BRITISH JOURNAL OF
HOSPITAL MEDICINE

SYSTEMATIC REVIEW

The Current State of Artificial Intelligence on Detecting Pulmonary Embolism via Computerised Tomography Pulmonary Angiogram: A Systematic Review

Moustafa Said Taman Ahmed Hassan[1], Mahmoud Abdrabo Mahmoud Elhotiby[2], Viraj Shah[3], Henry Rocha[4], Arian Arjomandi Rad[5,6,*], George Miller[6], Johann Malawana[6]

[1]Musgrove Park Hospital, Somerset NHS Foundation Trust, Taunton, UK
[2]Lincoln County Hospital, United Lincolnshire Hospitals NHS Trust, Lincoln, UK
[3]Department of Medicine, Faculty of Medicine, Imperial College London, London, UK
[4]School of Clinical Medicine, University of Cambridge, Cambridge, UK
[5]Department of Cardiothoracic Surgery, Oxford University NHS Foundation Trust, Oxford, UK
[6]Research Unit, The Healthcare Leadership Academy, London, UK
*Correspondence: arian.arjomandirad@gmail.com (Arian Arjomandi Rad)

Abstract

Aims/Background Pulmonary embolism (PE) is a life-threatening condition with significant diagnostic challenges due to high rates of missed or delayed detection. Computed tomography pulmonary angiography (CTPA) is the current standard for diagnosing PE, however, demand for imaging places strain on healthcare systems and increases error rates. This systematic review aims to assess the diagnostic accuracy and clinical applicability of artificial intelligence (AI)-based models for PE detection on CTPA, exploring their potential to enhance diagnostic reliability and efficiency across clinical settings.

Methods A systematic review was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Excerpta Medica Database (EMBASE), Medical Literature Analysis and Retrieval System Online (MEDLINE), Cochrane, PubMed, and Google Scholar were searched for original articles from inception to September 2024. Articles were included if they reported successful AI integration, whether partial or full, alongside CTPA scans for PE detection in patients.

Results The literature search identified 919 articles, with 745 remaining after duplicate removal. Following rigorous screening and appraisal aligned with inclusion and exclusion criteria, 12 studies were included in the final analysis. A total of three primary AI modalities emerged: convolutional neural networks (CNNs), segmentation models, and natural language processing (NLP), collectively used in the analysis of 341,112 radiographic images. CNNs were the most frequently applied modality in this review. Models such as AdaBoost and EmbNet have demonstrated high sensitivity, with EmbNet achieving 88–90.9% per scan and reducing false positives to 0.45 per scan.

Conclusion AI shows significant promise as a diagnostic tool for identifying PE on CTPA scans, particularly when combined with other forms of clinical data. However, challenges remain, including ensuring generalisability, addressing potential bias, and conducting rigorous external validation. Variability in study methodologies and the lack of standardised reporting of key metrics complicate comparisons. Future research must focus on refining models, improving peripheral emboli detection, and validating performance across diverse settings to realise AI's potential fully.

Key words: artificial intelligence; natural language processing; machine learning; pulmonary embolism; CT pulmonary angiogram scan; data science; medical technology

Submitted: 13 October 2024 Revised: 30 November 2024 Accepted: 13 December 2024

How to cite this article:
Hassan MSTA, Elhotiby MAM, Shah V, Rocha H, Rad AA, Miller G, Malawana J. The Current State of Artificial Intelligence on Detecting Pulmonary Embolism via Computerised Tomography Pulmonary Angiogram: A Systematic Review. Br J Hosp Med. 2025. https://doi.org/10.12968/hmed.2024.0757

Introduction

Pulmonary embolism (PE) is a life-threatening cardiovascular condition with significant global morbidity and mortality, ranking as the third most common acute cardiovascular event after myocardial infarction and stroke (Kearon et al, 2016). The estimated annual incidence of PE varies widely worldwide, with approximately 60–70 cases per 100,000 people in North America and Europe; significant regional variation has been observed, often related to healthcare accessibility and diagnostic practices (Vyas et al, 2024). Early and accurate detection of PE is essential, as untreated or undiagnosed cases can result in mortality rates of up to 30%, making prompt diagnosis critical to prevent adverse outcomes and improve survival rates (Curcio et al, 2022). However, the diagnosis of PE remains challenging, with high rates of missed diagnoses or delayed detection. Computed tomography pulmonary angiography (CTPA) is the current gold standard for diagnosing PE, offering high sensitivity and specificity, typically reported at approximately 80% for both metrics (Hirsh et al, 2008). Despite this, healthcare systems globally are burdened with an increasing demand for diagnostic imaging, while radiologists are tasked with interpreting complex CTPA scans within constrained timelines. This growing pressure has contributed to diagnostic errors, with some studies reporting missed diagnosis rates for PE on CTPA scans as high as 40% in certain centres (Raptis et al, 2022; Wang et al, 2023; Wichmann et al, 2015).

The diagnostic challenge of PE is further compounded by variability in radiologist experience and proficiency, particularly in detecting central and peripheral emboli, which may lead to inconsistencies in diagnostic accuracy (Wichmann et al, 2015). These issues underscore the urgent need for advanced diagnostic tools to improve the reliability, efficiency, and accuracy of PE detection, thus reducing the risk of missed or delayed diagnoses. In this context, artificial intelligence (AI) represents a promising adjunctive tool for enhancing the diagnostic process. In recent years, AI, particularly through deep learning (DL) methodologies like convolutional neural networks (CNNs), has demonstrated significant utility in medical imaging, with applications successfully developed to detect pulmonary nodules, pneumothorax, and intracranial hemorrhage, among other conditions (Esteva et al, 2019; Lakhani, 2017; Tang et al, 2019). For PE detection, AI models are being designed to automate the identification of emboli on CTPA scans, thereby improving diagnostic efficiency and workflow for radiologists and healthcare facilities.

Despite these advancements, a gap remains in understanding the full clinical impact and limitations of AI applications for PE detection on CTPA. While earlier studies, such as a review by Soffer et al (2021), have shown promising results in diagnostic accuracy with AI tools, the rapid evolution of machine learning (ML) and DL methodologies has led to increasingly sophisticated applications, often involving multiple imaging modalities and intricate segment analyses (Tang et al, 2019). Nevertheless, critical issues remain unresolved, including the management of false positives and false negatives and the overall effect of these technologies on patient outcomes and healthcare delivery when integrated into clinical settings (Hosny et al, 2018).

