

# Correlation of CT Findings with Pathology and Interobserver Agreement in Patients Undergoing Appendectomy for Suspected Acute Appendicitis: A Single-center Retrospective Study

✉ Erdem Özkan<sup>1</sup>, ✉ Abdullah Can<sup>1</sup>, ✉ Nurtaç Sarıkaş<sup>2</sup>, ✉ Yasin Alper Yıldız<sup>3</sup>

<sup>1</sup>Kastamonu Training and Research Hospital, Clinic of Radiology, Kastamonu, Türkiye

<sup>2</sup>Kastamonu University Faculty of Medicine, Department of Pathology, Kastamonu, Türkiye

<sup>3</sup>Kastamonu University Faculty of Medicine, Department of General Surgery, Kastamonu, Türkiye

## Abstract

**Objectives:** Computed tomography (CT) has become the cornerstone imaging modality for the evaluation of suspected acute appendicitis, yet real-world diagnostic performance data from single-centre experiences incorporating both CT-pathology correlation and interobserver agreement remain limited. This study aimed to evaluate the diagnostic performance of CT, using histopathology as the reference standard, to characterise individual CT imaging features, and to assess interobserver reproducibility in a consecutive adult cohort.

**Methods:** A retrospective analysis was conducted at our institution involving 195 adult patients who underwent appendectomy after preoperative contrast-enhanced abdominal CT for clinically suspected acute appendicitis between January 2023 and February 2026. Histopathological examination served as the reference standard. Diagnostic performance metrics (sensitivity, specificity, positive and negative predictive values, overall accuracy) were calculated from the 2×2 contingency table, with 95% confidence intervals (CIs) estimated using the Wilson score method. Odds ratios (ORs) with 95% CIs were computed for each CT feature using Fisher's exact test. Receiver operating characteristic (ROC) analysis with the Youden-optimal cut-point was performed for the appendiceal diameter. Interobserver agreement for CT features was assessed in a 40-patient subsample evaluated by two radiologists with differing levels of experience, using Cohen's kappa and the intraclass correlation coefficient (ICC).

**Results:** Acute appendicitis was histopathologically confirmed in 155 patients (79.5%). CT demonstrated a sensitivity of 98.7% (95% CI: 95.4-99.6%), a specificity of 87.5% (73.9-94.5%), a positive predictive value of 96.8% (92.8-98.6%), a negative predictive value of 94.6% (82.3-98.5%), and an overall accuracy of 96.4% (92.8-98.3%). Forty appendectomy specimens (20.5%) showed no histopathological evidence of acute appendicitis and were classified as the pathology-negative group (Path<sup>-</sup>). Periappendiceal fat stranding demonstrated the strongest independent association with confirmed appendicitis (OR: 173.83; 95% CI: 50.17-602.28; p<0.001), followed by contrast enhancement (OR: 120.00; 95% CI: 36.85-390.82; p<0.001) and wall thickening (OR: 49.23; 95% CI: 17.92-135.22; p<0.001). The presence of an appendicolith was not significantly associated with confirmed appendicitis (p=0.151). ROC analysis of the appendiceal diameter yielded an area under the ROC curve of 0.839, with an optimal cut point of 8.3 mm (sensitivity 86.5%, specificity 75.0%). All five false-positive cases were attributable to lymphoid hyperplasia. Both false-negative cases involved atypical presentations-one with early-stage disease and the other gangrenous appendicitis with an attenuated inflammatory response. Interobserver agreement for the overall CT diagnosis was perfect (Cohen's kappa coefficient=1.000), and appendiceal diameter measurement demonstrated excellent reproducibility (ICC=0.994).

**Conclusion:** CT achieves high diagnostic accuracy for acute appendicitis in real-world clinical practice. Periappendiceal fat stranding is the single most predictive CT finding, while the presence of an appendicolith is not independently statistically significant. Gangrenous appendicitis and early-stage disease are the principal causes of CT false-negative results. Interobserver agreement for the overall diagnosis is excellent, although experience-dependent variability persists for individual secondary signs.

**Keywords:** Acute appendicitis, computed tomography, diagnostic performance, interobserver agreement

**Cite this article as:** Özkan E, Can A, Sarıkaş N, Yıldız YA. Correlation of CT findings with pathology and interobserver agreement in patients undergoing appendectomy for suspected acute appendicitis: a single-center retrospective study. Adv Radiol Imaging. 2026;3(1):15-24



**Address for Correspondence:** Erdem Özkan MD, Kastamonu Training and Research Hospital, Clinic of Radiology, Kastamonu, Türkiye

**E-mail:** erdemozkan5454@gmail.com **ORCID ID:** orcid.org/0000-0001-8120-7051

**Received:** 08.03.2026 **Accepted:** 22.04.2026 **Published:** 30.04.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

## Introduction

Acute appendicitis is the most prevalent cause of emergency abdominal surgery worldwide, with an incidence of approximately 100 cases per 100,000 persons per year and a lifetime risk estimated at 7-9% in developed countries.<sup>1,2</sup> Despite centuries of clinical experience, the diagnosis of acute appendicitis remains a significant challenge in daily practice. Clinical evaluation, supported by laboratory findings and scoring systems such as the Alvarado score, does not reliably achieve sufficient accuracy to guide surgical decision-making, particularly in patients with atypical presentations, fertile women, elderly patients, and immunosuppressed individuals.<sup>3,4,5</sup>

Contrast-enhanced computed tomography (CT) of the abdomen and pelvis has transformed the preoperative evaluation of suspected acute appendicitis, and is currently recommended as the initial imaging modality of choice for non-pregnant adults by both the 2024 Infectious Diseases Society of America (IDSA) guidelines and the 2020 World Society of Emergency Surgery Jerusalem guidelines.<sup>6,7</sup> A landmark Cochrane systematic review encompassing 64 studies and 10,280 participants reported a pooled sensitivity of approximately 95% and specificity of approximately 94% for CT in this clinical context,<sup>1</sup> results broadly consistent with large single-centre experiences demonstrating sensitivities above 98% in high-volume multidetector CT practice.<sup>8</sup>

The fundamental CT findings associated with acute appendicitis have been extensively characterised and include appendiceal distension, wall thickening, contrast enhancement, periappendiceal fat stranding, free fluid, appendicolith, and CT-detected perforation.<sup>9,10,11</sup> Findings carry distinct diagnostic weight, and their clinical interpretation is further complicated by the well-recognised entity of lymphoid hyperplasia—a benign reactive condition capable of producing an appendiceal CT appearance indistinguishable from early appendicitis, particularly in younger patients.<sup>12,13</sup> The presence of an appendicolith has attracted renewed interest in the era of antibiotic-first management, as converging evidence suggests that appendicolith-associated appendicitis carries a significantly higher risk of complicated disease and failure of non-operative treatment.<sup>14,15,16</sup>

Notwithstanding the overall high accuracy of CT, residual diagnostic uncertainty persists. False-negative CT examinations, although uncommon, primarily occur in two settings: early-stage appendicitis with an appendiceal diameter below the commonly applied threshold and gangrenous appendicitis in which transmural necrosis paradoxically attenuates the inflammatory imaging response.<sup>17,18</sup> The negative appendectomy rate—reflecting cases in which appendectomy reveals a histologically normal appendix or an appendix without appendicitis—remains an important quality metric, with contemporary CT-guided series reporting rates of 5-20%.<sup>4,19</sup> CT-detected perforation is a critical but technically challenging diagnosis, with substantial variability in both its definition and the degree of experience-dependent interobserver agreement.<sup>20,21</sup>

Interobserver agreement for CT interpretation in suspected appendicitis has been a source of ongoing concern. Several studies have demonstrated experience-dependent variability in the identification of individual CT signs, whereas agreement for the overall categorical diagnosis of appendicitis tends to be higher.<sup>20</sup> Understanding the magnitude and sources of this variability has important implications for radiological training programmes and the safe implementation of

CT-based diagnostic pathways, particularly in centres where specialist radiological expertise is not available around the clock.

