

Original Article

Diagnostic accuracy of computed tomography for acute appendicitis: Comparison among in-training, junior, and senior radiologists

Puwitch Charoenchue, M.D., Ph.D.⁽¹⁾

Nartanong Jaijina, M.D.^(1, 2)

Suwalee Pojchamarnwiputh, M.D.⁽¹⁾

Natipat Jitmahawong, M.D.⁽¹⁾

Wittanee Na Chiangmai, M.D.⁽¹⁾

Amonlaya Amantakul, M.D., Ph.D.⁽¹⁾

From ⁽¹⁾Department of Radiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand,

⁽²⁾Department of Radiology, Srisangwan Hospital, Mae Hong Son, Thailand.

Address correspondence to A.A. (email: amonlaya.amantakul@cmu.ac.th)

Received 17 March 2026; revised 3 August 2026; accepted 9 September 2026
doi:10.46475/asean-jr.v27i3.1026

Abstract

Background: Computed tomography (CT) is widely used for diagnosing acute appendicitis, but the impact of radiologist experience on diagnostic accuracy remains unclear. This study compares the diagnostic performance of in-training, junior, and senior radiologists to assess whether experience influences accuracy and interobserver agreement.

Materials and Methods: A retrospective study analyzed CT scans of 196 patients suspected of having acute appendicitis from January to December 2020. Three radiologists—in training, junior, and senior—interpreted the scans independently. The diagnostic performance of each reader was compared based on pathological or clinical follow-up. Diagnostic accuracy indices were calculated for each reader, and interobserver agreement was assessed.

Results: CT demonstrated high diagnostic accuracy across all reader groups. The senior radiologist achieved the highest accuracy (94.4%; 95% CI, 90.2-97.2), followed by the junior radiologist (93.9%; 95% CI, 89.5-96.8) and the in-training radiologist (91.3%; 95% CI, 86.5-94.9), with no statistically significant differences. The AUROC values were 0.944

for the senior radiologist, 0.938 for the junior radiologist, and 0.912 for the in-training radiologist. Interobserver agreement was excellent, with Fleiss' kappa of 0.864 (95% CI, 0.806-0.921).

Conclusion: CT showed high diagnostic performance for acute appendicitis across different levels of radiologist experience in this cohort. Junior and senior radiologists demonstrated slightly higher numerical accuracy than the in-training radiologist, mainly due to fewer false-negative interpretations. These findings suggest comparable CT diagnostic results among in-training, junior, and senior radiologists, while emphasizing the importance of careful assessment of subtle or atypical cases.

Keywords: Acute appendicitis, Computed tomography, Diagnostic accuracy, Experience, Radiology training.

Introduction

Acute appendicitis is a frequently encountered cause of abdominal pain in emergency departments and requires prompt surgical intervention. Failure to diagnose or delays in diagnosis can lead to severe health complications and even death [1]. Worldwide, there are approximately 100 to 230 cases of acute appendicitis per 100,000 person-years, with the incidence rising in newly developed countries [2, 3]. Additionally, during the COVID-19 pandemic, the proportion of complicated appendicitis cases has increased [4].

Clinical evaluation remains a crucial first step in diagnosing acute appendicitis, particularly in cases with atypical presentations. The Alvarado score, introduced by Dr. Alfredo Alvarado in 1986, is a widely used clinical scoring system that helps assess the likelihood of appendicitis based on a combination of symptoms, signs, and laboratory findings [5, 6]. A score of 7 or higher suggests a high probability of appendicitis, often warranting surgical intervention. While the Alvarado score has a high sensitivity for ruling out appendicitis, its specificity is lower, necessitating imaging in ambiguous cases [6, 7].

CT has revolutionized the diagnosis of acute appendicitis, with reported sensitivity and specificity exceeding 95% and 94%, respectively [8-11]. Due to its superior diagnostic performance, CT is now included in national guidelines in multiple countries [12-14]. Key imaging findings for appendicitis include an appendiceal diameter >6 mm (93% sensitivity, 100% specificity), periappendiceal fat stranding (100% sensitivity, 80% specificity), and focal cecal thickening (69% sensitivity, 100% specificity) [15].

Radiologist experience is a crucial factor influencing the accuracy of CT interpretation. Several studies have shown that less experienced radiologists have lower sensitivity and specificity in diagnosing acute appendicitis [16]. In most medical centers, initial CT interpretations are conducted by in-training residents, particularly in emergency settings [17-19]. However, misinterpretation by less experienced radiologists may lead to false-negative or false-positive diagnoses, resulting in unnecessary surgeries or missed cases of appendicitis [20].

Studies have shown that structured training programs and continuous radiology education improve diagnostic accuracy among in-training radiologists [21]. The critical question remains: Does a radiologist's experience significantly impact the accuracy of diagnosing acute appendicitis? If expertise is a significant factor, hospitals should prioritize ensuring that board-certified radiologists interpret CT scans for suspected appendicitis.

This study aimed to compare the diagnostic performance and interobserver agreement of CT interpretation for acute appendicitis among in-training, junior, and senior radiologists.

Materials and methods

Study Design and Population

This retrospective, cross-sectional diagnostic accuracy study was conducted at a university hospital. The study compared the diagnostic accuracy of a second-year in-training radiologist with that of two board-certified abdominal radiologists, one with eight years of experience and the other with 20 years of experience. The authors assessed diagnostic accuracy by reanalyzing CT images stored in the hospital's Picture Archiving and Communication System (PACS).

The study encompassed patients aged ≥ 18 years who presented with acute abdominal pain and were referred for CT evaluation due to suspected acute appendicitis between January 2020 and December 2020. We identified cases using a structured search strategy that included the following terms: "abdominal," "pain," "appendicitis," "RLQ pain," or "lower abdominal pain" in the hospital's electronic medical records.

Patients were excluded if they had a history of appendectomy, known intra-abdominal malignancy, trauma, pregnancy, or insufficient post-treatment follow-up data for diagnostic validation. Figure 1 illustrates the study workflow, including patient selection, exclusion criteria, imaging interpretation, and the validation process.

