Research Article

Diagnostic Accuracy of CT, Ultrasound, and Clinical Scoring in Differentiating Acute Appendicitis from Gastrointestinal Mimickers

Hamna Sajid[1], Dr israr uddin[2*], Dr. Javed Memon[3], Dr. Sara Khalid Memon[4], Dr. Muhammad Rafique Pathan[5], Dr. Shahid Nazir Memon[6]

[1]Resident radiologist, Department of diagnostic radiology, CMH, Rawalpindi.
[2*]Associate Professor ENT Khyber teaching hospital. Khyber Medical College Peshawar.
[3]Assistant professor of surgery, Indus Medical College, Tando Muhammad Khan.
[4]Assistant professor of surgery, Indus Medical College, Tando Muhammad Khan.
[5]Associate professor of surgery, Indus Medical College, Tando Muhammad Khan.
[6]Assistant professor of surgery, Indus Medical College, Tando Muhammad Khan.

ABSTRACT

Background: Acute appendicitis is one of the most common surgical emergencies; however, differentiating it from gastrointestinal mimickers remains a clinical challenge, particularly in resource-limited settings.

Objective: To evaluate and compare the diagnostic accuracy of computed tomography (CT), ultrasonography (US), and clinical scoring systems in differentiating acute appendicitis from gastrointestinal mimickers.

Methods: This prospective cross-sectional diagnostic accuracy study was conducted at department of surgery indus medical college, Tando Muhammad Khan from 1st February 2025 to 31st January 2026, included 231 consecutive patients presenting with suspected acute appendicitis. Clinical assessment was performed using the Alvarado and Appendicitis Inflammatory Response (AIR) scores. All patients underwent ultrasonography, while CT was performed in cases with inconclusive or discordant findings.

Results: Acute appendicitis was confirmed in 156 (67.5%) patients, while 75 (32.5%) were diagnosed with gastrointestinal mimickers. The Alvarado score demonstrated sensitivity of 82.7% and specificity of 64.0%, whereas the AIR score showed sensitivity of 85.3% and specificity of 69.3%. Ultrasonography achieved sensitivity of 87.2%, specificity of 72.0%, and overall accuracy of 82.7%. CT demonstrated the highest diagnostic performance with sensitivity of 95.5%, specificity of 90.0%, overall accuracy of 93.3%, and AUC of 0.94. The negative appendectomy rate was 8.8%.

Conclusion: CT scan showed superior diagnostic accuracy in differentiating acute appendicitis from gastrointestinal mimickers, followed by ultrasound. Clinical scoring systems were effective for initial risk stratification but insufficient as standalone diagnostic tools.

Keywords: Acute Appendicitis, Computed Tomography, Ultrasonography, Alvarado Score, Diagnostic Accuracy.

INTRODUCTION

Acute appendicitis is still considered one of the most widespread surgical emergencies on the international level, and the share of its presentations in emergency departments is significant in the acute abdomen [1]. Although new diagnostic procedures have been developed, the issue of prompt and correct diagnosis of appendicitis remains a problem, especially in patients with non-characteristic signs or those having other gastrointestinal complications. Late diagnosis is more prone to perforation, abscesses, and sepsis, and overdiagnosis results in unnecessary appendectomies, which puts the patients at risk of unnecessary surgery-related complications

and healthcare expenditures [2]. Historical, physical examination, and laboratory findings have been the traditional components of clinical diagnosis of acute appendicitis. Well-described classic symptoms like periumbilical pain, which radiates to the right lower quadrant, anorexia, nausea, and local tenderness are not universal or specific [3]. A number of gastrointestinal diseases, such as mesenteric lymphadenitis, gastroenteritis, enteritis, an inflammatory bowel disease exacerbation, ovarian pathology, urinary tract infection, renal colic, and right-sided diverticulitis of the colon, can be clinically similar to appendicitis [4]. These gastrointestinal mimickers tend to show similar patterns of abdominal pain, fever, leukocytosis,

