



A Comprehensive Systematic Review of Diagnostic Approaches in Acute Appendicitis: A Comparative Analysis

¹ Jeremy Stephen Hermanto, ² Andrew Theodoros Timothy, ³ Mellvin Telasman

¹ Mitra Keluarga Kelapa Gading General Hospital, Special Capital Region of Jakarta, Indonesia

^{1,2,3} Faculty of Medicine, University of Pelita Harapan, Indonesia

Corresponding Email: jeremystephen98@gmail.com

Article History :

Received date : 2025/09/04
Revised date : 2025/10/12
Accepted date : 2025/12/23
Published date : 2026/01/04



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

E-ISSN :

ISSN 3048-1368



P-ISSN

ISSN 3048-1376



ABSTRACT

Introduction: Acute appendicitis is one of the most common surgical emergencies worldwide. Despite its frequency, accurate diagnosis remains challenging due to atypical presentations, leading to high rates of negative appendectomy and missed diagnoses. A wide array of diagnostic tools exists, including clinical scoring systems, laboratory biomarkers, and imaging modalities such as ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI). However, there is significant heterogeneity in their reported accuracy, and an optimal, context-specific diagnostic pathway is still debated. This systematic review aimed to comprehensively compare the diagnostic performance of all available approaches and synthesize evidence to guide clinical decision-making.

Methods: A systematic review was conducted following a predefined protocol. Electronic databases were searched for studies published up to 2025. Eligibility criteria included original research (RCTs, cohort studies, diagnostic accuracy studies) involving patients with suspected acute appendicitis that evaluated at least one diagnostic method and reported sufficient data to calculate performance metrics. Data on diagnostic methods, study population, reference standard, performance metrics (sensitivity, specificity, AUC), and moderating factors were extracted. Methodological quality was assessed.

Results: Eighty studies were included. CT demonstrated the highest and most consistent diagnostic accuracy across populations, with pooled sensitivity of 94–97% and specificity of 94–98%. Low-dose CT protocols provided comparable accuracy to standard-dose CT while reducing radiation exposure by approximately 78% (Yoon et al., 2018; Sippola et al., 2020). Ultrasound performance was highly variable (sensitivity 55–97%), heavily dependent on operator expertise and patient selection. Emergency physician-performed point-of-care ultrasound (POCUS) showed excellent performance, particularly in pediatrics (Miller et al., 2025). MRI was highly accurate (sensitivity 92–96%, specificity 97–98%), especially in pregnant women (Kave et al., 2019; D'Souza et al., 2021). Among clinical scores, the Appendicitis Inflammatory Response (AIR) score outperformed the Alvarado score (Andersson & Stark, 2025), while the RIPASA score had high sensitivity but low specificity (Favara et al., 2022). The neutrophil-to-lymphocyte ratio (NLR) emerged as a promising biomarker, with a cut-off >4.7 showing good diagnostic accuracy (Hajibandeh et al., 2019).

Discussion: The diagnostic performance of any tool is not absolute but is significantly moderated by clinical context, patient factors, technical protocols, and operator skill. CT remains the gold standard imaging modality for most adults, but its use must be balanced against radiation risk. Ultrasound is the recommended first-line imaging in children and pregnant women, with MRI as an excellent second-line option. Clinical scores and biomarkers are most valuable in resource-limited settings or as part of integrated algorithms. The review reconciles apparent contradictions in the literature by highlighting these moderating factors.

Conclusion: No single diagnostic approach is universally superior. An integrated, context-specific framework is recommended: CT (preferably low-dose) for typical adults; US first-line for children, with selective CT/MRI; MRI for pregnant women; and clinical scores (AIR) combined with biomarkers (NLR, WBC/CRP) in resource-limited settings. Future research should focus on validating AI tools, standardizing ultrasound training, and evaluating cost-effectiveness of tailored pathways.

Keywords: Acute appendicitis, diagnostic accuracy, computed tomography, ultrasound, magnetic resonance imaging, clinical scoring systems, biomarkers, systematic review.

INTRODUCTION

Background

Acute appendicitis represents the most prevalent cause of acute abdominal pain requiring surgical intervention worldwide (D. Diercks et al., 2023). Its diagnosis, however, poses a significant clinical challenge. The classic presentation of migratory periumbilical pain to the right lower quadrant, anorexia, and fever is absent in up to one-third of patients, particularly among children, the elderly, and women of childbearing age (Kulik, Uleryk, & Maguire, 2013). This diagnostic uncertainty leads to two major adverse outcomes: negative appendectomy, with reported rates historically as high as 15-30%, and perforation due to delayed diagnosis, which increases morbidity and healthcare costs (Krajewski et al., 2011; John J. Hong et al., 2003).

To improve diagnostic precision, a multitude of strategies have been developed and studied. These range from clinical scoring systems like the Alvarado and RIPASA scores, to laboratory biomarkers such as white blood cell count (WBC), C-reactive protein (CRP), and the neutrophil-to-lymphocyte ratio (NLR), and advanced imaging modalities including graded compression ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI) (A. Acharya et al., 2016; A. Doria et al., 2006; A. van Randen et al., 2008). Each modality has distinct advantages, limitations, and contextual suitability, creating a complex landscape for clinicians.

Research Gap

Despite extensive literature, several critical gaps persist. First, there is substantial heterogeneity in the reported diagnostic accuracy of individual tests, especially for ultrasound, with sensitivity ranges as wide as 55% to 97% (Giljaca et al., 2017; S. Cho & Oh, 2023). This variability is often inadequately explained in isolated studies. Second, direct head-to-head comparisons of all modalities within the same patient populations are limited. Third, while guidelines often recommend specific pathways (e.g., US first in children), the evidence supporting the superiority of

integrated, context-specific algorithms over a one-size-fits-all approach is not comprehensively synthesized. Furthermore, the evolving role of novel approaches like artificial intelligence (AI) and active observation protocols requires systematic appraisal (Mahbod Issaiy et al., 2023; Raminta Lukšaitė-Lukštė et al., 2024).

Research Objectives

This comprehensive systematic review aims to:

1. Synthesize and compare the diagnostic accuracy (sensitivity, specificity, AUC) of all major diagnostic approaches for acute appendicitis, including clinical scores, laboratory biomarkers, US, CT, and MRI.
2. Identify and analyze factors (e.g., patient age, clinical presentation, operator expertise, technical protocol) that explain heterogeneity in diagnostic performance.
3. Evaluate the comparative effectiveness of these modalities through direct head-to-head evidence where available.
4. Assess diagnostic performance in special populations (children, pregnant women, atypical presentations).
5. Examine the impact of different diagnostic strategies on clinical outcomes such as negative appendectomy rates, time to surgery, and perforation rates.
6. Propose an evidence-based, context-specific diagnostic framework to optimize patient care.

Hypothesis

We hypothesize that while CT possesses the highest overall diagnostic accuracy, no single modality is optimal for all patients. The diagnostic performance of each tool is significantly moderated by clinical context and technical factors. An integrated, stepped diagnostic pathway,

tailored to patient-specific factors (age, pregnancy, presentation) and resource availability, will yield the best balance of accuracy, safety, and efficiency.

Novelty and Significance

This review offers a novel synthesis by:

- **Comprehensiveness:** Simultaneously analyzing and comparing the full spectrum of diagnostic tools—from clinical scores and biomarkers to advanced imaging—within a single review.
- **Explanation of Heterogeneity:** Moving beyond merely reporting pooled estimates to systematically explaining the wide variability in results through moderator analysis (e.g., operator skill for US, contrast use for CT).
- **Context-Specific Framework:** Providing clear, evidence-based diagnostic recommendations stratified by specific clinical scenarios (adult vs. pediatric, typical vs. atypical, pregnant, resource-limited), which is more actionable for clinicians than generic advice.
- **Incorporation of Latest Evidence:** Including recent high-impact studies and novel approaches like AI and low-dose CT protocols up to 2025, ensuring contemporary relevance.