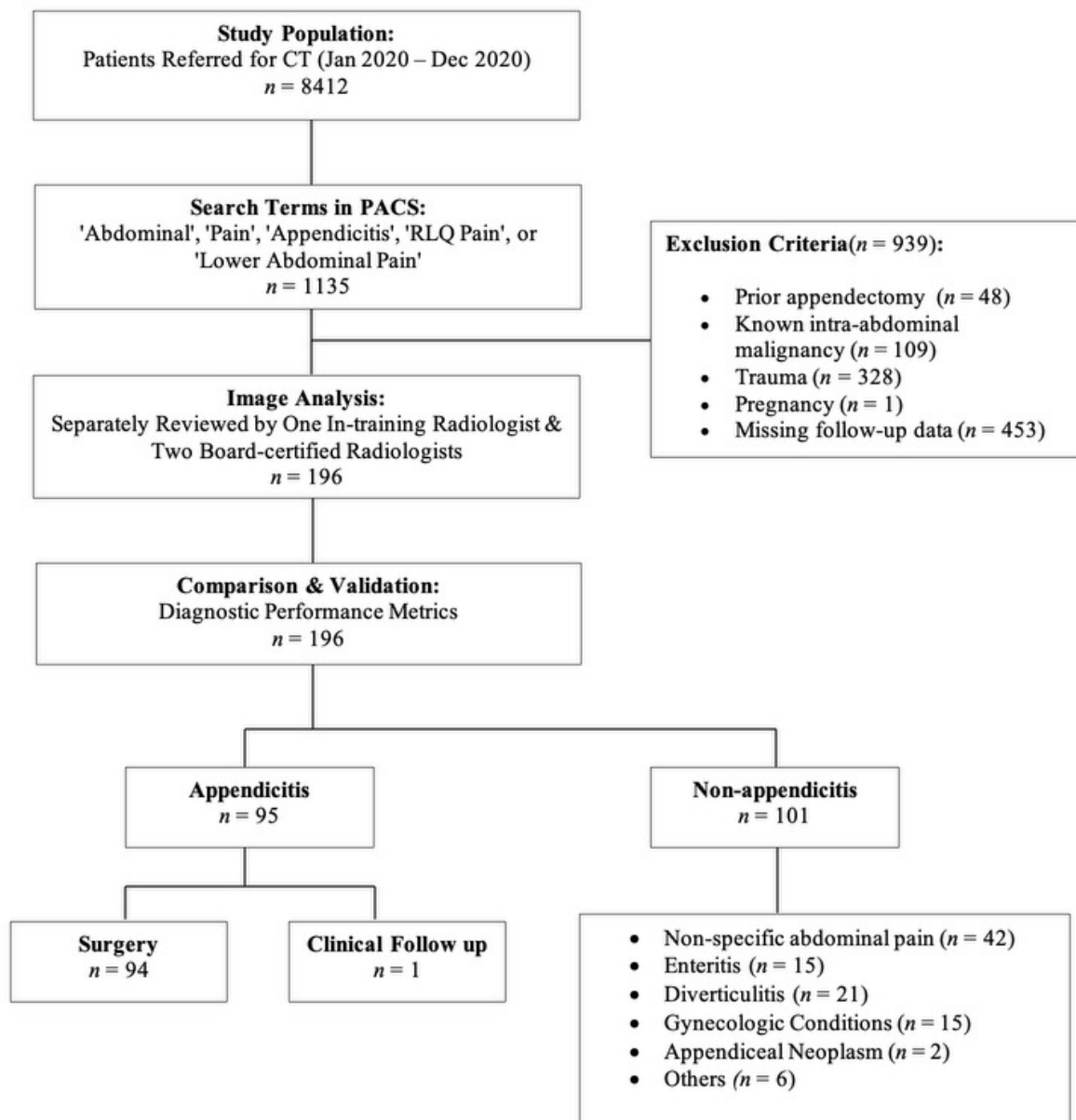


Figure 1. Study flow diagram outlining patient selection, exclusion criteria, imaging interpretation, and validation. Of 1,135 patients referred for abdomino-pelvic CT, 196 met the inclusion criteria, including 95 patients with appendicitis and 101 patients without appendicitis. Final diagnoses in the non-appendicitis group included non-specific abdominal pain (n = 42), enteritis (n = 15), diverticulitis (n = 21), gynecologic conditions (n = 15), appendiceal neoplasm (n = 2), and other diagnoses (n = 6). CT images were retrospectively analyzed, with diagnoses validated against pathology or clinical follow-up.

CT Imaging Protocol

All patients underwent contrast-enhanced abdominal CT scans using one of three multidetector CT systems: SOMATOM® Definition (Siemens Healthcare, Forchheim, Germany), SOMATOM® Force (Siemens Healthcare, Forchheim, Germany), or Philips IQon (Philips Healthcare, Best, The Netherlands). The standardized CT protocol included a plain scan of the upper abdomen (if clinically indicated for stone evaluation) and a portal venous phase post-contrast scan covering the entire abdomen.

Intravenous iodinated contrast agents were administered according to patient body weight, up to a maximum dose of 100 mL. The contrast agents included iodixanol (320 mg I/mL), iohexol (350 mg I/mL), ioversol (350 mg I/mL), and iopamidol (370 mg I/mL). Portal venous phase images were acquired approximately 60 seconds after contrast injection. No oral or rectal contrast material was used.

Image reconstruction involved axial images with 2-mm and 5-mm thicknesses, as well as coronal and sagittal images with 5-mm thicknesses. No delayed phase scans were performed. Although some scans were acquired using dual-energy spectral CT systems, only conventional-like (mixed-energy) images were analyzed to ensure the generalizability of the results across different CT systems.

Image Analysis

All CT scans were independently reviewed by a second-year radiology resident and two board-certified abdominal radiologists with 8 and 20 years of experience, respectively. The retrospective readers were blinded to the final diagnosis, histopathologic results, clinical outcome, and original CT report. The basic clinical indication was available to reflect routine clinical CT interpretation. Readers evaluated predefined CT features and assigned a four-level diagnostic confidence score:

- 1 = confidently negative for appendicitis, including a normal appendix or an alternative diagnosis without CT findings of appendicitis
- 2 = probably negative for appendicitis, including a visualized appendix without convincing inflammatory change or minimal equivocal findings considered insufficient for the diagnosis of appendicitis
- 3 = probably positive for appendicitis, including equivocal but suspicious appendiceal enlargement, wall enhancement, or periappendiceal inflammatory change
- 4 = confidently positive for appendicitis, based on typical CT findings of acute appendicitis, complicated appendicitis, abscess, or perforation

Before image interpretation, all readers received standardized written instructions describing the four-level confidence scoring system. Although scores 2 and 3 both represented equivocal cases, score 2 was assigned when the overall CT findings were considered insufficient to support appendicitis, whereas score 3 was assigned when the imaging findings were equivocal but the overall assessment favored appendicitis based on the reader's integrated assessment of all predefined CT features. This approach was intended to reflect routine clinical decision-making rather than strict quantitative imaging thresholds.