and elevated inflammatory markers, and can only be differentiated on clinical grounds [5]. The dilemma in diagnosis is even more acute in children, in elderly patients, and female patients of reproductive age [6]. Children can present with non-specific symptoms, and older patients often have a blunted inflammatory response, which can delay recognition. Gynecological pathologies in women, including ovarian torsion, ectopic pregnancy, or inflammatory disease of the pelvis, can cause right lower quadrant pain that is indistinguishable from appendicitis without imaging [7]. These differences highlight the limitations of relying on clinical assessment alone and underscore the need for objective diagnostic approaches. To enhance diagnostic accuracy, clinical scoring systems have been developed, including the Alvarado score and the Appendicitis Inflammatory Response (AIR) score. The tools are used to categorize patients into low-, intermediate-, and high-risk based on symptoms, signs, and laboratory results [8]. Clinical scoring systems also come in handy for triaging patients and identifying those who do not need imaging immediately. Nevertheless, they do not show the same performance in different populations and healthcare environments. Sensitivity can be sufficient to rule out disease in low-risk populations, whereas specificity can be inadequate to be sure that gastrointestinal mimickers are excluded. As a result, the use of clinical scores will either overlook abnormal appendicitis or cause unnecessary surgery [9].

The diagnostic pathway has thus been made to focus on imaging modalities. As a non-radiation imaging modality, ultrasonography is often suggested as the first-line test in children and pregnant women because it is less harmful and does not expose the subject to radiation [10]. The diagnosis is supported by the sonographic appearance of a non-compressible, blind-ending tubular structure with a diameter exceeding 6 mm, periappendiceal fluid, or increased vascularity [11]. Nevertheless, ultrasound can be very operator-dependent and can also produce inconclusive results in obese patients or with the presence of a retrocecal appendix that could be located retrocecal. It can thus exhibit a broad range of sensitivity with respect to technical know-how and patient factors [12]. Computed tomography (CT) has been recognized as the best imaging modality in suspected appendicitis. The inflammation of the appendix, stranding of fat periappendiceal, abscess, or perforation can be clearly outlined

with contrast-enhanced CT [13]. Notably, CT not only diagnoses appendicitis but also other diagnoses that cause abdominal pain and thus decreases the negative appendectomy rates. High sensitivity and specificity of CT in the diagnosis of appendicitis have been shown in many studies. Nevertheless, the risk of exposure to ionizing radiation, especially among younger individuals, and the issues of cost and access limit its uncontrolled use [14].

Objective
To evaluate and compare the diagnostic accuracy of computed tomography (CT), ultrasonography (US), and clinical scoring systems in differentiating acute appendicitis from gastrointestinal mimickers.

METHODOLOGY
This was a prospective cross-sectional study conducted at department of surgery indus medical college, Tando Muhammad Khan from 1st February 2025 to 31st January 2026. A total of 231 consecutive patients presenting to the emergency department with right lower quadrant abdominal pain and clinical suspicion of acute appendicitis were enrolled during the study period. Patients aged 15 years and above presenting with symptoms suggestive of acute appendicitis, including right lower quadrant abdominal pain, localized tenderness, rebound tenderness, fever, nausea or vomiting, and leukocytosis, were considered eligible for inclusion. Patients with a prior history of appendectomy, hemodynamic instability requiring immediate surgery without imaging, known inflammatory bowel disease, previously confirmed gynecological or urological pathology explaining the symptoms, pregnancy where CT was contraindicated, or refusal to provide informed consent were excluded from the study. Consecutive non-probability sampling was used to enroll eligible participants.

