

Deep learning reconstruction improves appendix visualization on pediatric MRI: a single-center experience

Sebastian Gallo-Bernal

Massachusetts General Hospital

Navaira Shoaib

Massachusetts General Hospital

Fedel Machado-Rivas

Massachusetts General Hospital

Teresa Victoria

Massachusetts General Hospital

Valeria Pena-Trujillo

UMass Memorial Medical Center

Rimaz Khadir

Massachusetts General Hospital

Eugene Milshteyn

General Electric (United States)

Arnaud Guidon

General Electric (United States)

John Conklin

Massachusetts General Hospital

Michael S Gee

MSGEE@MGH.HARVARD.EDU

Massachusetts General Hospital

Research Article

Keywords:

Posted Date: March 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8996746/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Pediatric Radiology on June 6th, 2026. See the published version at <https://doi.org/10.1007/s00247-026-06686-z>.

Abstract

Background

Thin-slice MRI may improve appendix visualization in children with suspected appendicitis, but reducing slice thickness decreases signal-to-noise ratio (SNR) and can degrade image quality. Deep learning reconstruction (DLR) enables denoising of thin-slice images and may mitigate these limitations.

Objective

To evaluate whether 2-mm T2-weighted MRI with DLR improves appendix visualization and diagnostic performance compared with conventional 4-mm MRI in pediatric appendicitis.

Materials and methods

In this prospective single-center study (January 2022-June 2025), 82 children (mean age 9.6 years) undergoing MRI for suspected appendicitis were included. MRI protocols contained both standard 4-mm and thin-slice 2-mm single-shot T2-weighted images; DLR was applied to 2-mm images (2-mm + DLR). Three blinded radiologists independently assessed paired series for appendix visualization, confidence, appendicitis diagnosis, ancillary findings, and overall image quality. Objective SNR was measured in psoas muscle, bladder, and fat. Diagnostic performance was calculated against surgical or clinical reference standards using paired statistical tests.

Results

DLR significantly increased appendix visualization compared with standard MRI (88.2% vs 83.3%, $p = 0.006$). Overall diagnostic quality scores were higher with 2-mm + DLR (median 5 vs 4, $p = 0.0004$), while confidence in appendix visualization and appendicitis diagnosis were similar (both medians 4, $p > 0.1$). SNR of 2-mm + DLR was significantly higher than non-DLR 2-mm images (all $p < 0.01$) and comparable to 4-mm images. Pooled diagnostic performance favored 2-mm + DLR (sensitivity 88% vs 80%, accuracy 91% vs 86%), although paired comparisons did not reach statistical significance.

Conclusion

DLR-enhanced 2-mm MRI improves appendix visualization and overall diagnostic image quality in pediatric appendicitis without reducing reader confidence or specificity. Thin-slice DLR MRI supports improved depiction of the appendix and alternative abdominal pathology in children.

INTRODUCTION

The selection of an appropriate imaging technique is crucial for the timely and accurate diagnosis of acute appendicitis in children[1]. The American College of Radiology (ACR) appropriateness criteria recommend abdominal and pelvic ultrasound (US) as the initial imaging modality for suspected appendicitis in pediatric patients [2]. If US results are inconclusive or unsatisfactory, cross-sectional imaging, such as computed tomography (CT) or magnetic resonance imaging (MRI), is recommended.

In institutions with pediatric imaging expertise, MRI is increasingly favored over CT due to its lack of ionizing radiation and superior soft tissue contrast[3, 4]. MRI provides high-resolution images and diagnostic accuracy comparable to or even superior to CT for appendicitis evaluation without the need for intravenous contrast [5–7]. Advances in MRI technology have

led to shorter scan times, decreased motion sensitivity, and improved spatial resolution, all of which have contributed to the increasing use of MRI in pediatric appendicitis protocols [8–11].

Despite these improvements, standard MRI protocols typically use 4–7 mm slice thicknesses to maintain short scan durations [5, 12–14]. This can impair visualization of the appendix, especially in pediatric patients, due to limited through-plane resolution and partial volume averaging effects. Reducing slice thickness to levels comparable to CT would likely improve appendix visualization but introduces other challenges: decreased signal-to-noise ratio (SNR) and longer scan times due to more slices required to cover the same anatomical area. These trade-offs can negatively affect image quality and diagnostic reliability.

To address this, recent advances in artificial intelligence (AI)-based denoising offer a promising solution [15–20]. Deep learning reconstruction techniques are trained to enhance both image sharpness and SNR, allowing for high-resolution thin-slice images without sacrificing quality [21, 22]. One such tool, AIR™ DL (GE HealthCare, Waukesha, WI), uses a deep convolutional neural network (CNN) trained on paired high- and low-SNR datasets. Integrated directly into the scanner's reconstruction pipeline, it enables real-time image enhancement and includes tunable noise reduction settings applicable across various MRI sequences [23]. The algorithm offers a tunable noise reduction factor and can be applied to a wide variety of pulse sequences.

This study aims to evaluate the impact of a scanner-integrated DLR algorithm on the diagnostic quality of 2-mm thin single-shot fast spin echo (SSFSE) T2-weighted MRI images for pediatric appendicitis in a tertiary pediatric emergency setting. We hypothesize that DLR-enhanced 2-mm slice images will offer superior appendix visualization and image quality comparable to, or better than, conventional 4-mm MR acquisitions without the typical compromises in SNR or scan time.

METHODS

Study population

This single-center prospective study was conducted at a large academic referral hospital with a dedicated pediatric emergency department (ED) from January 2022 to June 2025. The study protocol was approved by the Institutional Review Board, and all procedures were compliant with the Health Insurance Portability and Accountability Act. Eligible participants included pediatric patients (< 21 years old) who presented to the ED with clinical suspicion of appendicitis and subsequently underwent our institution's MRI appendix protocol. Verbal informed consent was obtained from parents, and assent was obtained from the pediatric participants when appropriate. Consistent with the Image Gently campaign and ACR recommendations, our institution follows a cross-departmental staged pediatric appendicitis imaging algorithm. US is the initial imaging modality, followed by unenhanced MRI or CT when US findings are equivocal. The decision to perform an MRI after inconclusive US results is made at the discretion of the referring provider.

During the study period, 1,035 pediatric patients underwent abdominal ultrasound for suspicion of appendicitis, and 126 of these were referred for additional imaging, including 83 patients who underwent abdominal MRI (Fig. 1). All MRIs included both 4-mm and 2-mm T2-weighted images (with and without DLR). For the initial 16 cases, the DLR was in its early stages and not yet a commercial product, requiring offline reconstruction. As such, only the conventional 4-mm T2-weighted images were used for clinical interpretation. After DLR was approved and integrated as a product on our scanners, the reconstruction process moved to the scanner's GPU, and the DLR images were automatically transmitted to the PACS and used for clinical interpretation alongside the conventional images.