Scores 2 and 3 were used to reflect equivocal or indeterminate real-world interpretations. For the primary diagnostic accuracy analysis, the four-level scores were dichotomized according to the clinically actionable diagnostic impression:

Scores 1-2 = negative for appendicitis

Scores 3-4 = positive for appendicitis

ROC curve analysis was performed using this dichotomized diagnostic output. Therefore, the reported AUROC values represent discrimination at the clinically relevant decision threshold used in routine practice. This approach reflects real-world reporting behavior, in which radiologists ultimately provide a binary impression for clinical management: appendicitis or no appendicitis.

The initial CT report was defined as the original clinical radiology report issued during routine patient care before retrospective image review. In our institutional workflow, emergency CT examinations were initially interpreted by the on-duty radiology resident and subsequently reviewed according to routine departmental practice. The initial CT report was analyzed separately as a real-world clinical comparator and was not included as one of the three blinded retrospective reader interpretations. Examples of images illustrating these criteria are shown in Figure 2.

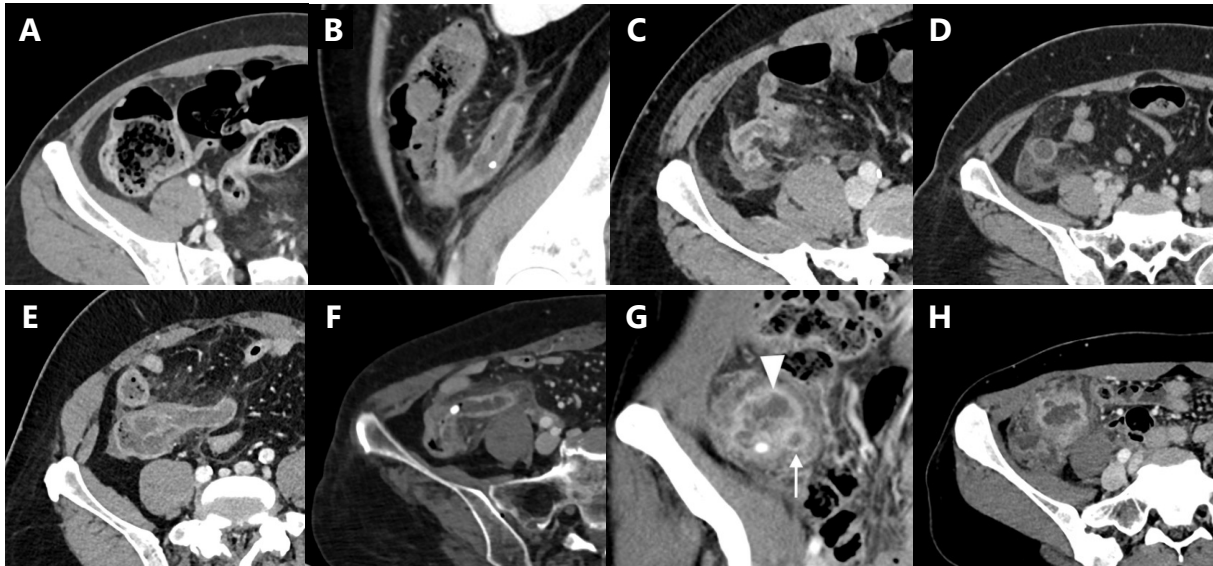


Figure 2. CT findings for diagnosing acute appendicitis. (A) Normal appendix with a normal size and wall thickness. (B) Acute suppurative appendicitis, confirmed by tissue pathology, showing an enlarged, fluid-filled appendix. (C) Appendiceal wall hyperenhancement, a key finding in acute appendicitis. (D) Periappendiceal fat stranding, indicative of inflammatory changes surrounding the appendix. (E) Cecal wall swelling, representing reactive changes in the adjacent cecum. (F) Appendicolith at the appendix base, a potential obstructive factor in appendicitis. (G) Ruptured appendix, showing a dilated appendix (arrow) with a fluid collection likely representing an abscess (arrowhead). (H) Another case of appendiceal abscess, demonstrating a thick-walled fluid collection adjacent to the appendix.

Reference Standard

For patients who underwent surgery, histopathologic examination of the resected appendix served as the definitive reference standard. For patients managed non-operatively, clinical follow-up through documented outpatient visits or inpatient observation was used to confirm the absence of appendicitis. All non-surgical patients had a minimum follow-up period of 7 days, during which one of the following clinical criteria was required to classify the case as not having appendicitis:

1. Complete resolution of abdominal pain without surgical intervention;
2. No subsequent diagnosis of appendicitis or related complications within the follow-up window; and
3. No readmission to the hospital with right lower quadrant pain.

Cases without sufficient clinical follow-up were excluded to maintain the reliability of the reference standard.

Statistical Analysis

Descriptive statistics were used to summarize demographic and clinical characteristics. Continuous variables were reported as means with standard deviations or medians with interquartile ranges, as appropriate. Categorical variables were reported as numbers and percentages. Comparisons between patients with and without appendicitis were performed using the Student's t-test or Mann-Whitney U test for continuous variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

Diagnostic performance was calculated for each reader and for the initial CT report, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), diagnostic accuracy, and the area under the receiver operating characteristic curve (AUROC). Because the primary diagnostic output was dichotomized as positive or negative for appendicitis, ROC analysis was performed using the binary diagnostic classification. Ninety-five percent confidence intervals (95% CIs) were calculated for diagnostic indices. Diagnostic accuracy values were compared among readers using the chi-square test.

Interobserver agreement among readers was assessed using percent agreement, pairwise Cohen's kappa, and multi-rater Fleiss' kappa. Kappa values were interpreted as follows: less than 0.20, poor agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, good agreement; and 0.81-1.00, excellent agreement.

All statistical analyses were performed using STATA version 19.0 (StataCorp, College Station, TX, USA). A p-value less than 0.05 was considered statistically significant.

Results

A total of 1,135 patients underwent CT evaluation for acute abdominal pain. After applying the inclusion and exclusion criteria, 196 patients were included, with 95 (48.5%) diagnosed with acute appendicitis and 101 (51.5%) classified as having no appendicitis. Among the 101 patients without appendicitis, the final diagnoses were nonspecific abdominal pain (n = 42), enteritis (n = 15), diverticulitis (n = 21), gynecologic conditions (n = 15), appendiceal neoplasm (n = 2), and other diagnoses (n = 6).

Patient Characteristics

The patient characteristics of the study population are summarized in Table 1. Age, sex distribution, underlying diseases, clinical symptoms, and physical examination findings were generally comparable between patients with and without appendicitis. The appendicitis group had significantly higher WBC counts, neutrophil percentages, and Alvarado scores. Urinalysis findings did not significantly differ between groups.