This systematic review addresses these gaps in the literature by examining the current state of AI applications specifically for PE detection on CTPA, evaluating the effectiveness, diagnostic accuracy, and workflow impact of these tools. This review provides a focused analysis of how these AI models perform across different clinical environments, such as emergency and non-emergency settings, to assess their applicability and potential integration into routine diagnostic pathways. Thus, the primary aims of this review are to evaluate the diagnostic accuracy of AI-based models for PE detection on CTPA, to determine the clinical applicability and effectiveness of these models in various healthcare settings, and to identify current limitations and potential areas for future research that may optimise AI applications in PE diagnosis. By addressing these key issues, this review aims to provide a comprehensive overview of AI's role in enhancing PE diagnosis, identifying both its potential and limitations in clinical practice.

Methods

Literature Search Strategy

A systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. A PRISMA checklist for this review is provided in **Supplementary material**. A literature search was conducted in Excerpta Medica Database (EMBASE), Medical Literature Analysis and Retrieval System Online (MEDLINE), Cochrane, PubMed, and Google Scholar from inception to September 2024. The full search terms can be found in Appendix Table 2. Additional articles were identified through a manual search of the references lists of articles found through the original search and use of the 'related articles' function on MEDLINE. The only limits used were the mentioned time frame and English language.

Study Inclusion and Exclusion Criteria

All original articles were included if they reported the use of AI or ML for PE detection via a CTPA scan. Studies were considered if they presented ML models in any aspect of a detection technique that made at least partial use of CTPA, whether as the mainstay or a supportive modality. There were no geographical restrictions. Studies were excluded from the review if there were inconsistencies in the data that impeded data extraction or the study was performed in an animal model. Reviews, editorials, case reports, abstracts from meetings, and preclinical studies were excluded. In line with the predefined inclusion and exclusion criteria, two reviewers (VS and HR) independently selected articles for further assessment following the title and abstract review. A third independent reviewer (AAR) resolved any disagreements between the two reviewers. Potentially eligible studies were then retrieved for full-text assessment. The software used for the described process was Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia, https://www.covidence.org/). A full list of the inclusion/exclusion criteria may be found in the Appendix Table 2, including the PRISMA screening process.

Data Extraction and Critical Appraisal of Evidence

All full texts of retrieved articles were read and reviewed by two authors (VS and HR) and a unanimous decision was made regarding the inclusion or exclusion of studies. When there was disagreement, the final decision was made by a third reviewer (AAR). Using a pre-established protocol, the following data was extracted: first author; study type and characteristics; number of patients; population demographics; study aims; category of ML method used; method of ML implemented and main reported outcomes. A data extraction sheet for this review was developed and pilot-tested using three randomly selected included studies and subsequently was refined accordingly. Data extraction was performed by two review authors (VS and HR) who carried out the process in duplicate on two separate extraction sheets. Correctness of the tabulated data was validated by a third author (AAR) who evaluated both extraction sheets and assessed full texts where incongruences existed. Due to the high heterogeneity of the included studies' methodology, the research group did not conduct a meta-analysis or include a quantitative risk-of-bias assessment. Nevertheless, a qualitative assessment of the quality of studies included was conducted, and is detailed within the Discussion below.

Results

Study Selection

The literature search identified 919 articles; following the removal of duplicates, 745 were screened. A total of 139 full-text articles were reviewed and assessed aligned with the inclusion and exclusion criteria. Following critical appraisal, 12 studies were included in this review. Fig. 1 outlines the study selection process. A summary of included articles and their respective designs, type of ML modalities, as well as the main reported outcomes are presented in Table 1.

Study Characteristics

Studies were published between 2017 and 2024, analysing a total of 341,112 radiographic images. All the eligible studies (n = 12/12) were retrospective in nature, and all of these employed radiologist interpretation as the reference comparison standard.

ML Modality Employed

Seven of the twelve studies used a supervised ML approach as their model in the form of a CNN, with two additional studies applying a semi-supervised model in the form of a segmentation model. In two cases, natural language processing (NLP) was employed as the modality of choice. In one study, the modality was unidentifiable. Within the studies that reported using a supervised learning method, there was strong homogeneity in the ML approach used, with all seven of these studies applying a CNN. The semi-supervised approach used a segmentation model that was based on a HRNet-based architecture; however, only two studies employed this modality. Seven studies were conducted in USA centres, which was the modal study location in this review.