MRI protocol

All studies were performed on a 1.5 T clinical scanner (Signa Artist, GE Healthcare) using a 30-channel phased-array torso coil (AIR™ Coil, GE Healthcare) without sedation/anesthesia or contrast media. The field of view and matrix size varied with

patient size and age, but all images were obtained from the inferior poles of the kidneys through the urinary bladder. The protocol consisted of the following sequences (Table 1): standard 4 mm axial and coronal T2-weighted SSFSE; thin-slice 2mm axial and coronal T2-weighted SSFSE; 4 mm axial and coronal T2-weighted SSFSE with fat suppression; axial fat-saturated T2-weighted periodically rotated overlapping parallel lines with enhanced reconstruction (PROPELLER, GE Healthcare); and axial diffusion-weighted imaging (DWI) for patients able to tolerate additional images. Conventional reconstruction strategies (Fourier transform and apodization of fully sampled k-space data) were used to reconstruct all the obtained images.

Table 1
MRI acquisition protocol

Parameter	Standard single-shot fast spin-echo non-fat suppressed	Thin-slice single-shot fast spin-echo non-fat suppressed	Standard single-shot fast spin-echo non-fat suppressed	Thin-slice single-shot fast spin-echo non-fat suppressed	Fat suppressed single-shot fast spin-echo	Fat suppressed single-shot fast spin-echo	Fat-saturated T2-weighted periodically rotated overlapping parallel lines with enhanced reconstruction
Imaging Plane	Axial	Axial	Coronal	Coronal	Axial	Coronal	Axial
TE (ms)	89–93	89–93	89–93	89–93	89–93	89–93	99–102
TR (ms)	400–2400	315–410	400–3000	350–410	500–3200	500–3500	3500–6000
Number of echoes	1	1	1	1	1	1	1
Field of view (mm)	300x300	300x300	350x350	300x300	300x300	350x350	280x280
Slice spacing (gap)	0	0	0	0	0	0	0
Slice thickness (mm)	4	2	4	2	4	4	4
Flip angle (degrees)	90	90	90	90	90	90	160
Number of excitations (NEX)	1	1	1	1	1	1	2

DLR algorithm

We used a vendor-supplied DLR algorithm (AIR™ DL) to denoise the 2-mm thin-slice SSFSE MRI images. Initially, the DLR technology evaluated in this study was a premarket prototype. However, during the course of the study, technology evolved and became a fully integrated product, receiving FDA clearance for clinical use after the conclusion of data collection and analysis. The DLR technique uses a deep convolutional neural network (CNN)-based algorithm embedded in the hardware reconstruction pipeline to generate reconstructed images in less than 2 seconds.

The DLR algorithm processes the raw k-space data from the 2-mm thin-slice images and produces high-fidelity, sharp, and denoised images with reduced ringing artifacts and improved in-plane resolution (Figs. 2–4). From now on, the 2-mm thin-slice DLR images will be referred to as 2mm + DLR. The software allows for manual adjustment of the noise reduction factor

between 0% and 100%, accommodating user preference. For this study, we set the noise reduction factor to 70%, a decision made empirically based on comparisons of images generated across a range of factor values.

The CNN was trained using a supervised learning approach based on pairs of near-perfect (high-resolution images with minimal ringing artifacts) and conventional MRI images. The training database included a broad range of image content, ensuring the algorithm's generalizability and robustness across all anatomies. The algorithm was trained exclusively by the vendor, and no data from our institution was used in its development. Additional information about this vendor-developed DLR algorithm has been reported elsewhere [23].

Image quality analysis: subjective and objective assessment

For the first 16 cases, a proof-of-concept quantitative analysis was performed to assess image quality. We calculated the SNR for the standard 4-mm, no-DLR 2-mm, and 2-mm + DLR images. Bilateral circular regions of interest (ROIs) of 100 ± 1.0 mm² (mean $\pm$ standard deviation [SD], or 133 pixels) were placed at the level of the psoas major muscle, urinary bladder, and abdominal fat. The mean signal intensity within each ROI was measured, and an average was calculated for each anatomic region. For further calculations, the standard deviation (SD) of signal intensity from a ROI placed in the background was used as the noise value. The SNR was defined as the mean signal intensity of each anatomic region divided by the noise.

The subjective image quality of the standard 4-mm and 2-mm + DLR images was evaluated by three independent, fellowship-trained radiologists (pediatric/abdominal, blank; pediatric, blank; and emergency, blank), each with > 5 years of experience interpreting appendicitis MRI. All examinations were de-identified and randomized, and readers were blinded to scan parameters, MRI protocol details, prior imaging, and clinical history. For each series, readers answered eight predefined items: Q1 - appendix visualized (visualized versus not visualized); Q2 - confidence in appendix visualization (Likert 1–5, where 1 = no confidence and 5 = high confidence); Q3 - study positive for appendicitis (determination based on the reader's overall imaging impression); Q4 - confidence in the appendicitis determination (Likert 1–5); Q5 - visualization of pathologic changes in the gallbladder (0 if the organ was not visible, otherwise 1–5 for diagnostic utility); Q6 - visualization of kidney stones or collecting-system changes (0 if not visible, otherwise 1–5); Q7 - visualization of mesenteric lymph-node abnormalities (0 if not visible, otherwise 1–5); and Q8 - overall diagnostic image quality (Likert 1–5).

Statistical Analysis:

Qualitative variables were reported as frequencies and percentages, while quantitative variables were reported as mean $\pm$ standard deviation (SD) or median $\pm$ interquartile range (IQR). Continuous variables, including SNR and signal intensities between protocols, were compared using either parametric Student's t-test or nonparametric Mann-Whitney U tests (assuming heterogeneous variance), depending on the distribution of the data. χ^2 tests were used for comparisons of categorical variables.

For the subjective image quality assessment, image confidence scores were compared using the Kruskal-Wallis test. To control family-wise error across these four endpoints, we applied a Holm-Bonferroni adjustment to the set of p-values from the primary Wilcoxon tests. Interobserver agreement was assessed using quadratic-weighted kappa, summarized across reader pairs separately for each reconstruction method, summarized across reader pairs separately for each reconstruction method. Interobserver agreement was classified as poor ($\kappa \leq 0.20$), fair ($\kappa, 0.21-0.40$), moderate ($\kappa, 0.41-0.60$), good ($\kappa, 0.61-0.80$), and excellent ($\kappa > 0.81$).

Diagnostic performance of standard and 2-mm + DLR was evaluated against the study ground truth (operative/anatomic confirmation or, when no surgery was performed, treatment with antibiotics as decided by the primary team). For each method, we calculated sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV), and accuracy from pooled reads. Pairwise method comparisons used McNemar's test on paired predictions overall, with additional stratified McNemar tests within ground-truth positive cases to compare sensitivities and within ground-truth

negative cases to compare specificities. PPV and NPV were reported descriptively without formal hypothesis testing. All tests were two-sided with $\alpha = 0.05$. Analyses were performed in R version 4.5.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

General Characteristics

From January 2022 to June 2025, a total of 82 pediatric patients (age range 5 to 20 years) who underwent MRI for clinically suspected appendicitis were enrolled in the study. The mean age $\pm$ SD was 9.55 years $\pm$ 3.69. 51.2% of the patients were female. Based on the study ground truth, 29 of 82 patients (35%) had appendicitis. The acquisition time for the standard 4 mm images was 1 minute 45 seconds, whereas the 2 mm required 2 minutes 22 seconds.