Table 1. Patient characteristics of the included patients evaluated for acute appendicitis.

Parameters	Appendicitis n=95 (48.5%)	Not appendicitis n=101 (51.5%)	p-value
Age (year, mean)	48.23 (19.49)	49.65(20.76)	0.62
Sex			
Male	34(35.79)	35(34.65)	0.88
Female	61(64.21)	66(65.35)	
Underlying			
DM	4(4.21)	11(10.89)	0.07
Hypertension	19(20.0)	25(24.75)	0.27
Immunocompromised	0(0%)	1(0.99)	0.52
Others	30(31.58)	57(56.44)	0.001
Symptoms			
RLQ pain	70(73.68)	66(65.35)	0.13
Fever	23(24.21)	28(27.72)	0.35
Nausea or vomiting	44(46.32)	48(47.52)	0.49
Diarrhea	21(22.11)	21(20.79)	0.48
Physical signs			
Body temperature (degree Celsius, median)	36.9(1)	36.8(1.1)	0.71
RLQ tenderness	93(97.89)	100(99.01)	0.48
Rebound tenderness	37(38.95)	34(33.66)	0.27
Blood test			
WBC count (cells per cc, median)	13,600(5910)	11,260(5365)	0.003*
Neutrophil (percent, median)	83(11.6)	79.9(13.65)	0.01*
Alvarado Score (median)	7(2)	6(3)	0.02*
Urinalysis			
WBC > 5-10 cell/HPF	15(15.79)	18(17.82)	0.43
RBC > 5-10 cell/HPF	21(22.11)	25(24.75)	0.40

Abbreviations: DM, Diabetes mellitus; RLQ, right lower quadrant; WBC, white blood cell; RBC, red blood cell

The numbers in parentheses represent the standard deviations (SD) for the means, interquartile ranges (IQR) for the medians, and proportional percentages for counting data.

Continuous variables with non-normal distributions were compared using the Mann-Whitney U test.

Diagnostic Performance

The diagnostic performance of the in-training radiologist, junior radiologist, senior radiologist, and initial CT report is summarized in Table 2. All readers demonstrated high diagnostic performance for acute appendicitis. The senior radiologist achieved the highest accuracy among the three retrospective readers (94.4%; 95% CI, 90.2-97.2), followed by the junior radiologist (93.9%; 95% CI, 89.5-96.8) and the in-training radiologist (91.3%; 95% CI, 86.5-94.9). The differences among the readers were not statistically significant ($p = 0.133$).

The in-training radiologist had 11 false-negative and 6 false-positive interpretations, the junior radiologist had 7 false-negative and 5 false-positive interpretations, and the senior radiologist had 6 false-negative and 5 false-positive interpretations. To evaluate the consistency of missed diagnoses across experience levels, a case-level analysis of false-negative interpretations was performed. All 6 false-negative cases identified by the senior radiologist were also misclassified as false negatives by both the junior and in-training radiologists. The junior radiologist missed 1 additional case (for a total 7 false-negative cases), which was also missed by the in-training radiologist. The in-training radiologist missed 4 unique cases that were correctly diagnosed as positive by both the junior and senior radiologists. All false-negative cases missed by the more experienced readers were also missed by the less experienced readers, suggesting that greater experience may improve the detection of subtle cases. Therefore, the modest numerical difference in performance was mainly attributable to fewer false-negative interpretations among the more experienced readers. The initial CT report showed 95 true-positive, no false-negative, 8 false-positive, and 93 true-negative interpretations, corresponding to an accuracy of 95.9% (95% CI, 92.1-98.2).

Table 2. *Diagnostic performance of CT for acute appendicitis according to reader group.*

Metric	In-Training Radiologist	Junior Radiologist	Senior Radiologist	Initial CT Report
True positive	84	88	89	95
False negative	11	7	6	0
False positive	6	5	5	8
True negative	95	96	96	93
Sensitivity (%)	88.4 (80.2-94.1)	92.6 (85.7-97.0)	93.7 (86.8-97.6)	100.0 (96.2-100.0)
Specificity (%)	94.1 (87.5-97.8)	95.0 (88.8-98.4)	95.0 (88.8-98.4)	92.1 (85.0-96.5)
PPV (%)	93.3 (86.1-97.5)	94.6 (87.9-98.2)	94.7 (88.0-98.3)	92.2 (85.3-96.6)
NPV (%)	89.6 (82.2-94.7)	93.2 (86.5-97.2)	94.1 (87.6-97.8)	100.0 (96.1-100.0)
Accuracy (%)	91.3 (86.5-94.9)	93.9 (89.5-96.8)	94.4 (90.2-97.2)	95.9 (92.1-98.2)
AUROC	0.912 (0.873-0.952)	0.938 (0.905-0.972)	0.944 (0.911-0.976)	0.960 (0.934-0.987)

The ROC analysis for each reader and the initial CT report is shown in Figure 3. The AUROC values were 0.912 for the in-training radiologist, 0.938 for the junior radiologist, 0.944 for the senior radiologist, and 0.960 for the initial CT report. Although the more experienced radiologists showed slightly higher numerical performance, the differences were not statistically significant. These findings suggest that CT diagnostic performance was comparable across different levels of radiologist experience in this cohort.