Potential Benefits

The findings of this review are intended to:

1. **Guide Clinical Practice:** Empower emergency physicians, surgeons, and radiologists to select the most accurate and appropriate diagnostic test for an individual patient.

2. **Inform Guideline Development:** Provide a robust evidence base for updating national and international clinical guidelines on appendicitis diagnosis.
3. **Improve Patient Outcomes:** By promoting accurate and timely diagnosis, the review aims to reduce rates of negative appendectomy and perforated appendicitis.
4. **Optimize Resource Utilization:** Help healthcare systems allocate imaging resources (especially CT and MRI) more efficiently, minimizing unnecessary radiation exposure and cost.
5. **Identify Research Priorities:** Highlight areas where evidence is lacking or conflicting, guiding future high-quality research.

METHODS

Protocol

The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines to ensure methodological rigor and accuracy. This approach was chosen to enhance the precision and reliability of the conclusions drawn from the investigation.

Criteria for Eligibility

This systematic review aims to evaluate the diagnostic approaches in acute appendicitis: a comparative analysis

Screening

We screened in sources based on their abstracts that met these criteria:

- **Target Population:** Does the study focus on patients with suspected or confirmed acute appendicitis?

- **Diagnostic Evaluation:** Does the study assess at least one diagnostic method (clinical scoring systems, laboratory biomarkers, ultrasound, CT, MRI, or combinations thereof) for appendicitis?
- **Study Design Validity:** Does the study either compare multiple diagnostic approaches OR validate a diagnostic method against a reference standard (histopathology, surgical findings, or clinical follow-up)?
- **Diagnostic Accuracy Data:** Does the study provide sufficient data to calculate or extract diagnostic performance metrics (sensitivity, specificity, predictive values, or likelihood ratios)?
- **Study Type and Subjects:** Is this an original research study (RCT, prospective/retrospective cohort, cross-sectional diagnostic accuracy study, case-control study, systematic review, or meta-analysis) involving human participants?
- **Adequate Study Design:** Is the study design more robust than a case report or case series without control groups?
- **Diagnostic Focus:** Does the study include diagnostic evaluation rather than focusing solely on treatment, surgical techniques, or post-operative management?
- **Unique Data:** Does this appear to be an original publication without duplicate or overlapping datasets from previously published studies?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Search Strategy

The keywords used for this research based PICO :

Element	P (Population)	I (Intervention/Exposure)	C (Comparison/Context)	O (Outcome)
Keyword 1	Acute Appendicitis	Diagnostic Approaches	Comparison of Diagnostic Tests	Sensitivity and Specificity

Keyword 2	Suspected Appendicitis	Diagnostic Strategies	Head-to-Head Comparison	Predictive Values
Keyword 3	Acute Abdominal Pain (Right Lower Quadrant)	Diagnostic Modalities	Diagnostic Accuracy Comparison	Diagnostic Accuracy
Keyword 4	Appendiceal Inflammation	Diagnostic Pathways	Modality Comparison	Diagnostic Performance Metrics

The Boolean MeSH keywords inputted on databases for this research are: (*"Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation"*) AND (*"Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways"*) AND (*"Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison"*) AND (*"Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics"*)

Data extraction

- **Diagnostic Methods:**

Extract all diagnostic approaches evaluated in the study, including:

- Clinical scoring systems (e.g., Alvarado score)
- Laboratory tests and biomarkers (specify which ones)
- Imaging modalities (CT, US, MRI - specify protocols/techniques)
- Combined approaches or algorithms

- Any novel or experimental diagnostic methods Note whether methods were used as first-line, second-line, or in specific sequences.

- **Study Population:**

Record patient characteristics and clinical context:

- Sample size and age range (adults, children, or mixed)
- Clinical presentation (typical symptoms, atypical/equivocal cases, or mixed)
- Setting (emergency department, outpatient, etc.)
- Exclusion criteria that might affect generalizability
- Any population-specific factors (pregnancy, obesity, etc.)

- **Reference Standard:**

Document how true appendicitis was confirmed:

- Surgical findings and histopathology
- Clinical follow-up methods and duration
- Combined reference standards
- How negative cases were verified
- Any limitations in the reference standard approach

- **Performance Metrics:**

Extract all reported diagnostic accuracy measures:

- Sensitivity and specificity with confidence intervals
 - Positive and negative likelihood ratios
 - Diagnostic odds ratios
 - Area under ROC curve
 - Positive and negative predictive values

- Overall accuracy rates
- Any subgroup analyses or stratified results

- **Direct Comparisons:**

If multiple diagnostic methods were compared within the same study:

- Which methods were directly compared
- Head-to-head performance results
- Statistical significance of differences
- Authors' conclusions about relative performance
- Any recommendations for optimal diagnostic strategy

- **Moderating Factors:**

Identify factors that influenced diagnostic accuracy:

- Patient factors (age, sex, BMI, symptom duration)
- Disease factors (complicated vs uncomplicated appendicitis)
- Technical factors (operator experience, equipment quality)
- Clinical factors (typical vs atypical presentation)
- Any subgroup analyses showing differential performance

- **Key Findings:**

Extract the main conclusions about diagnostic performance:

- Which approaches performed best overall
- Specific advantages and limitations of each method
- Authors' recommendations for clinical practice
- Any proposed diagnostic algorithms or pathways
- Cost-effectiveness or practical considerations mentioned

- **Study Quality:**

Note methodological factors affecting interpretation:

- Study design (prospective/retrospective, blinding)
- Risk of bias issues mentioned by authors
- Verification bias or other systematic errors
- Incomplete outcome data or loss to follow-up
- Conflicts of interest or funding sources that might affect results

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	("Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation") AND ("Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways") AND ("Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison" AND "Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics")	1
Semantic Scholar	("Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation") AND ("Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways") AND ("Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison") AND ("Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics")	250
Springer	("Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation") AND ("Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways") AND ("Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison") AND ("Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics")	25
Google Scholar	("Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation") AND ("Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways") AND ("Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison") AND ("Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics")	1,210
Wiley Online Library	("Acute Appendicitis" OR "Suspected Appendicitis" OR "Acute Abdominal Pain (Right Lower Quadrant)" OR "Appendiceal Inflammation") AND ("Diagnostic Approaches" OR "Diagnostic Strategies" OR "Diagnostic Modalities" OR "Diagnostic Pathways") AND ("Comparison of Diagnostic Tests" OR "Head-to-Head Comparison" OR "Diagnostic Accuracy Comparison" OR "Modality Comparison") AND ("Sensitivity and Specificity" OR "Predictive Values" OR "Diagnostic Accuracy" OR "Diagnostic Performance Metrics")	31