Emerging technologies, including artificial intelligence-assisted CT interpretation, hold promise for further improving diagnostic performance and reducing observer dependency,<sup>22</sup> yet robust local validation data remain essential before such approaches can supplant or substantively augment established radiological workflows.

Against this background, the present study was designed to evaluate the diagnostic performance of preoperative contrast-enhanced CT, with histopathology as the reference standard, in a consecutive adult cohort undergoing appendectomy at a single training and research hospital. Secondary objectives were to determine the diagnostic value of individual CT features, to characterise the clinical and imaging profiles of discordant cases (false-positive and false-negative examinations), and to assess interobserver agreement for CT features across radiologists with differing levels of clinical experience.

## Methods

### Study Design and Ethical Approval

This single-centre retrospective cross-sectional study was conducted at our hospital in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Kastamonu University Non-Interventional Clinical Research Ethics Committee (approval number: 2026-49, date: 19.03.2026). Given the retrospective nature of the study, the requirement for informed consent was waived. Patient data were anonymised prior to analysis, and access was restricted to the research team in compliance with applicable data protection regulations. The study was conducted without external funding; all research expenses were covered by the investigators.

### Study Population

Medical records and the Picture Archiving and Communication System were retrospectively reviewed from January 1, 2023, to February 1, 2026. Adult patients ( $\geq 18$  years) who underwent appendectomy following abdominal CT for clinically suspected acute appendicitis were screened for eligibility. The inclusion criteria were: (1) clinical suspicion of acute appendicitis resulting in surgical intervention; (2) availability of a preoperative contrast-enhanced abdominal CT examination; (3) availability of a postoperative histopathological report; and (4) age  $\geq 18$  years. Exclusion criteria were: unavailable or non-retrievable CT images; non-visualisation of the vermiform appendix on CT; inadequate image quality due to motion or other artefacts; and incomplete histopathological data. Ultimately, 195 patients met the study criteria and were included in the final analysis.

### CT Examination and Image Evaluation

All CT examinations were performed on a 128-Slice Revolution Maxima CT scanner (GE HealthCare) with intravenous contrast enhancement. CT images were independently evaluated by two radiologists with differing levels of experience (reader 1: 1 year; reader 2: 8 years), and were both blinded to patients' clinical information and histopathological results. The following imaging features were recorded using a standardised data collection form: (1) appendiceal diameter (outer-to-outer measurement, mm); (2) appendiceal wall thickening (present/absent); (3) appendiceal wall contrast enhancement (present/absent); (4) periappendiceal fat stranding (present/absent); (5) free fluid in the periappendiceal

region (present/absent); (6) appendicolith (present/absent); and (7) CT-detected perforation (present/absent). The overall CT diagnosis of acute appendicitis (positive or negative) was recorded for each case. In cases of disagreement between the two readers, a consensus decision, reached through joint re-evaluation, was used as the final CT diagnosis for the primary diagnostic performance analysis. Interobserver agreement for each imaging feature was assessed separately using data from the initial independent readings.

**Reference Standard**

Histopathological examination of the resected appendix specimen served as the reference standard. Acute appendicitis was confirmed when the pathology report demonstrated transmural or mucosal neutrophilic infiltration consistent with acute inflammation. Confirmed cases were further classified as complicated—defined as the presence of perforation and/or gangrene—or uncomplicated. For the purposes of this study, the pathology-negative group (Path<sup>-</sup>) was defined as appendectomy specimens showing no histopathological evidence of acute appendicitis. This classification was used for CT-histopathology correlation and should not be interpreted as an institutional negative appendectomy rate.

**Statistical Analysis**

Statistical analyses were performed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). A two-tailed p value of less than 0.05 was considered statistically significant. The normality of continuous variables was assessed using the Shapiro-Wilk test. Since both age and appendiceal diameter deviated significantly from normality, they are reported as medians with interquartile ranges (IQR). Categorical variables are expressed as absolute numbers and percentages. Between-group comparisons of continuous variables were performed using the Mann-Whitney U test. Associations between categorical CT

findings and histopathological outcome were evaluated using Fisher’s exact test. The chi-square test was used for sex-stratified comparisons of pathological confirmation rates. Diagnostic performance metrics—sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy—were calculated from the 2x2 contingency table, with 95% confidence intervals (CI) derived using the Wilson score method. Receiver operating characteristic (ROC) curve analysis was performed for the appendiceal diameter; the area under the curve (AUC) was reported, and the optimal diameter threshold was determined by the Youden index (J=sensitivity + specificity -1). For each individual CT feature, odds ratios (ORs) with 95% CIs were calculated from the contingency table using the Haldane-Anscombe correction where applicable. Interobserver agreement for categorical CT features was quantified using Cohen’s kappa coefficient (κ), and interobserver agreement for the continuous appendiceal diameter measurement was quantified using the intraclass correlation coefficient (ICC, two-way mixed model, absolute agreement). Kappa values were interpreted according to the scale proposed by Landis and Koch: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect.

**Results**

**Patient Demographics and General Characteristics**

A total of 195 patients met the inclusion criteria and were enrolled in the study (Table 1). The cohort comprised 99 female patients (50.8%) and 96 male patients (49.2%). The median age was 33 years (IQR: 25.5-44.5; range: 18-73). Both age and appendiceal diameter deviated significantly from normality (Shapiro-Wilk test; p<0.001 for both); consequently, non-parametric methods were used for all subsequent analyses. The overall median appendiceal diameter was 10.3 mm (IQR: 8.1-13.2).

**Table 1. Demographic characteristics and distribution of histopathological findings (N=195)**

Variable	Female (n=99)	Male (n=96)	Total (n=195)	p value
<b>Demographics</b>				
Age - median (IQR), years	33 (26-44)	33 (25-45)	33 (25.5-44.5)	NS
Age range, years	18-73	18-71	18-73	-
Appendiceal diameter - median (IQR), mm	10.1 (7.8-13.0)	10.5 (8.6-13.5)	10.3 (8.1-13.2)	NS
Normality (Shapiro-Wilk)	-	-	Non-normal (p<0.001)	-
<b>Histopathological outcomes</b>				
<b>Confirmed appendicitis (Path<sup>+</sup>)</b>	71 (71.7%)	84 (87.5%)	155 (79.5%)	0.011
Complicated (perforation and/or gangrene)	-	-	18 (11.6%) <sup>a</sup>	-
Isolated perforation	-	-	9 (5.8%) <sup>a</sup>	-
Gangrene	-	-	9 (5.8%) <sup>a</sup>	-
Abscess/phlegmon	-	-	53 (34.2%) <sup>a</sup>	-
Uncomplicated	-	-	137 (88.4%) <sup>a</sup>	-
<b>Pathology-negative appendectomy specimens (Path<sup>-</sup>)</b>	28 (28.3%)	12 (12.5%)	40 (20.5%)	-
Normal appendix histology	-	-	16 (40%) <sup>b</sup>	-
Lymphoid hyperplasia	-	-	14 (35.0%) <sup>b</sup>	-
Fibrous obliteration	-	-	9 (22.5%) <sup>b</sup>	-
Congestion	-	-	1 (2.5%) <sup>b</sup>	-

