Abdominal Radiology
https://doi.org/10.1007/s00261-024-04578-0

RESEARCH

Quality assessment of expedited AI generated reformatted images for ED acquired CT abdomen and pelvis imaging

Daniel Freedman[1] · Barun Bagga[1] · Kira Melamud[1] · Thomas O'Donnell[2] · Emilio Vega[1] · Malte Westerhoff[3] · Bari Dane[1]

Received: 8 July 2024 / Revised: 6 September 2024 / Accepted: 9 September 2024

Abstract

Purpose Retrospectively compare image quality, radiologist diagnostic confidence, and time for images to reach PACS for contrast enhanced abdominopelvic CT examinations created on the scanner console by technologists versus those generated automatically by thin-client artificial intelligence (AI) mechanisms.

Methods A retrospective PACS search identified adults who underwent an emergency department contrast-enhanced abdominopelvic CT in 07/2022 (Console Cohort) and 07/2023 (Server Cohort). Coronal and sagittal multiplanar reformatted images (MPR) were created by AI software in the Server cohort. Time to completion of MPR images was compared using 2-sample t-tests for all patients in both cohorts. Two radiologists qualitatively assessed image quality and diagnostic confidence on 5-point Likert scales for 50 consecutive examinations from each cohort. Additionally, they assessed for acute abdominopelvic findings. Continuous variables and qualitative scores were compared with the Mann-Whitney U test. A $p < .05$ indicated statistical significance.

Results Mean[SD] time to exam completion in PACS was 8.7[11.1] minutes in the Console cohort ($n = 728$) and 4.6[6.6] minutes in the Server cohort ($n = 892$), $p < .001$. 50 examinations in the Console Cohort (28 women 22 men, 51[19] years) and Server cohort (27 women 23 men, 57[19] years) were included for radiologist review. Age, sex, CTDIvol, and DLP were not statistically different between the cohorts (all $p > .05$). There was no significant difference in image quality or diagnostic confidence for either reader when comparing the Console and Server cohorts (all $p > .05$).

Conclusion Examinations utilizing AI generated MPRs on a thin-client architecture were completed approximately 50% faster than those utilizing reconstructions generated at the console with no statistical difference in diagnostic confidence or image quality.

Keywords Image reformatting · Emergency radiology · CT abdomen/pelvis imaging · Radiology workflow

✉ Daniel Freedman
daniel.freedman@nyulangone.org

Barun Bagga
barun.bagga@nyulangone.org

Kira Melamud
kira.melamud@nyulangone.org

Thomas O'Donnell
tom.odonnell@siemens-healthineers.com

Emilio Vega
emilio.vega@nyulangone.org

Malte Westerhoff
mwesterhoff@visageimaging.com

Bari Dane
bari.dane@nyulangone.org

1 New York University Langone Medical Center, New York, USA

2 Siemens Healthineers (United States), Malvern, USA

3 Visage Imaging (Germany), Berlin, Germany

Published online: 18 September 2024

Introduction

Spiral CT permits the direct reconstruction of images in any orientation (e.g., coronal, sagittal, oblique) from raw data *at the scanner console* (Workstream4D™ Siemens Healthineers, Forchheim, Germany). While these images have very high quality, and can be pushed directly to the Picture Archiving and Communication System (PACS) from the scanner, they are created on the console using the scanner's Image Reconstruction System (IRS). Unfortunately, this process is time consuming. Time spent creating one study's (optional) raw data reformatted images takes resources away from another study's creation of (primary) axial images. Additionally, this is time consuming for the technologist initiating the process and may also delay subsequent patients from being scanned, particularly if there is only one scanner console [1].

When throughput is not an issue, reconstructing using raw data is generally preferable. However, time is a crucial factor in the workup and ultimate treatment of critically ill emergency department (ED) patients. CT must provide a fast and effective diagnostic assessment of patients presenting with emergent symptoms. In addition to prompt imaging, in the ED setting it is critical for images to reach PACS as quickly as possible to ensure timely diagnosis of potentially acute findings [1–6].