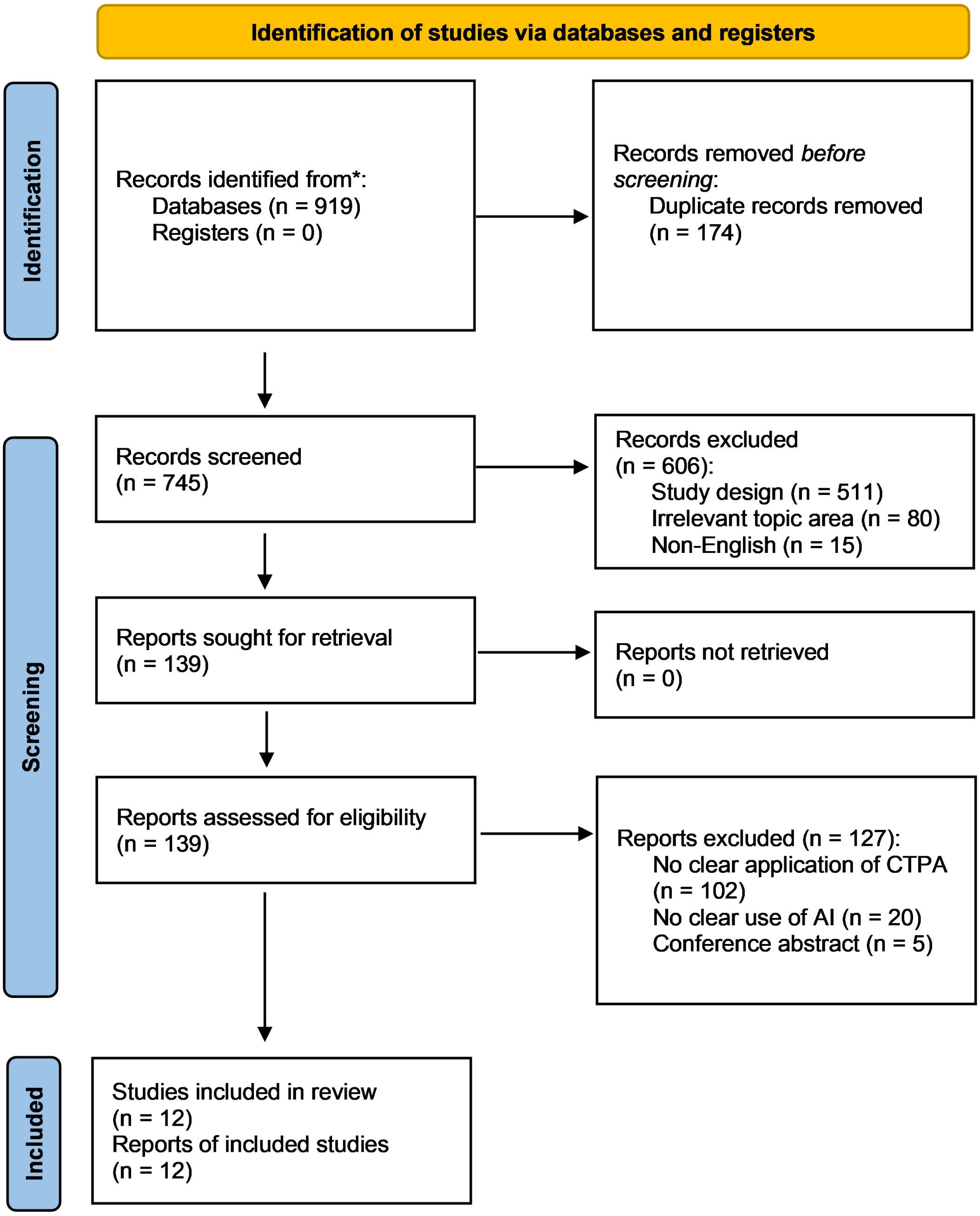

Fig. 1. A flow summary of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology adopted for this review. "*": Databases. CTPA, computed tomography pulmonary angiography; AI, artificial intelligence.

Outcome in NLP Techniques

Of the two studies that employed NLP, the range of sensitivities observed was between 86.0% and 97.1%, while specificities observed were between 98.6% and 100% (Amin et al, 2024; Swartz et al, 2017).

Table 1. A summary table of the main reported outcomes of all studies that met review's inclusion criteria.

Study & Year	Study design	Country (s)	Number of datapoints analysed (n)	Machine learning modality employed	Main reported outcomes
(Swartz et al, 2017)	R, NA	USA	2000	Natural language processing	• The novel NLP tool demonstrated high accuracy in classifying venous thromboembolism (VTE) reports, achieving a sensitivity of 95.7% (UTZ) and 97.1% (CTPA), with both modalities having a specificity of 100% and 98.6%, respectively. • The diagnostic yield was evaluated at the individual provider level, leading to the development of an imaging dashboard for enhanced data presentation.
(Tajbakhsh et al, 2019)	R, M	USA	121	Convoluted neural network	• Vessel-oriented image representation (VOIR) enhances pulmonary embolism (PE) detection, achieving 0 mm localisation error in the PE challenge and enabling more accurate diagnoses by both radiologists and computer-aided design (CAD) systems. • The novel image representation accelerates model training and improves accuracy, outperforming 3D CNNs while requiring fewer training samples. • VOIR's compact and consistent design facilitates multi-view visualisation of vessels, supporting confident diagnosis and demonstrating robust performance across various computed tomography (CT) scanners.
(Wittenberg et al, 2012)	R, M	Netherlands	129	Segmentation model	• Patient groups were similar regarding age (p = 0.22), accompanying lung disease (p = 0.12), and inpatient/outpatient ratio (p = 0.67), with no significant differences in sensitivity (100%, 97%, and 92%) or specificity (18%, 15%, and 13%) across institutions (p = 0.21 and p = 0.820, respectively). • While the mean number of false positive (FP) findings varied significantly (4.5, 6.2, and 3.7; p = 0.02 and p = 0.03), the median numbers (2, 3, and 3) were comparable, with image quality parameters showing a significant association with the number of FPs ($p < 0.05$), but not with sensitivity. • The performance of CAD systems was found to be independent of the type of scanner used, emphasising that image quality and scanning protocols are critical factors influencing CAD effectiveness.

Table 1. Continued.

Study & Year	Study design	Country (s)	Number of datapoints analysed (n)	Machine learning modality employed	Main reported outcomes
(Weikert et al, 2020)	R, M	Switzerland	29,465	Convoluted neural network	● The AI prototype algorithm demonstrated a high degree of diagnostic accuracy, identifying 215 of 232 pulmonary embolism (PE) positive exams [sensitivity 92.7%; 95% confidence interval (CI) 88.3–95.5%)] and 1178 of 1233 negative exams (specificity 95.5%; 95% CI 94.2–96.6%). ● On a per-finding basis, the algorithm correctly marked 1174 of 1352 identified emboli as true positives, with false positives primarily attributed to contrast agent-related flow artefacts, pulmonary veins, and lymph nodes. ● This algorithm can aid clinicians by automatically prioritising exams with high suspicion of PE, serving as a secondary reading tool, thereby expediting the diagnostic and therapeutic processes while minimising the risk of false negatives.
(Huang et al, 2020a)	R, NM	USA	1837	Convoluted neural network	● This study developed and evaluated various multimodal fusion model architectures that integrate pixel data from volumetric Computed Tomography Pulmonary Angiography (CTPA) scans and clinical patient data from electronic medical records (EMR) for the automatic classification of pulmonary embolism (PE) cases. ● The top-performing model, a late fusion approach, achieved an area under the receiver operating characteristic curve (AUROC) of 0.947 (95% CI: 0.946–0.948) on the entire held-out test set. ● This late fusion model significantly outperformed both imaging-only and EMR-only single modality models, demonstrating the efficacy of combining multimodal data for enhanced PE classification.

Table 1. Continued.

Study & Year	Study design	Country (s)	Number of datapoints analysed (n)	Machine learning modality employed	Main reported outcomes
(Liu et al, 2020)	R, M	China	878	Convoluted neural network	● The study found that the U-Net deep learning model demonstrated a sensitivity of 94.6% and specificity of 76.5% for detecting clots at a probability threshold of 0.1, with an area under the curve (AUC) of 0.926 (95% CI: 0.884–0.968). ● Additionally, the clot burden quantified using U-Net was significantly correlated with established clinical scores, such as the Qanadli score (r = 0.819, $p < 0.001$) and the Mastora score (r = 0.874, $p < 0.001$), as well as with right ventricular functional parameters assessed via CTPA. ● These findings suggest that the deep learning convolutional neural network (DL-CNN) can effectively detect acute pulmonary embolism (APE) and quantitatively assess clot burden, potentially alleviating clinician workload in the diagnostic process.
(Huang et al, 2020b)	R, M	USA	1997	Convoluted neural network	● The PENet model achieved an AUROC of 0.84 (95% CI: 0.82–0.87) on an internal dataset and 0.85 (95% CI: 0.81–0.88) on an external dataset, demonstrating robust generalisability for detecting pulmonary embolism (PE). ● PENet outperformed current state-of-the-art 3D convolutional neural networks, effectively diagnosing PE without extensive preprocessing. ● This model has the potential to function as a triage tool, enabling automatic identification of clinically significant PEs and improving prioritisation for diagnostic interpretation.
(Shapiro et al, 2024)	R, M	USA	NA	NA	● Implementation of the Viz.ai PE Module at TriHealth Bethesda Hospital significantly reduced time-to-consult from an average of 240.45 minutes (pre-AI) to 6.72 minutes (post-AI), highlighting a dramatic improvement in evaluation times for potential pulmonary embolism (PE) cases. ● The average time for pulmonary embolism response team (PERT) activation was notably faster in the post-AI cohort, with a marked decrease in in-hospital mortality from 8.4% (pre-AI) to 2.2% (post-AI). ● Overall, the AI-powered platform improved scan-to-assessment time from 318.42 minutes to 5.47 minutes and facilitated faster anticoagulant administration for PERT-activated cases, demonstrating a significant enhancement in clinical workflow and patient outcomes.