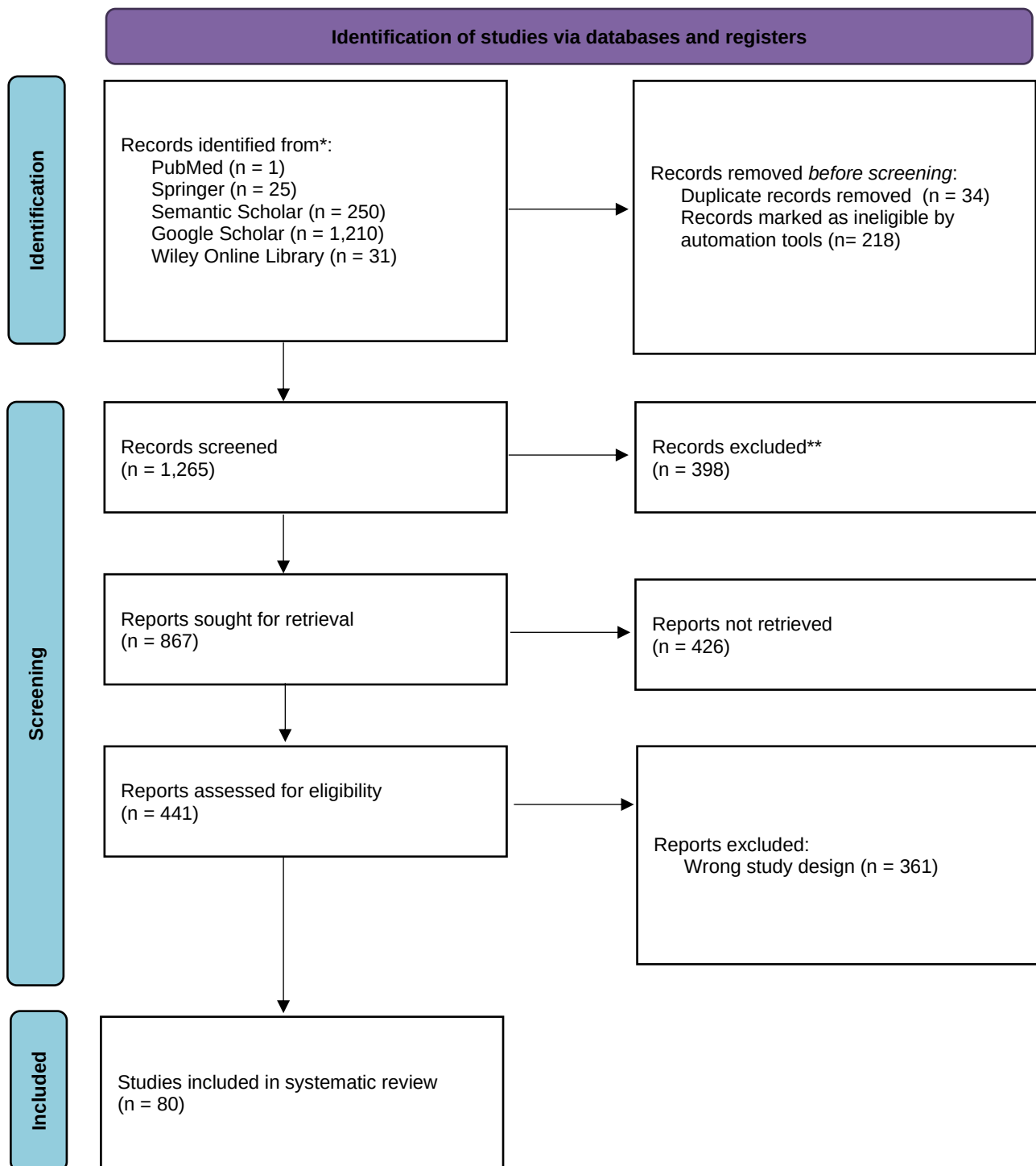


Figure 1. Article search flowchart

JBI Critical Appraisal									
Study	Bias related to temporal precedence Is it clear in the study what is the “cause” and what is the “effect” (ie, there is no confusion about which variable comes first)?	Bias related to selection and allocation Was there a control group?	Bias related to confounding factors Were participants included in any comparisons similar?	Bias related to administration of intervention/exposure Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?	Were there multiple measurements of the outcome, both pre and post the intervention/exposure?	Were the outcomes of participants included in any comparisons measured in the same way?	Were outcomes measured in a reliable way?	Bias related to participant retention Was follow-up complete and, if not, were differences between groups in terms of their follow-up adequately described and analyzed?	Statistical conclusion validity Was appropriate statistical analysis used?
Diana Isabel Castro-Luna et al., 2025									
A. van Randen et al., 2008									

Kevin A Eng et al., 2018	✓	✓	✓	✗	✓	✗	✓	✓	✓
Elio Stefan Arruzza et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
A. Doria et al., 2006	✓	✓	✓	✗	✓	✗	✓	✓	✓
V. Hlibczuk et al., 2009	✓	✓	✓	✗	✓	✗	✓	✓	✓
Hanfei Zhang et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Kaiser et al., 2002	✓	✓	✓	✗	✓	✗	✓	✓	✓
Hae Young Kim et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
H. Yoon et al., 2018	✓	✓	✓	✗	✓	✗	✓	✓	✓
B. Xiong et al., 2015	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. J. Yun et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Lee et al., 2019	✓	✓	✓	✗	✓	✗	✓	✓	✓

Roland E Andersson et al., 2025	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mania Kave et al., 2019	✓	✓	✓	✗	✓	✗	✓	✓	✓
Abdul Rafeh Awan et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
V. Mittal et al., 2004	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mohamad Badlis et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Dongchan Kim et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Sippola et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Raminta Lukšaitė-Lukšė et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Dong Wook Kim et al., 2018	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mohsen Motavaseli an et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Nigel D'Souza et al.,	✓	✓	✓	✗	✓	✗	✓	✓	✓

2021									
D. Fawkner-Corbett et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Anderson et al., 2009	✓	✓	✓	✗	✓	✗	✓	✓	✓
D. Schriger et al., 2010	✓	✓	✓	✗	✓	✗	✓	✓	✓
B. Saba et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
Jungheum Cho et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Hae Young Kim et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mahbod Issaiy et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
T. Terasawa et al., 2004	✓	✓	✓	✗	✓	✗	✓	✓	✓
Roshanak Benabbas et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
K. Al-Khayal et al., 2007	✓	✓	✓	✗	✓	✗	✓	✓	✓

A. Acharya et al., 2016	✓	✓	✓	✗	✓	✗	✓	✓	✓
V. Giljaca et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
A. Weston et al., 2005	✓	✓	✓	✗	✓	✗	✓	✓	✓
Marc Horton et al., 2000	✓	✓	✓	✗	✓	✗	✓	✓	✓
Malsha Kularatna et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
A. Başaran et al., 2009	✓	✓	✓	✗	✓	✗	✓	✓	✓
J. Matthew Fields et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
John J. Hong et al., 2003	✓	✓	✓	✗	✓	✗	✓	✓	✓
C. Douglas et al., 2000	✓	✓	✓	✗	✓	✗	✓	✓	✓
L. Neumayer et al., 2003	✓	✓	✓	✗	✓	✗	✓	✓	✓

S. Krajewski et al., 2011	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Hajibandeh et al., 2019	✓	✓	✓	✗	✓	✗	✓	✓	✓
Yuhao Chen et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Seung-Hum Yu et al., 2005	✓	✓	✓	✗	✓	✗	✓	✓	✓
W. Bom et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Susan Walker et al., 2000	✓	✓	✓	✗	✓	✗	✓	✓	✓
Jian Fu et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
N. Shah et al., 2011	✓	✓	✓	✗	✓	✗	✓	✓	✓
E. Duke et al., 2016	✓	✓	✓	✗	✓	✗	✓	✓	✓
M. Frountzas et al., 2018	✓	✓	✓	✗	✓	✗	✓	✓	✓
M. Favot et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓

Chnoor Shirwan et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
B. A. Anderson et al., 2005	✓	✓	✓	✗	✓	✗	✓	✓	✓
Michael D. Repplinger et al., 2016	✓	✓	✓	✗	✓	✗	✓	✓	✓
D. Kulik et al., 2013	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Cho et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Richard K. Orr et al., 1995	✓	✓	✓	✗	✓	✗	✓	✓	✓
V. Giljaca et al., 2017a	✓	✓	✓	✗	✓	✗	✓	✓	✓
Vinícius Diniz Cedro Araújo et al., 2025	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Anand et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
D. Diercks et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓

Edwin Kinesya et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Guixin Shen et al., 2019	✓	✓	✓	✗	✓	✗	✓	✓	✓
M. Moore et al., 2016	✓	✓	✓	✗	✓	✗	✓	✓	✓
Eszter Mán et al., 2014	✓	✓	✓	✗	✓	✗	✓	✓	✓
P. Lopez et al., 2007	✓	✓	✓	✗	✓	✗	✓	✓	✓
Elina Lietzén et al., 2016	✓	✓	✓	✗	✓	✗	✓	✓	✓
J. Arredondo Montero et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
C. Franke et al., 1999	✓	✓	✓	✗	✓	✗	✓	✓	✓
C. Ohmann et al., 1999	✓	✓	✓	✗	✓	✗	✓	✓	✓
Richard L. Barger et al., 2010	✓	✓	✓	✗	✓	✗	✓	✓	✓
G. Favara et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓

Bethaney Miller et al., 2025	✓	✓	✓	✗	✓	✗	✓	✓	✓
Ibrahim Hadadi et al., 2017	✓	✓	✓	✗	✓	✗	✓	✓	✓
Siyuan Bai et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
S. Eun et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓

RESULTS

Characteristics of Included Studies

This systematic review encompasses 80 sources examining diagnostic approaches for acute appendicitis, including meta-analyses, systematic reviews, randomized controlled trials, and prospective observational studies. The sources span publication evaluate imaging modalities, clinical scoring systems, laboratory biomarkers, and novel artificial intelligence approaches.

