


ORIGINAL ARTICLE

Open Access



Retrospective batch analysis to evaluate the diagnostic accuracy of a clinically deployed AI algorithm for the detection of acute pulmonary embolism on CTPA

Eline Langius-Wiffen^{1*} , Pim A. de Jong², Firdaus A. Mohamed Hoesein², Lisette Dekker², Andor F. van den Hoven^{2,3}, Ingrid M. Nijholt¹, Martijn F. Boomsma^{1,4} and Wouter B. Veldhuis²

Abstract

Purpose To generate and extend the evidence on the clinical validity of an artificial intelligence (AI) algorithm to detect acute pulmonary embolism (PE) on CT pulmonary angiography (CTPA) of patients suspected of PE and to evaluate the possibility of reducing the risk of missed findings in clinical practice with AI-assisted reporting.

Methods Consecutive CTPA scan data of 3316 patients referred because of suspected PE between 24-2-2018 and 31-12-2020 were retrospectively analysed by a CE-certified and FDA-approved AI algorithm. The output of the AI was compared with the attending radiologists' report. To define the reference standard, discordant findings were independently evaluated by two readers. In case of disagreement, an experienced cardiothoracic radiologist adjudicated.

Results According to the reference standard, PE was present in 717 patients (21.6%). PE was missed by the AI in 23 patients, while the attending radiologist missed 60 PE. The AI detected 2 false positives and the attending radiologist 9. The sensitivity for the detection of PE by the AI algorithm was significantly higher compared to the radiology report (96.8% vs. 91.6%, $p < 0.001$). Specificity of the AI was also significantly higher (99.9% vs. 99.7%, $p = 0.035$). NPV and PPV of the AI were also significantly higher than the radiology report.

Conclusion The AI algorithm showed a significantly higher diagnostic accuracy for the detection of PE on CTPA compared to the report of the attending radiologist. This finding indicates that missed positive findings could be prevented with the implementation of AI-assisted reporting in daily clinical practice.

Critical relevance statement

Missed positive findings on CTPA of patients suspected of pulmonary embolism can be prevented with the implementation of AI-assisted care.

Key points

- The AI algorithm showed excellent diagnostic accuracy detecting PE on CTPA.
- Accuracy of the AI was significantly higher compared to the attending radiologist.

*Correspondence:

Eline Langius-Wiffen
elinelangius@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

- Highest diagnostic accuracy can likely be achieved by radiologists supported by AI.
- Our results indicate that implementation of AI-assisted reporting could reduce the number of missed positive findings.

Keywords Artificial intelligence, Pulmonary embolism, Computed tomography angiography, Retrospective studies

Graphical abstract

Retrospective batch analysis to evaluate the diagnostic accuracy of a clinically deployed AI algorithm for the detection of acute pulmonary embolism on CTPA.



- The AI algorithm showed excellent diagnostic accuracy detecting pulmonary embolism on CTPA.
- Accuracy of the AI was significantly higher compared to the attending radiologist.
- Highest diagnostic accuracy can likely be achieved by radiologists supported by AI.

		Reference		
		+	-	
Radiology report	+	657	9	666
	-	60	2590	2650
		717	2599	3316

		Reference		
		+	-	
AI algorithm	+	694	2	696
	-	23	2597	2620
		717	2599	3316

Missed positive findings on CTPA of patients suspected of pulmonary embolism can be prevented with the implementation of AI assisted care.

Insights Imaging (2023) Langius-Wiffen et al. DOI: 10.1186/s13244-023-01454-1

Background

Workload for radiologists during regular working hours and on-call hours has dramatically increased in the last decades. CT scans during on-call hours were reported to have increased with 500% between 2006 and 2020 and the number of CT pulmonary angiography (CTPA) to detect pulmonary embolism (PE) even showed an increase of 1360% [1]. As many studies continue to show the added value of medical imaging in patient care and technical developments lead to the acquisition of more complex and larger datasets, the workload is expected to increase even further in the future [2].