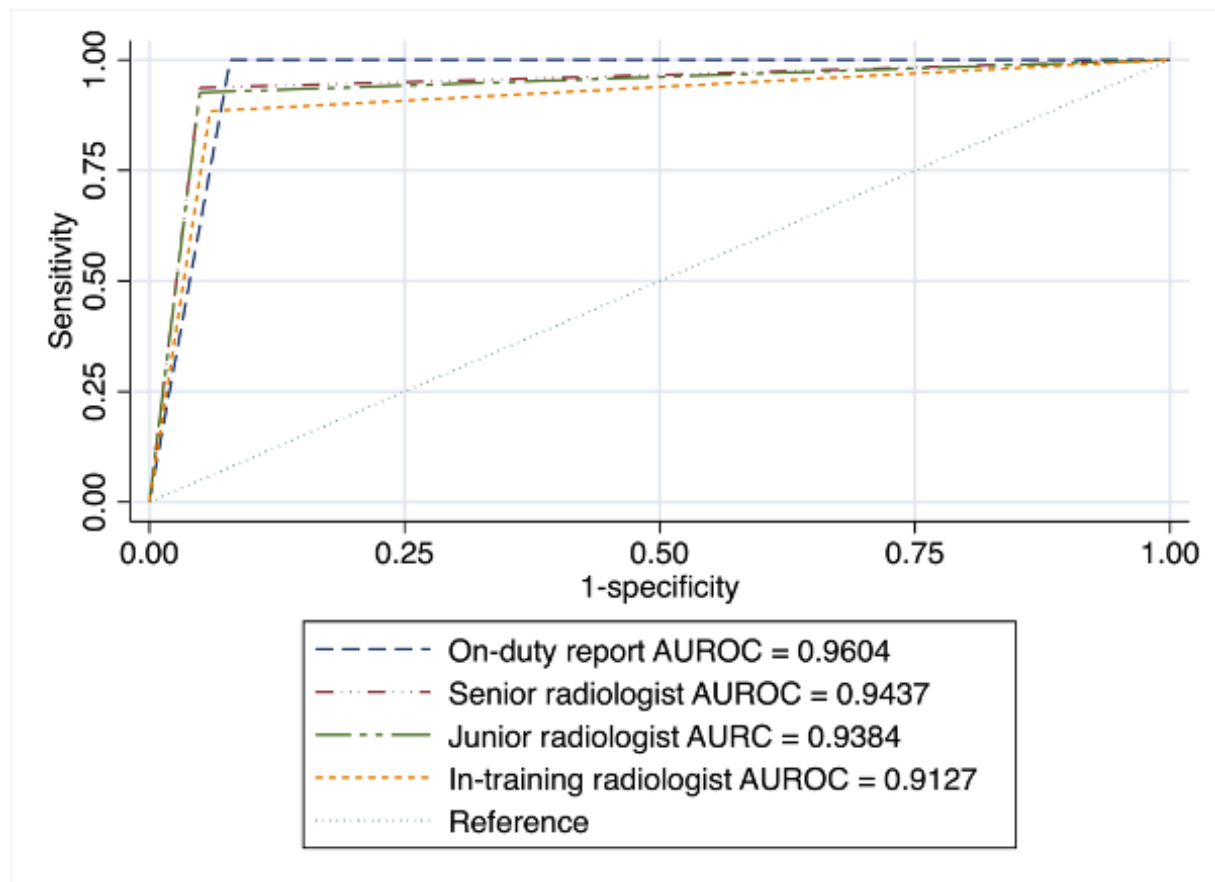


Figure 3. Receiver operating characteristic (ROC) curves for diagnosing acute appendicitis using CT. The AUROC values are shown for each interpretation: initial CT report (0.960), senior radiologist (0.944), junior radiologist (0.938), and in-training radiologist (0.912).

Interobserver Agreement

Interobserver agreement among the three readers was excellent. The overall percent agreement was 93.2% (95% CI, 90.4-96.1%). The multi-rater agreement analysis showed a Fleiss' kappa of 0.864 (95% CI, 0.806-0.921), indicating excellent reproducibility of CT interpretation for acute appendicitis. Pairwise Cohen's kappa values were also excellent: 0.867 (95% CI, 0.727-1.000) for the in-training versus junior radiologists, 0.857 (95% CI, 0.717-0.997) for the in-training versus senior radiologists, and 0.867 (95% CI, 0.727-1.000) for the junior versus senior radiologists.

Discussion

This study showed high diagnostic performance of CT for acute appendicitis across different levels of radiologist experience. Among the three retrospective readers, the senior radiologist achieved the highest accuracy, followed by the junior radiologist and the in-training radiologist. However, the differences were not statistically significant. Therefore, our findings should not be interpreted as evidence that radiologist experience has no effect, but rather that CT interpretation showed comparable diagnostic performance among in-training, junior, and senior radiologists in this selected cohort. The high overall performance observed in this study is consistent with the established role of CT as a reliable imaging modality for suspected acute appendicitis. Current ACR Appropriateness Criteria support CT as an important imaging test in adult patients with right lower quadrant pain and suspected appendicitis, particularly when clinical findings are equivocal or when alternative diagnoses need to be assessed [22]. A Cochrane review also reported high diagnostic performance of CT for acute appendicitis in adults, supporting its use as a robust diagnostic tool in routine clinical practice [23]. The 2020 WSES Jerusalem guidelines elaborate the diagnostic workup of acute appendicitis as a combined clinical, laboratory, and radiological process, reflecting the real-world need to integrate imaging with clinical judgment [24].

The prevalence of appendicitis in our cohort was 48.5%, which likely reflects the retrospective search strategy and case selection rather than the prevalence in the general emergency department population. This is relevant because PPV and NPV are affected by disease prevalence. Therefore, the predictive values in this study should be interpreted in the context of a selected group of patients who underwent CT for suspected appendicitis, rather than as estimates for all patients presenting with abdominal pain.

Clinical findings, laboratory markers, and clinical scoring systems remain useful for initial triage, but they were not the primary focus of this study. In our cohort, WBC count, neutrophil percentage, and Alvarado score were higher in patients with appendicitis, but overlap between the appendicitis and non-appendicitis groups remained. This supports the practical role of CT in patients with equivocal or overlapping clinical presentations. The WSES guidelines similarly state that clinical scores can guide risk stratification and imaging decisions, but the diagnostic workup still relies on integration of clinical, laboratory, and imaging findings [24].

Although overall accuracy was high for all readers, the numerical differences between experience levels were mainly due to false-negative interpretations. The in-training radiologist had 11 false-negative interpretations, compared with 7 for the junior radiologist and 6 for the senior radiologist. This pattern suggests that experience may

be most relevant in cases with subtle or early inflammatory changes rather than in typical appendicitis.

Analysis of the overlap in false-negative cases showed that all 6 cases missed by the senior radiologist were also missed by the junior and in-training radiologists. In contrast, 4 additional cases missed by the in-training radiologist were correctly identified by both board-certified radiologists. These findings suggest that greater reader experience may improve the detection of subtle or early appendicitis cases that are more likely to be missed by less-experienced readers, rather than resulting in readers missing completely different cases. The two cases shown in Figure 4 illustrate this finding, in which subtle features, such as minimal wall thickening, limited periappendiceal inflammation, or subtle enhancement without marked dilatation, were missed by the in-training and junior radiologists but correctly identified by the senior radiologist. Educational efforts for residents should therefore focus on equivocal cases, including mildly enlarged appendices, subtle wall thickening or enhancement, minimal periappendiceal fat stranding, tip appendicitis, and atypical appendix locations.

False-positive interpretations were less frequent and were similar among the three retrospective readers, with 6 false-positive interpretations by the in-training radiologist and 5 each by the junior and senior radiologists. Two examples are shown in Figure 5. These cases highlight the importance of considering alternative diagnoses when appendiceal wall thickening, luminal dilatation, or cecal involvement is present, particularly when the imaging findings or clinical presentation are not typical for uncomplicated appendicitis.