Objective image quality:

When compared to the non-DLR images, the 2-mm + DLR images yielded statistically significantly higher SNR at the level of the urinary bladder (26.93 ± 11.2 versus 13.45 ± 4.9 ; $P < 0.01$), psoas major muscle (5.04 ± 1.8 versus 2.62 ± 0.9 ; $P < 0.01$), and abdominal fat (26.29 ± 6.15 versus 15.57 ± 4.0 ; $P < 0.01$) (FIGURE 5). Compared to the standard 4-mm images, there were no significant SNR differences (all $P > 0.05$) in the same anatomic regions (26.93 ± 11.2 versus 29.61 ± 11.9 ; 5.04 ± 1.8 versus 5.23 ± 1.9 ; 26.29 ± 6.1 versus 29.78 ± 10.1).

Subjective image quality:

Across 82 patients and three reviewers (246 paired reads), 2-mm + DLR improved appendix visualization while confidence and positivity remained similar on paired testing (Table 2). Appendix visualization was higher with 2-mm + DLR than standard (88.2% versus 83.3%; $p = 0.006$) (Fig. 6). Confidence in visualization did not demonstrate a statistically significant difference (both medians 4 [IQR 3–5]; Wilcoxon $p = 0.915$). Confidence in appendicitis determination likewise showed no statistically significant difference (both medians 4 [IQR 3–5]; Wilcoxon $p = 0.141$). Interobserver agreement, summarized across reader pairs, was substantial for the appendicitis call and more variable for visualization and confidence. For standard images, κ ranges were: Q1 0.51–0.906, Q2 weighted 0.122–0.604, Q3 0.661–0.771, Q4 weighted 0.246–0.936; for DL images, κ ranges were: Q1 0.1–0.921, Q2 weighted 0.1–0.864, Q3 0.738–0.949, Q4 weighted 0.7–0.959.

Table 2
Subjective image quality and ancillary visualization (paired within reviewer–patient)

Outcome	Measure	Standard 4-mm	2-mm + DLR	Paired test p-value
Appendix visualized	Percent yes	82.9%	88.2%	0.004
Confidence of appendix visualization	Median [IQR]	4 [3–5]	4 [3–5]	0.915
Study positive for appendicitis	Percent yes	35.4%	37.0%	0.658
Confidence in appendicitis determination	Median [IQR]	4 [3–5]	4 [3–5]	0.141
Gallbladder pathology visible	Percent yes	100.0%	100.0%	1.000
Gallbladder pathology visualization score	Median [IQR]	4 [4–5]	5 [4–5]	0.0019
Collecting system or stones visible	Percent yes	100.0%	100.0%	1.000
Collecting system or stones visualization score	Median [IQR]	4 [4–4]	4 [4–5]	0.0217
Mesenteric lymph node pathology visualization score	Median [IQR]	4 [4–5]	4 [4–5]	0.4447
Overall diagnostic quality	Median [IQR]	4 [4–5]	5 [4–5]	0.000039

For gallbladder pathology visualization (Q5), 2-mm + DLR showed higher scores than standard on paired ordinal testing (median 5 [IQR 4–5] vs 4 [4–5]; $p = 0.0019$). Visibility as a binary metric was saturated for both methods (100% visible for standard and DL), so the paired test for “visible > 0” was non different ($p = 1.00$). For the collecting system or stone visualization (Q6), 2-mm + DLR modestly improved scores (median 4 [4–5] vs 4 [4–4]; $p = 0.0217$), again with 100% visibility for both methods; $p = 1.00$. For mesenteric lymph node pathology visualization (Q7), paired medians were similar (both 4 [4–5]) and the difference did not demonstrate statistical significance ($p = 0.083$). For overall diagnostic quality (Q8), 2-mm + DLR outperformed standard (median 5 [4–5] vs 4 [4–5]; Wilcoxon $p = 0.0004$). Interobserver agreement (quadratic-weighted κ) for Q5–Q8 ranged from slight to substantial and was generally similar to or higher with DL. For standard images, κ ranges were: Q5 0.1–0.49, Q6 0.1–0.65, Q7 0.1–0.61, Q8 0.1–0.70. For DL images, κ ranges were: Q5 0.1–0.65, Q6 0.1–0.85, Q7 0.2–0.84, Q8 0.2–0.79.

Performance

Against ground truth, pooled diagnostic point estimates favored 2-mm + DLR. Standard yielded sensitivity 80%, specificity 88%, PPV 77%, NPV 94%, and accuracy 86%. 2-mm + DLR yielded sensitivity 88%, specificity 89%, PPV 79%, NPV 96%, and accuracy 91%. In paired inference, the overall distribution of positive versus negative calls did not differ ($p = 0.883$). Sensitivity was numerically higher with 2-mm + DLR among ground-truth positive cases, demonstrating a trend toward improvement that did not reach statistical significance ($p = 0.061$). Specificity was not significantly different between techniques ($p = 0.296$). Paired accuracy numerically favored the DLR images (difference = + 5.7%; discordant standard-correct/DL-incorrect = 16 versus DL-correct/standard-incorrect = 30), again showing a trend that did not reach statistical significance ($p = 0.055$).

Discussion

Achieving high-quality MRI images of the pediatric abdomen is challenging, given the trade-offs between image noise, contrast, scan time, and spatial resolution. Nevertheless, recent advancements in computational sciences have expanded the applications of artificial intelligence in pediatric radiology beyond lesion detection or organ segmentation and have

improved our ability to acquire better and faster MRI studies [8, 24–27]. With this in mind, we assessed the clinical utility of DLR for improving through-plane spatial resolution in an ED pediatric appendicitis MRI protocol. The decision to acquire thinner slices was based on the hypothesis that, in pediatric patients, improving through-plane resolution could enhance visualization of the appendix, similar to the thin-slice approach used in CT appendicitis protocols. Our results support this hypothesis. Compared with standard 4-mm images, 2-mm DLR images increased the frequency of appendix visualization, improved overall diagnostic quality scores, and yielded near statistically significant gains in sensitivity for appendicitis. Specificity and reader positivity rates remained similar, and reader confidence scores were comparable between methods. To our knowledge, this is the first study evaluating the diagnostic impact of a DLR algorithm applied to thin-slice T2-weighted MRI in pediatric appendicitis assessment in the ED.

Recent studies have demonstrated the effectiveness of DLR in enhancing spatial resolution while preserving diagnostic image quality in multiple clinical contexts [28–31]. For instance, Kim and colleagues reported that 1-mm contrast-enhanced T1-weighted pituitary images reconstructed with DLR had higher diagnostic performance than standard 3-mm images for evaluating postoperative residual tumor and cavernous sinus invasion[15, 16]. Similarly, Kakigi et al. showed that 1-mm 2D FS-PDWI images of the knee, reconstructed with a DLR-based technique, improved diagnostic consistency and confidence when compared with 3D FS-PDWI[19]. The authors also noted that the acquisition time for the thin-slice protocol was shorter than that of the 3D sequence. These results complement our findings and support the use of thin-slice DLR across diverse imaging applications.

Thin slices improved direct visualization of the normal appendix. By minimizing partial volume effects and providing more complete sampling of the appendiceal length, readers were more frequently able to identify a continuous, well-defined lumen and wall. This made confident “not inflamed” interpretations faster and easier, even though the analysis did not show a statistically significant increase in specificity. The practical implication is that thin-slice DLR facilitates more reliable negative studies without contributing to false positive overcalls.