Unfortunately, the increased workload is not without risk to the quality of radiologists’ reports, as it has been shown to result in radiologists developing reading fatigue, which may cause diagnostic errors [3, 4], and additionally in an increased rate of burn-out among radiologists and residents [5]. Missed PE can influence patient care and outcome, as it is a potentially life threatening condition with risk of developing severe

complications such as pulmonary hypertension. Radiologists reading CTPA have shown excellent performance to detect PE [6, 7], but the continued increase in workload could lead to additional missed findings in the future.

Artificial intelligence (AI) applications are increasingly evaluated in studies to determine their value in supporting radiologists through workflow improvement and assisted diagnosis [8].

Various requirements have to be met in order to guarantee successful clinical implementation of AI in daily practice [9, 10]. Besides technical validity, clinical validity needs to be assessed in diagnostic accuracy studies. A limited number of studies were already reported on the diagnostic performance of AI algorithms to detect PE on CTPA [11–16]. A recent meta-analysis reviewed five studies and showed variability in performance of the different AI algorithms in detecting PE in research settings [17]. Possible reasons for the discrepancy in results could be that not all of the algorithms had been thoroughly

externally validated for use in clinical practice and/or the large variation in sample sizes and PE prevalence [18]. A commercially available FDA-approved and CE-marked AI algorithm to detect PE on CTPA has shown a promising diagnostic accuracy to detect PE with a sensitivity and specificity of over 90% [12, 15]. Cheikh et al. showed that this AI detected 19 PE that were missed by the initial report in a sample of 1202 CTPA with a PE prevalence of 15.8%, resulting in a higher sensitivity of the AI than the initial report. However, the AI had a significantly lower specificity leading to more false positives. Thus, the AI could cause overdiagnosis when radiologists would rely on the algorithm, which raises the question whether widespread implementation is currently beneficial and safe [9].

The goal of this comparative diagnostic accuracy study was to generate and extend the evidence on the performance of this FDA-approved and CE-marked AI algorithm to detect PE on a large sample of > 3000 CTPA and to determine the additional value of using AI assistance in daily clinical practice.

Methods

Patient inclusion

This comparative diagnostic accuracy study was approved by the institutional review board, and informed consent was waived (no. 2021-78371). CTPA scans of all consecutive patients ≥ 18 years referred to the Radiology department between 24-2-2018 and 31-12-2020 because of suspected PE were retrospectively included.

Artificial intelligence algorithm

All anonymised scans were analysed by an FDA-approved and CE-marked PE (dedicated CTPA) AI algorithm (Aidoc Medical). The architecture of the algorithm has been recently described in detail by Petry et al. [19]. The algorithm was deployed on a virtual machine and integrated with our PACS (SECTRA). The CTPAs within our set time-frame were automatically analysed by the algorithm, where after we received AI activation maps that provided possible PE findings. Aidoc Medical was not involved in study design or analysis of the AI output they provided. No financial support was provided for this study. Their algorithm has not been trained on data originating from our hospital.

Local reading and CT scanning protocol

In our hospital, CT scans are reviewed and reported just once by either a radiologist or by a radiology resident with an Entrusted Professional Activity level 3, 4 or 5 to read CTPA. During the study period, there was no algorithmic support for PE detection in usage.

The hospital had access to the Somaton Force (Siemens Healthineers), Brilliance 64 (Philips Healthcare), iCT (Philips Healthcare) and the IQon (Philips Healthcare). All CT scans were acquired and reconstructed with thin slices of about 1 mm.

The standard scanning protocol for CTPA included a bolus tracking in the pulmonary trunk with a trigger at 200 Hounsfield Units. The post threshold delay was 5 s (or minimal possible). The patients were instructed to maintain a regular breathing pattern and avoid deep inspiration. We used a flow of 5 mL/sec and inject 55–65 mL of contrast depending on body weight and 50 mL of NaCl, except for the spectral detector CT (IQon) where we used a kVp as low as possible to boost the iodine.