Table 1. Continued.

Study & Year	Study design	Country (s)	Number of datapoints analysed (n)	Machine learning modality employed	Main reported outcomes
(Somani et al, 2021)	R, NM	USA	320,746	Convoluted neural network	● In a cohort of 21,183 patients with moderate to high suspicion of pulmonary embolism (PE), a Fusion model combining clinical data and electrocardiogram (ECG) waveform representation achieved an AUROC of 0.81, outperforming ECG (0.59) and electronic health record (HER) models (0.65). ● The Fusion model showed superior specificity (0.18) and performance (AUROC 0.84) compared to four clinical scores (AUROC 0.50–0.58) and maintained consistent performance across sex (0.81) and racial/ethnic groups (0.77–0.84). ● This study demonstrates that integrating ECG waveforms with clinical data enhances PE detection specificity in patients with moderate to high suspicion of PE.
(Cheng et al, 2023)	R, NM	Taiwan	33	Segmentation model	● Training a semi-supervised learning model on an open-source dataset and unlabeled data from National Cheng Kung University Hospital achieved mean intersection over union (mIOU), dice score, and sensitivity of 0.3510, 0.4854, and 0.4253, respectively. ● Fine-tuning the model with a limited set of unlabeled pulmonary embolism CTPA images from China Medical University Hospital resulted in improvements in mIOU (from 0.2344 to 0.3721), dice score (from 0.3325 to 0.5113), and sensitivity (from 0.3151 to 0.4967). ● These findings indicate that the semi-supervised model enhances accuracy across datasets while minimizing labelling efforts by requiring only a small number of unlabeled images for fine-tuning.

Table 1. Continued.

Study & Year	Study design	Country (s)	Number of datapoints analysed (n)	Machine learning modality employed	Main reported outcomes
(Amin et al, 2024)	R, M	USA	12,183	Natural language processing	● The natural language processing model demonstrated high accuracy in identifying pulmonary embolism (PE) from radiology reports, achieving an area under the receiver operating characteristic curve (ROC) of 0.99, sensitivity of 0.86 to 0.87, and specificity of 0.99 in internal and temporal validation cohorts. ● Among PE-CTPA studies, 10.5% were positive for PE, with 22.2% classified as low-risk by the simplified Pulmonary Embolism Severity Index (sPESI) score; notably, 74.3% of low-risk PE patients were admitted for inpatient management. ● This study highlights the model's effectiveness in real-time identification of PE and its potential as a clinical decision support tool for outpatient management of low-risk patients in the emergency department when combined with the validated sPESI risk stratification score.
(Zhu et al, 2024)	R, M	China	555	Convoluted neural network	● The EmbNet achieved per-scan sensitivity of 90.9% (internal validation) and 88.0% (external validation), with specificity at 75.4% and 70.5%, respectively. Positive predictive value (PPV) was 48.4% (internal) and 42.7% (external), while negative predictive value (NPV) was 97.0% and 95.9%. ● At the per-embolus level, sensitivity was 86.0% (internal) and 83.5% (external), with PPV at 61.3% and 57.5%. Sensitivity for central emboli was 89.7% (internal) and 94.4% (external), while lobar emboli sensitivity was 95.2% and 93.5%. Peripheral emboli sensitivity was 82.6% (internal) and 80.2% (external). ● The average false positive rate was 0.45 (internal) and 0.69 (external) false emboli per scan, indicating that EmbNet provides high sensitivity across different embolus locations, supporting its potential use for initial screening in clinical practice.

Note: R, Retrospective; M, Multicentre; NM, Not Multicentre; NLP, natural language processing.

Outcome in Segmentation Model Techniques

Of the two studies that employed segmentation models, the range of sensitivities observed was between 31.51% and 100%, and specificities observed were between 13% and 18% (Wittenberg et al, 2012; Cheng et al, 2023).

Outcome in CNN Techniques

Of the seven studies that employed CNNs, the range of sensitivities observed was between 83.5% and 95.2%, while specificities observed ranged from 70.5% to 95.5% (Weikert et al, 2020; Huang et al, 2020a; Liu et al, 2020; Huang et al, 2020b; Somani et al, 2021; Zhu et al, 2024; Tajbakhsh et al, 2019).

Discussion

The use of AI in detecting PE on CTPA imaging has made significant strides. This systematic review has identified significant heterogeneity inherent in these approaches, including NLP and DL models. While the review has found clear potential for all AI modalities identified to enhance diagnostic accuracy, evidence of their utility in clinical workflows remains seldom considered. This remains challenging given several key themes still persist in radiology, including time constraints, accuracy, and scalability.

AI models, such as those integrating clinical data with ML, provide a substantial improvement over traditional clinical scoring systems like the Wells and Geneva scores. For instance, in the Exarchos et al (2020) pilot study and follow-on Huang et al (2020a) application, the AdaBoost classification coupled with the Wrapper feature selection technique has demonstrated higher accuracy and sensitivity compared to traditional methods, indicating the potential for AI to enhance early and accurate diagnosis of PE. This is particularly crucial as PE continues to be underdiagnosed despite advancements in imaging, as aforementioned. Such models offer the potential to mitigate this by reducing over-reliance on CTPA and improving diagnostic yields (Amin et al, 2024).

NLP in Clinical Decision Support for Low-Risk PEs

NLP has emerged as a promising approach in assisting clinicians with PE diagnosis and risk stratification. Several studies, notably Swartz et al (2017), who pioneered the integration of NLP with clinical decision support systems (CDSS), have demonstrated the utility of NLP in analysing CTPA radiology reports to identify low-risk PE cases in real-time. These tools, including those developed by Amin et al (2024), are designed to support outpatient management by identifying low-risk patients, thus reducing unnecessary hospitalisations and associated healthcare costs. The incorporation of NLP allows for the automated extraction of essential clinical and operational terms—like "right heart strain" or "result discussed with Dr XYZ"—that help stratify patient risk.