<sup>a</sup>: Percentage of pathology-positive cases (n=155). <sup>b</sup>: Percentage of Path<sup>-</sup> cases (n=40). Sex-stratified comparison of pathological confirmation rates performed using the chi-square test. Data are expressed as n (%) unless stated otherwise  
 Path<sup>+</sup>: Histopathologically confirmed appendicitis, Path<sup>-</sup>: Negative appendectomy, IQR: Interquartile range, NS: Not significant (Mann-Whitney U test)

### Histopathological Findings

Histopathological examination confirmed acute appendicitis in 155 patients (79.5%), whereas 40 patients (20.5%) had appendectomy specimens without histopathological evidence of acute appendicitis and were therefore classified as the pathology-negative group (Path<sup>-</sup>) (Table 1). Among the 40 pathology-negative cases, 18 (11.6%) were classified as complicated appendicitis: 9 (5.8%) had isolated perforation and 9 (5.8%) had gangrene. Abscess or phlegmon formation was identified in 53 pathology-positive cases (34.2%). Among the 40 negative appendectomy cases, histopathological findings included: a normal appendix in 16 cases (40%), lymphoid hyperplasia in 14 cases (35.0%), fibrous obliteration in 9 cases (22.5%), and congestion in 1 case (2.5%). The rate of histopathologically confirmed appendicitis was significantly higher among male patients (84/96, 87.5%) than among female patients (71/99, 71.7%) ( $\chi^2$  test,  $p=0.011$ ).

### Overall CT Diagnostic Performance

The overall diagnostic performance of CT for the detection of acute appendicitis is summarised in Table 2. CT yielded a sensitivity of 98.7% (95% CI: 95.4-99.6%), a specificity of 87.5% (95% CI: 73.9-94.5%), a PPV of 96.8% (95% CI: 92.8-98.6%), an NPV of 94.6% (95% CI: 82.3-98.5%), and an overall accuracy of 96.4% (95% CI: 92.8-98.3%).

Of 158 CT-positive examinations, 153 (96.8%) were confirmed as true-positive cases. Five CT-positive cases (3.2%) were false positives, representing patients with imaging findings suggestive of appendicitis but without histopathological confirmation of acute appendicitis. Of the 37 CT-negative examinations, 35 (94.6%) were true-negatives. Two CT-negative cases (5.4%) were false-negatives (FNs), representing missed cases of appendicitis.

### Diagnostic Value of Individual CT Findings

The diagnostic performance of individual CT features is detailed in Table 3. All primary inflammatory CT signs—periappendiceal fat stranding, contrast enhancement, and wall thickening—were significantly associated with histopathologically confirmed appendicitis (Fisher’s exact test,  $p<0.001$  for all three).

Periappendiceal fat stranding showed the strongest independent association with pathological appendicitis (OR: 173.83, 95% CI: 50.17-602.28;  $p<0.001$ ), and was present in 149 of 155 pathology-positive cases (96.1%) but in only 5 of 40 pathology-negative cases (12.5%), yielding a sensitivity of 96.1%, a specificity of 87.5%, and a PPV of 96.8%. Contrast enhancement demonstrated the highest sensitivity among all individual features (96.8%), with an OR of 120.00 (95% CI: 36.85-390.82;  $p<0.001$ ). Wall thickening was present in 89.7% of pathology-positive cases, with an OR of 49.23 (95% CI: 17.92-135.22;  $p<0.001$ ).

**Table 2. Overall CT diagnostic performance for acute appendicitis (N=195)**

	Path <sup>+</sup> (n=155)	Path <sup>-</sup> (n=40)	Total
CT positive	TP=153	FP=5	158
CT negative	FN=2	TN=35	37
<b>Total</b>	155	40	195
<b>Diagnostic performance metrics</b>			
	Value	95% CI (lower)	95% CI (upper)
Sensitivity	98.7%	95.4%	99.6%
Specificity	87.5%	73.9%	94.5%
Positive predictive value	96.8%	92.8%	98.6%
Negative predictive value	94.6%	82.3%	98.5%
Overall accuracy	96.4%	92.8%	98.3%

Sensitivity=TP/(TP+FN), Specificity=TN/(TN+FP), PPV=TP/(TP+FP), NPV=TN/(TN+FN), Accuracy=(TP+TN)/N  
 CT: Computed tomography, Path<sup>+</sup>: Histopathologically confirmed appendicitis, Path<sup>-</sup>: Absence of histopathological evidence of acute appendicitis, TP : True-positive, FP: False-positive, FN: False-negative, TN: True-negative, CI: Confidence interval (Wilson score method)

**Table 3. Diagnostic performance of individual CT findings for the detection of acute appendicitis**

CT feature	Path <sup>+</sup> n/N (%)	Path <sup>-</sup> n/N (%)	Sens.	Spec.	PPV	NPV	OR (95% CI) p value
Wall thickening	139/155 (89.7%)	6/40 (15.0%)	89.7%	85.0%	95.9%	68.0%	49.23 (17.92-135.22) $p<0.001$
Contrast enhancement	150/155 (96.8%)	8/40 (20.0%)	96.8%	80.0%	94.9%	86.5%	120.00 (36.85-390.82) $p<0.001$
Periappendiceal fat stranding	149/155 (96.1%)	5/40 (12.5%)	96.1%	87.5%	96.8%	85.4%	173.83 (50.17-602.28) $p<0.001$
Free fluid	74/155 (47.7%)	2/40 (5.0%)	47.7%	95.0%	97.4%	31.9%	17.36 (4.05-74.48) $p<0.001$
Appendicolith†	41/155 (26.5%)	6/40 (15.0%)	26.5%	85.0%	87.2%	23.0%	2.04 (0.80-5.21) $p=0.151$
CT-detected perforation†	14/155 (9.0%)	0/40 (0.0%)	9.0%	100.0%	100.0%	22.1%	$\infty$ $p=0.078$
Appendiceal diameter (ROC; AUC=0.839)*	11.2 mm (9.1-13.6)	6.8 mm (5.5-8.3)	86.5%*	75.0%*	-	-	AUC 0.839 $p<0.001$

\*: Sensitivity and specificity for appendiceal diameter correspond to the Youden-optimal cut-point of 8.3 mm. †: Features that did not reach statistical significance ( $p\geq 0.05$ )- CT-detected perforation: 14/155 cases in Path<sup>+</sup> group, 0/40 in Path<sup>-</sup> group. Values for categorical features represent frequency (n/N, %). Diameter values represent median (IQR). OR: Odds ratio calculated using Fisher’s exact test (two-tailed); 95% CI for OR computed with Haldane-Anscombe correction where applicable  
 Sens.: Sensitivity, Spec.: Specificity, PPV: Positive predictive value, NPV: Negative predictive value, AUC: Area under the receiver operating characteristic curve, CI: Confidence interval, CT: Computed tomography

Free fluid was identified in 74 pathology-positive cases (47.7%) and in only 2 pathology-negative cases (5.0%), reflecting a high specificity of 95.0% and a PPV of 97.4% (OR: 17.36, 95% CI: 4.05-74.48;  $p < 0.001$ ).