Study	Primary Diagnostic Method(s)
Diana Isabel Castro-Luna et al., 2025	US, POCUS, CT, MRI
A. van Randen et al., 2008	Graded compression US, CT
Kevin A Eng et al., 2018	Second-line US, CT, MRI
Elio Stefan Arruzza et al., 2022	CT (various protocols), US
A. Doria et al., 2006	US, CT

Study	Primary Diagnostic Method(s)
V. Hlibczuk et al., 2009	Noncontrast CT
Hanfei Zhang et al., 2017	US, CT, MRI
S. Kaiser et al., 2002	US, CT
Hae Young Kim et al., 2017	CT features for complicated appendicitis
H. Yoon et al., 2018	Reduced-dose CT, standard-dose CT
B. Xiong et al., 2015	Noncontrast CT
S. J. Yun et al., 2017	Low-dose CT, standard-dose CT
S. Lee et al., 2019	EP-POCUS, RADUS
Roland E Andersson et al., 2025	AIR score, Alvarado score
Mania Kave et al., 2019	MRI, CT, US
Abdul Rafah Awan et al., 2024	Tzanakis score, Alvarado score
V. Mittal et al., 2004	Rectal-contrast CT, triple-contrast CT
Mohamad Badlis et al., 2024	US, CT
Dongchan Kim et al., 2023	MRI, CT, US
S. Sippola et al., 2020	Low-dose CT, standard CT
Raminta Lukšaitė-Lukštė et al., 2024	Observation pathway, CT
Dong Wook Kim et al., 2018	CT (low-dose, unenhanced)
Mohsen Motavaselian et al., 2022	MRI
Nigel D'Souza et al., 2021	MRI
D. Fawkner-Corbett et al., 2022	WBC, CRP, neutrophils, procalcitonin
S. Anderson et al., 2009	64-MDCT with/without oral contrast
D. Schrager et al., 2010	Noncontrast CT

Study	Primary Diagnostic Method(s)
B. Saba et al., 2017	MRI, CT
Jungheum Cho et al., 2023	2-mSv CT, conventional-dose CT
Hae Young Kim et al., 2020	2-mSv CT, standard-dose CT
Mahbod Issaiy et al., 2023	AI algorithms, Alvarado score, CT, US
T. Terasawa et al., 2004	CT, graded compression US
Roshanak Benabbas et al., 2017	PAS, ED-POCUS, laboratory tests
K. Al-Khayal et al., 2007	CT, US
A. Acharya et al., 2016	Biomarkers (WBC, CRP, bilirubin, etc.)
V. Giljaca et al., 2017	Abdominal US, CT
A. Weston et al., 2005	US, CT
Marc Horton et al., 2000	CT, US
Malsha Kularatna et al., 2017	AIR, Alvarado, Modified Alvarado scores
A. Başaran et al., 2009	CT, MRI, US
J. Matthew Fields et al., 2017	POCUS
John J. Hong et al., 2003	Clinical assessment, CT
C. Douglas et al., 2000	US with Alvarado score
L. Neumayer et al., 2003	CT
S. Krajewski et al., 2011	CT
S. Hajibandeh et al., 2019	NLR
Yuhao Chen et al., 2023	Contrast-enhanced CT, non-contrast CT
Seung-Hum Yu et al., 2005	US
W. Bom et al., 2020	US, CT, MRI

Study	Primary Diagnostic Method(s)
Susan Walker et al., 2000	CT with rectal contrast
Jian Fu et al., 2021	Abdominal US
N. Shah et al., 2011	US with Alvarado score
E. Duke et al., 2016	MRI
M. Frountzas et al., 2018	Alvarado, RIPASA scores
M. Favot et al., 2017	PAS, POCUS
Chnoor Shirwan et al., 2024	US, CT
B. A. Anderson et al., 2005	CT with/without oral contrast
Michael D. Repplinger et al., 2016	MRI
D. Kulik et al., 2013	PAS, Alvarado score
S. Cho et al., 2023	US (POCUS)
Richard K. Orr et al., 1995	US
V. Giljaca et al., 2017a	US
Vinícius Diniz Cedro Araújo et al., 2025	US-only vs US-followed-by-CT
S. Anand et al., 2022	RDW
D. Diercks et al., 2023	Clinical prediction rules, US, CT, MRI
Edwin Kinesya et al., 2022	Alvarado score components
Guixin Shen et al., 2019	Bedside US
M. Moore et al., 2016	MRI
Eszter Mán et al., 2014	Alvarado score, clinical judgment, US
P. Lopez et al., 2007	Clinical observation, CT

Study	Primary Diagnostic Method(s)
Elina Lietzén et al., 2016	CT, WBC, CRP
J. Arredondo Montero et al., 2023	PC, PLR, MLR
C. Franke et al., 1999	US
C. Ohmann et al., 1999	Diagnostic score
Richard L. Barger et al., 2010	MRI
G. Favara et al., 2022	RIPASA, Alvarado scores
Bethaney Miller et al., 2025	POCUS
Ibrahim Hadadi et al., 2017	US, CT
Siyuan Bai et al., 2023	Alvarado score
S. Eun et al., 2021	NLR, WBC, CRP, ANC

The included sources predominantly utilized surgical findings and histopathology as the reference standard , with clinical follow-up employed for negative cases in many studies . Several authors noted methodological limitations including verification bias, inadequate blinding, and retrospective designs .

Diagnostic Performance of Imaging Modalities

Computed Tomography

CT imaging demonstrated consistently high diagnostic accuracy across multiple protocols and populations.

Study	CT Protocol	Sensitivity (95% CI)	Specificity (95% CI)	Additional Metrics
Diana Isabel	Standard CT	96% (93-97%)	98% (96-98%)	DOR: 864.43

Study	CT Protocol	Sensitivity (95% CI)	Specificity (95% CI)	Additional Metrics
Castro-Luna et al., 2025				
Elio Stefan Arruzza et al., 2022	All CT	97.2% (95.8-98.1%)	95.6% (94.1-96.7%)	—
Elio Stefan Arruzza et al., 2022	IV + Oral contrast	99.2% (96.5-99.8%)	97.4% (93.6-99.0%)	—
Elio Stefan Arruzza et al., 2022	Low-dose CT	93.4% (88.5-96.3%)	93.7% (91.1-95.5%)	—
A. Doria et al., 2006	Mixed	Children: 94% (92-97%)	Children: 95% (94-97%)	Adults: 94% sensitivity
V. Hlibczuk et al., 2009	Noncontrast	92.7% (89.5-95.0%)	96.1% (94.2-97.5%)	LR+: 24, LR-: 0.08
H. Yoon et al., 2018	Reduced-dose	96% (93-98%)	94% (92-95%)	78% dose reduction
B. Xiong et al., 2015	Noncontrast	90% (86-92%)	94% (92-97%)	DOR: 162.76
S. J. Yun et al., 2017	Low-dose	96.25% (91.88-98.31%)	93.22% (88.75-96.00%)	—
Dong Wook Kim et al.,	Mixed pediatric	95% (93-97%)	94% (90-96%)	AUC: 0.98

Study	CT Protocol	Sensitivity (95% CI)	Specificity (95% CI)	Additional Metrics
2018				
Yuhao Chen et al., 2023	Contrast-enhanced	95% (93-96%)	94% (93-96%)	DOR: 305.31
Yuhao Chen et al., 2023	Non-contrast	85% (82-87%)	93% (92-95%)	DOR: 80.98
L. Neumayer et al., 2003	Mixed	77-100%	83-100%	Accuracy: 88-98%

CT with intravenous plus oral contrast demonstrated statistically significantly greater diagnostic accuracy than CT with intravenous contrast alone . Low-dose CT protocols achieved comparable diagnostic performance to standard-dose protocols while reducing radiation exposure by approximately 78% . The median effective radiation dose of reduced-dose CT was 1.8 mSv compared to standard protocols .

