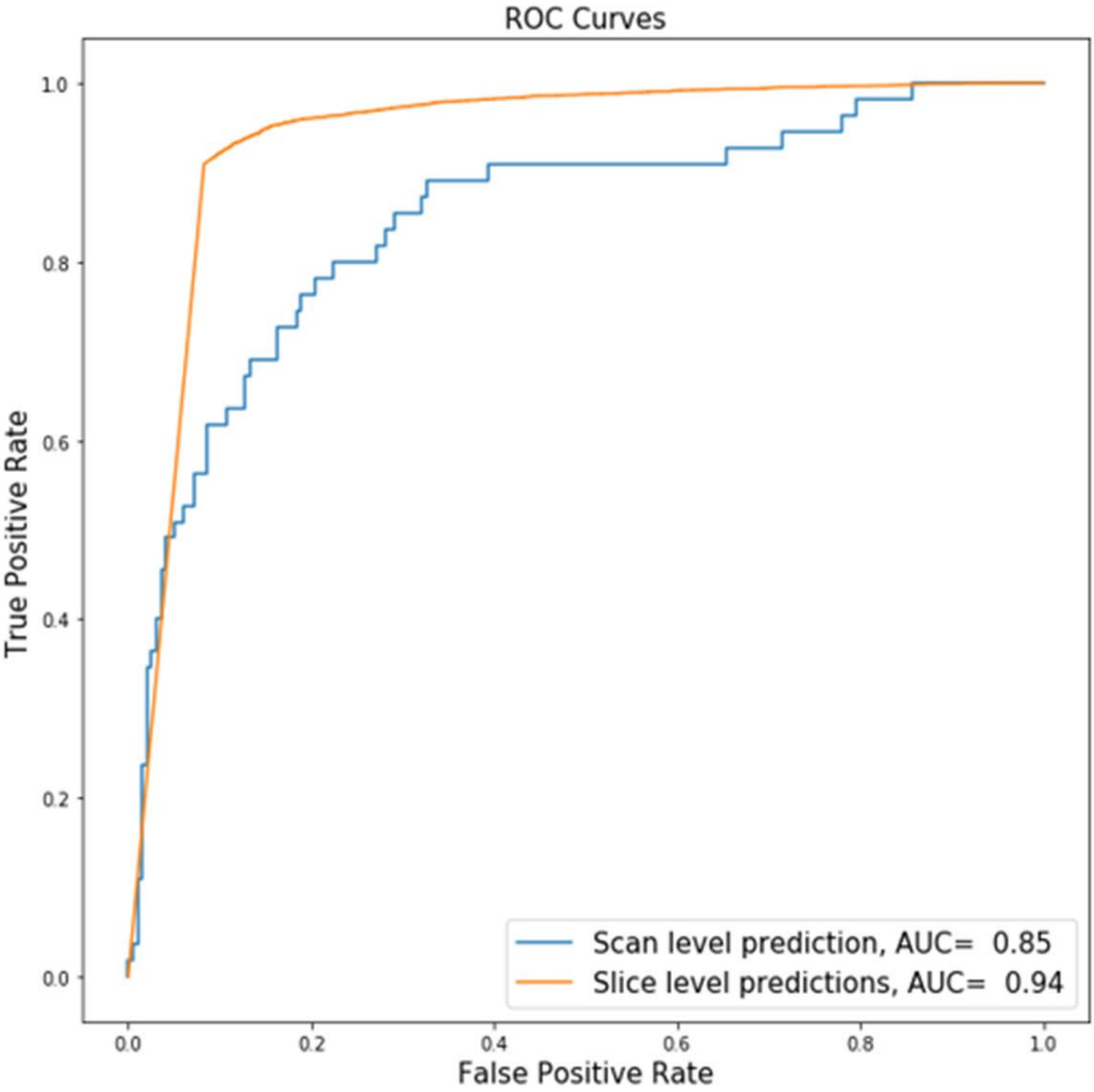

Figure 4

Performance of 2D segmentation U-Net model. The AUROC for the external clinical test set.