While the performance of NLP-based models is comparable to other PE detection models, their scalability, ease of integration into existing healthcare infrastructures, and lower development costs make them particularly attractive (Amin et al, 2024; Soffer et al, 2021). Moreover, the ability of NLP models to operate in real-time aligns well with clinical needs, ensuring prompt intervention and follow-up. Nevertheless, the current literature indicates that clinical decision support systems incorporating NLP are most beneficial when applied to low-risk patient groups—as identified by Amin et al (2024), where 26.7% of such patients are routinely outpatient cases. As the evidence is focused on this demographic, questions persist about their generalisability across different risk categories and clinical settings.

DL and Image-Based AI

The introduction of DL models, particularly CNNs, has significantly improved the accuracy of PE detection on CTPA scans. Models such as EmbNet (Zhu et al, 2024) have demonstrated high sensitivity in both internal and external validations, achieving sensitivity rates of 88% to 90.9% per scan and 83.5% to 86% per embolus. Notably, these models can be enhanced by preprocessing techniques like vessel segmentation, which help focus detection efforts on pulmonary arteries, thereby reducing false positives—this has previously been applied to retinal vasculature (Zhu et al, 2024). Compared to traditional models, which often yield high false positive rates—up to 2.5 false positives per scan—AI-based systems have achieved notable reductions, bringing rates down to as low as 0.45 false positives per scan, as identified by Zhu et al (2024). This improvement not only aids radiologists in prioritising cases but also reduces the cognitive load associated with manually reviewing false positives.

However, despite these advancements, there are inherent challenges in DL models, particularly concerning the trade-off between sensitivity and positive predictive value (PPV). While models like EmbNet (Zhu et al, 2024) have optimised sensitivity to ensure high recall of PE cases, their relatively lower PPV (42.7%–48.4% per scan) highlights the need for radiologist oversight in balancing the detection of true positives against potential over-detection. Furthermore, as Lynch and Suriya (2022) portray in their study, CNN-based models often face difficulties in detecting peripheral emboli, with sensitivity rates for such cases ranging lower from 80.2% to 82.6%. This issue underscores the importance of developing AI models that can accurately detect both central and peripheral emboli without compromising sensitivity.

Combining CTPA and Non-Imaging Data via AI

A novel aspect of recent AI applications is their ability to integrate non-imaging data, such as raw ECG waveforms, alongside traditional clinical variables like CTPA data (Somani et al, 2021). This multi-modal approach appears to be more effective in detecting PE, especially in cases with moderate-to-high clinical suspicion, compared to conventional clinical scores. For example, in Somani et al (2021) cohort study, ECG waveform embeddings contributed significantly to model predictions, identifying subtle electrocardiographic markers that may signal early PE or predisposition to thrombosis, which might not be visible to clinicians using standard diagnostic tools. This aligns with existing literature suggesting that subtle cardiac abnormalities could correlate with higher PE risk (Abdulaal et al, 2024).

AI's ability to process complex data inputs from multiple sources has a clear efficiency advantage, potentially improving diagnostic accuracy and enhancing clinical workflows. By integrating diverse data types, these AI models can refine patient triage decisions, ensuring that those with the highest likelihood of PE are prioritised for imaging, while others may avoid unnecessary CTPA exposure. This not only helps reduce resource strain on imaging departments but also enables clinicians to make more targeted and data-informed decisions in complex cases where clinical symptoms alone may not be definitive.

Enhanced Diagnostic Specificity and Eliminating CTPA Overuse

One of the most critical contributions of AI in this space is its capacity to refine patient selection for CTPA, aiming to reduce unnecessary imaging. Traditional clinical scoring systems often overpredict the need for CTPA, leading to increased patient exposure to radiation and contrast agents, which is especially concerning in older populations. AI-driven models have demonstrated superior specificity, reducing reliance on CTPA and potentially improving the diagnostic efficiency of PE workups. For instance, fusion models that incorporate ECG and clinical data alongside AI have achieved higher specificity and discrimination power than clinical scores alone, effectively reducing unnecessary imaging in patients where PE is unlikely—a trend observed as early as 2021 (Soffer et al, 2021). Based on the current evidence, the use of AI-driven ECG and clinical data models could potentially allow for the selective omission of CTPA in patients with low pre-test probability for PE, such as those with low Wells scores and negative D-dimer results (Chen et al, 2020).

However, while these AI-enhanced diagnostic pathways have shown promise in reducing CTPA utilisation, they do introduce some risk of misdiagnosis if CTPA is omitted. Misdiagnosis rates vary across models and depend on patient selection criteria; however, misdiagnosis rates in studies employing AI-driven ECG fusion models have generally ranged from 3% to 8% in cases with a low clinical probability for PE (Somani et al, 2021). Although these rates are promising, they highlight the need for further validation before CTPA can be safely omitted in broader patient populations. Thus, while AI can potentially reduce the frequency of CTPA, careful patient selection and external validation of AI models are essential to avoid compromising diagnostic accuracy. This approach not only improves patient safety but also enhances resource allocation within clinical settings, addressing inefficiencies in diagnostic workflows. Future research should continue to evaluate the misdiagnosis risk across patient demographics to refine selection criteria and ensure safe, effective use of AI as an adjunct to traditional diagnostic methods.

Generalisability and External Validation

The generalisability of AI models across diverse clinical settings remains a critical concern. Although many AI-based systems demonstrate high accuracy in internal validations, their performance tends to decrease during external validations. For example, models like EmbNet experienced a modest decline in sensitivity (1.1–5.7%) and an increase in false positives (FPs) when validated externally, reflecting the discrepancies between controlled environments and real-world clinical contexts (Somani et al, 2021). These differences can be attributed to the presence of concomitant diseases such as atelectasis and pleural effusion, which were excluded from internal datasets but present in real-world cases (Abdulaal et al, 2024). Consequently, AI models trained on homogeneous datasets may struggle to replicate their performance in broader, more varied clinical populations. The performance of AI models must be evaluated across diverse patient populations, particularly as ML models infamously reflect the biases present in the data used to train them (Belkouchi et al, 2023).

Encouragingly, recent studies, such as Amin et al (2024), suggest that AI models for PE detection have shown consistent performance across different demographic groups, including varied racial and gender subgroups, which indicates a positive step towards reducing healthcare disparities.