Contrast-enhanced CT showed significantly higher sensitivity compared to non-contrast CT (95% vs 85%, $p < 0.001$) , though both approaches maintain clinical utility. Noncontrast CT was deemed adequate for clinical decision-making in emergency department settings .

Ultrasound

Ultrasound performance varied considerably across studies, with notable differences between conventional ultrasound and point-of-care approaches.

Study	US Type	Sensitivity (95% CI)	Specificity (95% CI)
Diana Isabel Castro-	Conventional	93% (87-96%)	89% (80-95%)

Study	US Type	Sensitivity (95% CI)	Specificity (95% CI)
Luna et al., 2025			
Diana Isabel Castro-Luna et al., 2025	POCUS	80% (61-91%)	93% (83-98%)
Kevin A Eng et al., 2018	Second-line	91.3% (83.8-95.5%)	95.2% (91.8-97.3%)
Elio Stefan Arruzza et al., 2022	Mixed	82.1% (73.8-88.2%)	85.9% (72.7-93.3%)
A. Doria et al., 2006	Mixed	88% (86-90%)	94% (92-95%)
S. Lee et al., 2019	EP-POCUS	84% (72-92%)	91% (85-95%)
V. Giljaca et al., 2017	Abdominal	69% (59-78%)	81% (73-88%)
Seung-Hum Yu et al., 2005	Graded compression	86.7% (85.4-88.0%)	90.0% (88.9-91.2%)
Jian Fu et al., 2021	Abdominal	77.2% (75.4-78.9%)	60% (58-62%)
S. Cho et al., 2023	ED-based	81% (78-83%)	87% (85-88%)
Guixin Shen et al., 2019	Bedside	90% (82-95%)	95% (89-98%)
J. Matthew Fields et al., 2017	POCUS	91% (83-96%)	97% (91-99%)
Bethaney Miller et al., 2025	POCUS	85.6% (68.9-94.1%)	90.2% (86.5-93.0%)

Emergency physician-performed POCUS demonstrated excellent diagnostic performance, particularly in pediatric patients where sensitivity reached 95% (75-99%) and

specificity 95% (85-98%) . Direct comparison revealed no significant differences between EP-POCUS and radiologist-performed ultrasound ($p=0.18-0.85$) . Using a 7-mm appendix diameter cut-off value and having the emergency physician as the initial operator were associated with improved diagnostic accuracy .

Ultrasound studies excluding equivocal findings demonstrated significantly greater diagnostic values than those including all results ($p<0.0001$) , suggesting potential overestimation of accuracy when indeterminate cases are excluded. Operator experience significantly impacted ultrasound accuracy, with diagnostic radiologists showing better performance (sensitivity 84.9%, specificity 93.0%) compared to other professionals .

Magnetic Resonance Imaging

MRI demonstrated high diagnostic accuracy across all populations, including pregnant women and children where radiation avoidance is paramount.

Study	Population	Sensitivity (95% CI)	Specificity (95% CI)	Additional Metrics
Diana Isabel Castro-Luna et al., 2025	Pediatric	96% (94-97%)	98% (96-99%)	DOR: 1,030.42
Kevin A Eng et al., 2018	Children	97.4% (85.8-100%)	97.1% (92.1-99.0%)	—
Kevin A Eng et al., 2018	Adults	89.9% (84.8-93.5%)	93.6% (90.9-95.5%)	—
Hanfei Zhang et al., 2017	Pediatric	—	—	AUC: 0.995
Mania Kave et	Pregnant	91.8% (87.7-	97.9% (97.2-	DOR: 177.60

Study	Population	Sensitivity (95% CI)	Specificity (95% CI)	Additional Metrics
al., 2019		94.9%)	100%)	
Mohsen Motavaselian et al., 2022	Pregnant	92% (88-95%)	98% (97-98%)	AUC: 0.99
Nigel D'Souza et al., 2021	Mixed	95% (94-97%)	96% (95-97%)	—
Dongchan Kim et al., 2023	Adults	96% (93-97%)	93% (80-98%)	—
E. Duke et al., 2016	General	96% (95-97%)	96% (95-97%)	—
E. Duke et al., 2016	Pregnant	94% (87-98%)	97% (96-98%)	—
B. Saba et al., 2017	Pediatric	96% (94-98%)	97% (96-98%)	—
Michael D. Repplinger et al., 2016	Adults	96.6% (92.3-98.5%)	95.8% (89.4-98.4%)	LR+: 20, LR-: 0.03
Richard L. Barger et al., 2010	Adults	97% (92-99%)	95% (94-99%)	DOR: 299.85
M. Moore et al., 2016	Pediatric	96.5% (94.3-97.8%)	96.1% (93.5-97.7%)	NPV: 98.3%

MRI appears highly accurate for diagnosing acute appendicitis regardless of protocol . In pregnant patients, MRI demonstrated sensitivity of 91.8% and specificity of 97.9%, making it an excellent imaging option that avoids ionizing radiation . Half-acquisition single-shot fast spin-echo (SSFSE) pulse sequences are considered crucial for diagnostic protocols, while gadolinium enhancement provides uncertain additional benefit . Protocol duration ranged from 2.26 to 30 minutes across studies, with four-sequence protocols commonly utilized .

Head-to-Head Imaging Comparisons

Direct comparisons between imaging modalities revealed important differences in diagnostic performance.

Study	Methods Compared	Key Finding	Statistical Significance
A. van Randen et al., 2008	CT vs US	CT superior (LR+ 9.29 vs 4.50)	p=0.011
A. Doria et al., 2006	CT vs US	CT higher sensitivity in children and adults	Significant
S. Kaiser et al., 2002	US alone vs US+CT	Combined approach: 99% sensitivity vs 86%	Not reported
Marc Horton et al., 2000	CT vs US	CT: 97% sensitive, 100% specific; US: 76% sensitive, 90% specific	p=0.018
B. Saba et al.,	MRI vs CT	No significant	Not significant

Study	Methods Compared	Key Finding	Statistical Significance
2017		difference (p=0.11 sensitivity, p=0.06 specificity)	
Diana Isabel Castro-Luna et al., 2025	US vs CT vs MRI	No statistically significant differences (p=0.07)	Not significant
Ibrahim Hadadi et al., 2017	CT vs US	CT more sensitive (95% vs 75%); similar specificity	Not reported

CT consistently demonstrated superior sensitivity compared to ultrasound in head-to-head comparisons , with positive likelihood ratios approximately twice as high . However, all four major imaging modalities (conventional US, POCUS, CT, MRI) demonstrated high diagnostic accuracy overall . The choice of modality should consider radiation exposure concerns, particularly in children and pregnant women .

Diagnostic Performance of Clinical Scoring Systems

Clinical scoring systems provide objective assessment tools for appendicitis diagnosis, particularly valuable in resource-limited settings.