We propose to significantly speed up image throughput by *automating* image space reformatting on a remote server using Artificial Intelligence (AI) techniques by leveraging two new technologies. Used together, they have the potential to replace raw data reconstruction at the console when time is of the essence. "Auto Post-Process and Archive" technology (Rapid Results™, Siemens Healthineers, Forchheim, Germany) is configured as part of a scan protocol to automatically push a specified DICOM series to a post-processing server with instructions on how to process the images (e.g., reformat to a sagittal series) and also automatically push the results from the post-processing server to PACS. "Landmark Detection" technology (ALPHA™ - Automatic Landmark Parsing of Human Anatomy, Siemens Healthineers) is capable of employing AI to determine the coronal and sagittal planes *with respect to the patient* (as opposed to the scanner coordinate system) [7, 8]. This is particularly valuable for patients not aligned with the table because they are unable to lie flat. Through the detection of anatomic landmarks, AI determines where the optimal sagittal and coronal planes are for each particular patient.

Automating image space reformatting using the approach described above has the potential to allow images to reach PACS faster for imaging diagnosis, while reducing the time technologists spend generating reconstructions. Prior studies have shown that generating reconstructions from a separate console can facilitate increased throughput in the ED, particularly for trauma imaging [1–6, 9]. A prior study showed improved throughput using an AI tool for generating reconstructions in trauma pan-scans [6]. However, these prior studies did not assess image quality, diagnostic confidence, or diagnostic ability using artificial intelligence generated reconstructions. Therefore, the purpose of our study was to retrospectively compare the image quality, diagnostic confidence, and time for images to reach PACS for contrast enhanced abdominopelvic CT examinations by reconstructing sagittal and coronal images at the console versus employing AI to automatically acquire these reformats in image space on a remote server.

Methods

Patients

This study was Institutional Review Board approved and Health Insurance Portability and Accountability Act compliant. A retrospective PACS search identified adults (≥18 years) who underwent a CT Abdomen and Pelvis with Oral and IV contrast in the ED from 07/01/2022 through 07/31/2022 and 07/01/2023 through 07/31/2023. No patients were excluded for time analysis. For image quality assessment, patients with extensive surgical hardware, significant motion-related artifact, or lack of visible positive oral contrast were excluded. Clinical indication, age, and sex were recorded from the electronic medical record.

Imaging technique

CT Abdomen and Pelvis examinations were performed with positive oral and intravenous contrast during the portal venous phase on a dual-source dual-energy CT (SOMATOM Definition Force, Siemens Healthineers, Forchheim, Germany). Images were acquired in the axial plane from the xiphoid through the pubic symphysis during the portal venous phase 70-90 s after the administration of weight-based intravenous contrast (Isovue 300 mg Iodine per milliliter, Bracco Diagnostics, Princeton NJ) at a rate of 3–4 cc per second.

For CT examinations performed in 2022 all reconstructions were created by a CT technologist at the ED CT scanner console (Console cohort). This includes 4 mm axial, 3 mm coronal, 3 mm sagittal, and 0.75 mm axial images from the xiphoid to the pubic symphysis. For CT examinations performed in 2023, the axial images (0.75 mm and 4 mm slice thickness) were created by the technologist whereas the coronal and sagittal reformatted images were generated using automatic image space reformatting (Server cohort) from the 0.75 mm thin axial slices.

The CT dose index volume (CTDIvol) and dose length product (DLP) were recorded.

Server-based Reformatting

ALPHA™ was employed to automatically determine the patient-centric coronal and sagittal planes. This technology relies on a redundant set of anatomy detectors which are trained to learn local appearance cues. These detectors capture group-wise spatial configurations and together are highly robust to even significant distortion or occlusion. In addition, these detectors are invoked in a hierarchical fashion which significantly improves overall accuracy [7, 8].

Reconstruction Time determination

The time from topogram to coronal reconstruction generation and from topogram to sagittal reconstruction creation were obtained from PACS (Visage Imaging, Inc., San Diego, CA, United States) DICOM information for all examinations in the cohort. This time incorporates scan range selection, intravenous contrast injection, CT acquisition, reformat generation, and time for images to reach PACS for imaging review. Venous access was established for all patients prior to topogram acquisition and is therefore not reflected in the time. Contrast injection delay was 70–90 s for most patients imaged in this cohort.