Beyond the appendix, DLR also enhanced the visualization of other structures relevant to pediatric abdominal pain. Readers gave higher paired scores for gallbladder visualization and for identifying collecting system pathology or stones. Overall diagnostic quality improved, while mesenteric lymph node visualization remained similar between methods. In the ED setting, these differences are clinically important[32, 33]. Clearer gallbladder imaging can guide diagnosis toward cholecystitis or biliary colic. Better visualization of the kidneys and ureters supports evaluation for nephrolithiasis or hydronephrosis. Reliable lymph node assessment can assist in recognizing or excluding mesenteric adenitis. These findings suggest that thin-slice DLR adds clinically meaningful discriminative detail that can guide decision-making when appendicitis is not the cause of symptoms.

Although the improvements in sensitivity and overall accuracy observed with 2-mm DLR images did not reach statistical significance, this could be due to limited statistical power. Our study included a relatively small number of ground-truth positive cases. In this setting, paired statistical tests are influenced by a small number of discordant pairs, which introduce wide uncertainty and result in p-values that lie near the significance threshold. Nonetheless, the direction of the effect consistently favored the DLR-enhanced images, suggesting that with a larger sample size, statistical significance might be achieved. From a mechanistic perspective, we hypothesize that increased through-plane spatial resolution improves sensitivity for appendicitis by reducing partial volume averaging and allowing better depiction of tip appendicitis and short-segment disease. Two-millimeter sections sample the appendix more densely along its length than 4-mm sections. This increased sampling makes small inflammatory foci more conspicuous and allows clearer identification of early or localized appendiceal changes, such as mural thickening, focal T2 hyperintensity, tiny intraluminal fluid levels, or limited periappendiceal fat stranding. In these cases, thinner slices enable abnormalities to be visible across multiple contiguous images, rather than appearing as a subtle or ambiguous finding on a single thick slice. This improves the likelihood of a true positive diagnosis without decreasing specificity.

This single-center study, conducted in a pediatric emergency department setting, may limit the broader applicability of the findings. The number of confirmed appendicitis cases was relatively small, which reduced the statistical power of paired comparisons and made the results sensitive to a few discordant cases. During the course of the study, the deep learning reconstruction method transitioned from an offline prototype to a commercially available scanner-integrated product. This evolution introduced potential sources of confounding related to both time and clinical workflow. Because images generated during the product phase were used in clinical decision-making, they may have influenced diagnostic thresholds, particularly for subtle or early presentations such as tip appendicitis. This could have introduced verification bias, management bias, and learning effects as radiologists adapted to the new image characteristics. Although readers were blinded to image type, deep learning reconstructions may have had a visually recognizable appearance that posed a risk of inadvertent unblinding. Additionally, the reference standard for appendicitis included a combination of surgical, imaging, and clinical treatment outcomes. While this approach reflects real-world decision-making, it may have led to misclassification in cases that were equivocal or managed nonoperatively.

Conclusion

Deep learning reconstruction is increasingly used across imaging modalities, and its role in abdominal MRI is growing. Our findings show that 2-mm T2-weighted MRI with DLR improves appendix visualization and overall diagnostic quality in a real-world pediatric ED setting, without compromising specificity or reader confidence. While gains in sensitivity were not statistically significant, the radiologists consistently favored DLR in their image quality assessment. DLR also improved visualization of alternative causes of abdominal pain, supporting its broader clinical utility. These results provide new evidence for incorporating DLR into thin-slice pediatric abdominal MRI protocols.

Declarations

Author Contribution

Sebastian Gallo-Bernal, MD: Study conception and design; data collection and curation; statistical analysis; drafting of the manuscript; figure and table preparation; critical revision. Navaira Shoaib, MD: Data collection and curation; statistical analysis; drafting of the manuscript; critical revision. Fedel Machado-Rivas, MD: Data collection and curation; statistical analysis; drafting of the manuscript; critical revision. Teresa Victoria, MD: Study conception and design; supervision; project administration; critical revision. Valeria Peña-Trujillo, MD: Data collection and curation; statistical analysis; drafting of the manuscript; critical revision. Rimaz Khadir: Data collection and curation; technical support; critical revision. Eugene Milshteyn: Image processing and technical methodology; data analysis support; critical revision. Arnaud Guidon: Image processing and technical methodology; data analysis support; critical revision. John Conklin, MD, PhD: Study conception and design; methodology; supervision; critical revision. Michael S. Gee, MD, PhD: Study conception and design; methodology; supervision; funding acquisition; critical revision.

References

1. van Rijn RR (2018) Imaging Pediatric Appendicitis: Time to Put the Discussion to Bed? *Radiology* 286:1030–1032. <https://doi.org/10.1148/radiol.2017172706>
2. Garcia EM, Camacho MA, Karolyi DR et al (2018) ACR Appropriateness Criteria® Right Lower Quadrant Pain-Suspected Appendicitis. *J Am Coll Radiol* 15:S373–S387. <https://doi.org/10.1016/j.jacr.2018.09.033>
3. Moore MM, Kulaylat AN, Hollenbeak CS et al (2016) Magnetic resonance imaging in pediatric appendicitis: a systematic review. *Pediatr Radiol* 46:928–939. <https://doi.org/10.1007/s00247-016-3557-3>
4. Covelli JD, Madireddi SP, May LA et al (2019) MRI for Pediatric Appendicitis in an Adult-Focused General Hospital: A Clinical Effectiveness Study—Challenges and Lessons Learned. *Am J Roentgenol* 212:180–187. <https://doi.org/10.2214/AJR.18.19825>