CT-detected perforation, when present, carried a PPV of 100% for pathological appendicitis; all 14 CT-reported perforation cases (9.0% of pathology-positive group) were pathologically confirmed. However, the overall association with histopathologically complicated appendicitis did not reach statistical significance ( $p = 0.078$ ), reflecting the limited sensitivity of CT for detecting perforation compared with the histopathological complication rate.

In contrast, an appendicolith was present in 26.5% of pathology-positive and 15.0% of pathology-negative cases, and did not demonstrate a statistically significant association with confirmed appendicitis (OR: 2.04, 95% CI: 0.80-5.21;  $p = 0.151$ ).

**Appendiceal Diameter Analysis**

The median appendiceal diameter was significantly larger in pathology-positive cases compared to pathology-negative cases [11.2 mm (IQR: 9.1-13.6) vs. 6.8 mm (IQR: 5.5-8.3); Mann-Whitney U test,  $p < 0.001$ ] (Table 3). ROC curve analysis demonstrated an AUC of 0.839, indicating good discriminatory performance. The Youden-optimal cut-point was 8.3 mm, corresponding to a sensitivity of 86.5% and a specificity of 75.0%.

**Analysis of Discordant Cases**

The characteristics of the five false-positive (FP) and two FN cases are presented in detail in Table 4a, 4b.

Among the five FP cases, four were female and one was male (median age: 33 years; range: 19-49). All five cases demonstrated periappendiceal fat stranding on CT. Histopathological findings in FP cases included lymphoid hyperplasia in four patients (80.0%) and combined lymphoid hyperplasia with fibrous obliteration in one patient (20.0%). In two cases, the appendiceal diameter exceeded 13 mm, and all secondary CT signs were positive, suggesting that the CT appearance was highly suggestive of appendicitis despite a benign histopathological substrate-likely reflecting reactive periappendiceal inflammatory changes secondary to lymphoid hyperplasia.

Both FN cases were young patients (aged 20 and 25 years). The first case was a 20-year-old male with an appendiceal diameter of 5.4 mm and no secondary CT signs (wall thickening, contrast enhancement, fat stranding, free fluid, or appendicolith), consistent with early-stage appendicitis that was below the CT detection threshold (Figure 1). The second was a 25-year-old female whose diameter was 7.9 mm and who had gangrenous changes on pathology, in the absence of convincing CT findings—a pattern consistent with atypical or gangrenous appendicitis, in which transmural necrosis may paradoxically reduce or attenuate the inflammatory imaging response (Figure 2). Both FN patients were ultimately taken to the operating theatre based on persistent clinical suspicion, and appendicitis was confirmed histopathologically in both cases.

**Interobserver Agreement**

Interobserver agreement was assessed in a subset of 40 patients who were independently evaluated by two radiologists with differing levels of experience (Reader 1: 1 year; Reader 2: 8 years). Results

**Table 4a. False-positive cases-CT positive/pathology negative (n=5)**

Case	Sex	Age (year)	Diameter (mm)	Wall thickening	Contrast enhancement	Fat stranding	Free fluid	Appendicolith	Histopathology/remark
1	F	19	8.2	-	-	+	-	-	Lymphoid hyperplasia; fat stranding only; reactive change
2	M	19	10.0	-	+	+	-	+	Lymphoid hyperplasia; appendicolith + reactive enhancement
3	F	32	8.0	+	+	+	-	-	Lymphoid hyperplasia + fibrous obliteration; dual pathology; convincing CT pattern
4	F	49	13.1	+	+	+	+	+	Lymphoid hyperplasia; all secondary signs positive; reactive substrate
5	F	33	13.3	+	+	+	+	+	Lymphoid hyperplasia; dilated diameter + full secondary sign complex

All five false-positive cases were managed with appendectomy; lymphoid hyperplasia was the most frequent underlying histopathological finding (4/5, 80%)  
 +: Present, -: Absent, F: Female, M: Male

**Table 4b. False-negative cases-CT negative/pathology positive (n=2)**

Case	Sex	Age (year)	Diameter (mm)	Wall thickening	Contrast enhancement	Fat stranding	Free fluid	Appendicolith	Histopathology/reason for miss
1	M	20	5.3	-	-	-	-	-	Acute appendicitis (mild); below-threshold diameter; no secondary signs → early-stage disease
2	F	25	7.1	-	-	-	-	-	Acute appendicitis + gangrene; atypical presentation; transmural necrosis attenuated inflammatory CT response

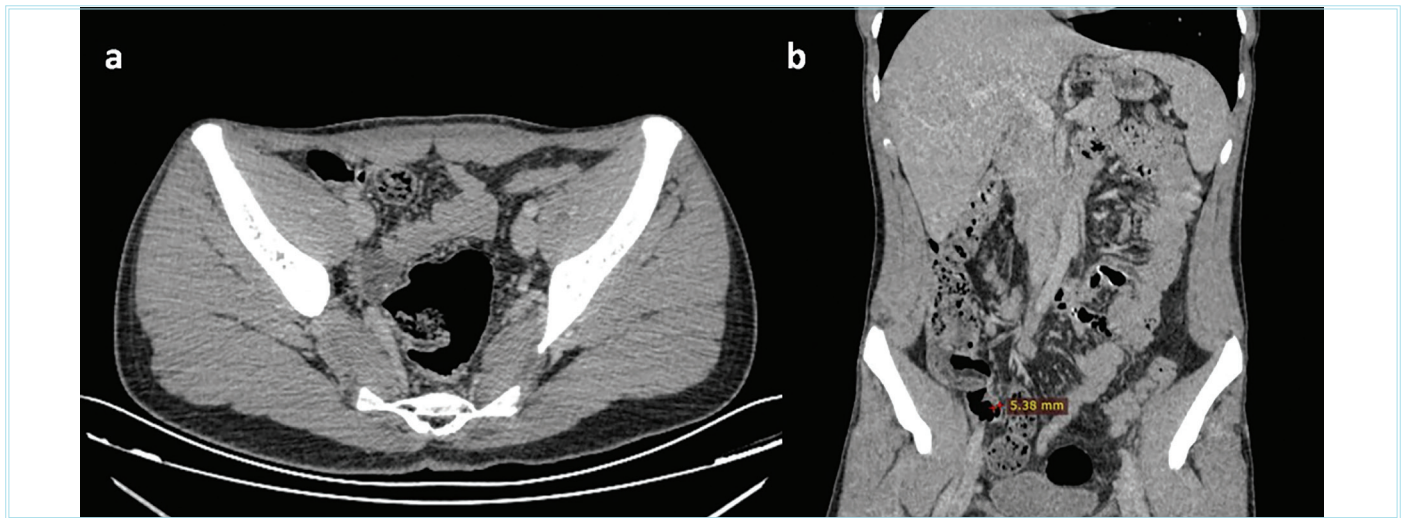
Both false-negative patients underwent appendectomy based on persistent clinical suspicion; histopathological diagnosis was confirmed postoperatively in both cases  
 CT: Computed tomography

are summarised in Table 5. Agreement on the overall CT diagnosis of appendicitis was perfect ( $\kappa=1.000$ ; 100% raw concordance), indicating that both readers arrived at identical final diagnoses in all 40 cases. Among individual CT features, periappendiceal fat stranding and appendicolith each demonstrated almost perfect agreement ( $\kappa=0.805$ , 95% CI: 0.541-1.068). Contrast enhancement and wall thickening showed substantial agreement ( $\kappa=0.754$ , 95% CI: 0.486-1.022;  $\kappa=0.630$ , 95% CI: 0.326-0.933, respectively). Agreement for free fluid was moderate ( $\kappa=0.504$ , 95% CI: 0.237-0.770); the majority of disagreements (7 of 10) were attributable to reader 1 scoring free fluid as present when reader 2 did not. CT-detected perforation yielded a near-zero kappa ( $\kappa=-0.034$ ); however, raw agreement was 92.5% (37/40), and this discrepancy reflects the well-recognised

prevalence-dependency of Cohen's kappa (the “kappa paradox”): the very low prevalence of CT-detected perforation in the subsample ( $\leq 10\%$ ) renders the statistic artificially low despite high observed concordance. Appendiceal diameter measurements demonstrated excellent reproducibility, with an ICC of 0.994 (95% CI: 0.988-0.997), a mean inter-reader difference of -0.01 mm, and Bland-Altman limits of agreement of -0.79 to +0.76 mm.