Image Quality Assessment

Two fellowship-trained abdominal radiologists with 7 and 3 years of post-fellowship experience retrospectively evaluated deidentified and anonymized CT abdomen and pelvis examinations on PACS, blinded to the method of multiplanar reformat, date of acquisition, and imaging report. Scan labels and markings, including the date of examination acquisition, were removed before images were sent to research PACS for imaging review. The sequence of scans to review was assigned by a random number generator. Readers scored the image quality for axial, coronal, and sagittal images separately on a 5-point Likert scale (1 = nondiagnostic, 2 = poor – significant limitation in diagnostic ability, 3 = acceptable – moderate limitation in diagnostic quality, 4 = good – mild limitation in diagnostic quality, 5 = no significant limitation in diagnostic quality). Readers scored diagnostic confidence for axial, coronal, and sagittal images separately on a 5-point Likert scale (1 = very unconfident, 2 = slightly unconfident, 3 = neutral, 4 = slightly confident, 5 = very confident). Additionally, readers recorded the presence or absence of acute findings as well as any explanation for abdominal pain. 50 scans were included in each cohort for retrospective radiologist review such that the study would be powered to detect a difference of 0.3 in scores.

Statistical analysis

Continuous variables including patient age, CTDIvol, and DLP were compared between the Console and Server cohorts using the Mann-Whitney U test, and sex was compared using Pearson Chi-Square. Time to completion of multiplanar reformat creation was compared between the cohorts using 2-sample t-tests. Qualitative scores for image quality and diagnostic confidence were compared between the Console and Server cohorts using the Mann-Whitney U test. Axial images were reconstructed on the console by the IRS in both cohorts; however, the coronal and sagittal reformats varied in their reconstruction method. Since axial images in both cohorts were generated on the IRS, this allowed the axial scores from each reader to serve as an internal control for comparison with the coronal and sagittal qualitative evaluations, and were compared using the Wilcoxon Signed-Rank Test. In other words, comparing the axial image quality between the cohorts allowed for a comparison of technologist related factors so that any difference in coronal or sagittal image quality could be attributed to the reconstruction method. Presence of acute findings and explanation for abdominal pain were compared between the cohorts using the Pearson Chi-Square. Cohen's kappa was used for inter-reader agreement for the presence of acute findings and explanation for abdominal pain. Data from the two readers were analyzed individually. The significance level was set at 0.05. Kappa value of ≤ 0 is considered no agreement; 0.01–0.20, none to slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and ≥ 0.81 is considered almost perfect agreement [10]. The analysis was performed using SPSS statistical software (SPSS Statistics Version 28, IBM).

Results

Patients

728 CT abdomen and pelvis examinations on an ED CT scanner were identified from 07/01/2022 through 07/31/2022 and 892 were identified from 07/01/2023 through 07/31/2023. All of these examinations were included for reconstruction time analysis.

For image quality assessment in the 2022 Console cohort, 50 consecutive patients were selected after 12 patients without oral contrast and 1 patient with excessive motion were excluded. In the 2023 Server cohort, 50 consecutive patients

Table 1 Patient demographics

	Console Cohort	Server Cohort	*p*-value
Age (years)	51[19]*	57[19]*	0.08
Sex	28 women, 22 men	27 women, 23 men	0.84
CTDIvol (mGy)	13.4[7.2]*	13.9[7.7]*	0.87
DLP (mGycm)	674.7[397.5]*	697.9[393.6]*	0.74
Frequency of acute findings (n, %)	18, 36%	16, 32%	0.67
Indication for CT (n, %)	Pain: 45, 90% Suspected Bowel Obstruction: 3, 6% Fever: 1, 2% Hematuria: 1, 2%	Pain: 40, 80% Suspected Bowel Obstruction: 3, 6% Oncologic: 3, 6% Fever: 2, 4% Post-surgical: 2, 4%	

*mean[SD]

Table 2 Reconstruction time comparison

Cohort	Total CT examinations included	Time from topogram to sagittal reconstruction (minutes)	Time from topogram to coronal reconstruction (minutes)
Console cohort	728	8.7[11.1]*	6.6[4.8]*
Server cohort	892	4.5[6.6]*	4.1[6.2]*
p-value		< 0.001	< 0.001

*mean[SD]

were selected after 9 patients without oral contrast, 2 with extensive beam hardening artifact emanating from spinal surgical hardware, and 1 with large body habitus extending outside of the imaged field were excluded. This resulted in a final cohort of 100 patients for qualitative assessment: 50 patients from the Console cohort in 2022 (28 women and 22 men, mean[SD] age: 51[19] years) and 50 patients from Server cohort in 2023 (27 women and 23 men, 57[19] years. Age and sex did not differ between the groups ($p = .08$ and $p = .84$, respectively). Table 1 shows patient demographics.