The initial CT report showed the highest sensitivity and negative predictive value, with no false-negative interpretations, but had more false-positive interpretations than the retrospective readers. This finding likely reflects real-world emergency reporting behavior, in which radiologists may favor sensitivity to avoid missing acute appendicitis. While this approach is clinically understandable, it may reduce specificity and increase false-positive diagnoses. The initial CT report should therefore be interpreted as a real-world clinical comparator rather than a direct measure of individual reader expertise.

Interobserver agreement among the three retrospective readers was excellent, with an overall percent agreement of 93.2% and a Fleiss' kappa of 0.864. This supports the reproducibility of CT-based diagnosis of acute appendicitis when standardized CT criteria are applied. However, excellent agreement does not eliminate the importance of experience. Rather, it suggests that most cases in this cohort had sufficiently clear CT findings to allow consistent interpretation across readers, while subtle cases remained more challenging for less experienced readers.

This study has several limitations. First, the retrospective design may have introduced selection bias. Clinical management may also have been influenced by the initial CT report, and mild uncomplicated appendicitis treated conservatively or cases not clinically recognized as appendicitis may not have been captured. Second, the sample size was relatively small, and no formal sample size calculation was performed before the study. Third, the reference standard was mixed, using histopathology in surgical cases and clinical follow-up in non-operative cases. Among the non-appendicitis group, 8 patients had normal appendiceal pathology, whereas the remaining cases were classified based on clinical follow-up and/or alternative CT-based diagnoses. Fourth, although final diagnostic categories were available for the non-appendicitis group, detailed clinical outcomes and diagnostic confirmation methods for each alternative diagnosis were limited by the retrospective nature of the study. The original 1-4 confidence score distribution and the time interval between CT and surgery were also not uniformly available for re-analysis. Finally, the institutional imaging pathway was not strictly CT-first or ultrasound-first; CT was performed at the discretion of the attending physician. This may limit generalizability to institutions with different appendicitis workup pathways.

Overall, CT showed high and comparable diagnostic performance for acute appendicitis among in-training, junior, and senior radiologists in this cohort. Experienced radiologists demonstrated a small numerical advantage, mainly due to fewer false-negative interpretations. These findings support the role of resident participation in emergency CT interpretation when appropriate supervision is available, while emphasizing the need for focused teaching on subtle and atypical presentations of appendicitis.

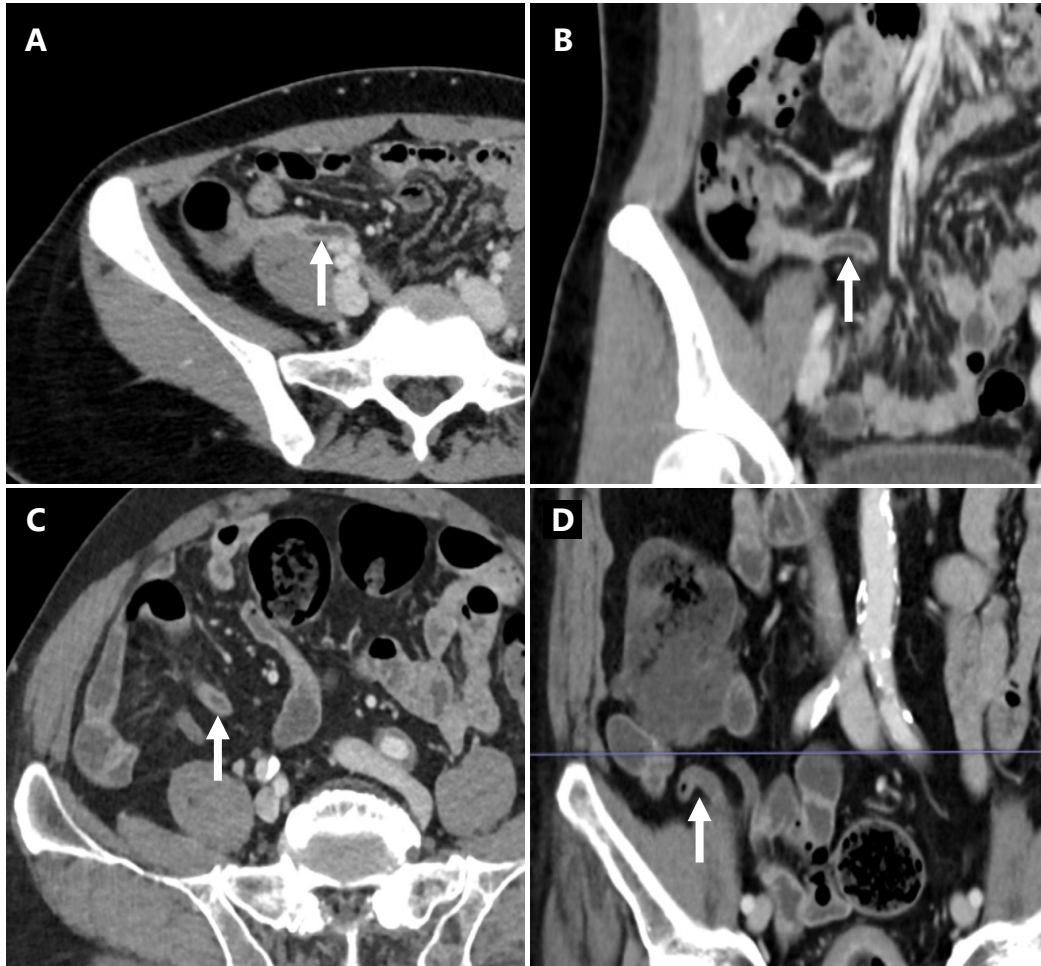


Figure 4. Two examples of false-negative CT Diagnoses of Acute Appendicitis.

Case 1 (A&B): An adult woman presented with right lower quadrant (RLQ) pain. Axial (A) and coronal (B) contrast-enhanced CT images demonstrate a normal-caliber appendix (0.5 cm in diameter) with subtle mural thickening and mild hyperenhancement (arrow), which were overlooked by the in-training and junior radiologists but correctly identified by the senior radiologist. Histopathological examination confirmed acute interstitial appendicitis.

Case 2 (C&D): An adult man presented with right lower quadrant (RLQ) pain. Axial (C) and coronal (D) contrast-enhanced CT images demonstrate a non-dilated appendix with subtle mural thickening and mild hyperenhancement (arrow), without significant periappendiceal inflammatory change. These subtle findings were missed by the in-training and junior radiologists but correctly identified by the senior radiologist. Histopathological examination confirmed acute interstitial appendicitis.

Both cases represent false-negative CT interpretations by the in-training and junior radiologists.