5. Mushtaq R, Desoky SM, Morello F et al (2019) First-Line Diagnostic Evaluation with MRI of Children Suspected of Having Acute Appendicitis. *Radiology* 291:170–177. <https://doi.org/10.1148/radiol.2019181959>
6. Kinner S, Pickhardt PJ, Riedesel EL et al (2017) Diagnostic Accuracy of MRI Versus CT for the Evaluation of Acute Appendicitis in Children and Young Adults. *Am J Roentgenol* 209:911–919. <https://doi.org/10.2214/AJR.16.17413>
7. Eng KA, Abadeh A, Ligocki C et al (2018) Acute Appendicitis: A Meta-Analysis of the Diagnostic Accuracy of US, CT, and MRI as Second-Line Imaging Tests after an Initial US. *Radiology* 288:717–727. <https://doi.org/10.1148/radiol.2018180318>
8. Gallo-Bernal S, Bedoya MA, Gee MS, Jaimes C (2022) Pediatric magnetic resonance imaging: faster is better. *Pediatr Radiol* 53:1270–1284. <https://doi.org/10.1007/s00247-022-05529-x>
9. Kozak BM, Jaimes C, Kirsch J, Gee MSMRI Techniques to Decrease Imaging Times in Children. *Radiographics* 40:485–502. <https://doi.org/10.1148/rg.2020190112>
10. Jaimes C, Kirsch JE, Gee MS (2018) Fast, free-breathing and motion-minimized techniques for pediatric body magnetic resonance imaging. *Pediatr Radiol* 48:1197–1208. <https://doi.org/10.1007/s00247-018-4116-x>
11. Jaimes C, Gee MS (2016) Strategies to minimize sedation in pediatric body magnetic resonance imaging. *Pediatr Radiol* 46:916–927. <https://doi.org/10.1007/s00247-016-3613-z>
12. Orth RC, Guillerman RP, Zhang W et al (2014) Prospective Comparison of MR Imaging and US for the Diagnosis of Pediatric Appendicitis. *Radiology* 272:233–240. <https://doi.org/10.1148/radiol.14132206>
13. Dillman JR, Gadepalli S, Sroufe NS et al (2016) Equivocal Pediatric Appendicitis: Unenhanced MR Imaging Protocol for Nonsedated Children—A Clinical Effectiveness Study. *Radiology* 279:216–225. <https://doi.org/10.1148/radiol.2015150941>
14. Moore MM, Gustas CN, Choudhary AK et al (2012) MRI for clinically suspected pediatric appendicitis: an implemented program. *Pediatr Radiol* 42:1056–1063. <https://doi.org/10.1007/s00247-012-2412-4>
15. Kim M, Kim HS, Kim HJ et al (2021) Thin-Slice Pituitary MRI with Deep Learning–based Reconstruction: Diagnostic Performance in a Postoperative Setting. *Radiology* 298:114–122. <https://doi.org/10.1148/radiol.2020200723>
16. Kim M, Kim HS, Park JE et al (2022) Thin-Slice Pituitary MRI with Deep Learning–Based Reconstruction for Preoperative Prediction of Cavernous Sinus Invasion by Pituitary Adenoma: A Prospective Study. *Am J Neuroradiol* 43:280–285. <https://doi.org/10.3174/ajnr.A7387>
17. Chaudhari AS, Fang Z, Kogan F et al (2018) Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med* 80:2139–2154. <https://doi.org/10.1002/mrm.27178>
18. Almansour H, Herrmann J, Gassenmaier S et al (2022) Combined Deep Learning-based Super-Resolution and Partial Fourier Reconstruction for Gradient Echo Sequences in Abdominal MRI at 3 Tesla: Shortening Breath-Hold Time and Improving Image Sharpness and Lesion Conspicuity. <https://doi.org/10.1016/j.acra.2022.06.003>. *Acad Radiol*
19. Kakigi T, Sakamoto R, Tagawa H et al (2022) Diagnostic advantage of thin slice 2D MRI and multiplanar reconstruction of the knee joint using deep learning based denoising approach. *Sci Rep* 12:10362. <https://doi.org/10.1038/s41598-022-14190-1>
20. Tajima T, Akai H, Yasaka K et al (2022) Clinical feasibility of an abdominal thin-slice breath-hold single-shot fast spin echo sequence processed using a deep learning-based noise-reduction approach. *Magn Reson Imaging* 90:76–83. <https://doi.org/10.1016/j.mri.2022.04.005>
21. Koch KM, Sherafati M, Arpinar VE et al (2021) Analysis and Evaluation of a Deep Learning Reconstruction Approach with Denoising for Orthopedic MRI. <https://doi.org/10.1148/ryai.2021200278>. *Radiol Artif Intell* 3:
22. Li Z, Wang Y, Yu J (2017) Reconstruction of Thin-Slice Medical Images Using Generative Adversarial Network. pp 325–333
23. Lebel RM (2020) Performance characterization of a novel deep learning-based MR image reconstruction pipeline

24. Zucker EJ, Sandino CM, Kino A et al (2021) Free-breathing Accelerated Cardiac MRI Using Deep Learning: Validation in Children and Young Adults. *Radiology* 300:539–548. <https://doi.org/10.1148/radiol.2021202624>
25. Bash S, Wang L, Airriess C et al (2021) Deep Learning Enables 60% Accelerated Volumetric Brain MRI While Preserving Quantitative Performance: A Prospective, Multicenter, Multireader Trial. *Am J Neuroradiol* 42:2130–2137. <https://doi.org/10.3174/ajnr.A7358>
26. Recht MP, Zbontar J, Sodickson DK et al (2020) Using Deep Learning to Accelerate Knee MRI at 3 T: Results of an Interchangeability Study. *Am J Roentgenol* 215:1421–1429. <https://doi.org/10.2214/AJR.20.23313>
27. Bash S, Johnson B, Gibbs W et al (2022) Deep Learning Image Processing Enables 40% Faster Spinal MR Scans Which Match or Exceed Quality of Standard of Care: A Prospective Multicenter Multireader Study. <https://doi.org/10.1007/s00062-021-01121-2>. *Clin Neuroradiol* 32:
28. Pocepcova V, Zellner M, Callaghan F et al (2025) Deep learning-based denoising image reconstruction of body magnetic resonance imaging in children. *Pediatr Radiol* 55:1235–1244. <https://doi.org/10.1007/s00247-025-06230-5>
29. Rai P, Mark IT, Soni N et al (2025) Deep Learning-Based Acceleration in MRI: Current Landscape and Clinical Applications in Neuroradiology. *Am J Neuroradiol* ajnr A8943. <https://doi.org/10.3174/ajnr.A8943>
30. Zormpas-Petridis K, Tunariu N, Curcean A et al (2021) Accelerating Whole-Body Diffusion-weighted MRI with Deep Learning–based Denoising Image Filters. *Radiol Artif Intell* 3:e200279. <https://doi.org/10.1148/ryai.2021200279>
31. Choi JW, Cho YJ, Lee SB et al (2025) Accelerated brain magnetic resonance imaging with deep learning reconstruction: a comparative study on image quality in pediatric neuroimaging. *Pediatr Radiol* 55:1903–1914. <https://doi.org/10.1007/s00247-025-06314-2>
32. Epstein KN, Trout AT, Debnath P et al (2023) Rapid, free-breathing non-contrast MRI for first-line imaging evaluation of ovarian torsion in the emergency department. *Pediatr Radiol*. <https://doi.org/10.1007/s00247-023-05827-y>
33. Gallo-Bernal S, Gee MS (2024) Commentary: Redefining the role of emergency department MRI in pediatric ovarian torsion evaluation. *Pediatr Radiol* 54:236–238. <https://doi.org/10.1007/s00247-023-05850-z>