CT Radiation exposure

The mean[SD] CTDIvol was 13.4[7.2] mGy in the Console cohort and 13.9[7.7] mGy in the Server cohort ($p = .87$). The mean[SD] DLP was 674.7[397.5] mGycm in the Console cohort and 697.9[393.6] mGycm in the Server cohort ($p = .74$).

Reconstruction Time comparison

The mean[SD] time from topogram to sagittal reconstruction was 525[665] seconds (8.7[11.1] minutes) in the Console cohort ($n = 728$) and 273[394] seconds (4.6[6.6] minutes) in the Server cohort ($n = 892$), $p < .001$. The mean time from topogram to coronal reconstruction was 393[286] seconds (6.6[4.8] minutes) in the Console cohort and 247[373] seconds (4.1[6.2] minutes) in the Server cohort ($p < .001$). Table 2 shows the comparison of reconstruction times.

Image quality comparison

There was no significant difference in image quality or diagnostic confidence scores for reader 1 or reader 2 when comparing the technologist reconstructions (Console cohort) with the AI created coronal and sagittal reformatted images (Server cohort), all $p > .05$. Table 3 details image quality and diagnostic confidence score comparisons. There was no significant difference in intra-reader image quality or diagnostic confidence scores when comparing axial images with coronal and sagittal images created in the same year (all $p > .05$). Figure 1 shows sample images created by the technologist at the scanner console and Fig. 2 shows AI-based thin-client reformatted images.

There was almost perfect inter-reader agreement for the presence of acute findings (kappa = 0.83, IQR 0.72–0.94) and strong agreement for findings explaining abdominal pain (kappa = 0.75, IQR 0.62–0.88).

Table 3 Qualitative image comparison scores (mean[SD]). *mean[SD]

		Reader 1			Reader 2		
		Console cohort	Server cohort	p-value	Console cohort	Server cohort	p-value
Image quality	Axial	4.8[0.4]*	4.9[0.3]*	0.19	5.0[0.2]*	4.9[0.3]*	0.14
	Coronal	4.8[0.4]*	4.9[0.4]*	0.70	5.0[0.2]*	4.9[0.3]*	0.14
	Sagittal	4.8[0.4]*	4.9[0.3]*	0.19	5.0[0.2]*	4.9[0.3]*	0.14
	p-value (axial/coronal, axial/sagittal)	0.99, 0.99	0.99, 0.99		0.99, 0.99	0.99, 0.99	
Diagnostic confidence	Axial	4.9[0.2]*	4.9[0.3]*	0.21	5.0[0.2]*	4.9[0.3]*	0.14
	Coronal	4.9[0.2]*	4.9[0.3]*	0.70	5.0[0.2]*	4.9[0.3]*	0.14
	Sagittal	4.9[0.2]*	4.9[0.3]*	0.70	5.0[0.2]*	4.9[0.3]*	0.14
	p-value (axial/coronal, axial/sagittal)	0.99, 0.99	0.32, 0.99		0.99, 0.99	0.99, 0.99	