## Discussion

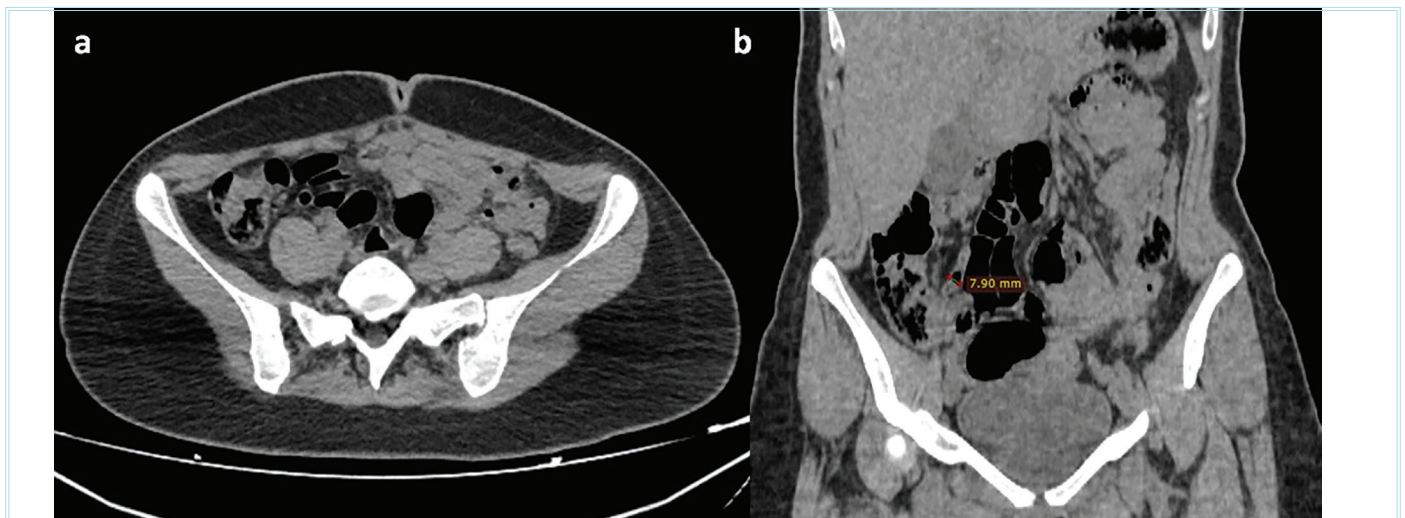
The present study evaluated the real-world diagnostic performance of preoperative contrast-enhanced CT, with histopathology as the gold standard, in a consecutive cohort of adults undergoing appendectomy for clinically suspected acute appendicitis. Our findings confirm that



**Figure 1.** False-negative CT in early-stage appendicitis in a 20-year-old male patient

(a) Axial and (b) coronal CT images demonstrate a non-dilated appendix in the right lower quadrant measuring 5.4 mm, without wall thickening or secondary inflammatory findings such as periappendiceal fat stranding, free fluid, or appendicolith. Although the radiological findings were not typical of appendicitis, histopathological examination revealed early inflammatory changes

CT: Computed tomography



**Figure 2.** False-negative CT in gangrenous appendicitis in a 25-year-old female patient

(a) Axial and (b) coronal CT images show an appendix measuring 7.9 mm without convincing secondary CT findings of appendicitis. Histopathological analysis revealed gangrenous appendicitis, despite the absence of radiological features typically suggestive of a gangrenous process

CT: Computed tomography

**Table 5. Interobserver agreement for CT imaging features (n=40)**

CT feature	R1 n (%)	R2 n (%)	Agreement n/40	Agreement (%)	Disagr. R1+/R2-	Disagr. R1-/R2+	κ (95% CI)	Strength
Wall thickening	33 (82.5%)	29 (72.5%)	35/40	87.5%	4	1	0.630 (0.326-0.933)	Substantial
Contrast enhancement	35 (87.5%)	33 (82.5%)	37/40	92.5%	2	1	0.754 (0.486-1.022)	Substantial
Periappendiceal fat stranding	34 (85.0%)	32 (80.0%)	38/40	95.0%	2	0	0.805 (0.541-1.068)	Almost perfect
Free fluid	17 (42.5%)	13 (32.5%)	30/40	75.0%	7	3	0.504 (0.237-0.770)	Moderate
Appendicolith	8 (20.0%)	10 (25.0%)	38/40	95.0%	0	2	0.805 (0.541-1.068)	Almost perfect
CT-detected perforation†	3 (7.5%)	4 (10.0%)	37/40	92.5%	1	2	-0.034 (-1.160-1.091)	Poor† (kappa paradox)
Overall CT diagnosis	34 (85.0%)	34 (85.0%)	40/40	100.0%	0	0	1.000 (1.000-1.000)	Perfect

Continuous Variable					
Variable	R1 Median (IQR), mm	R2 Median (IQR), mm	Mean diff. (R1-R2)	LoA (±1.96 SD)	ICC (95% CI)
Appendiceal diameter	10.1 (8.3-13.0)	10.1 (8.1-12.6)	-0.01 mm	-0.79 to +0.76 mm	0.994 (0.988-0.997)

95% CI computed using the asymptotic standard error method for κ and F-distribution-based method for ICC. Strength of agreement for κ: <0.00 poor; 0.00-0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; 0.81-1.00 almost perfect (Landis & Koch, 1977). †: The near-zero kappa for CT-detected perforation reflects the kappa paradox (prevalence-dependency): raw agreement was 92.5% (37/40), but the very low prevalence of perforation (≤10%) renders kappa artificially low despite high observed concordance  
κ: Cohen's kappa coefficient, ICC: Intraclass correlation coefficient (two-way mixed model, absolute agreement), R1: Reader 1 (1 year of experience), R2: Reader 2 (8 years of experience), IQR: Interquartile range, LoA: Bland-Altman limits of agreement, CI: Confidence

CT achieves high diagnostic accuracy in this setting, with an overall sensitivity of 98.7%, specificity of 87.5%, and accuracy of 96.4%. These results are broadly consistent with the pooled estimates from large meta-analytic and multicentre benchmarking data,<sup>1,8</sup> yet provide important granularity regarding the relative diagnostic weight of individual CT signs, the imaging and histopathological characteristics of discordant cases, and the extent of interobserver variability across experience levels.