Reference standard

The reference standard was established using the report by the attending radiologist or resident with adequate Entrusted Professional Activity level, the AI output and an evaluation of discordant cases. Cases classified as PE positive by both the radiology report and the AI were considered positive according to the reference standard. Similarly, cases classified as PE negative by both the radiology report and the AI were considered negative. Discordant cases were independently reviewed by a chest radiologist (P.A.J., > 10 years of experience) and a medical doctor (L.A.) who had access to the images, the radiology report and the AI output. When these readers agreed on the presence or absence of PE, the reference standard was set. In case of discrepancies a third adjudicator (a cardiothoracic radiologist (F.M.H., > 10 years of experience)) was consulted, who was given access to all available information, to determine the reference standard.

Statistical analysis

For data analysis R version 4.2.0 was used (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-sided. A p value < 0.05 was considered statistically significant.

Diagnostic accuracy measures of the AI algorithm were compared with the diagnostic accuracy measures of the radiology report using the DTComPair package of R. McNemar test was used to compare sensitivity and specificity. Relative predictive values were used to compare positive and negative predictive values.

Results

Patient inclusion

We included CT scans of 3316 patients suspected of PE (1615 females, 48.7%). The mean \pm SD age was 58.6 ± 16.1 years. Given the retrospective nature of our study and the absence of informed consent, we did not

Table 1 Diagnostic accuracy measures of radiology report and AI algorithm on CTPA

	Radiology report	AI algorithm	<i>p</i> value
Sensitivity in % (95% CI)	91.6 (89.6–93.7)	96.8 (95.5–98.1)	<i>p</i> < 0.001
Specificity in % (95% CI)	99.7 (99.4–99.9)	99.9 (99.8–100.0)	<i>p</i> = 0.035
PPV in % (95% CI)	98.6 (97.8–99.5)	99.7 (99.3–100.0)	<i>p</i> = 0.030
NPV in % (95% CI)	97.8 (97.2–98.3)	99.1 (98.8–99.5)	<i>p</i> < 0.001

PPV = positive predictive value, NPV = negative predictive value, CI = confidence interval, *p* values concern the comparison between the diagnostic measures of the radiology report versus the AI algorithm

have other clinical variables available in this study. All images were analysed by the AI.

Radiology report versus AI output

In 633 patients, both the radiology report and the AI output were positive for PE. In 2588 patients, both the radiology report and the AI output were negative for PE. In 95 patients, the diagnosis was discrepant.

According to the reference standard, after re-evaluation by two readers and if needed adjudication, 717 CT patients were positive for PE, resulting in a prevalence of 21.6% (95% confidence interval 20.2–23.1%). Of those, 60 (8.4%) cases of PE were not reported by the attending radiologist and 23 (3.2%) were not detected by the AI algorithm. The cases of missed PE by the attending radiologist concerned two central/lobar, 12 segmental and 46 subsegmental PE. Solely peripheral PE were missed by the AI algorithm (7 segmental, 16 subsegmental). The attending radiologist reported 9 false positive findings, while the algorithm marked 2 false positives.

Overall, the algorithm showed significantly higher diagnostic accuracy measures compared to the radiology reports with sensitivity of 96.8% versus 91.6%, respectively, and specificity of 99.9% versus 99.7%. PPV and NPV of the AI algorithm were also significantly higher than of the radiology report (Table 1).

Discussion

Our study showed that both radiologists and the FDA-approved and CE-marked AI algorithm have an excellent performance on CTPA, with a significantly higher diagnostic accuracy for the AI algorithm to detect PE on a large sample of CTPA of patients suspected of PE compared to the radiology report.

Diagnostic accuracy AI algorithm

Only two previous studies reported on the performance of the FDA-approved and CE-marked automated integrated workflow AI algorithm to detect PE on CTPA used in our

study. Weikert et al. reported that this AI algorithm had a sensitivity of 92.7% and specificity of 95.5% when analysing a sample of 1465 consecutive patients suspected of PE with a PE prevalence of 18.5% [12]. The initial report, reviewed by two physicians served as reference standard. Contrary to our study, they could not compare the performance of the AI to the initial report. A similar approach to ours was used by Cheikh et al., who compared the performance of the same AI algorithm to the initial report of the radiologist using a sample of 1202 consecutive patients suspected of PE from 3 different hospitals with a PE prevalence of 15.8%. They found that the sensitivity of the AI was slightly higher than the radiology report, albeit not significantly (92.6% versus 90%). However, radiologists showed a significantly higher specificity compared to the AI (99.1% vs. 95.8%) [15]. It is not uncommon that the performance of a single AI algorithm varies in different clinical settings [18]. Their scanning protocol was fairly similar to our scanning protocol, but the reference standard was obtained using one radiologist with access to the initial report and the AI output who, in case of doubt, could request the judgement of another senior radiologist. The difference in sample size and prevalence of PE may also at least in part explain the discrepancies between our study and the study of Cheikh et al.