Figures

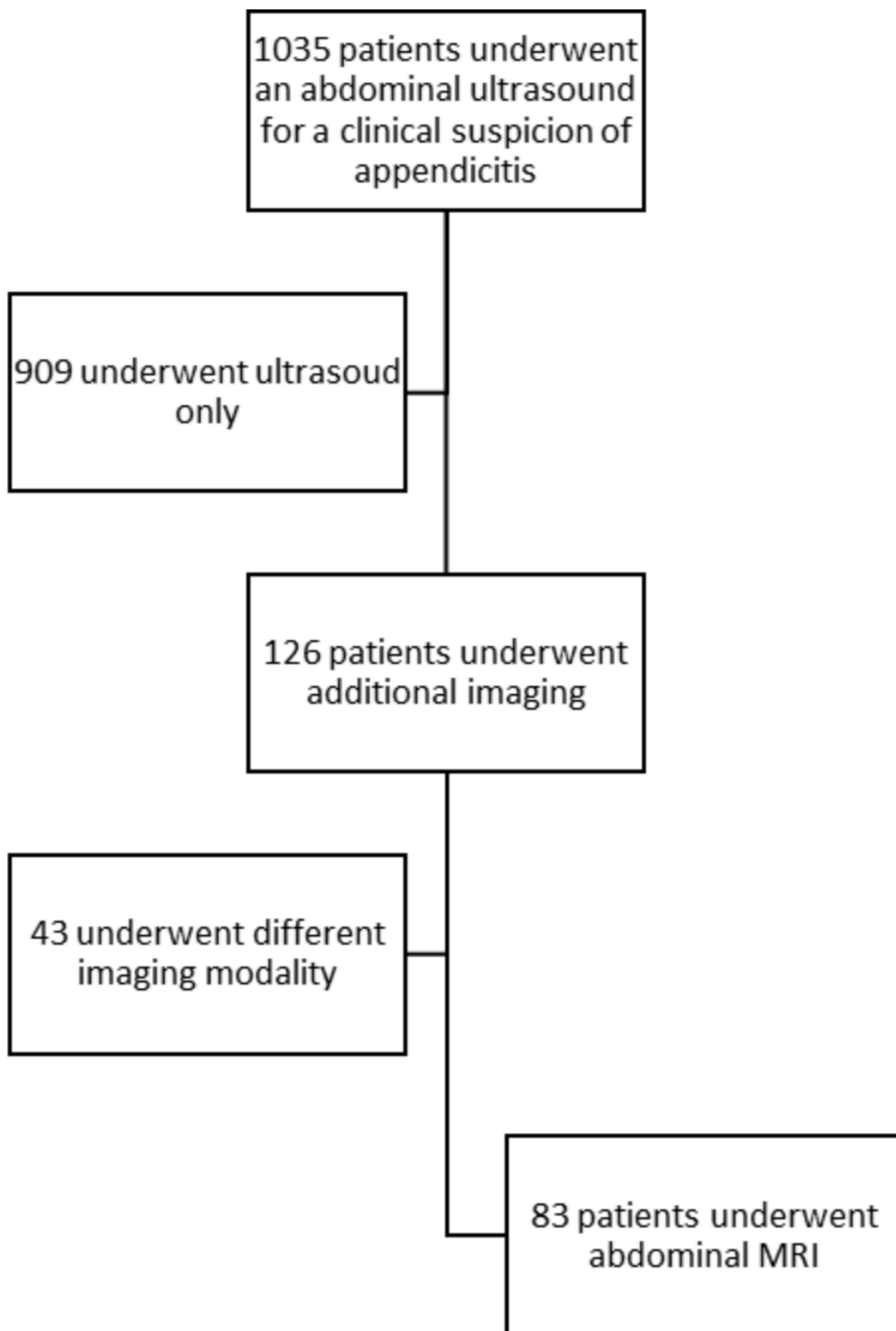


Figure 1

Flowchart shows patient selection process

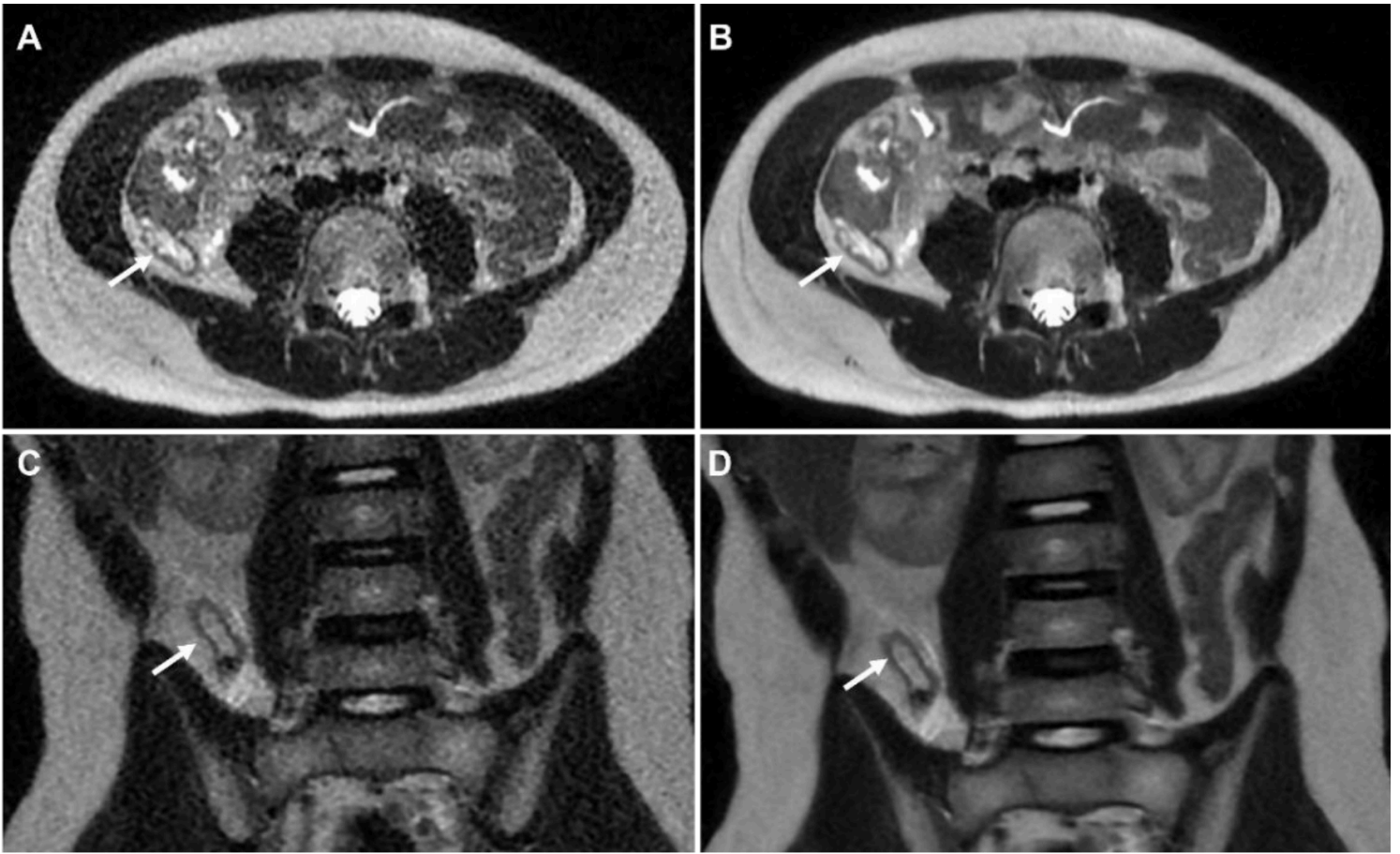


Figure 2

Deep learning-based reconstruction for 2 mm images in a 12-year-old girl with right lower quadrant pain and clinical suspicion of appendicitis. Axial and Coronal single-shot T2-weighted images before (A and C, respectively) and after (B and D, respectively) deep learning reconstruction. Note the significant improvement in noise, signal-to-noise ratio, and noise distribution. Images showed an edematous and enlarged appendix (white arrow), consistent with appendicitis.