**Overall CT Diagnostic Performance and Comparison with the Literature**

The sensitivity of 98.7% achieved in the present series is comparable to and, in some comparisons, exceeds figures reported in seminal large-scale multidetector CT studies. Pickhardt et al.<sup>8</sup> reported a sensitivity of 98.5% and a specificity of 98.0% in a retrospective analysis of 2,871 adults spanning a decade of multidetector CT practice at a single academic centre. The Cochrane systematic review by Rud et al.<sup>1</sup>, incorporating 64 studies and 10,280 participants, reported a pooled sensitivity of approximately 95% and specificity of 94% for CT in suspected appendicitis. The slightly lower specificity observed in our series (87.5%) may be partly explained by the proportion of pathology-negative appendectomy specimens within this selected surgical cohort (40/195, 20.5%). Importantly, this figure represents the proportion of patients without histopathological evidence of acute appendicitis among surgically treated patients included in the study, rather than an institutional negative appendectomy rate. Therefore, it should be interpreted in the context of the study design and patient-selection criteria. This discrepancy may be attributable to the tertiary-referral patterns in a training hospital setting, in which higher-risk or diagnostically challenging cases are concentrated, and to a patient-selection effect inherent to centres where CT is routinely employed for all clinically suspected appendicitis cases, irrespective of pre-test probability.

The 2024 IDSA guideline update, representing the most current evidence-based imaging recommendation, conditionally supports CT as the initial imaging modality in non-pregnant adults with suspected acute appendicitis, acknowledging that its accuracy justifies direct CT use without necessitating additional imaging studies.<sup>6</sup> Our data support

this recommendation: CT correctly excluded appendicitis in 35 of 37 CT-negative cases, and the two false negatives were both detected clinically through persistent symptom evaluation, independent of imaging.

**Individual CT Signs: Relative Diagnostic Weight**

Periappendiceal fat stranding emerged as the single most powerful CT predictor of histopathologically confirmed appendicitis in our series (OR: 173.83), present in 96.1% of pathology-positive cases. This finding is consistent with the broader radiology literature, which identifies periappendiceal inflammatory changes as the imaging hallmark of appendiceal inflammation.<sup>9,11</sup> Fat stranding mechanistically reflects exudative oedema and vascular engorgement in the mesoappendix and adjacent fat, and its identification is generally robust across experience levels—as confirmed by the almost perfect interobserver agreement (κ=0.805) in our subsample.

Contrast enhancement demonstrated the highest sensitivity among individual CT features (96.8%) and had a strong independent association (OR: 120.00). Appendiceal wall enhancement reflects intact mural vascularity in the setting of active inflammation and is an earlier and more reliable sign than luminal distension alone.<sup>9,10</sup> Wall thickening, although less specific (specificity 85.0%) than fat stranding, was present in nearly 90% of confirmed cases (OR: 49.23), which makes it a sensitive but not highly discriminatory feature when considered in isolation.

Free fluid reached a high specificity of 95.0% and a PPV of 97.4%, rendering it a powerful confirmatory sign when present, notwithstanding its moderate sensitivity (47.7%). The predominantly periappendiceal distribution of free fluid in our series suggests that when present, it reliably indicates an active periappendiceal inflammatory process rather than coincidental findings such as ovarian follicular fluid or nonspecific pelvic fluid.

**The Appendicolith: A Radiologically Identified but Statistically Non-significant Finding**

The absence of a statistically significant association between appendicolith and confirmed appendicitis in the present series (OR:

2.04;  $p=0.151$ ) warrants specific attention. This finding may appear discordant with the rapidly expanding literature on appendicolith-associated appendicitis. However, it should be interpreted in the context of the study design: This analysis assessed whether the presence of an appendicolith predicts histopathologically confirmed appendicitis among patients who have already undergone surgery-not whether it predicts complicated disease or failure of conservative management. Both populations (appendicitis-confirmed and negative appendectomy) harboured appendicoliths at different frequencies (26.5% vs. 15.0%), although sample size limitations prevented the difference from reaching statistical significance.

The clinical relevance of appendicolith detection, however, lies not in confirming the diagnosis of appendicitis per se, but in stratifying disease severity and guiding management decisions in the antibiotic-first era. A large multicentre Finnish study by Sula et al.<sup>14</sup> demonstrated that, among 3,085 patients with CT-confirmed appendicitis, the presence of an appendicolith was associated with a markedly elevated risk of complicated appendicitis (47.1% vs. 21.5%;  $p<0.001$ ), with larger appendicolith diameter, a base-of-appendix location, and heterogeneous mural enhancement further amplifying this risk. Similarly, Oktay et al.<sup>15</sup> found that CT-detected appendicoliths in paediatric appendicitis were associated with significantly larger appendiceal diameters (10 mm vs. 8 mm;  $p=0.001$ ), and Weitzner et al.<sup>16</sup> demonstrated that appendicolith presence and size independently predicted CT-histopathological discordance and failure of conservative management. Taken together, the available evidence supports systematic documentation of the presence and characteristics of appendicoliths in CT reports-not as a binary diagnostic sign, but as a prognostically meaningful imaging biomarker that shapes decision-making about non-operative management.

**CT-Detected Perforation: The Kappa Paradox and Diagnostic Limitations**

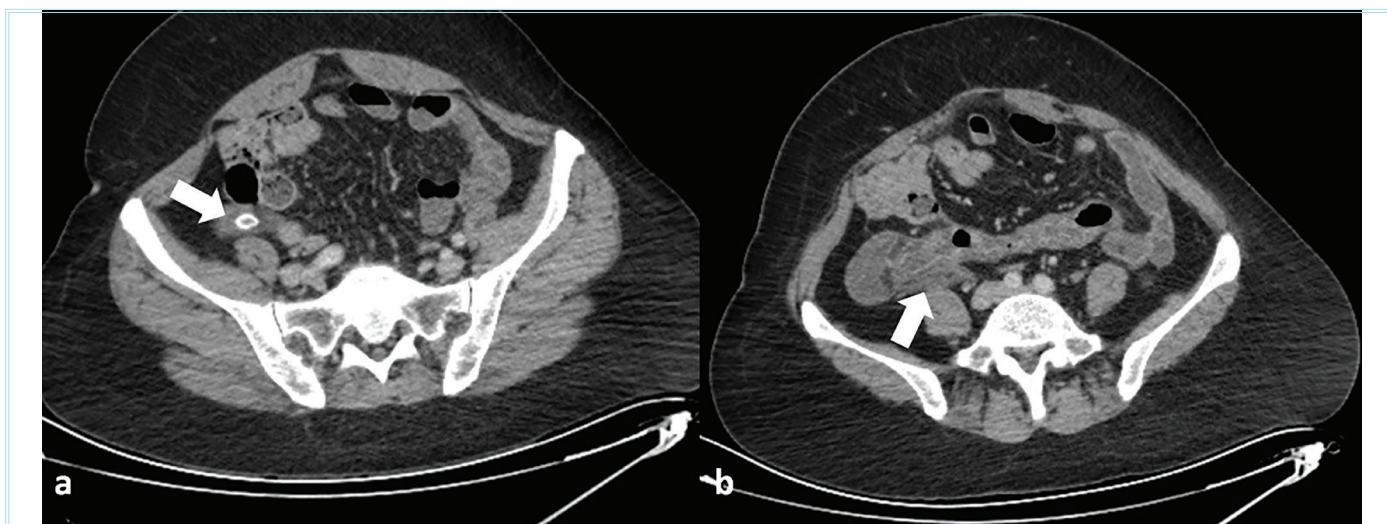
CT-detected perforation demonstrated a PPV of 100% in the present series (all 14 CT-reported perforations were pathologically confirmed),

yet it failed to reach statistical significance ( $p=0.078$ ) as a predictor of histopathologically complicated disease and yielded a near-zero kappa for interobserver agreement ( $\kappa=-0.034$ ). Both observations are readily explained by well-documented phenomena. The lack of statistical significance reflects the limited sensitivity of CT for detecting perforation, which leads to underestimation of the true histopathological perforation rate; perforation identified on CT represents a subset of cases with macroscopic or extraluminal gas/fluid collections, whereas histopathological assessment detects subtler transmural involvement. The near-zero kappa reflects the well-known kappa paradox: when prevalence is very low, Cohen’s kappa may substantially underestimate agreement beyond chance, even when observed concordance is high, as widely discussed in the methodological literature. The 92.5% observed concordance for perforation status in our subsample illustrates this phenomenon clearly.<sup>24,25</sup>