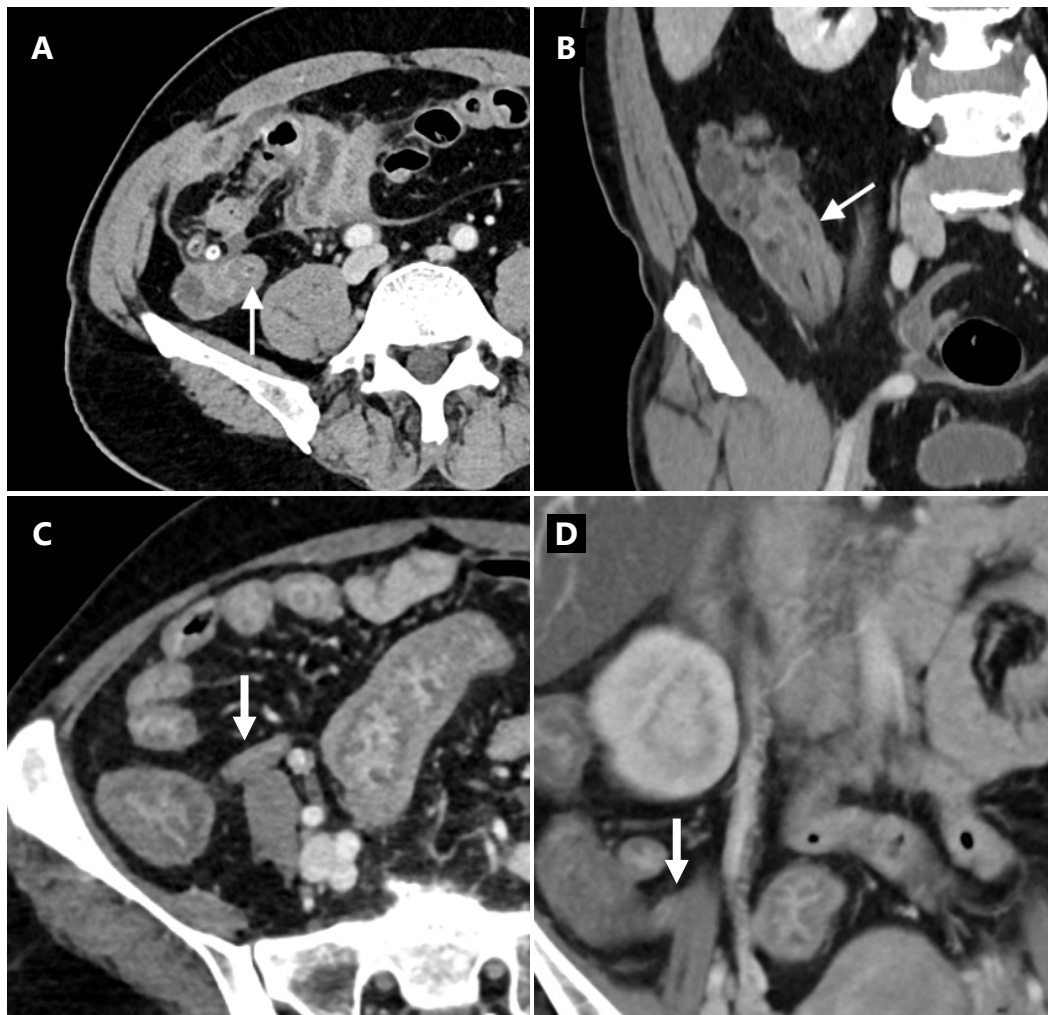


Figure 5. Two examples of false-positive CT Diagnoses of Acute Appendicitis.

Case 1 (A&B): An adult man presented with right lower quadrant (RLQ) abdominal pain. Axial (A) and coronal (B) contrast-enhanced CT images demonstrate appendiceal wall thickening and hyperenhancement (arrow) with associated cecal wall thickening, mimicking acute appendicitis. However, pathological examination confirmed a high-grade appendiceal mucinous neoplasm (HAMN), representing a false-positive CT diagnosis of acute appendicitis.

Case 2 (C&D): An adult woman presented with generalized abdominal pain. Axial (C) and coronal (D) contrast-enhanced CT images demonstrate mild appendiceal wall thickening and hyperenhancement (arrow) without appendiceal dilatation (a normal appendiceal caliber). Diffuse long-segment colonic wall thickening is also present, consistent with colitis. The appendiceal findings were considered secondary inflammatory involvement rather than primary acute appendicitis. Following treatment for colitis, follow-up CT demonstrated a normal appendix.

Conclusion

CT demonstrated high overall diagnostic performance for acute appendicitis across all levels of radiologist experience, with no statistically significant differences in accuracy. However, case-level analysis of false-negative interpretations showed that more experienced radiologists identified some subtle or early appendicitis cases that were missed by the in-training radiologist. These findings support the feasibility of independent resident interpretation of emergency CT examinations under appropriate supervision, while emphasizing the continued value of experienced radiologists in challenging or equivocal cases.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

Ethics Approval

This study was approved by the Research Ethics Committee, Faculty of Medicine, Chiang Mai University (Approval No. 354/2022, Study Code: RAD-2565-09224; Project Code: RAD-2569-0877). The initial approval was granted with expedited review on October 18, 2022. The study approval was subsequently extended, with continued approval granted from August 17, 2026 to August 16, 2027. An amendment was approved on August 10, 2026 to update the investigator/author information. The requirement for informed consent was waived because the study was retrospective. All data were anonymized before analysis. The study was conducted in accordance with the Declaration of Helsinki and ICH-GCP guidelines.

Author Contributions

P.C. contributed to study design, statistical analysis, data interpretation, manuscript preparation, and literature search. N.J. contributed to study design, data collection, data interpretation, manuscript preparation, and literature search. S.P. contributed to study design, data collection, and manuscript preparation. N.Ji. contributed to data collection. W.N.C. contributed to data interpretation and manuscript revision. A.A. contributed to data collection, data interpretation, manuscript revision, and supervision. All authors read and approved the final manuscript.

Data Availability Statement

The data used in this study are not publicly available under the PDPA Act of Thailand but can be accessed upon official request to the corresponding author.

Acknowledgments

The authors thank Tanin Lertsiriladakul, M.S., and Kriangkamon Chaisak, B.Sc., statistical officers, for their assistance with statistical consultation and research documentation. The authors also sincerely thank the staff and personnel of the Abdominal Radiology Unit, Department of Radiology, Faculty of Medicine, Chiang Mai University, for their support throughout the study.