Score	Study	Sensitivity	Specificity	AUC	Population
AIR score	Roland E Andersson et al., 2025	91% (>4 points)	98% (>8 points)	0.86	Mixed
AIR score	Roland E	95% (>4	99% (>8	0.93	Mixed

Score	Study	Sensitivity	Specificity	AUC	Population
(advanced AA)	Andersson et al., 2025	points)	points)		
AIR score	Malsha Kularatna et al., 2017	92%	63%	0.84-0.97	Adults
Alvarado	Abdul Rafeh Awan et al., 2024	67% (65-69%)	74% (69-79%)	0.74	Mixed
Alvarado	M. Frountzas et al., 2018	69% (67-71%)	77% (74-80%)	0.79	Mixed
Alvarado	G. Favara et al., 2022	72% (66-77%)	77% (70-82%)	—	Adults
Alvarado	Siyuan Bai et al., 2023	76% (74-78%)	—	—	Pediatric
Tzanakis	Abdul Rafeh Awan et al., 2024	86% (84-87%)	73% (69-78%)	0.93	Mixed
RIPASA	M. Frountzas et al., 2018	94% (92-95%)	55% (51-55%)	0.94	Mixed
RIPASA	G. Favara et al., 2022	95% (92-97%)	71% (60-80%)	—	Adults
PAS	Roshanak Benabbas et al., 2017	—	—	LR+ 5.26 (≥9)	Pediatric

The AIR score demonstrated better diagnostic capacity than the Alvarado score, with higher ROC areas for both all patients and advanced appendicitis . The Tzanakis scoring system showed higher sensitivity (86% vs 67%) and AUC (0.93 vs 0.74) compared to the Alvarado score, though the Alvarado score had marginally better specificity .

RIPASA demonstrated the highest sensitivity (94-95%) among scoring systems but had notably low specificity (55-71%) , requiring supplementary diagnostic means due to its high false-positive rate . Clinical judgment demonstrated better specificity and sensitivity than the Alvarado score alone in one prospective trial (clinical judgment AUC: 0.933 vs Alvarado AUC: 0.749) .

Neither the Pediatric Appendicitis Score nor the Alvarado score met current performance benchmarks in validation studies . None of the history, physical examination, or laboratory findings alone could reliably rule in or rule out acute appendicitis in patients with suspected disease .

Diagnostic Performance of Laboratory Biomarkers

Laboratory tests serve as adjuncts to clinical assessment and imaging in appendicitis diagnosis.

Biomarker	Study	Sensitivity	Specificity	AUC	Cut-off
WBC ($\geq 10,000$)	D. Fawkner- Corbett et al., 2022	85% (80- 89%)	—	—	Pediatric
WBC	A. Acharya et al., 2016	79% (78- 81%)	55% (54- 57%)	0.75	Mixed
ANC ($\geq 7,500$)	D. Fawkner- Corbett et	90% (85- 94%)	—	—	Pediatric

Biomarker	Study	Sensitivity	Specificity	AUC	Cut-off
	al., 2022				
CRP (≥ 50)	D. Fawkner-Corbett et al., 2022	—	87% (80-91%)	—	Pediatric
CRP	A. Acharya et al., 2016	76% (75-78%)	50% (48-52%)	0.80	Mixed
NLR (>4.7)	S. Hajibandeh et al., 2019	88.89%	90.91%	0.96	Appendicitis
NLR (>8.8)	S. Hajibandeh et al., 2019	76.92%	100%	0.91	Complicated
NLR	S. Eun et al., 2021	82% (79-85%)	76% (69-81%)	0.86	Pediatric
Bilirubin	A. Acharya et al., 2016	51% (50-52%)	78% (76-80%)	0.72	Mixed
Procalcitonin	A. Acharya et al., 2016	36% (31-40%)	88% (83-91%)	0.82	Mixed
PLR	J. Arredondo Montero et al., 2023	Significant differences	—	—	Pediatric
RDW	S. Anand et al., 2022	Not significant	Not significant	—	Mixed

The neutrophil-to-lymphocyte ratio (NLR) emerged as a promising biomarker with cut-off values of 4.7 for appendicitis diagnosis (sensitivity 88.89%, specificity 90.91%) and 8.8 for complicated appendicitis (specificity 100%). NLR may be particularly useful for prioritizing surgical cases and monitoring conservatively treated patients, especially in pregnant or pediatric patients who may not routinely undergo CT imaging.

Combining WBC and CRP increased sensitivity to 97% (93-99%) for ruling out appendicitis. When both WBC count and CRP are normal, acute appendicitis is very unlikely. The best-performing single tests for ruling out pediatric appendicitis were WBC (sensitivity 85%) and ANC (sensitivity 90%).

Red cell distribution width (RDW) showed no statistically significant difference between appendicitis and non-appendicitis cases, providing no evidence of diagnostic utility. Similarly, platelet count and monocyte-to-lymphocyte ratio did not show significant differences between groups.

Diagnostic Accuracy for Complicated Appendicitis

Distinguishing complicated from uncomplicated appendicitis is crucial for treatment planning.

Modality	Study	Sensitivity	Specificity	Notes
CT	W. Bom et al., 2020	78% (64-88%)	91% (85-99%)	NPV: 93%
CT	Mohamad Badlis et al., 2024	62-68.4%	81-92.4%	Pediatric
US	Mohamad Badlis et al.,	33.9-51.5%	68.8-95%	Pediatric

Modality	Study	Sensitivity	Specificity	Notes
	2024			
CT features	Hae Young Kim et al., 2017	14-59%	74-100%	Individual features

Ten informative CT features for differentiating complicated appendicitis were identified, nine of which showed high specificity (74-100%) but limited sensitivity (14-59%) . Periappendiceal fat stranding was the only feature showing high sensitivity (94%) but low specificity (40%) . CT has far from perfect sensitivity for complicated appendicitis, though its negative predictive value is high (93%) .

NLR values >8.8 predicted complicated appendicitis with 76.92% sensitivity and 100% specificity , suggesting potential utility in predicting disease severity without imaging.

Special Populations

Pregnant Women

MRI represents the optimal imaging modality for pregnant patients with suspected appendicitis, demonstrating sensitivity of 91.8-94% and specificity of 97-99% . MRI does not expose the fetus or mother to ionizing radiation and should be used when ultrasonography results are inconclusive . CT maintains diagnostic utility with sensitivity of 85.7% and specificity of 97.4% when MRI is unavailable .

Pediatric Patients

In children, MRI demonstrated the highest area under the ROC curve (0.995) compared to US (0.987) and CT (0.982), though differences were not statistically significant ($p>0.05$) . US is recommended as the first-line imaging modality in pediatric patients due to

its favorable specificity, non-invasiveness, and lack of radiation exposure . CT should be reserved for ambiguous cases where US results are inconclusive .

The combination of US as primary examination followed by CT for non-diagnostic US yielded the highest sensitivity (100%), specificity (85%), and accuracy (92%) in pediatric populations . However, a US-only pathway was associated with significantly lower odds of negative appendectomy compared to a US-followed-by-CT pathway (OR 0.44; 95% CI 0.21-0.90) .

Atypical Presentations

In patients with atypical or equivocal presentations, CT demonstrated 100% specificity and 97% sensitivity, compared to 90% specificity and 76% sensitivity for US . CT was found superior in diagnostic accuracy and reliability for atypical patients . Both CT and US should be used as part of the initial assessment for patients presenting with equivocal findings .

Novel Diagnostic Approaches

Artificial Intelligence

Artificial neural networks (ANNs) were the most commonly employed algorithm for appendicitis diagnosis, with accuracy rates often exceeding 80% and AUC values peaking at 0.985 . AI models demonstrated promising results in predicting postoperative outcomes such as sepsis risk and ICU admission . Selection bias and lack of internal validation were identified as common issues affecting the quality of AI diagnostic studies .