Nonetheless, future AI models must undergo rigorous external validation with similarly diverse datasets to corroborate this, ensuring the models perform reliably outside controlled development datasets. Furthermore, datasets must include complex cases, such as post-pulmonary surgery patients and chronic PE cases. Expanding training datasets to include asymptomatic PE cases, which are often underrepresented due to selection bias in current CTPA use, a fact well-understood from the Soffer et al (2021) review, could enhance model sensitivity and mitigate the risk of false-negative readings. Without such advancements, the clinical utility of AI in PE detection may remain limited to high-risk and symptomatic patient cohorts.

Clinical Applications, Strengths, and Limitations of AI Models for PE Detection

The use of AI models for detecting PE in clinical settings shows significant promise across emergency, inpatient, and outpatient contexts, each benefiting from specific AI strengths while facing particular limitations.

In emergency settings, where rapid response is crucial, DL models, especially CNNs, can swiftly identify high-risk PE cases on CTPA, facilitating prompt treatment. Models such as EmbNet demonstrate high sensitivity, achieving detection rates of up to 88–90.9% per scan, helping prioritise critical cases (Zhu et al, 2024). However, the associated lower positive predictive value (PPV) contributes to FPs, requiring radiologists to review non-significant findings, which can increase workload. This trade-off underscores the need for optimising PPV without sacrificing essential sensitivity.

In inpatient scenarios, AI's ability to integrate clinical and imaging data supports targeted monitoring of high-risk patients, such as post-surgical or immobilised individuals. Models like the AdaBoost classifier with Wrapper feature selection outperform traditional scoring systems by improving accuracy and sensitivity, reducing unnecessary CTPA reliance and limiting patient exposure to radiation and contrast (Somani et al, 2021). However, patient heterogeneity within inpatient settings may challenge the consistency of AI performance, highlighting the need for external validation to maintain reliability across diverse patient profiles.

In outpatient and low-risk settings, NLP models offer an efficient, scalable option for identifying cases where immediate imaging may not be necessary. NLP systems, demonstrated by Swartz et al (2017) and Amin et al (2024), streamline PE management by extracting key terms from radiology reports and identifying low-risk cases suitable for outpatient follow-up (Chen et al, 2020). Although cost-effective and easily integrated, these models primarily target low-risk cases, limiting their generalisability for higher-risk patients.

The integration of CTPA with non-imaging data, such as ECG waveforms, represents a recent AI innovation. For instance, ECG embeddings in models like those studied by Somani et al (2021) help detect early PE indicators, refining patient selection for CTPA and reducing unnecessary imaging. However, incorporating such diverse data into routine workflows poses logistical challenges, emphasising the importance of robust clinical infrastructure.

Overall, each AI model displays unique strengths within its respective clinical context. Emergency and inpatient settings benefit from the sensitivity of CNNs and multi-modal models, while NLP is best suited for low-risk outpatient management. Rigorous validation across diverse datasets remains critical to optimising these models' effectiveness across patient populations.

Future Directions

The integration of AI in clinical respiratory practice, particularly for PE detection on CTPA, holds significant promise for enhancing diagnostic workflows and reducing the burden on radiologists. However, the continued development of AI applications in this field requires a strategic focus on hybrid and multimodal fusion models, which combine the complementary strengths of NLP and DL. Hybrid models that integrate NLP and CNN algorithms could harness the interpretability of NLP for analysing radiology reports alongside the advanced image recognition capabilities of CNNs, resulting in more comprehensive diagnostic systems. Such an approach could enhance model accuracy by enabling systems to synthesise both imaging data and textual insights, thus supporting a more holistic analysis of PE presentations.

To improve model robustness and generalisability, future research should also prioritise semi-supervised learning methods, as demonstrated in recent segmentation approaches included in this review. Semi-supervised learning could help AI models adapt to a wider array of datasets with limited labelling, which is especially pertinent for detecting PE given its often subtle radiological presentations. Another important avenue for enhancing model adaptability is the use of multicentre datasets for model training and validation. Validating models across multiple healthcare centres and varied populations would improve generalisability and ensure applicability across diverse clinical settings, increasing the feasibility of AI deployment in everyday practice.

Additionally, iterative advancements in algorithmic architecture should be explored to improve diagnostic sensitivity and reduce false positives, especially as PE detection requires distinguishing between central and peripheral emboli—tasks that can be challenging for traditional algorithms. Thus, the development of more sophisticated segmentation algorithms and fusion models is essential for AI to reach a clinically reliable standard, paving the way for AI models that can be adapted across multiple diagnostic workflows and patient demographics.

Operational and Ethical Considerations

The application of AI in medical imaging also necessitates addressing crucial operational, ethical, and data privacy concerns to ensure sustainable and responsible integration within healthcare systems. Patient privacy, for example, remains paramount; AI models require access to substantial amounts of clinical data, which introduces potential vulnerabilities. Healthcare systems implementing AI tools must establish stringent data security protocols, incorporating anonymisation and secure data storage practices, to protect patient confidentiality and comply with regulations such as the General Data Protection Regulation (GDPR). Furthermore, transparency in data handling processes can help build trust with both patients and clinicians, facilitating AI adoption.

Another key consideration is mitigating data bias, which can arise if models are trained on datasets that do not adequately represent diverse populations. Data bias not only limits generalisability but also poses ethical risks, as biased models may deliver suboptimal or inaccurate diagnoses for underrepresented groups. To address this, future AI models should be trained and validated on inclusive datasets that reflect the diversity of real-world patient populations. This effort could be bolstered by collaborations between institutions to aggregate data from varied sources, thus reducing model biases and ensuring fairness and equity in AI-assisted diagnosis.

Finally, the adaptability and feasibility of AI within various healthcare systems require careful planning. Many AI tools, including those discussed in this review, are currently in the early stages of development and may require substantial computational resources and technical infrastructure to be integrated effectively. Resource limitations, particularly in under-resourced healthcare systems, may affect the feasibility of widespread AI adoption. Developing AI models that are both computationally efficient and scalable can help mitigate these challenges, as can training programmes to enhance clinician understanding and acceptance of AI tools. By proactively addressing these ethical, privacy, and feasibility concerns, AI applications in PE detection and beyond can be deployed responsibly and effectively, maximising clinical impact while safeguarding patient rights and promoting equitable healthcare outcomes.

Limitations of Studies

Despite the promising results, there remain several limitations to the current AI models for PE detection. One key issue is the retrospective nature of many studies, where data timeliness and completeness can influence model accuracy. Additionally, models tend to perform well in patients with moderate-to-high suspicion of PE but require further refinement to address low-risk populations that are often excluded from CTPA. Future work should focus on improving model calibration for patients with low clinical suspicion, enabling better decision-making across the full spectrum of PE risk. Moreover, embedding institutional or provider-specific diagnostic biases within models poses another challenge. Continued refinement, feature ablation studies, and validation against more diverse datasets will be vital to ensure the robustness and ethical deployment of AI in real-world settings. Given the marked heterogeneity of studies, it was not possible to perform a pooled analysis of all outcomes for a particular AI modality.

