



First Evaluation of AI-Assisted Preliminary Report Drafting in Abdomen–Pelvis CT: Effects on Agreement, Confidence, and Mental Demand

Pranav Rajpurkar, PhD, a2z Radiology AI; Deepanshu Jindal; Julian N. Acosta, MD; Siddhant Dogra, MD; Pooja Jain, MD; Quirin D. Strotzer, MD; Lukas Müller, MD, PhD; Lucas Bijmens, MD; Srivardhan Vanka, MD, MBBS; Viswanathan Murugaiah, MD; Su Hwan Kim, MD; Divya Buchireddygar, MBBS; João Abrantes, MD; Michael Moritz, MD; Jenanan Vairavamurthy, MD; John Fahrner, MD; Samir Rajpurkar, MS

Introduction/Background

Abdomen–pelvis CT (CTAP) is one of the highest-volume CT examinations in the United States. Interpreting these volumetric studies and generating complete reports remains time-intensive. We built an integrated AI system that detects 63 findings and auto-generates preliminary structured text. To our knowledge, this is the first study evaluating AI-assisted preliminary report drafting for volumetric CT interpretation—specifically abdomen–pelvis CT—and its impact on pathology-level agreement, reader confidence, and perceived mental demand.

Methods/Intervention

An integrated AI system, trained on over 280,000 CTAP exams, was developed to detect 63 CTAP findings and auto-generate preliminary draft text within a structured report template. Seven radiologists each interpreted 50 CTAP cases containing at least one acute finding. Readers completed their assignments at their own pace within the study window. For each reader, the case sequence alternated in five-case blocks between AI-assisted drafting and standard (unassisted) drafting (block size = 5). The initial condition was randomized per reader to mitigate systematic order effects. After editing each report, readers rated mental demand and confidence on 3-point Likert scales. Agreement was assessed at the pathology level. For each acute pathology within a case, the reader's determination (present/absent) was compared to a consensus for that pathology derived from the remaining readers' reports. Positive percent agreement (PPA) and negative percent agreement (NPA) between assisted and unassisted conditions were compared using chi-square tests. Confidence and mental demand scores were compared using mixed-effects models with random intercepts for reader and case to account for intra-reader and intra-case correlation.

Results/Outcome

AI assistance increased PPA (0.86 vs 0.63; χ^2 test, $p < 0.001$), while NPA remained comparable (0.98 vs 0.97; χ^2 test, $p > 0.05$). Confidence scores were higher with assistance (mean 2.63 vs 2.27; mixed-effects model, $p < 0.001$), and perceived mental demand was lower (mean 1.51 vs 1.94; mixed-effects model, $p < 0.001$). Overall pathology-level agreement with consensus across 16 acute conditions did not differ significantly between conditions ($p > 0.05$).

Conclusion

AI-assisted CTAP reports improve confidence and decrease mental demand while improving agreement for present acute pathologies.

Statement of Impact

AI-assisted report generation can help radiologists meet workflow demands.

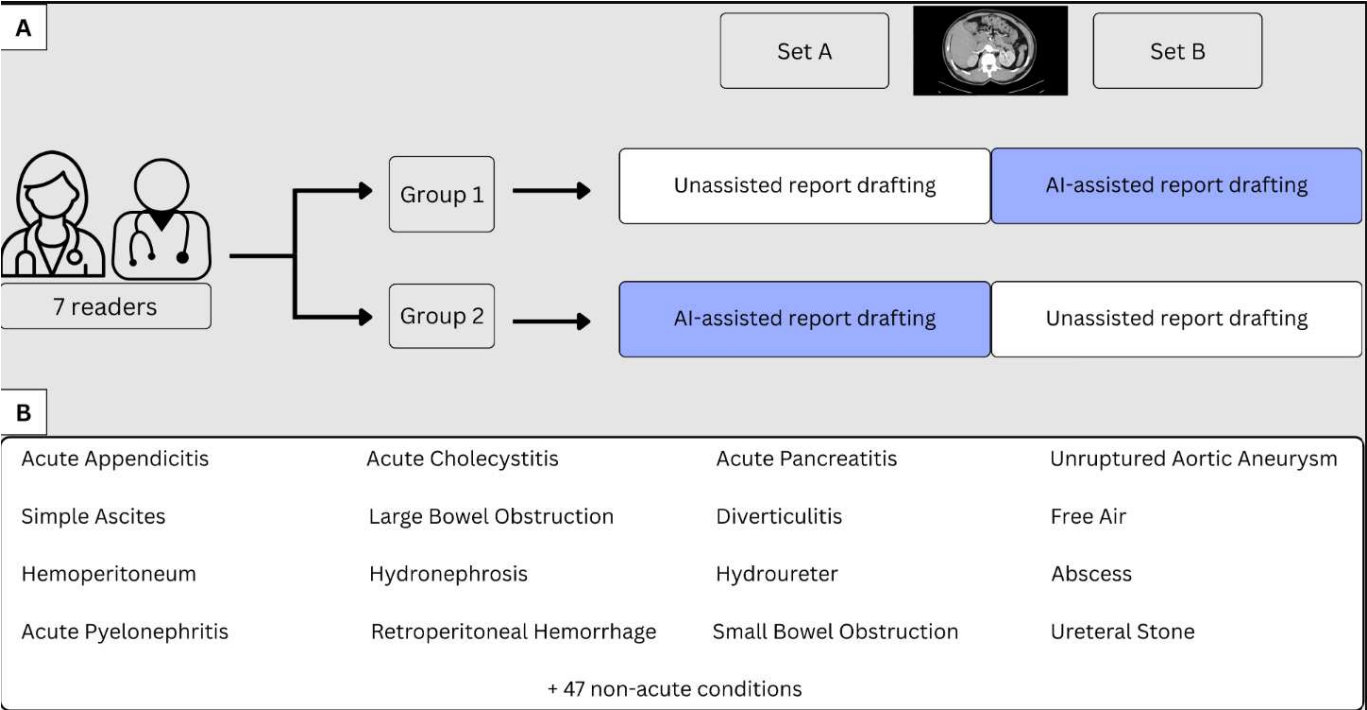


Figure 1. Study workflow. Seven radiologists evaluated 50 abdomen-pelvis CT exams, alternating in five-case blocks between AI-assisted drafts and unassisted drafts, with the initial condition randomized per reader (A). Each case included at least one of a select set of 16 acute pathologies (B); AI-assisted reports detected these 16 findings plus 47 non-acute findings.

Keywords

Artificial Intelligence; AI-Assisted Report Drafting; CT Abdomen Pelvis; Radiologist Workflow