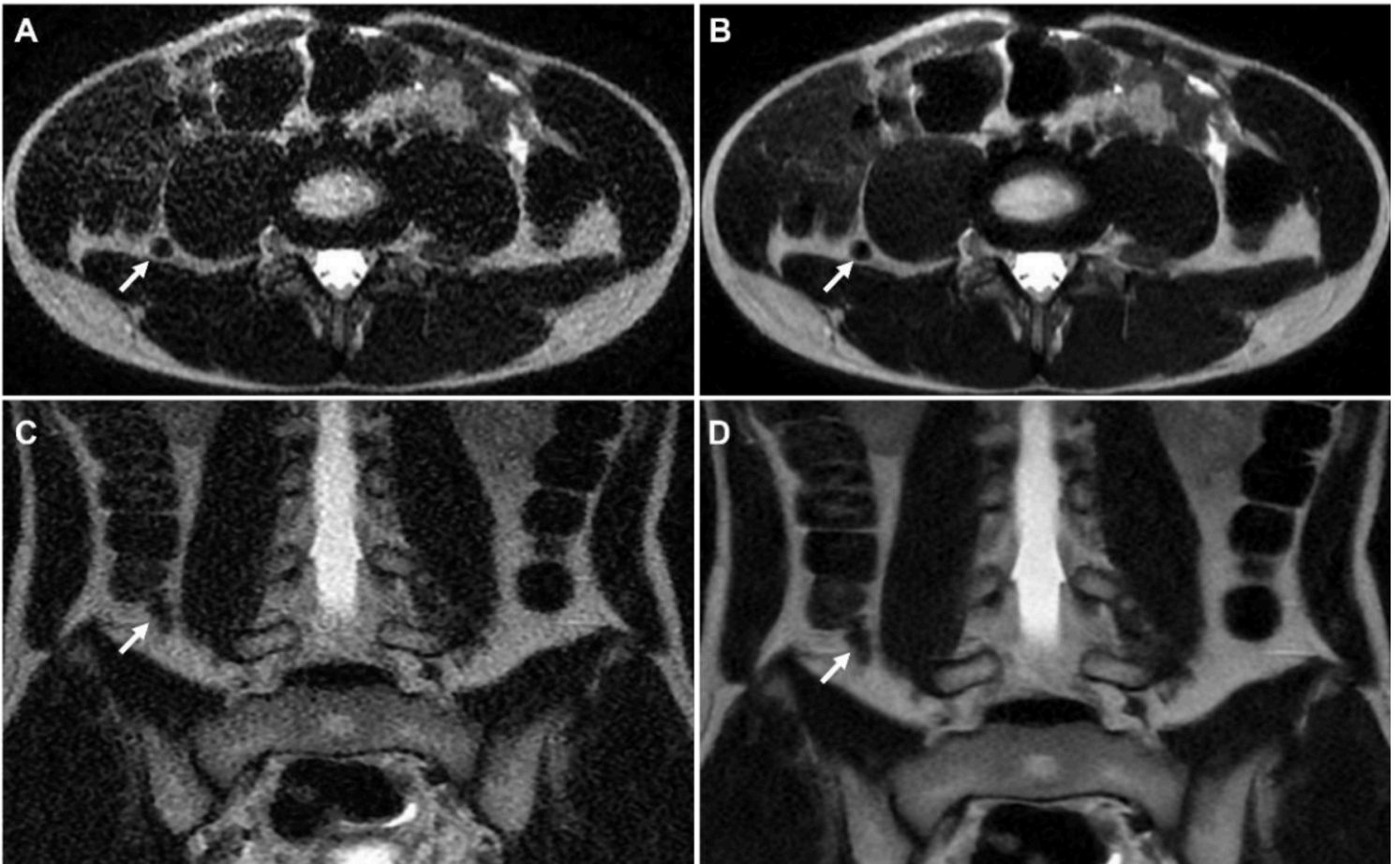


Figure 3

Deep learning-based reconstruction for acquiring 2 mm (thin-sliced) images in a 15-year-old boy with right lower quadrant pain. Axial and Coronal single-shot T2-weighted images before (A and C, respectively) and after (B and D, respectively) deep learning reconstruction. Images showed a normal-appearing appendix without significant mural thickening

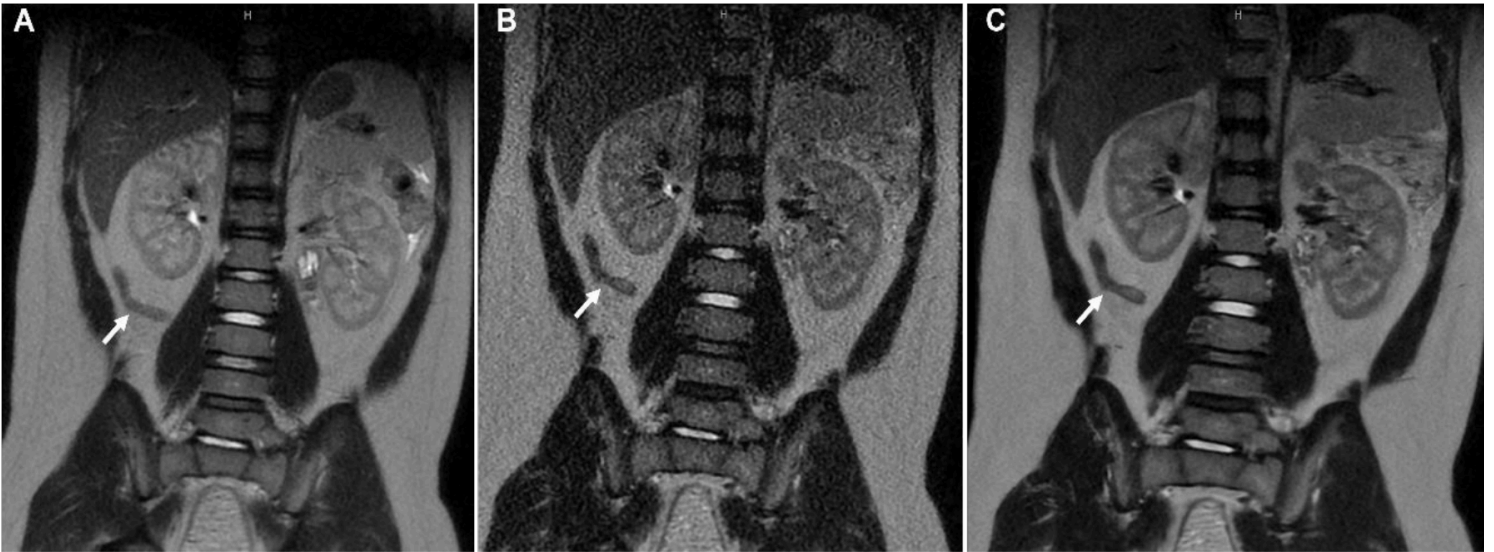


Figure 4

MRI images in a 6-year-old with abdominal pain and equivocal ultrasound findings. A) standard 4-mm, B) conventional 2-mm (no-DLR), and C) deep learning reconstructed 2-mm (2MM+DLR) coronal images.