Observation Protocols

Active observation with repeated clinical and laboratory evaluation significantly reduced CT scan utilization from 100% to 36.3% without compromising diagnostic accuracy

in medium-to-low-risk patients . Both conditional CT and observation pathways resulted in high sensitivity and specificity (97.7% and 94.6% vs 96.7% and 95.8%) .

Clinical Outcomes Associated with Diagnostic Strategies

Study	Comparison	Negative Appendectomy Rate	Time to Surgery	Perforation Rate
S. Krajewski et al., 2011	CT vs clinical evaluation	8.7% vs 16.7% (p<0.001)	Increased with CT	23.4% vs 16.7% (NS)
S. Kaiser et al., 2002	US+CT vs US alone	3.7% overall	—	21%
C. Douglas et al., 2000	US protocol vs clinical	9% vs 11% (NS)	7.0 vs 10.2 hours (p=0.016)	3% vs 1% (NS)
John J. Hong et al., 2003	CT vs clinical assessment	Similar	10.6 vs 19.0 hours (p<0.01)	6% vs 9% (NS)

CT imaging was associated with significantly lower negative appendectomy rates compared to clinical evaluation alone (8.7% vs 16.7%, p<0.001) . However, CT use may delay surgery in the absence of expedited imaging protocols . This delay does not appear to increase appendiceal perforation rates (23.4% vs 16.7%, p=0.15) .

Ultrasound protocols with Alvarado scoring led to significantly shorter time to operation (7.0 vs 10.2 hours, p=0.016) but did not significantly reduce hospital stay or prevent adverse outcomes .

Synthesis

Explaining Heterogeneity in Diagnostic Performance

Substantial heterogeneity in reported diagnostic accuracy exists across studies, which can be explained through several mechanisms:

Population and Clinical Context Distinctions:

Diagnostic performance varies systematically by population characteristics. In children, ultrasound sensitivity ranges from 55.6% to 97.56% , while CT maintains consistently high performance (88.1-97.8% sensitivity) . This variation reflects both operator dependency of ultrasound and the anatomical challenges in pediatric imaging . Emergency physician-performed POCUS achieved higher sensitivity in pediatric cases (95%) compared to overall performance (84%) , suggesting that targeted training in pediatric ultrasound improves outcomes.

In pregnant women, MRI demonstrated consistently excellent performance (sensitivity 91.8-94%, specificity 97-99%) across studies, whereas CT data are more limited due to radiation concerns. The consistent MRI performance in pregnancy reflects standardized protocols and experienced interpretation in this carefully studied population.

For atypical presentations, CT superiority over ultrasound is more pronounced (sensitivity 97% vs 76%) compared to mixed populations, indicating that complex clinical scenarios benefit more from cross-sectional imaging's comprehensive visualization.

Technical and Methodological Factors:

Ultrasound heterogeneity is substantially explained by operator experience. Studies performed by diagnostic radiologists demonstrated higher sensitivity (84.9%) compared to studies with mixed operators (sensitivity as low as 55%) . The wide sensitivity range for

ultrasound (55-97%) largely reflects training and experience differences rather than inherent modality limitations.

CT protocol variations account for significant performance differences. Contrast-enhanced CT showed significantly higher sensitivity than non-contrast CT (95% vs 85%, $p<0.001$), while CT with combined intravenous and oral contrast achieved the highest accuracy (sensitivity 99.2%, specificity 97.4%). Low-dose CT maintained comparable diagnostic performance to standard-dose protocols (sensitivity 96% vs 96%, $p=0.71$), indicating that dose reduction does not compromise accuracy.

Reference Standard Variability:

Studies using combined surgical and histopathological reference standards with adequate clinical follow-up for negative cases demonstrated more robust estimates. Verification bias occurred in studies where positive test results preferentially led to surgical confirmation. Studies excluding equivocal imaging results showed inflated sensitivity and specificity estimates.

Integrated Diagnostic Framework

Based on the synthesized evidence, diagnostic approach selection should be tailored to clinical context:

General Adult Population: CT represents the highest-performing individual imaging modality with pooled sensitivity of 94-97% and specificity of 94-98%. Low-dose CT protocols are appropriate first-line options, achieving 78% radiation reduction without compromising accuracy. Contrast-enhanced protocols provide incremental benefit in diagnostic certainty.

Pediatric Population: Ultrasound should be the initial imaging modality due to radiation concerns and high specificity (90-95%). When ultrasound is non-diagnostic,

selective CT or MRI provides excellent accuracy . A US-only pathway may reduce negative appendectomy rates compared to routine CT follow-up .

Pregnant Women: MRI is the preferred imaging modality with sensitivity of 92% and specificity of 98% . Ultrasound remains appropriate for initial evaluation, with MRI recommended when ultrasound is inconclusive .

Resource-Limited Settings: Clinical scoring systems, particularly the AIR score (AUC 0.86-0.93) , combined with laboratory markers (NLR cut-off 4.7, sensitivity 89%, specificity 91%) provide reasonable diagnostic accuracy when imaging is unavailable. The combination of WBC and CRP has high sensitivity (97%) for ruling out appendicitis .

Atypical or Equivocal Presentations: CT imaging is recommended for patients with atypical presentations, where it demonstrates superior accuracy compared to ultrasound or clinical assessment alone . Active observation protocols with serial clinical evaluation can safely reduce CT utilization in medium-to-low-risk patients .

Performance Hierarchy by Context

The evidence supports context-specific diagnostic strategies rather than universal recommendations:

Context	Recommended Primary Approach	Expected Sensitivity	Expected Specificity
Adults, typical presentation	CT (low-dose acceptable)	94-97%	93-96%
Adults, atypical presentation	CT with contrast	95-99%	94-100%
Children, initial evaluation	Ultrasound	88-93%	89-95%

Context	Recommended Primary Approach	Expected Sensitivity	Expected Specificity
Children, US non-diagnostic	CT or MRI	94-97%	94-97%
Pregnant women	MRI	92-96%	97-98%
Resource-limited settings	AIR score + NLR	89-91%	90-98%

The apparent contradictions in literature regarding ultrasound performance (sensitivity ranging from 55% to 97%) are reconciled by recognizing the critical role of operator expertise , patient selection , and handling of equivocal results . When performed by experienced sonographers with appropriate protocols, ultrasound approaches CT performance in selected populations .

DISCUSSION

This comprehensive systematic review of 80 studies provides a detailed landscape of the diagnostic accuracy for acute appendicitis. The central finding is that diagnostic performance is not an intrinsic property of a test but is profoundly influenced by a complex interplay of patient characteristics, disease factors, technical execution, and study methodology. The discussion will interpret the results, reconcile apparent contradictions, and propose a pragmatic clinical framework.

Hierarchy of Imaging Modalities and the Central Role of Context

CT consistently emerged as the imaging modality with the highest and most robust diagnostic accuracy for the general adult population, with pooled sensitivity of 94–97% and specificity of 94–98% (Elio Stefan Arruzza et al., 2022; Yuhao Chen et al., 2023). Its superiority over ultrasound in head-to-head comparisons, particularly in adults and those with atypical presentations, is well-documented (A. van Randen et al., 2008; Marc Horton et al., 2000). However,

declaring CT the universal winner is an oversimplification. The significant radiation dose, especially concerning in children and young adults, mandates a risk-benefit analysis. The advent of low-dose and reduced-dose CT protocols, which reduce radiation exposure by approximately 78% without compromising diagnostic accuracy, is a pivotal advancement that mitigates this concern and supports its broader, more responsible use (H. Yoon et al., 2018; S. J. Yun et al., 2017; S. Sippola et al., 2020).