Data Collection
A standardized clinical assessment was done to all patients at the time of presentation. There was registration of demographic information, duration of the symptoms, physical examination and laboratory tests (complete blood count and C-reactive protein). Each patient was calculated using predetermined requirements to calculate the clinical scoring systems, the Alvarado score, and the Appendicitis Inflammatory Response (AIR) score. Using the set cutoff values, patients were grouped into low-, intermediate-, or high-risk cases of acute appendicitis. As the first modality of imaging, all the enroll patients

underwent ultrasonography with the help of a high-frequency linear transducer. The sonographic features of acute appendicitis were a non-compressible blind end tubular structure whose outermost diameter measures 6 mm and above, the accumulation of peri-appendiceal fluid, the augmentation of echogenicity in the surrounding fat, and the hyperemia on the Doppler images. Where the ultrasound results were unconclusive or did not match the clinical observation, contrast-enhanced CT of the abdomen and pelvis was done. CT features of appendicitis such as appendiceal diameter more than 6 mm, thickening of the wall, peri appendiceal fat stranding, appendicolith, abscess and perforation were used as CT criteria of appendicitis. Clinical scoring results were blinded to radiologists who were interpreting radiological imaging studies in order to reduce observer bias. This was ultimately determined by the histopathological study of patients who had appendectomy to confirm the definitive diagnosis of acute appendicitis. Gastrointestinal mimickers were determined by imaging results, course and resolution of symptoms on two-four weeks follow-up of patients who were managed conservatively. To analyze the patients in terms of appendicitis and non- appendicitis, they were categorized into this composite reference standard. The most important was the diagnostic accuracy of CT, ultrasound, and clinical scoring system in the distinction between acute appendicitis and gastrointestinal mimickers. The parameters of diagnostic performance were calculated, such as sensitivity, specificity, positive predictive value,

negative predictive value, and overall accuracy of each modality. Negative appendectomy rate and alternative diagnosis finding on imaging became secondary outcomes.

Statistical Analysis

Data were analyzed using SPSS version 26.0 (IBM Corp., Armonk, NY). Continuous variables were expressed as mean ± standard deviation, while categorical variables were presented as frequencies and percentages. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated using 2×2 contingency tables. Receiver Operating Characteristic (ROC) curve analysis was performed to determine the Area under the Curve (AUC) for CT, ultrasound, and clinical scoring systems. A p-value less than 0.05 was considered statistically significant.

RESULTS

Data were collected from 231 patients, 156 (67.5%) were diagnosed with acute appendicitis and 75 (32.5%) with gastrointestinal mimickers. The mean age was slightly lower in the appendicitis group compared to the mimicker group (26.9 ± 10.8 vs 29.1 ± 12.3 years). A higher proportion of males was observed among appendicitis patients (65.4%) compared to mimickers (48.0%). The mean duration of abdominal pain was longer in appendicitis cases (29.6 ± 11.9 hours) than in mimickers (25.8 ± 13.4 hours). Clinical severity was reflected by significantly higher mean Alvarado (7.8 ± 1.2 vs 5.3 ± 1.5) and AIR scores (8.2 ± 1.4 vs 5.6 ± 1.7) in the appendicitis group.

Table 1: Baseline Demographic and Clinical Characteristics (N = 231)

Variable	Appendicitis (n=156)	Mimickers (n=75)
Age (years), mean ± SD	26.9 ± 10.8	29.1 ± 12.3
Male, n (%)	102 (65.4%)	36 (48.0%)
Duration of pain (hours), mean ± SD	29.6 ± 11.9	25.8 ± 13.4
Alvarado score, mean ± SD	7.8 ± 1.2	5.3 ± 1.5
AIR score, mean ± SD	8.2 ± 1.4	5.6 ± 1.7

Among the 75 patients diagnosed with gastrointestinal mimickers, acute gastroenteritis was the most frequent alternative diagnosis, accounting for 28.0% of

cases. Gynecological causes represented 24.0% of diagnoses, followed by mesenteric lymphadenitis (21.3%), ileitis (14.7%), and right-sided diverticulitis (12.0%).

Table 2: Distribution of Gastrointestinal Mimickers (N = 75)

Diagnosis	n (%)
Acute gastroenteritis	21 (28.0%)
Mesenteric lymphadenitis	16 (21.3%)
Ileitis	11 (14.7%)
Right-sided diverticulitis	9 (12.0%)

Gynecological causes	18 (24.0%)

In diagnostic performance analysis, CT scan demonstrated the highest sensitivity (95.5%), specificity (90.0%), and overall accuracy (93.3%), with statistically significant association (p < 0.001). Ultrasound showed good sensitivity (87.2%) and moderate specificity (72.0%), with overall accuracy of 82.7%. Among clinical scoring systems, the AIR score performed better than the Alvarado score, showing sensitivity of 85.3% and specificity of 69.3% compared to 82.7% and 64.0%, respectively.