Our results might be suggestive of a general statement that AI could replace radiologists as it reached higher diagnostic accuracy than the initial report of the attending radiologist. This has for example been previously suggested for digital mammography screening [20].

However, standalone use of AI algorithms in reading CTPA is currently not warranted as reading CTPAs doesn't solely include analysis of possible occlusion of the pulmonary arteries. Radiologists can identify other pathology and thus remain responsible for good patient care [21]. Aside from the possibility of the AI missing clinically relevant PE without the additional reading of the radiologist, current algorithms focused on the detection of PE will miss other relevant findings. This might include thrombus in the right atrial appendage, signs of significant pulmonary hypertension and right-ventricular pressure overload, and a variety of additional clinically relevant findings of infectious, oncological or cardiovascular aetiology [22–25].

A strength of our study was the use of a large consecutive cohort of >3000 patients in an academic medical centre with a PE prevalence of 21.6%, reflecting every day clinical practice [26, 27]. As the AI found 60 additional cases of PE in our study population of 3316 patients, to detect a single additional acute PE case would require a number-needed-to-analyse by AI of 56 patients. Due to the low rate of false positive AI findings, evaluating the AI output could be considered minimal additional effort for radiologists.

Limitations

This study comes with some limitations. We did not re-evaluate all images of patients that were classified positive by both the radiology report and the AI output and negative by both the radiology report and AI algorithm. It has been observed that for small PE there can be substantial disagreement between readers depending on expertise level and that small PE are often overdiagnosed in routine practice [28]. However, missed or false positive cases by both would not have affected the outcome of the comparison between the radiology report and the AI output.

The retrospective nature of our study did not allow a direct comparison of conventional care and AI-assisted care, which is considered the ideal design to evaluate the additional value of AI [29]. This would require a preferably prospective study design with large paired or parallel groups in a hospital where AI is implemented. However, the seamless workflow integration of the algorithm complicates the conduct of such a prospective clinical study design, since the AI output is readily available to the radiologist. Another possibility would be to compare diagnostic accuracy of radiologists before and after the implementation of AI, which has the limitation of the time interval between the evaluation before and after implementation. Moreover, to acquire an acceptable reference standard for such a design, several radiologists would have to do a consensus reading of a large number of scans. It could be argued whether such a large investment in effort and costs is still required with the current excellent diagnostic accuracy results of this AI in mind.

Future perspectives

Our results combined with those of previous studies call out for clinical utility studies to determine the cost-effectiveness and the impact of AI assistance on healthcare quality and efficiency, as a next step towards reimbursement and clinical adoption of AI assistance for the detection of PE in CTPA [30–32].

Conclusion

In conclusion, this study showed that the AI algorithm had a significantly higher diagnostic accuracy to detect PE on CTPA of patients suspected of PE compared to the report of the attending radiologist. This indicates that missed positive findings can be prevented with the implementation of AI-assisted care in daily clinical practice.

Abbreviations

AI	Artificial intelligence
CI	Confidence interval
CTPA	Computed tomography pulmonary angiography
PE	Pulmonary embolism

Acknowledgements

The authors would like to thank Aidoc Medical for facilitating analysis of the data.

Author contributions

WBV was the guarantor of integrity of the entire study. ELW wrote the draft manuscript with major support from IN and MFB. All authors contributed substantially to the draft manuscript by providing critical feedback. All authors contributed towards study design and data analysis. Data acquisition was conducted by LD, PAJ and FMH. IN was responsible for statistical analysis. ELW, IN, MFB and PAJ provided literature research. All authors read and approved the final manuscript.