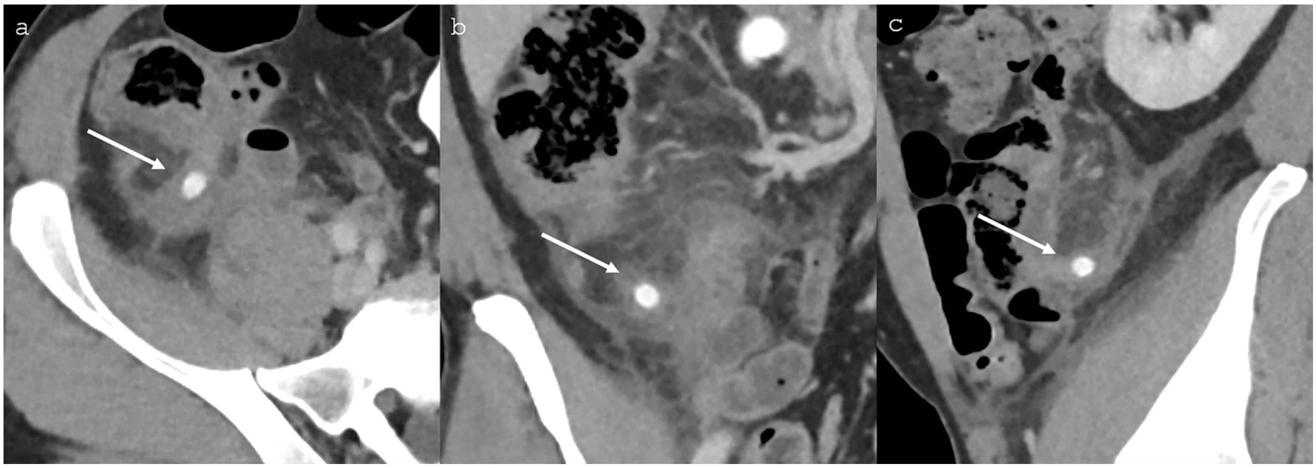

Fig. 1 (**a**) axial, (**b**) coronal, and (**c**) sagittal reformatted CT images generated by the technologist at the console (Console Cohort) in a 47-year-old man with abdominal pain, demonstrate a dilated, fluid-filled appendix containing an appendicolith with adjacent fat stranding (arrow) compatible with acute appendicitis

Fig. 2 (**a**) axial, (**b**) coronal, and (**c**) sagittal server generated reformatted CT images (Server Cohort) in a 22-year-old male with abdominal pain demonstrate a dilated, fluid filled appendix with adjacent fat stranding (arrow), compatible with acute appendicitis

Discussion

Our study compared multiplanar reconstructions created at the console by the technologist with AI generated reformatted images. Examinations utilizing AI generated reformatted images were complete in PACS in mean 4.5 min compared with 8.7 min when created by the technologist ($p < .001$). There was no statistically significant difference in image quality or radiologist diagnostic confidence when utilizing AI generated compared with console reconstructions.

With growing CT volume throughout the country, technologic advances that assist CT technologist workflow without compromising diagnostic ability are increasingly valuable [11]. Utilizing AI to generate coronal and sagittal reformatted images reduces the time technologists spend generating reconstructions while also allowing these images to reach PACS faster for diagnostic interpretation. Additionally, generating the multiplanar reformatted images on a thin-client server reduces the computational load necessary at the CT scanner console, allowing for earlier subsequent imaging. In the ED in particular, faster time for images to reach PACS could allow for earlier life-saving diagnoses to be made.

Prior studies have employed additional consoles to expedite image reconstruction. Yu et al. evaluated a similar artificial intelligence tool for automated reformatted images in trauma pan-scans, saving a mean 5.7 min per pan-scan which was largely attributable to a 4.7 min decrease in mean torso reconstruction time. However, this study focused primarily on scanner throughput and reconstruction time, and did not evaluate the post-processed image quality [6].

Itri et al. compared technologist workflow and the source of time savings in an automated system. Surprisingly, in their study the manual reconstruction process often was not the largest contributor, rather much of the delay was associated with the prioritization of preparation for and performance of a subsequent scan on the next patient in the ED setting [1]. The automated reconstructions generated in our study allow for more prompt scanner console readiness for subsequent patients. There was also a significant time reduction in image availability on PACS in our large cohort.

In addition to diagnostic confidence and image quality, it is also important that there not be differences in diagnosis rendered when using AI generated reconstructions. In our study, two radiologists had almost perfect agreement for acute findings (kappa = 0.83) and strong agreement for explanation of abdominal pain (kappa = 0.75) using both AI generated and console reconstructions. Notably, during routine clinical practice, there were no reported quality tickets or suboptimal image quality in radiology reports when utilizing these AI rendered reconstructions.

Limitations include that this is a single center, retrospective study. Additionally, this study was performed using a single vendor CT scanner in the ED setting and utilized the reconstruction methods of one vendor, so the results may not be generalizable to other AI-based reconstruction tools. The axial images in all examinations were generated by the technologist to serve as an internal control between the Console and Server generated reconstructions, however, this precluded comparison of axial images between the two cohorts. In addition, 1 patient with motion artifact, 2 patients with streak artifact emanating from surgical hardware, and 1 large patient extending beyond the field of view were excluded from qualitative assessment so as to not bias the image quality assessment because image quality assessment of these patients would not be reflective of the reconstruction mechanism, but rather instead it would reflect image quality challenges subject to all examinations.