**False-positive Cases: Lymphoid Hyperplasia as the Principal Mimic**

Lymphoid hyperplasia accounted for 80% (4/5) of false-positive cases in our series; one of these cases also demonstrated concurrent fibrous obliteration. This finding is consistent with the established radiological literature identifying lymphoid hyperplasia as the most important CT mimic of acute appendicitis, particularly in younger patients.<sup>12,13</sup> Lymphoid hyperplasia produces reactive follicular expansion in the lamina propria, leading to appendiceal wall thickening and-critically-periappendiceal inflammatory changes, secondary to reactive enlargement of regional mesenteric lymph nodes and oedematous mesenteric fat. In cases where multiple secondary CT signs are present (as in our two largest-diameter false-positive cases, in which all secondary CT parameters were positive), a false-positive diagnosis is virtually inevitable on CT alone (Figure 3).

Ultrasound-based discrimination between appendicitis and lymphoid hyperplasia has been investigated, with a lamina propria thickness below 1 mm identified as the most effective ultrasound parameter for this differentiation.<sup>12,13</sup> Our findings reinforce the value of the systematic



**Figure 3.** False-positive CT in a 47-year-old female patient with lymphoid hyperplasia mimicking acute appendicitis (a, b) Axial CT images show an appendicolith in panel (a) and an appendix measuring 13 mm in maximal diameter, with associated periappendiceal inflammatory changes highly suggestive of acute appendicitis. Despite the convincing radiological appearance, histopathological examination demonstrated no evidence of appendicitis and confirmed lymphoid hyperplasia

CT: Computed tomography

integration of clinical pre-test probability with imaging interpretation; in young patients—particularly those aged under 25 years without fever, leukocytosis, or typical pain migration—a degree of diagnostic restraint may be warranted before committing to surgical intervention based solely on CT findings. This argument is also supported by risk-stratification approaches incorporating clinical scores.<sup>3</sup>

### FN Cases: Early-stage and Gangrenous Appendicitis

Both FN cases in our series illustrate important and well-described CT diagnostic pitfalls. The first—a 20-year-old male with an appendiceal diameter of 5.4 mm and no secondary signs—represents the classic early-stage appendicitis below the CT detection threshold, in which luminal obstruction has not yet produced sufficient luminal distension or secondary inflammatory changes detectable on cross-sectional imaging (Figure 1). Appendiceal diameters below 6–7 mm in the clinical context of right lower quadrant pain do not reliably exclude early appendicitis; such patients may require close clinical observation, serial examinations, or repeat imaging.<sup>23</sup>

The second case—a 25-year-old woman with an appendiceal diameter of 7.9 mm and gangrenous pathology—exemplifies the paradox of gangrenous appendicitis: transmural necrosis of the appendiceal wall can selectively destroy the mucosal and submucosal layers that mediate the inflammatory and enhancement responses on CT, thereby producing an imaging appearance that underrepresents the severity of the underlying disease (Figure 2). Published case series have demonstrated that gangrenous appendicitis may occasionally present with a poorly enhancing, thin-walled appendix that does not fulfil conventional diagnostic criteria, with the diagnosis ultimately established on clinical and intraoperative grounds.<sup>26</sup> That both FN patients were operated on because of sustained clinical suspicion underscores the principle that a negative or equivocal CT examination does not mandate clinical inaction when other disease indicators remain convincing.

### Interobserver Agreement

The perfect interobserver agreement for the overall CT diagnosis of appendicitis ( $\kappa=1.000$ ) in our subsample of 40 patients is an encouraging finding, suggesting that even with one year of experience, a trainee radiologist can make accurate categorical CT diagnoses of appendicitis when evaluating a properly performed contrast-enhanced examination in a blinded manner. This result aligns with data from Hof et al.<sup>20</sup>, who demonstrated that the specificity of CT for diagnosing appendicitis was consistent across radiologists with lower and intermediate experience (94% in both groups), whereas sensitivity improved with greater experience.

At the level of individual CT features, a consistent pattern emerges: objective, well-defined morphological features (appendicolith:  $\kappa=0.805$ ; periappendiceal fat stranding:  $\kappa=0.805$ ) achieve higher agreement than subjective or prevalence-dependent features (free fluid:  $\kappa=0.504$ ; CT-detected perforation:  $\kappa=-0.034$ , reflecting the kappa paradox). This hierarchy has important implications for structured CT reporting: standardised reporting templates that explicitly require a binary assessment of each primary inflammatory sign—rather than a gestalt impression—may reduce the experience-dependent variability observed in secondary features, such as free fluid.

The excellent reproducibility of appendiceal diameter measurement ( $ICC=0.994$ ; Bland-Altman limits of agreement  $-0.79$  to  $+0.76$  mm) is

clinically relevant, particularly given the emerging evidence supporting appendiceal diameter as a prognostic biomarker for complicated disease and failure of non-operative management. Near-perfect diameter reproducibility across experience levels indicates that this metric can be reliably incorporated into clinical decision algorithms and risk-stratification tools without concern about measurement variability.

### Sex-specific Differences

A statistically significant sex-based difference in histopathological confirmation rates was observed: 87.5% of male patients had confirmed appendicitis, compared with 71.7% of female patients ( $p=0.011$ ). This differential likely reflects the greater diagnostic challenge posed by fertile women, in whom gynaecological conditions such as ovarian cysts, adnexitis, and ectopic pregnancy can closely mimic appendicitis both clinically and on CT. The higher proportion of pathology-negative appendectomy specimens among women is a well-established observation in the appendicitis literature and underscores the diagnostic challenge posed by gynaecological mimics in women of reproductive age.<sup>6,7</sup>

### Pathology-negative Appendectomy Specimens and Clinical Implications

In this selected cohort of patients who underwent appendectomy after preoperative CT for clinically suspected acute appendicitis, 40 of 195 patients (20.5%) had no histopathological evidence of acute appendicitis. This proportion should be interpreted as the pathology-negative subgroup within the study population, rather than as a direct institutional quality metric or a general negative appendectomy rate. Several factors may have contributed to the presence of pathology-negative cases, including the training hospital setting, a heterogeneous referral base, the predominance of young patients in whom lymphoid hyperplasia may mimic acute appendicitis, and the tendency to proceed to surgery when clinical suspicion persists despite diagnostic uncertainty. Structured risk-stratification pathways incorporating clinical scoring systems such as the Appendicitis Inflammatory Response (AIR) score may help refine decision-making and should be prospectively evaluated at our institution.<sup>27,28</sup>

### Study Limitations

Several limitations of this study should be acknowledged. First, the retrospective single-centre design may limit the generalisability of the findings to other clinical settings. Second, the relatively small numbers of complicated appendicitis cases ( $n=18$ ) and discordant cases ( $n=7$ ) limited the statistical power of subgroup analyses. Third, clinical scoring data, including Alvarado and AIR scores, were not systematically available for the entire cohort, precluding formal evaluation of the incremental diagnostic value of CT beyond clinical assessment. Finally, although a consensus CT diagnosis served as the primary CT outcome, the consensus process itself may have introduced bias in borderline cases.