Table 3: Diagnostic Performance of Clinical Scoring Systems and Imaging Modalities

Modality	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	P-Value*
Alvarado Score (≥7)	82.7	64.0	82.1	65.3	76.2	<0.001
AIR Score (≥8)	85.3	69.3	85.8	68.4	80.1	<0.001
Ultrasound	87.2	72.0	90.1	66.7	82.7	<0.001
CT Scan	95.5	90.0	94.6	91.8	93.3	<0.001

Receiver Operating Characteristic (ROC) curve analysis demonstrated superior discriminatory ability of CT scan with an AUC of 0.94 (95% CI: 0.89–0.98). Ultrasound showed good performance with an AUC of 0.81 (95% CI: 0.73–0.89), while AIR (AUC 0.79) and Alvarado (AUC 0.76) scores demonstrated moderate discrimination. All AUC values were statistically significant (p < 0.001).

Table 4: Roc Curve Comparison

Modality	AUC	95% CI	P-Value (AUC Vs. 0.5)
CT Scan	0.94	0.89 – 0.98	<0.001
Ultrasound	0.81	0.73 – 0.89	<0.001
Alvarado Score	0.76	0.68 – 0.84	<0.001
AIR Score	0.79	0.71 – 0.87	<0.001

Comparative ROC Curves
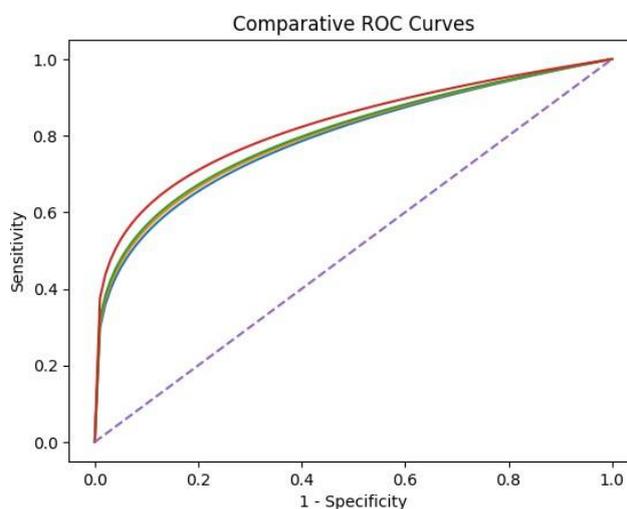
DISCUSSION

The current paper compared and contrasted the diagnostic accuracies of computed tomography (CT), ultrasonography (US), and clinical scoring systems for distinguishing acute appendicitis from gastrointestinal mimickers at a tertiary care hospital in Pakistan. 67.5% of the patients were diagnosed with acute appendicitis, and 32.5% were diagnosed with other gastrointestinal or gynecological complications. Such results represent the clinical dilemma usually faced in emergency departments, especially in under-resourced environments, where reliance on clinical opinion alone can lead to incorrect ascertainment. Clinical scoring systems were found to have a reasonable sensitivity but moderate specificity. The Alvarado score has a