Funding

The authors state that this work has not received any funding.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Institutional Review Board approval was obtained.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. The Department of Radiology, Isala, has established a strategic partnership with Aidoc Medical. However, this study was not part of this partnership.

Author details

¹Department of Radiology, Isala Hospital, Dr. van Heesweg 2, 8025 AB Zwolle, The Netherlands. ²Department of Radiology, University Medical Centre Utrecht, Utrecht, The Netherlands. ³Department of Nuclear Medicine, St. Antonius Hospital, Nieuwegein, The Netherlands. ⁴Division of Imaging and Oncology, University Medical Centre Utrecht, Utrecht, The Netherlands.

Received: 9 January 2023 Accepted: 17 May 2023

Published online: 06 June 2023

References

1. Bruls RJM, Kwee RM (2020) Workload for radiologists during on-call hours: Dramatic increase in the past 15 years. *Insights Imaging* 11:121. <https://doi.org/10.1186/s13244-020-00925-z>
2. Kwee TC, Kwee RM (2021) Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. *Insights Imaging* 12:88–94. <https://doi.org/10.1186/s13244-021-01031-4>
3. Ruutiainen AT, Durand DJ, Scanlon MH, Itri JN (2013) Increased error rates in preliminary reports issued by radiology residents working more than 10 consecutive hours overnight. *Acad Radiol* 20:305–311. <https://doi.org/10.1016/j.jacr.2012.09.028>
4. Hanna TN, Lamoureux C, Krupinski EA, Weber S, Johnson J (2018) Effect of shift, schedule, and volume on interpretive accuracy: a retrospective analysis of 2.9 million radiologic examinations. *Radiology* 287:205–212. <https://doi.org/10.1148/radiol.2017170555>