References

1. Von Tittle SN, McCabe CJ, Ottinger LW. Delayed appendectomy for appendicitis: Causes and consequences. *Am J Emerg Med* 1996;14:620–2. doi: 10.1016/S0735-6757(96)90074-1.
2. Buckius MT, McGrath B, Monk J, Grim R, Bell T, Ahuja V. Changing epidemiology of acute appendicitis in the United States: Study period 1993-2008. *J Surg Res* 2012;175:185–90. doi: 10.1016/j.jss.2011.07.017.
3. Guan L, Liu Z, Pan G, Zhang B, Wu Y, Gan T, et al. The global, regional, and national burden of appendicitis in 204 countries and territories, 1990-2019: A systematic analysis from the Global Burden of Disease Study 2019. *BMC Gastroenterol* 2023;23:44. doi: 10.1186/s12876-023-02678-7.
4. Kariya A, Krutsri C, Singhatas P, Sumritpradit P, Thampongsa T, Lertsitthichai P, et al. Incidence of complicated appendicitis during the COVID-19 pandemic: A systematic review and meta-analysis. *Int J Surg Open* 2022;45:100512. doi: 10.1016/j.ijso.2022.100512.
5. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557–64. doi: 10.1016/s0196-0644(86)80993-3.
6. Ohle R, O'Reilly F, O'Brien KK, Fahey T, Dimitrov BD. The Alvarado score for predicting acute appendicitis: A systematic review. *BMC Med* 2011;9:139. doi: 10.1186/1741-7015-9-139.
7. Kinesya E, Cintya EP, Dorothy MJ, Ennaidi NN, Rusti HF, Mannagalli Y, et al. Diagnostic accuracy of Alvarado score components in patients with appendicitis: Systematic review and meta-analysis approach. *Health Sci Rev* 2022;2:100018. doi: 10.1016/j.hsr.2022.100018.
8. Lane MJ, Katz DS, Ross BA, Clautice-Engle TL, Mindelzun RE, Jeffrey RB Jr. Unenhanced helical CT for suspected acute appendicitis. *AJR Am J Roentgenol* 1997;168:405–9. doi: 10.2214/ajr.168.2.9016216.
9. Lane MJ, Liu DM, Huynh MD, Jeffrey RB Jr, Mindelzun RE, Katz DS. Suspected acute appendicitis: Nonenhanced helical CT in 300 consecutive patients. *Radiology* 1999;213:341–6. doi: 10.1148/radiology.213.2.r99nv44341.

10. Rao PM, Rhea JT, Novelline RA, McCabe CJ, Lawrason JN, Berger DL, et al. Helical CT technique for the diagnosis of appendicitis: Prospective evaluation of a focused appendix CT examination. *Radiology* 1997;202:139–44. doi: 10.1148/radiology.202.1.8988203.
11. Rao PM, Rhea JT, Novelline RA, Mostafavi AA, Lawrason JN, McCabe CJ. Helical CT combined with contrast material administered only through the colon for imaging of suspected appendicitis. *AJR Am J Roentgenol* 1997;169:1275–80. doi: 10.2214/ajr.169.5.9353441.
12. Bakker OJ, Go PM, Puylaert JB, Kazemier G, Heij HA; Werkgroep richtlijn Diagnostiek en behandeling van acute appendicitis. [Guideline on diagnosis and treatment of acute appendicitis: Imaging prior to appendectomy is recommended]. *Ned Tijdschr Geneesk* 2010;154:A303.
13. Smith MP, Katz DS, Lalani T, Carucci LR, Cash BD, Kim DH, et al. ACR appropriateness criteria(R) right lower quadrant pain--suspected appendicitis. *Ultrasound Q* 2015;31:85–91. doi: 10.1097/RUQ.0000000000000118.
14. Wazzan M, Abduljabbar A, Khizindar H, Alzahrani A, Aljohani RM, Nahas R, et al. Up-to-Date diagnostic CT standards for acute appendicitis: Wall thickness and intraluminal fluid thickness. *Cureus* 2023;15:e48154. doi: 10.7759/cureus.48154.
15. Rao PM, Rhea JT, Novelline RA. Sensitivity and specificity of the individual CT signs of appendicitis: Experience with 200 helical appendiceal CT examinations. *J Comput Assist Tomogr* 1997;21:686–92. doi: 10.1097/00004728-199709000-00002.
16. Yang HK, Ko Y, Lee MH, Woo H, Ahn S, Kim B, et al. Initial performance of radiologists and radiology residents in interpreting low-dose (2-mSv) appendiceal CT. *AJR Am J Roentgenol* 2015;205:W594–611. doi: 10.2214/AJR.15.14513.
17. Hunter TB, Taljanovic MS, Krupinski E, Ovitt T, Stubbs AY. Academic radiologists' on-call and late-evening duties. *J Am Coll Radiol* 2007;4:716–9. doi: 10.1016/j.jacr.2007.06.012.
18. Lietzen E, Salminen P, Rinta-Kiikka I, Paajanen H, Rautio T, Nordstrom P, et al. The accuracy of the computed tomography diagnosis of acute appendicitis: Does the experience of the radiologist matter? *Scand J Surg* 2018;107:43–7. doi: 10.1177/1457496917731189.

19. Strub WM, Vagal AA, Tomsick T, Moulton JS. Overnight resident preliminary interpretations on CT examinations: Should the process continue? *Emerg Radiol* 2006; 13:19–23. doi: 10.1007/s10140-006-0498-4.
20. Levine CD, Aizenstein O, Lehavi O, Blachar A. Why we miss the diagnosis of appendicitis on abdominal CT: Evaluation of imaging features of appendicitis incorrectly diagnosed on CT. *AJR Am J Roentgenol* 2005;184:855–9. doi: 10.2214/ajr.184.3.01840855.
21. Cho J, Kim HY, Lee S, Park JH, Lee KH. Radiology residents' independent diagnosis of appendicitis using 2-mSv computed tomography: A secondary analysis of a large pragmatic randomized trial. *Korean J Radiol* 2023;24:529–40. doi: 10.3348/kjr.2023.0015.
22. Expert Panel on Gastrointestinal Imaging; Kambadakone AR, Santillan CS, Kim DH, Fowler KJ, Birkholz JH, Camacho MA, et al. ACR appropriateness criteria® right lower quadrant pain: 2022 update. *J Am Coll Radiol* 2022;19(11S):S445–61. doi: 10.1016/j.jacr.2022.09.011.
23. Rud B, Vejborg TS, Rappeport ED, Reitsma JB, Wille-Jørgensen P. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev* 2019;CD009977. doi: 10.1002/14651858.CD009977.pub2.
24. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg* 2020;15:27. doi: 10.1186/s13017-020-00306-3.