sensitivity of 82.7% and a specificity of 64.0%, while the AIR score has slightly better diagnostic performance, with a sensitivity of 85.3% and a specificity of 69.3%. These tools are useful for initial risk stratification, but their specificity is low; thus, they cannot reliably distinguish appendicitis from mimickers [15]. This is especially relevant in our population, where inflammatory diseases such as gastroenteritis and mesenteric lymphadenitis are prevalent and can increase leukocyte counts, thereby inflating clinical scores. The results of these studies provide the rationale for considering clinical scoring more as a supplementary instrument for diagnosis than as a conclusive outcome [16]. The ultrasonography showed positive sensitivity (87.2) but relatively low specificity (72.0). This is in line with prior regional and international research showing that ultrasound has a good capacity to detect appendicitis, but challenges in ruling out other pathologies that mimic it. Continuous dependencies of operators, the absence of visualization of the appendix, and factors associated with patients, like obesity or excessive bowel gas, probably played a role in inconclusive results in about 9 percent of cases [17]. Nevertheless, ultrasound can be effectively used as the first-line imaging modality in situations where radiation or financial constraints preclude CT. CT scan had the best diagnostic performance among the modalities tested, with a sensitivity of 95.5, specificity of 90.0, and overall accuracy of 93.3. The AUC at 0.94 is another indication of its excellent discriminative power [18]. CT was able to both prove appendicitis and effectively pick up alternate diagnoses, thus minimizing unnecessary surgeries. Our study has a negative appendectomy rate of 8.8, which is relatively low, and this can be attributed to increased use of imaging before surgery. This rate aligns with modern evidence that cross-sectional imaging significantly reduces negative appendectomy rates compared with clinical examination alone [19,20]. The superiority of CT over ultrasound and clinical scoring systems in our cohort indicates the value of imaging-based confirmation, especially in cases of equivocality. Its use, however, needs to be weighed against cost and radiation exposure considerations. Financial constraints can limit universal use of CT in the Pakistani healthcare setting, where it can be seen that borderline patients have been operated on without CT because of financial reasons. This highlights the importance of context-sensitive diagnostic algorithms that integrate clinical evaluation with selective imaging. A number of limitations must be mentioned. This was a single-center study, which may limit generalizability. Sampling can be biased due to non-probability sampling. Also, not every patient underwent CT imaging, which could affect comparative estimates. In spite of these shortcomings, the study reflects actual practice in an emergency department in a developing country and provides clinically significant data.

CONCLUSION

It is concluded that computed tomography demonstrated the highest diagnostic accuracy in differentiating acute appendicitis from gastrointestinal mimickers in patients presenting with right lower quadrant abdominal pain. Ultrasonography showed good sensitivity and remains a valuable first-line imaging modality, particularly in resource-limited settings, while clinical scoring systems such as the Alvarado and AIR scores were useful for initial risk stratification but lacked sufficient specificity for definitive diagnosis.

REFERENCES

1. Iqbal J, Sayani R, Tahir M, et al. (March 01, 2018) Diagnostic Efficiency of Multidetector Computed Tomography in the Evaluation of Clinically Equivocal Cases of Acute Appendicitis with Surgical Correlation. Cureus 10(3): e2249. DOI 10.7759/cureus.2249

2. Callahan MJ, Rodriguez DP, Taylor GA. CT of appendicitis in children. Radiology. 2002; 224(2):325-332. doi:10.1148/radiol.2242010998

3. Cotton D, Vinson DR, Vazquez-Benitez G, et al. Validation of the Pediatric Appendicitis Risk Calculator (PARC) in a community emergency department setting. Ann Emerg Med. 2019; 74(4):471-480. doi:10.1016/j.annemergmed.2019.04.023

4. Pereira JM, Sirlin CB, Pinto PS, Jeffrey RB Jr, Stella DL, Casola G. Disproportionate fat stranding: a helpful CT sign in patients with acute abdominal pain. Radiographics. 2004; 24(3):703-715. doi:10.1148/rg.243035084

5. Puylaert JB. Acute appendicitis: US evaluation using graded compression. Radiology. 1986; 158(2):355-360. doi:10.1148/radiology.158.2.2934762

6. Doherty GM, Way LW. Current surgical diagnosis & treatment. 12th ed. New

York: McGraw-Hill; 2006. ISBN:007142315X

7. Pinto Leite N, Pereira JM, Cunha R, Pinto P, Sirlin C. CT evaluation of appendicitis and its complications: imaging techniques and key diagnostic findings. AJR Am J Roentgenol. 2005; 185(2):406-417. doi:10.2214/ajr.185.2.01850406