Actual values	Predicted values	
	PE	No PE
PE	4818	366
No PE	15881	125102

A. Slice level

Actual values	Predicted values	
	PE	No PE
PE	44	11
No PE	50	146

B. Scan level

Figure 5

Confusion matrix of PE detection at (a) slice level and (b) scan level.

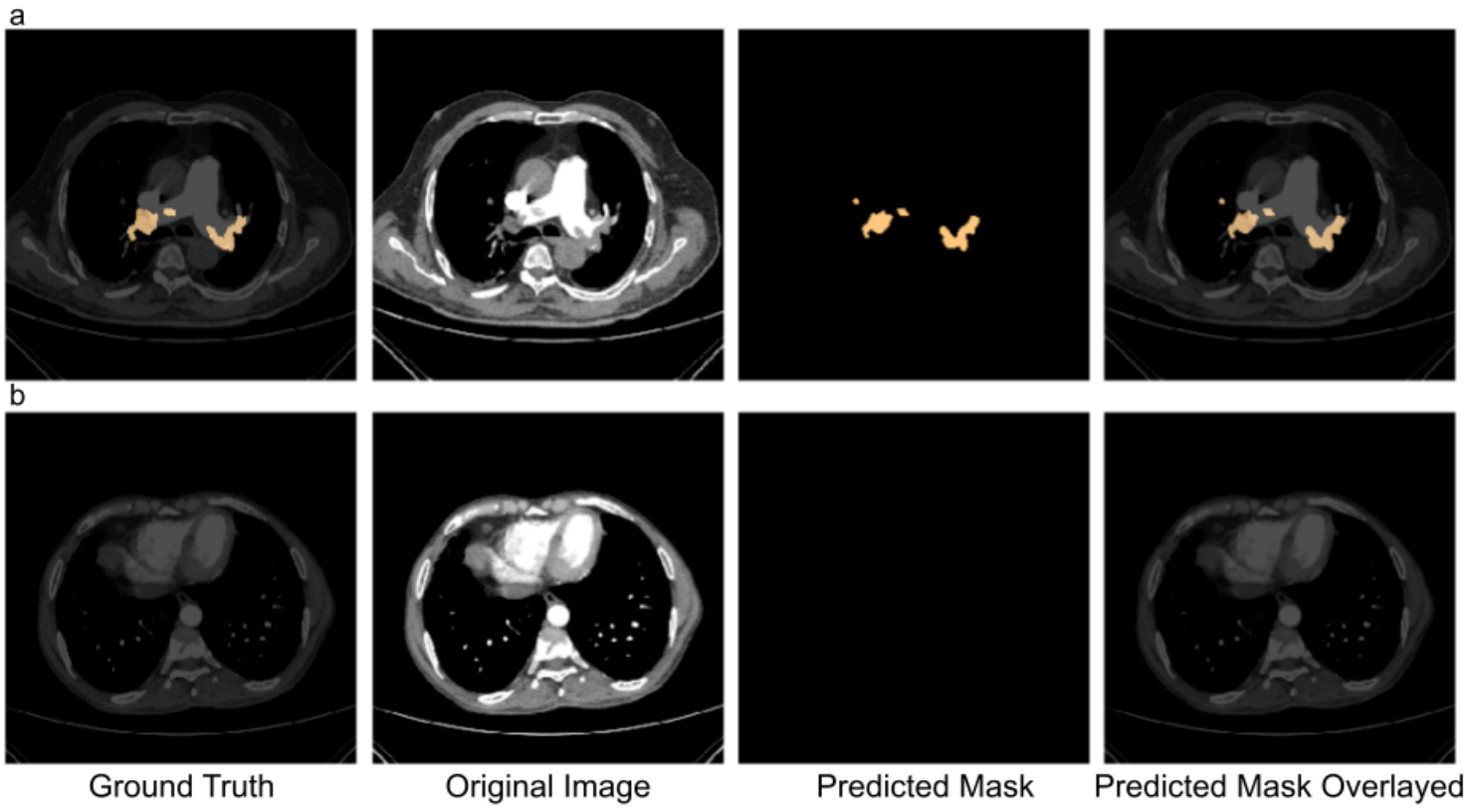

Figure 6

Example of model prediction on a CT slice with **(a)** embolism present and **(b)** no embolism.