5. Harolds JA, Parikh JR, Bluth EI, Dutton SC, Recht MP (2016) Burnout of radiologists: frequency, risk factors, and remedies: a report of the ACR commission on human resources. *J Am Coll Radiol* 13:411–416. <https://doi.org/10.1016/j.jacr.2015.11.003>
6. Kligerman S, Lahiji K, Weihe E et al (2015) Detection of pulmonary embolism on computed tomography: Improvement using a model-based iterative reconstruction algorithm compared with filtered back projection and iterative reconstruction algorithms. *J Thorac Imaging* 30:60–68. <https://doi.org/10.1097/RTI.0000000000000122>
7. Konstantinides SV, Torbicki A, Agnelli G et al (2014) 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism: the task force for the diagnosis and management of acute pulmonary embolism of the european society of cardiology (ESC) endorsed by the european respiratory society (ERS). *Eur Heart J* 35:3033–3080. <https://doi.org/10.1093/eurheartj/ehu283>
8. Neri E, de Souza N, Brady A et al (2019) What the radiologist should know about artificial intelligence—an ESR white paper. *Insights Imaging* 10:44. <https://doi.org/10.1186/s13244-019-0738-2>
9. Daye D, Wiggins WF, Lungren MP et al (2022) Implementation of clinical artificial intelligence in radiology: Who decides and how? *Radiology* 305:555–563. <https://doi.org/10.1148/radiol.212151>
10. Vasey B, Nagendran M, Campbell B et al (2022) Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* 28:924–933. <https://doi.org/10.1038/s41591-022-01772-9>
11. Liu Z, Yuan H, Wang H (2022) CAM-wnet: an effective solution for accurate pulmonary embolism segmentation. *Med Phys* 49:5294–5303. <https://doi.org/10.1002/mp.15719>
12. Weikert T, Winkel DJ, Bremerich J et al (2020) 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>
13. Huang S, Kothari T, Banerjee I et al (2020) 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>
14. Buls N, Watté N, Nieboer K, Ilsen B, de Mey J (2021) Performance of an artificial intelligence tool with real-time clinical workflow integration—detection of intracranial hemorrhage and pulmonary embolism. *Phys Med* 83:154–160. <https://doi.org/10.1016/j.ejmp.2021.03.015>
15. Cheikh AB, Gorincour G, Nivet H et al (2022) How artificial intelligence improves radiological interpretation in suspected pulmonary embolism. *Eur Radiol* 32:5831–5842. <https://doi.org/10.1007/s00330-022-08645-2>
16. Huhtanen H, Nyman M, Mohsen T, Virkki A, Karlsson A, Hirvonen J (2022) 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>
17. Soffer S, Klang E, Shimon O et al (2021) Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Sci Rep* 11:15814–15823. <https://doi.org/10.1038/s41598-021-95249-3>
18. Yu AC, Mohajer B, Eng J (2022) External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell* 4:e210064. <https://doi.org/10.1148/ryai.210064>
19. Petry M, Lansky C, Chodakiewicz Y, Maya M, Pressman B (2022) Decreased hospital length of stay for ICH and PE after adoption of an artificial intelligence-augmented radiological worklist triage system. *Radiol Res Pract* 2022:2141839. <https://doi.org/10.1155/2022/2141839>
20. Romero-Martín S, Elías-Cabot E, Raya-Povedano J, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M (2022) Stand-alone use of artificial intelligence for digital mammography and digital breast tomosynthesis screening: a retrospective evaluation. *Radiology* 302:535–542. <https://doi.org/10.1148/radiol.211590>
21. Geis JR, Brady AP, Wu CC et al (2019) Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. *Radiology* 293(2):436–440. <https://doi.org/10.1148/radiol.2019191586>
22. Eskandari A, Narayanasamy S, Ward C et al (2022) Prevalence and significance of incidental findings on computed tomography pulmonary angiograms: a retrospective cohort study. *Am J Emerg Med* 54:232–237. <https://doi.org/10.1016/j.ajem.2022.01.064>
23. Hall WB, Truitt SG, Scheunemann LP et al (2009) The prevalence of clinically relevant incidental findings on chest computed tomographic angiograms ordered to diagnose pulmonary embolism. *Arch Internal Med* 169:1961–1965. <https://doi.org/10.1001/archinternmed.2009.360>
24. Richman PB, Courtney DM, Friese J et al (2004) Prevalence and significance of nonthromboembolic findings on chest computed tomography angiography performed to rule out pulmonary embolism: a multicenter study of 1025 emergency department patients. *Acad Emerg Med* 11:642–647
25. Devaraj A, Sayer C, Sheard S, Grubnic S, Nair A, Vlahos I (2015) Diagnosing acute pulmonary embolism with computed tomography: Imaging update. *J Thorac Imaging* 30:176–192. <https://doi.org/10.1097/RTI.0000000000000146>
26. van der Hulle T, Cheung WY, Kooij S et al (2017) Simplified diagnostic management of suspected pulmonary embolism (the YEARS study): a prospective, multicentre, cohort study. *Lancet* 390:289–297. [https://doi.org/10.1016/S0140-6736\(17\)30885-1](https://doi.org/10.1016/S0140-6736(17)30885-1)
27. van der Pol LM, Dronkers CEA, van der Hulle T et al (2018) The YEARS algorithm for suspected pulmonary embolism: shorter visit time and reduced costs at the emergency department. *J Thromb Haemost* 16:725–733. <https://doi.org/10.1111/jth.13972>
28. Hutchinson BD, Navin P, Marom EM, Truong MT, Bruzzi JF (2015) Overdiagnosis of pulmonary embolism by pulmonary CT angiography. *AJR Am J Roentgenol* 205:271–277. <https://doi.org/10.2214/AJR.14.13938>
29. Park SH, Han K, Jang HY et al (2022) Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. *Radiology*. <https://doi.org/10.1148/radiol.220182>
30. Bossuyt PMM, Reitsma JB, Linnik K, Moons KGM (2012) Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem* 58:1636–1643. <https://doi.org/10.1373/clinchem.2012.182576>
31. Grutters JPC, Govers T, Nijboer J, Tummers M, van der Wilt G, Jan RMM (2019) Problems and promises of health technologies: the role of early health economic modeling. *Int J Health Policy Manag* 8:575–582. <https://doi.org/10.15171/ijhpm.2019.36>
32. Wolff J, Pauling J, Keck A, Baumbach J (2020) The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res* 22:e16866. <https://doi.org/10.2196/16866>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)