In conclusion, AI generated coronal and sagittal reformations were generated approximately 50% faster than console reconstructions. In addition, there was no statistical difference in diagnostic confidence or subjective image quality for console created versus AI generated multiplanar reformatted images. Utilizing AI generated multiplanar reformatted images from the thin-client reduces the time technologists spend creating reconstructions while ensuring more rapid examination completion in PACS for imaging review.

Acknowledgements Kun Qian for statistical consultation.

Author contributions DF - data curation, chart review, manuscript preparation; BB - statistical analysis, image review; KM - image review; EV - manuscript review, protocol development; TO - manuscript preparation and review; MW - statistical analysis; BD - oversee research, manuscript preparation, manuscript review.

Funding None

Data availability Available.

Code availability N/A.

Declarations

Ethics approval This study was institutional review board approved and Health Insurance Portability and Accountability Act compliant.

Consent to participate Waiver of the requirement for informed consent.

Consent for publication Yes.

Competing interests Thomas O'Donnell is a Siemens Healthineers employee, Malte Westerhoff is a Visage employee, Bari Dane received speaker honorarium from Siemens Healthineers. The other authors have no conflict of interest to report.

References

1. Itri JN, Boonn WW. Use of a Dedicated Server to Perform Coronal and Sagittal Reformations in Trauma Examinations. J Digit Imaging. Jun 2011;24(3):494–9. doi:https://doi.org/10.1007/s10278-010-9296-3
2. Jeong DK, Lee KH, Kim BH, et al. On-the-fly Generation of Multiplanar Reformation Images Independent of CT Scanner Type. *Journal of Digital Imaging*. 2008/09/01 2008;21(3):306–311. doi:https://doi.org/10.1007/s10278-007-9032-9
3. Körner M, Krötz M, Kanz KG, Pfeifer KJ, Reiser M, Linsenmaier U. Development of an accelerated MSCT protocol (Triage MSCT) for mass casualty incidents: comparison to MSCT for single-trauma patients. *Emergency Radiology*. 2006/07/01 2006;12(5):203–209. doi:https://doi.org/10.1007/s10140-006-0485-9
4. Mueck FG, Wirth K, Muggenthaler M, et al. Radiological mass casualty incident (MCI) workflow analysis: single-centre data of a mid-scale exercise. The British Journal of Radiology. 2016/05/01 2016;89(1061):20150918. doi:https://doi.org/10.1259/bjr.20150918
5. Soto JA, Anderson SW. Multidetector CT of Blunt Abdominal Trauma. Radiology. 2012;265(3):678–693. doi:https://doi.org/10.1148/radiol.12120354
6. Yu H, Bay CP, Czajkowski B, Sodickson AD. Automated CT reformations reduce time and variability in trauma panscan exam completion. Emerg Radiol. Jun 2022;29(3):461–469. doi:https://doi.org/10.1007/s10140-022-02031-7
7. Zhan Y, Dewan M, Harder M, Krishnan A, Zhou XS. Robust Automatic Knee MR Slice Positioning Through Redundant and Hierarchical Anatomy Detection. IEEE Transactions on Medical Imaging. 2011;30(12):2087–2100. doi:https://doi.org/10.1109/TMI.2011.2162634
8. Zhou XS, Zhan Y, Raykar VC, Hermosillo G, Bogoni L, Peng Z. Mining anatomical, physiological and pathological information from medical images. SIGKDD Explor Newsl. 2012;14(1):25–34. doi:https://doi.org/10.1145/2408736.2408741
9. Crönlein M, Holzapfel K, Beirer M, et al. Evaluation of a new imaging tool for use with major trauma cases in the emergency department. *BMC Musculoskeletal Disorders*. 2016/11/17 2016;17(1):482. doi:https://doi.org/10.1186/s12891-016-1337-8
10. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.

11. Salastekar NV, Richard Duszak J, Santavicca S, et al. Utilization of Chest and Abdominopelvic CT for Traumatic Injury From 2011 to 2018: Evaluation Using a National Commercial Database. American Journal of Roentgenology. 2023;220(2):265–271. doi:https://doi.org/10.2214/ajr.22.27991