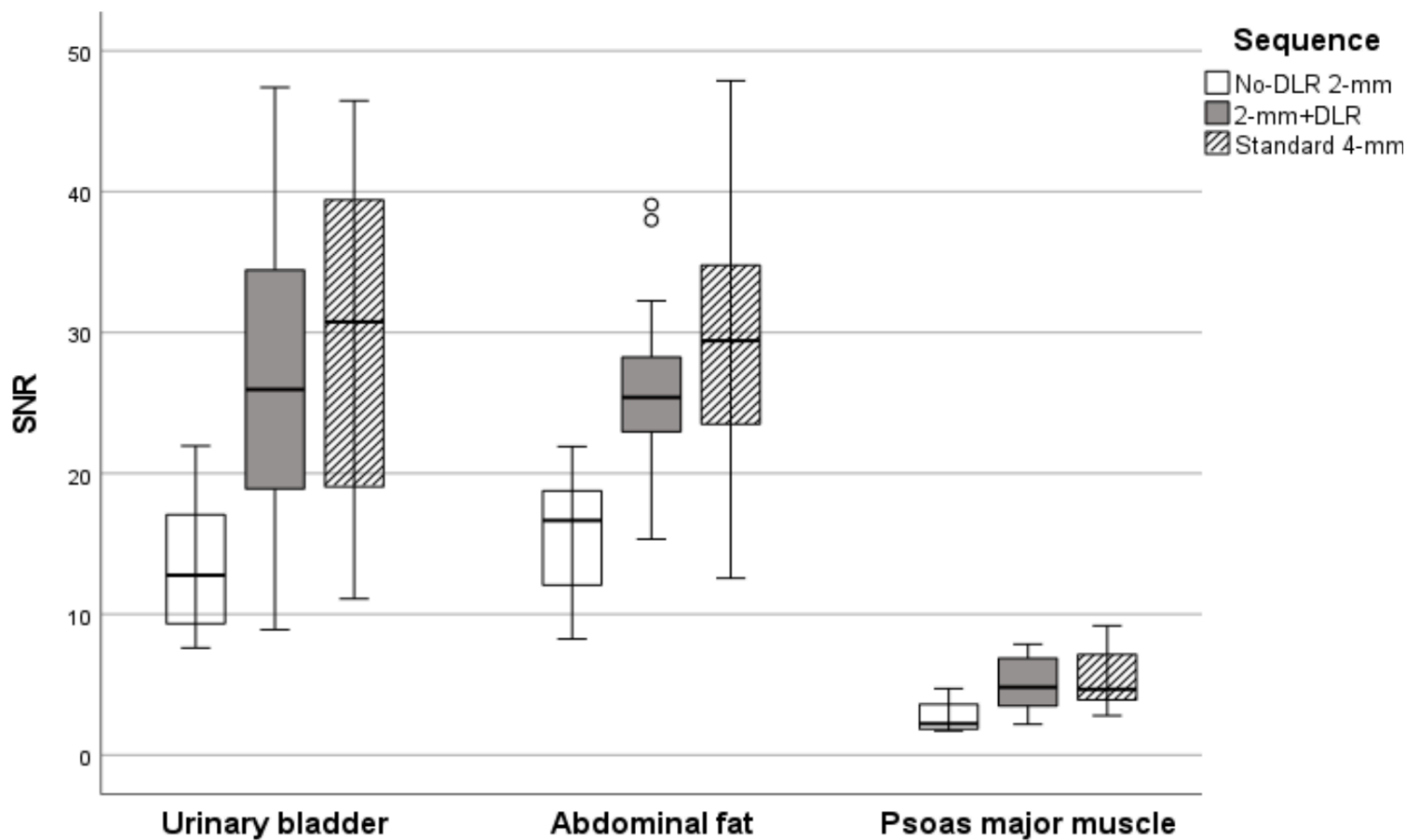


Figure 5

Box and whisker plot comparing the standard 4-mm, the no-DLR 2-mm, and the 2-mm+DLR images for the computed signal-to-noise at the level of the urinary bladder, abdominal fat, and psoas major muscle. The upper and lower borders represent the interquartile range (25th and 75th percentiles), the solid line within each box represents the median, and the whiskers represent the minimum and maximum values within 1.5 times the interquartile range. Circles denote outlying values ($>$ upper quartile + 1.5 times the interquartile range).

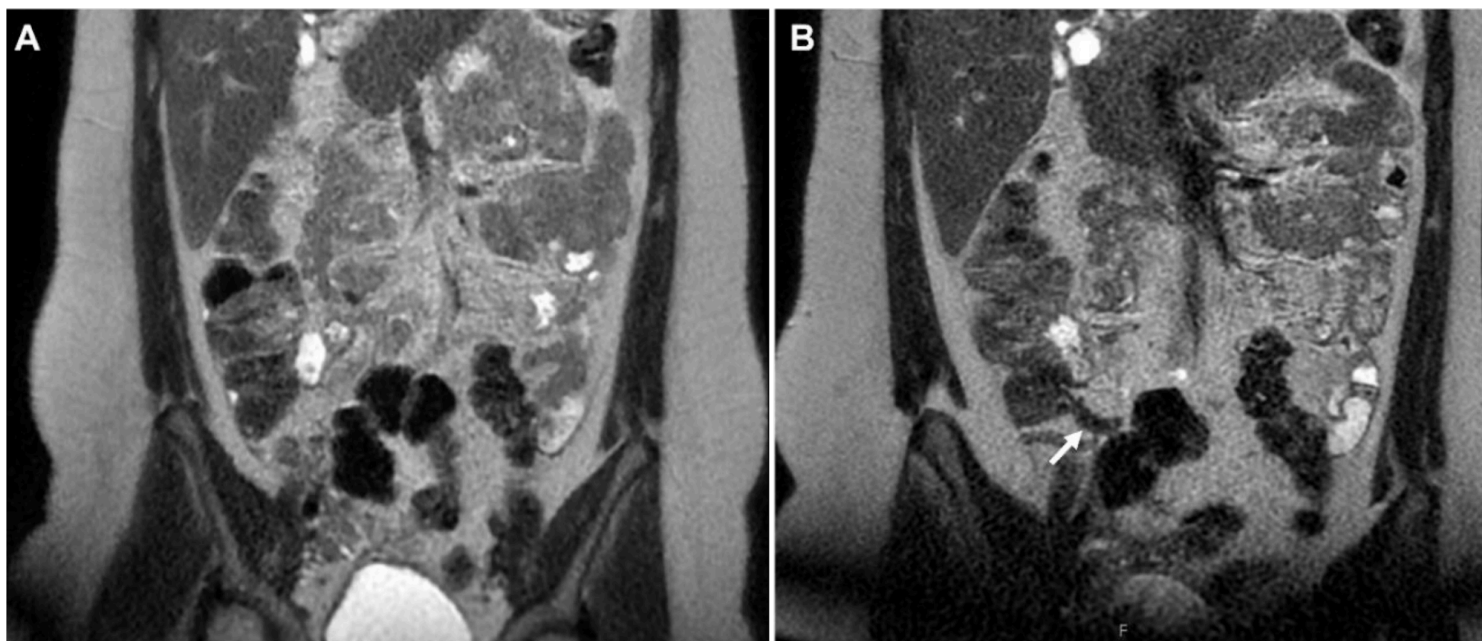


Figure 6

MRI images in a 17-year-old female patient with leukocytosis and right lower quadrant pain. A) Standard 4-mm and B) 2-mm (thin-slice) deep learning reconstructed images (2mm+DLR) in the coronal plane. The appendix was only identified in the thin-slice images (arrow in B).