Limitations of Review

This systematic review is subject to several inherent limitations that must be carefully considered when interpreting its findings. Given the novelty of the AI models investigated, this review identified only 12 studies, all of which were proof-of-concept or pilot studies. A significant limitation is the small sample sizes present across these studies, with the majority (>6 studies) having fewer than 100 cases each. This limitation in sample size may increase the propensity for both type I and type II errors, thereby limiting the statistical power and generalisability of the review's conclusions.

Additionally, the generalisability of these findings is further constrained by data biases that exist in the reviewed studies. These biases stem from imbalanced datasets, where certain patient demographics or PE characteristics are underrepresented. Such biases may affect the performance and applicability of AI models when applied to a broader, more diverse clinical population, as models trained on biased datasets may not generalise well across varying patient presentations. Moreover, the majority of the studies included were retrospective, introducing a time lag that may not reflect AI's most recent advancements and real-time capabilities in CTPA-based PE detection.

To address these limitations, future studies should focus on increasing sample sizes and ensuring dataset diversity to reduce data bias and enhance generalisability. Prospective studies, rather than retrospective analyses, are recommended to provide a more current and realistic evaluation of AI's clinical utility. As AI in this field is advancing rapidly, particular attention should be given to more recent studies within the review's time horizon, as these are likely to offer the most relevant insights into the present capabilities of AI for PE detection on CTPA.

Conclusion

The integration of AI into PE detection on CTPA scans presents significant advancements in diagnostic accuracy, with both NLP and DL models showing distinct strengths. NLP models are particularly beneficial for streamlining the management of low-risk patients, while DL models offer improved sensitivity and reduced FP rates, positioning them as valuable tools for radiologists. However, limitations related to generalisability, the trade-off between sensitivity and PPV, and the practical challenges of operational integration must be addressed to fully realise the potential of AI in this context. Future research should prioritise hybrid models that combine the strengths of NLP and CNNs, alongside rigorous external validation across diverse patient populations, to maximise the clinical utility of AI in PE detection.

Key Points

- The review collectively highlights significant advancements in the diagnosis and management of pulmonary embolism (PE), via artificial intelligence (AI) and machine learning techniques. There is clear evidence supporting the effectiveness of using deep learning algorithms in improving the accuracy and efficiency of PE detection through CT pulmonary angiography (CTPA).
- The integration of natural language processing (NLP) tools is emphasised for enhancing clinical workflows and supporting nationwide surveillance of venous thromboembolism, ultimately streamlining diagnostic processes.
- Research also addresses the importance of developing reliable prediction models and diagnostic tools, aiming to optimise patient outcomes while mitigating challenges such as overtesting and inconsistencies in diagnostic approaches.
- Collectively, these advancements signify a notable shift towards utilising AI to support clinicians in more effectively managing and diagnosing pulmonary embolism via CTPA.

Availability of Data and Materials

Data collection form and search results are available on enquiry to the corresponding author.

Author Contributions

MSTAH, MAME, VS, HR: conception, manuscript writing, revision, visualisation, data collection, data analysis, validation. AAR, GM, JM: conception, manuscript writing, revision, supervision, validation. MSTAH and MAME contributed equally to the manuscript. All authors contributed to important editorial changes of important content in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgement

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

Arian Arjomandi Rad is serving as one of the Editorial Board Members of this journal. We declare that Arian Arjomandi Rad had no involvement in the review of this article and has no access to information regarding its review. Other authors declared no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://www.magonlinelibrary.com/doi/suppl/10.12968/hmed.2024.0757.

References

Abdulaal L, Maiter A, Salehi M, Sharkey M, Alnasser T, Garg P, et al. A systematic review of artificial intelligence tools for chronic pulmonary embolism on CT pulmonary angiography. Frontiers in Radiology. 2024; 4: 1335349. https://doi.org/10.3389/fradi.2024.1335349

Amin KD, Weissler EH, Ratliff W, Sullivan AE, Holder TA, Bury C, et al. Development and Validation of a Natural Language Processing Model to Identify Low-Risk Pulmonary Embolism in Real Time to Facilitate Safe Outpatient Management. Annals of Emergency Medicine. 2024; 84: 118–127. https://doi.org/10.1016/j.annemergmed.2024.01.036

Belkouchi Y, Lederlin M, Ben Afia A, Fabre C, Ferretti G, De Margerie C, et al. Detection and quantification of pulmonary embolism with artificial intelligence: The SFR 2022 artificial intelligence data challenge. Diagnostic and Interventional Imaging. 2023; 104: 485–489. https://doi.org/10.1016/j.diii.2023.05.007

Chen C, Millar FR, Jones A. Outpatient management of pulmonary emboli: when to ambulate. British Journal of Hospital Medicine. 2020; 81: 1–10. https://doi.org/10.12968/HMED.2019.0370

Cheng TW, Chua YW, Huang CC, Chang J, Kuo C, Cheng YC. Feature-enhanced adversarial semi-supervised semantic segmentation network for pulmonary embolism annotation. Heliyon. 2023; 9: e16060. https://doi.org/10.1016/j.heliyon.2023.e16060

Curcio R, Gandolfo V, Alcidi R, Giacomino L, Campanella T, Casarola G, et al. Vaccine-induced massive pulmonary embolism and thrombocytopenia following a single dose of Janssen Ad26.COV2.S vaccination. International Journal of Infectious Diseases. 2022; 116: 154–156. https://doi.org/10.1016/j.ijid.2021.12.345

Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. A guide to deep learning in healthcare. Nature Medicine. 2019; 25: 24–29. https://doi.org/10.1038/S41591-018-0316-Z

Exarchos K, Aggelopoulou A, Bartziokas K, Tsina E, Tagkas C, Drouvis V, et al. Machine learning algorithm for the diagnosis of pulmonary embolism: a proof of concept study. European Respiratory Journal. 2020; 56: 4972. https://doi.org/10.1183/13993003.CONGRESS-2020.4972

Hirsh J, Guyatt G, Albers GW, Harrington R, Schünemann HJ. Antithrombotic and thrombolytic therapy: American College of Chest Physi-cians evidence-based clinical practice guidelines (8th Edition). Chest. 2008; 133: 110S–112S. https://doi.org/10.1378/chest.08-0652

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nature Reviews Cancer. 2018; 18: 500–510. https://doi.org/10.1038/s41568-018-0016-5