8. Fritsch H, Kühnel W. Color atlas of human anatomy. 5th ed. Stuttgart: Thieme; 2008. ISBN:9783135334059

9. Shin LK, Jeffrey RB Jr, Berry GJ, Olcott EW. Spectral Doppler waveforms for diagnosis of appendicitis: potential utility of point peak systolic velocity and resistive index values. Radiology. 2017; 285(3):990-998. doi:10.1148/radiol.2017162251

10. Gaitini D, Beck-Razi N, Mor-Yosef D, et al. Diagnosing acute appendicitis in adults: accuracy of color Doppler sonography and MDCT compared with surgery and clinical follow-up. AJR Am J Roentgenol. 2008; 190(5):1300-1306. doi:10.2214/ajr.07.2955

11. Connolly JA, Dean AJ, Hoffmann B, et al. Emergency point-of-care ultrasound. Hoboken: Wiley-Blackwell; 2017. ISBN:9780470657577

12. Avanesov M, Wiese NJ, Karul M, et al. Diagnostic prediction of complicated appendicitis by combined clinical and radiological appendicitis severity index (APSI). Eur Radiol. 2018; 28(9):3601-3610. doi:10.1007/s00330-018-5339-9

13. Monsonis B, Mandoul C, Millet I, Taourel P. Imaging of appendicitis: tips and tricks. Eur J Radiol. 2020; 130:109165. doi:10.1016/j.ejrad.2020.109165

14. Bahrami M, Mirgaloyebayat H, Mohajeri Z, Mohammadi H, Afshari SA, Fazeli P, Masaeli D, Nourian SMA. The diagnostic value of the computed tomography scan and ultrasonography in acute appendicitis. Am J Nucl Med Mol Imaging. 2023 Feb 15; 13(1):11-17. PMID: 36923598; PMCID: PMC10009470.

15. Zhao Y, Liu H, Guo D. Diagnostic efficacy of ultrasound and computed tomography for acute appendicitis: A single center retrospective study. Medicine (Baltimore). 2025 Mar 28; 104(13):e41968. doi: 10.1097/MD.0000000000041968. PMID: 40153755; PMCID: PMC11957627.

16. Noori IF, Jabbar AS, Noori AF. Clinical scores (Alvarado and AIR scores) versus imaging (ultrasound and CT scan) in the diagnosis of equivocal cases of acute appendicitis: a randomized controlled study. Ann Med Surg (Lond). 2023 Mar 14; 85(4):676-683. doi: 10.1097/MS9.0000000000000270. PMID: 37113930; PMCID: PMC10129239.

17. Koroth A, Basheer S, Abdul Rasheed MF, Ali Usman A, Sadanandan A. Association Between Raja Isteri Pengiran Anak Saleha Appendicitis (RIPASA) Scoring and Pathological Severity of Acute Appendicitis: A Cross-Sectional Study. Cureus. 2024 Mar 14; 16(3):e56166. doi: 10.7759/cureus.56166. PMID: 38618466; PMCID: PMC11015883.

18. Ozturk MO, Resorlu M, Aydin S, Memis KB. Use of the vertebrae and iliac bone as references for localizing the appendix vermiformis in computed tomography. World J Radiol. 2024 Nov 28; 16(11):629-637. doi: 10.4329/wjr.v16.i11.629. PMID: 39635315; PMCID: PMC11612806.

19. Yoon KW, Yoo K, Choi K, Gil E, Park CM, Lee D. The implementation of the acute care surgery model in the management of patients with acute appendicitis - A 5-year single-center, retrospective experience: An observational study. Medicine (Baltimore). 2024 Jul 19; 103(29):e38927. doi: 10.1097/MD.0000000000038927. PMID: 39029014; PMCID: PMC11398796.

20. Lu, Y.-T., Chen, P.-C., Huang, Y.-H., & Huang, F.-C. (2020). Making a Decision between Acute Appendicitis and Acute Gastroenteritis. *Children*, *7*(10), 176. https://doi.org/10.3390/children7100176