### Conclusion

This single-centre, retrospective study confirms that contrast-enhanced CT has high diagnostic performance for acute appendicitis in an adult surgical population, with a sensitivity of 98.7%, specificity of 87.5%, and overall accuracy of 96.4%. Periappendiceal fat stranding represents the strongest individual CT predictor (OR: 173.83), while a CT-detected appendicolith does not independently reach diagnostic significance for confirming appendicitis, although its prognostic value for complicated disease is well supported by the wider literature. Gangrenous appendicitis

with an attenuated inflammatory response on CT and early-stage disease below the luminal threshold are the principal sources of FN diagnoses. Interobserver agreement for the overall CT diagnosis is perfect across experience levels, whereas individual secondary signs—particularly free fluid and perforation assessment—demonstrate experience-dependent variability. These findings support the current evidence-based recommendation for CT as the primary diagnostic imaging modality in non-pregnant adult patients with suspected acute appendicitis and highlight the continued importance of integrated clinical assessment for CT-negative cases with persistent symptomatology.

## Ethics

**Ethics Committee Approval:** Ethical approval was obtained from the Kastamonu University Non-Interventional Clinical Research Ethics Committee (approval number: 2026-49, date: 19.03.2026).

**Informed Consent:** Retrospective Study.

## Footnotes

### Authorship Contributions

Surgical and Medical Practices: N.S., Y.A.Y., Concept: E.Ö., Design: E.Ö., Data Collection or Processing: E.Ö., A.C., N.S., Y.A.Y., Analysis or Interpretation: E.Ö., A.C., Literature Search: E.Ö., A.C., N.S., Y.A.Y., E.Ö., Writing: E.Ö.

**Conflict of Interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declared that this study received no financial support.

## References

- 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;2019:CD009977.
- Börner N, Kappenberger AS, Weber S, Scholz F, Kazmierczak P, Werner J. The acute abdomen: structured diagnosis and treatment. *Dtsch Arztebl Int.* 2025;122:137-44.
- Deboni VS, Rosa MI, Lima AC, Graciano AJ, Garcia CE. The appendicitis inflammatory response score for acute appendicitis: is it important for early diagnosis? *Arq Bras Cir Dig.* 2022;35:e1686.
- Chen KC, Arad A, Chen KC, Storrar J, Christy AG. The clinical value of pathology tests and imaging study in the diagnosis of acute appendicitis. *Postgrad Med J.* 2016;92:611-9.
- Andre JB, Sebastian VA, Ruchman RM, Saad SA. CT and appendicitis: evaluation of correlation between CT diagnosis and pathological diagnosis. *Postgrad Med J.* 2008;84:321-4.
- Bonomo RA, Tamma PD, Abrahamian FM, et al. 2024 Clinical practice guideline update by the Infectious Diseases Society of America on complicated intra-abdominal infections: diagnostic imaging of suspected acute appendicitis in adults, children, and pregnant people. *Clin Infect Dis.* 2024;79(Suppl 3):S94-103.
- Di Saverio S, Podda M, De Simone B, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg.* 2020;15:27.
- Pickhardt PJ, Lawrence EM, Pooler BD, Bruce RJ. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. *Ann Intern Med.* 2011;154:789-96.
- Aydın S, Karavas E, Şenbil DC. Imaging of acute appendicitis: advances. *World J Gastrointest Surg.* 2022;14:370-3.
- Gurian MS, Kovanlikaya A, Beneck D, Baron KT, John M, Brill PW. Radiologic-pathologic correlation in acute appendicitis: can we use it as a quality measure to assess interpretive accuracy of radiologists? *Clin Imaging.* 2011;35:421-3.
- Simianu VV, Shamitoff A, Hippe DS, et al. The reliability of a standardized reporting system for the diagnosis of appendicitis. *Curr Probl Diagn Radiol.* 2017;46:267-74.
- Aydın S, Tek C, Ergun E, Kazci O, Kosar PN. Acute appendicitis or lymphoid hyperplasia: how to distinguish more safely? *Can Assoc Radiol J.* 2019;70:354-60.
- Tanabe M, Maeda K, Kuninaka H, et al. An infant autopsy case of acute appendicitis with lymphoid hyperplasia. *Pediatr Rep.* 2025;17:96.
- Sula S, Paananen T, Tammilehto V, et al. Impact of an appendicolith and its characteristics on the severity of acute appendicitis. *BJS Open.* 2024;8:zrae093.
- Oktay C, Goksu M, Yavuz S. Prevalence of appendicolith in children with acute appendicitis and its correlation with disease severity. *North Clin Istanbul.* 2023;10:631-5.
- Weitzner ZN, Chung A, Naini BV, Graham D, Livingston EH. Correlation of computed tomography, pathological findings, and clinical outcomes for appendicoliths in appendicitis. *Ann Surg Open.* 2023;4:e280.
- Zhang D, Wang S, Li H, et al. Retrospective analysis of 331 acute appendicitis patients: how appendicolith and CT features aid in differentiating complicated vs. uncomplicated appendicitis. *BMC Gastroenterol.* 2025;26:40.
- Coutureau J, Millet I, Taourel P. CT of acute abdomen in the elderly. *Insights Imaging.* 2025;16:95.
- Sula S, Kujala M, Tammilehto V, et al. Prognostic CT-imaging findings for complicated acute appendicitis: a prospective cohort study. *Scand J Surg.* 2026;115:42-9.
- in't Hof KH, Krestin GP, Steijgerberg EW, et al. Interobserver variability in CT scan interpretation for suspected acute appendicitis. *Emerg Med J.* 2009;26:92-4.
- Alsayaf Alghamdi AG, Alzhirani SM, Fayraq A, Alzhirani SA. Predictive value of clinical and CT scan findings for complicated appendicitis: a retrospective analysis. *Cureus.* 2025;17:e88948.
- Issaiy M, Zarei D, Saghadzadeh A. Artificial intelligence and acute appendicitis: a systematic review of diagnostic and prognostic models. *World J Emerg Surg.* 2023;18:59.
- Moris D, Paulson EK, Pappas TN. Diagnosis and management of acute appendicitis in adults: a review. *JAMA.* 2021;326:2299-311.
- Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43:543-9.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46:423-9.
- Suzuki T, Matsumoto A, Sugiki D, Akao T, Matsumoto H. Clinical prediction model for gangrenous appendicitis: a retrospective single-center study. *Scand J Surg.* 2025;114:210-7.
- Von-Mühlen B, Franzon O, Beduschi MG, Kruehl N, Lupselo D. AIR score assessment for acute appendicitis. *Arq Bras Cir Dig.* 2015;28:171-3.
- Baştürk T, Duran M, Baştürk S. Evaluation of computed tomography (CT) appendicitis score and laboratory parameters in acute appendicitis with and without CT-detected appendicolith. *TJTES.* 2025;31:651-60.