Huang SC, Kothari T, Banerjee I, Chute C, Ball RL, Borus N, et al. PENet-a scalable deep-learning model for automated diagnosis of pulmo-nary embolism using volumetric CT imaging. NPJ Digital Medicine. 2020a; 3: 61. https://doi.org/10.1038/s41746-020-0266-y

Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Scientific Reports. 2020b; 10: 22147. https://doi.org/10.1038/s41598-020-78888-w

Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, et al. Antithrombotic therapy for VTE disease: CHEST guideline and expert panel report. Chest. 2016; 149: 315–52. https://doi.org/10.1016/j.chest.2015.11.026

Lakhani P. Deep convolutional neural networks for endotracheal tube position and X-ray image classification: Challenges and opportunities. Journal of Digital Imaging. 2017; 31: 356–65. https://doi.org/10.1007/s10278-017-9980-7

Liu W, Liu M, Guo X, Zhang P, Zhang L, Zhang R, et al. Evaluation of acute pulmonary embolism and clot burden on CTPA with deep learn-ing. European Radiology. 2020; 30: 3567–3575. https://doi.org/10.1007/s00330-020-06699-8

Lynch D, Suriya M. PE-DeepNet: A deep neural network model for pulmonary embolism detection. International Journal of Intelligent Net-works. 2022; 3: 176–180. https://doi.org/10.1016/J.IJIN.2022.10.001

Raptis CA, Steinbrecher KL, Marquis KM. Perspectives on the Evolution of Diagnosis and Management of Acute Pulmonary Embolism. Ca-nadian Association of Radiologists Journal. 2022; 73: 17–18. https://doi.org/10.1177/08465371211026530

Shapiro J, Reichard A, Muck PE. New Diagnostic Tools for Pulmonary Embolism Detection. Methodist DeBakey Cardiovascular Journal. 2024; 20: 5–12. https://doi.org/10.14797/mdcvj.1342

Soffer S, Klang E, Shimon O, Barash Y, Cahan N, Greenspana H, et al. Deep learning for pulmonary embolism detection on computed tomog-raphy pulmonary angiogram: a systematic review and meta-analysis. Scientific Reports. 2021; 11: 15814. https://doi.org/10.1038/s41598-021-95249-3

Somani SS, Honarvar H, Narula S, Landi I, Lee S, Khachatoorian Y, et al. Development of a machine learning model using electrocardiogram signals to improve acute pulmonary embolism screening. European Heart Journal. Digital Health. 2021; 3: 56–66. https://doi.org/10.1093/ehjdh/ztab101

Swartz J, Koziatek C, Theobald J, Smith S, Iturrate E. Creation of a simple natural language processing tool to support an imaging utilization quality dashboard. International Journal of Medical Informatics. 2017; 101: 93–99. https://doi.org/10.1016/j.ijmedinf.2017.02.011

Tajbakhsh N, Shin JY, Gotway MB, Liang J. Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation. Medical Image Analysis. 2019; 58: 101541. https://doi.org/10.1016/j.media.2019.101541

Tang H, Chen X, Liu Y, Lu Z, You J, Yang M, et al. Clinically applicable deep learning framework for organs at risk delineation in CT images. Nature Machine Intelligence. 2019; 1: 480–491. https://doi.org/10.1038/s42256-019-0099-z

Vyas V, Sankari A, Goyal A. Acute Pulmonary Embolism. 2024. Available at: https://www.ncbi.nlm.nih.gov/books/NBK560551/ (Accessed: 8 October 2024).

Wang R, Chaudhari P, Davatzikos C. Bias in machine learning models can be significantly mitigated by careful training: Evidence from neu-roimaging studies. Proceedings of the National Academy of Sciences of the United States of America. 2023; 120: e2211613120. https://doi.org/10.1073/pnas.2211613120

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. European Radiology. 2020; 30: 6545–6553. https://doi.org/10.1007/s00330-020-06998-0

Wichmann JL, Hu X, Kerl JM, Schulz B, Frellesen C, Bodelle B, et al. 70 kVp computed tomography pulmonary angiography: potential for reduction of iodine load and radiation dose. Journal of Thoracic Imaging. 2015; 30: 69–76. https://doi.org/10.1097/RTI.0000000000000124

Wittenberg R, Peters JF, Weber M, Lely RJ, Cobben LPJ, Prokop M, et al. Stand-alone performance of a computer-assisted detection prototype for detection of acute pulmonary embolism: a multi-institutional comparison. The British Journal of Radiology. 2012; 85: 758–764. https://doi.org/10.1259/bjr/26769569

Zhu H, Tao G, Jiang Y, Sun L, Chen J, Guo J, et al. Automatic detection of pulmonary embolism on computed tomography pulmonary angio-gram scan using a three-dimensional convolutional neural network. European Journal of Radiology. 2024; 177: 111586. https://doi.org/10.1016/j.ejrad.2024.111586

Appendix

See Table 2.

Table 2a. Inclusion and exclusion criteria for review.

1.	((pulmonary embol* or PE or lung embol* or respiratory embol* or respiratory clot* or pulmonary clot* or lung clot* ('*' indicates a shortening for a search term)) and (Artificial Intelligence or artificial intelligence or computational intelligence or machine intelligence or ambient intelligence or computer reasoning or automated reasoning or machine reasoning or computer inference or automated inference or machine inference or computer vision system or multi-criteria decision or multicriteria decision or multiple criteria decision or machine learning or transfer learning or deep learning or hierarchical learning or semi-supervised learning or semi supervised learning or semisupervised learning or support vector or heuristic or hyperheuristic or metaheuristic or expert system or fuzzy logic or knowledge base or natural language processing or neural network or perceptron or connectionist model or big data or computer algorithm)).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kf, fx, dq, bt, nm, ox, px, rx, ui, sy, ux, mx]
2.	computed tomography pulmonary angiogram or CTPA or CT angio* or CT or CT scan* or CT imag*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kf, fx, dq, bt, nm, ox, px, rx, ui, sy, ux, mx]
3.	1 and 2
4.	limit 3 to english language

Table 2b. Inclusion and exclusion criteria for review.

Criteria	Inclusion	Exclusion
Study Type	Original articles	Reviews, editorials, case reports, abstracts from meetings, preclinical studies
Focus of Study	Use of artificial intelligence and/or machine learning for PE detection via CTPA scans	Studies not reporting AI or ML for PE detection
Methodology	Machine learning models that use CTPA as a main or supportive modality	Studies with inconsistencies in data that impede data extraction
Geographical Restrictions	No geographical restrictions	N/A
Population	Studies involving human participants	Animal models
Language	Articles published in English	Articles published in languages other than English
Data Consistency	Data presented is consistent and suitable for extraction	Inconsistent data that impedes extraction