SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

Estimating the Causal Effect of Photon counting detector CT on AI Performance in an Outpatient Oncology Setting Using Counterfactual Comparisons with Dual Energy CT

Yin Xi, PhD, UT Southwestern Medical Center; Kiran Batra, MD; Michael Long, MS; Xinhui Duan, PhD; Ron Peshock, MD

Introduction/Background

A high-pitch, photon counting detector CT (PCD-CT) has been shown to yield different image characteristics when compared to dual energy CT (DECT). Characteristics of CT scan images can potentially affect the performance of an AI algorithm. We investigated the performance of an AI algorithm for detection of pulmonary embolism in patients scanned using PCD-CT using counterfactual comparisons with Dual Energy CT.

Methods/Intervention

All CTPA exams performed between 2022-12-21 and 2025-05-14 at the outpatient cancer center, where the PCD-CT is located, were included. A DECT scanner (Force, Siemens) is also available at the same location. Structured reports and vendor NLP served as reference standards. Observed sensitivity, specificity, and CT dose on the PCD-CT were compared to the counterfactual performance, had the same cohort been scanned on DECT, using inverse probability of treatment weighting (IPTW). The estimated differences reflect the average effect of treatment (ATT). IPTW was estimated based on the probability of being scanned on PCD-CT given age, sex, BMI, race/ethnicity and exam priority. Bootstrap 95% confidence intervals were calculated for the absolute percentage point difference in sensitivity and specificity, as well as the percent difference in average CTDIvol.

Results/Outcome

Overall, 756 CTPA exams were included in the analysis, with 345 scanned on PCD-CT and 411 scanned on DECT (Table 1). PCD-CT had lower sensitivity of AI-PE detection by 26 percentage points (95% bootstrap CI: 7%-45%) compared to DECT; while specificity was not significantly affected (Table 2). The percentage reduction in average radiation dose in CTDIvol in PCD-CT Compared to DECT exam was 30.9% mGy (95% CI: 25.9%-35.5%).

Conclusion

The AI PE algorithm showed inferior sensitivity in CTPA exams performed on PCD-CT compared to DECT, with comparable specificity. A significant reduction in radiation dose was observed with PCD-CT.

Statement of Impact

The performance of commercial AI image algorithms may degrade with new imaging technologies.

Characteristic	Unweighted			with IPTW		
	DECT N = 411[1]	PCCT N = 345[1]	p-value[2]	DECT N = 345[1]	PCCT N = 345[1]	p-value[3]
age_group			0.7			>0.9
18-39	59/411 (14%)	46/345 (13%)		46/345 (13%)	46/345 (13%)	
40-64	193/411 (47%)	155/345 (45%)		155/345 (45%)	155/345 (45%)	
65+	159/411 (39%)	144/345 (42%)		143/345 (42%)	144/345 (42%)	
Sex			0.4			>0.9
Female	251/411 (61%)	222/345 (64%)		220/345 (64%)	222/345 (64%)	
Male	160/411 (39%)	123/345 (36%)		124/345 (36%)	123/345 (36%)	
BMI_group			0.2			>0.9
<18.5	20/411 (4.9%)	9/345 (2.6%)		9/345 (2.7%)	9/345 (2.6%)	
18.5-24.9	140/411 (34%)	103/345 (30%)		103/345 (30%)	103/345 (30%)	
25-29.9	122/411 (30%)	106/345 (31%)		106/345 (31%)	106/345 (31%)	
30+	129/411 (31%)	127/345 (37%)		127/345 (37%)	127/345 (37%)	
RaceEthnicity			>0.9			>0.9
Hispanic	47/411 (11%)	41/345 (12%)		40/345 (12%)	41/345 (12%)	
NHW	235/411 (57%)	198/345 (57%)		199/345 (58%)	198/345 (57%)	
NHB	55/411 (13%)	51/345 (15%)		52/345 (15%)	51/345 (15%)	
Asian	27/411 (6.6%)	20/345 (5.8%)		20/345 (5.7%)	20/345 (5.8%)	
Other/Mixed/Unknown	47/411 (11%)	35/345 (10%)		34/345 (9.9%)	35/345 (10%)	
trueExamPriority			0.2			>0.9
STAT	154/411 (37%)	149/345 (43%)		149/345 (43%)	149/345 (43%)	
URGENT	43/411 (10%)	40/345 (12%)		39/345 (11%)	40/345 (12%)	
ROUTINE	214/411 (52%)	156/345 (45%)		157/345 (45%)	156/345 (45%)	

[1] n/N (%)

[2] Pearson's Chi-squared test

[3] Pearson's X^2: Rao & Scott adjustment

Table 1. Patient characteristics by scanner, before and after inverse probability of treatment weighting (IPTW).

Endpoints	Unweighted			with IPTW		
	DECT N = 411[1]	PCCT N = 345[1]	p-value[2]	DECT N = 345[1]	PCCT N = 345[1]	p-value[3]
Sensitivity	39/44 (89%)	21/34 (62%)	0.005	34/38 (88%)	21/34 (62%)	0.008
Specificity	363/367 (99%)	308/311 (99%)	>0.9	303/306 (99%)	308/311 (99%)	>0.9
CTDIvol Body (mGy) per exam	11.4 ± 5.8	8.2 ± 3.6	<0.001	11.9 ± 5.9	8.2 ± 3.6	<0.001

[1] n/N (%); Mean ± SD

[2] Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum test

[3] Pearson's X^2: Rao & Scott adjustment; Design-based KruskalWallis test

Table 2. Comparison of scanner sensitivity, specificity, and CT dose, before and after inverse probability of treatment weighting (IPTW).

Keywords

Photon counting detector CT; dual-energy CT; artificial intelligence; pulmonary embolism; diagnostic performance