The performance of ultrasound exhibited the greatest variability across studies, with sensitivity ranging from 55% to 97% (Giljaca et al., 2017; S. Cho & Oh, 2023). This wide range is not random error but is largely explained by two key moderators: **operator expertise** and **handling of equivocal results**. Studies conducted by dedicated radiologists or trained emergency physicians using point-of-care ultrasound (POCUS) reported sensitivities above 90% and specificities above 95% (J. Matthew Fields et al., 2017; S. Lee & S. J. Yun, 2019; Bethaney Miller et al., 2025). Conversely, studies with less experienced operators or in mixed settings reported significantly lower figures. Furthermore, studies that excluded indeterminate scans from accuracy calculations artificially inflated performance metrics ($p < 0.0001$). When performed by experts and with indeterminate results considered negative or triggering further imaging, ultrasound's accuracy approaches that of CT in select populations, particularly children. This makes it the unequivocal recommended first-line imaging modality in pediatrics and a valuable first step in pregnant women, primarily due to the absence of ionizing radiation (Diana Isabel Castro-Luna et al., 2025; Hanfei Zhang et al., 2017; D. Diercks et al., 2023).

MRI has cemented its role as the imaging modality of choice for pregnant women with suspected appendicitis, demonstrating excellent sensitivity (91.8–94%) and specificity (97–99%) without fetal radiation exposure (Mania Kave et al., 2019; Mohsen Motavaselian et al., 2022; E. Duke et al., 2016). Its high accuracy in the general adult and pediatric populations is also established, making it a powerful alternative to CT when radiation must be avoided, though accessibility and cost remain barriers (Michael D. Repplinger et al., 2016; B. Saba, 2017).

Clinical Scores and Biomarkers: Adjuncts with Specific Niches

Clinical scoring systems provide an objective, low-cost starting point. This review confirms that the AIR score has superior diagnostic capacity compared to the classic Alvarado score (Roland E Andersson & J. Stark, 2025; Malsha Kularatna et al., 2017). The RIPASA score, while highly sensitive, suffers from low specificity, leading to a high false-positive rate that limits its standalone use (M. Frountzas et al., 2018; G. Favara et al., 2022). No clinical score alone is sufficient to rule in or rule out appendicitis reliably, but they are invaluable for risk stratification, helping to identify low-risk patients who may be suitable for observation and high-risk patients who require expedited imaging.

Among laboratory biomarkers, the neutrophil-to-lymphocyte ratio (NLR) stands out. With a cut-off >4.7 demonstrating good sensitivity (88.89%) and specificity (90.91%), and a cut-off >8.8 showing high specificity (100%) for complicated appendicitis, NLR is a promising tool (S. Hajibandeh et al., 2019). It may be particularly useful for triage and monitoring, especially in populations where immediate CT is not desired (e.g., pediatrics, early pregnancy). The combination of WBC and CRP, when both are normal, has a high sensitivity for ruling out appendicitis (97%), providing a simple and effective "rule-out" strategy in the emergency department (D. Fawkner-Corbett et al., 2022; Elina Lietzén et al., 2016).

Synthesizing an Integrated Diagnostic Framework

Based on the synthesized evidence, a rigid, universal protocol is inferior to a flexible, context-aware algorithm. The following framework is proposed:

1. **Initial Assessment:** For all patients, begin with history, physical exam, and basic labs (WBC, CRP). Calculate a clinical score (AIR score preferred).
2. **Adult, Typical Presentation (Medium/High Risk):** Proceed directly to CT imaging. A low-dose protocol is recommended as the first-line CT technique to minimize radiation.

3. **Adult, Atypical/Equivocal Presentation:** CT (with contrast for optimal accuracy) is the imaging modality of choice due to its superior reliability in complex cases (A. van Randen et al., 2008).
4. **Children:** Ultrasound, performed by a skilled operator, is the mandatory first-line imaging. If diagnostic (positive or clearly negative), proceed accordingly. If non-diagnostic, use a selective, step-up approach with either MRI (preferred) or CT, based on availability and institutional expertise (Kevin A Eng et al., 2018; Diana Isabel Castro-Luna et al., 2025).
5. **Pregnant Women:** Begin with ultrasound. If inconclusive or negative with high clinical suspicion, MRI is the definitive next test and should be readily available (Mania Kave et al., 2019; Nigel D'Souza et al., 2021).
6. **Resource-Limited Settings:** Rely on a combination of the AIR score and biomarkers (NLR, WBC/CRP). A patient with a high AIR score and elevated NLR has a high probability of appendicitis. A patient with a low AIR score and normal WBC/CRP can often be safely observed or discharged with clear instructions.
7. **The Role of Active Observation:** For patients classified as medium-to-low risk after initial assessment, a period of active observation with serial clinical re-evaluation is a safe and effective strategy that significantly reduces unnecessary CT scans without increasing adverse outcomes (Raminta Lukšaitė-Lukštė et al., 2024; P. Lopez et al., 2007).

Impact on Clinical Outcomes and Future Directions

The choice of diagnostic strategy has tangible consequences. CT use is associated with a significantly lower negative appendectomy rate compared to clinical assessment alone (8.7% vs. 16.7%) (S. Krajewski et al., 2011). However, it can introduce delays if not rapidly available. Streamlined imaging protocols are essential. Ultrasound-first pathways can shorten time to decision-making without increasing adverse events (C. Douglas et al., 2000).

Future research must address several key areas: First, standardized training and certification for POCUS to maximize its potential. Second, large-scale implementation studies of the proposed integrated frameworks to validate their real-world effectiveness and cost-effectiveness. Third, robust external validation of promising AI algorithms to assess their role as decision-support tools (Mahbod Issaiy et al., 2023). Finally, continued refinement of low-dose CT and rapid MRI protocols to make these accurate, lower-risk modalities more accessible.

The diagnosis of acute appendicitis has evolved from reliance on clinical acumen alone to a sophisticated, multi-tool enterprise. The evidence compels a shift from asking "Which test is best?" to "Which test is best *for this specific patient in this specific setting?*" By adopting the context-sensitive framework outlined, clinicians can achieve the ultimate goal: accurate, timely, and safe diagnosis for every patient.

CONCLUSION AND RECOMMENDATIONS

Conclusion

This systematic review demonstrates that the diagnostic landscape for acute appendicitis is rich with options, each with distinct strengths and limitations. Computed Tomography (CT) maintains the highest overall diagnostic accuracy and is the reference standard for non-pregnant adults, especially with atypical presentations, with low-dose protocols now offering a safer profile. Ultrasound is the cornerstone of pediatric diagnosis and first-line imaging in pregnancy, but its accuracy is inextricably linked to operator skill. Magnetic Resonance Imaging (MRI) is the modality of choice for pregnant women and an excellent radiation-free alternative for all patients. Clinical scoring systems, particularly the AIR score, and biomarkers like the Neutrophil-to-Lymphocyte Ratio (NLR) are valuable for initial risk stratification and in resource-constrained environments. The substantial heterogeneity in reported accuracies, especially for ultrasound, is not a flaw in the modality itself but a reflection of variable study populations, technical execution, and

methodological quality. Therefore, an effective diagnostic strategy cannot be monolithic but must be dynamically tailored to the clinical context.

Recommendations

1. For Clinical Practice:

- **Adopt a Stepped, Context-Specific Approach:** Implement the integrated diagnostic framework proposed in the discussion, selecting initial imaging based on patient age, gender, clinical presentation, and pregnancy status.
- **Promote Ultrasound Expertise:** Invest in standardized training and credentialing for emergency physicians and radiologists in appendiceal ultrasound to harness its full potential and reduce reliance on CT, particularly in children.
- **Utilize Low-Dose CT:** Advocate for and adopt low-dose or reduced-dose CT protocols as the default for most adults to minimize population radiation exposure without sacrificing diagnostic certainty.
- **Incorporate Biomarkers:** Use the combination of WBC and CRP to help rule out appendicitis in low-risk patients, and consider NLR as an adjunct for assessing likelihood and predicting complexity.

2. For Hospital and Health System Management:

- **Ensure MRI Access:** Develop and streamline 24/7 access to MRI for pregnant women with suspected appendicitis to avoid diagnostic delays and unnecessary CT radiation.
- **Develop Clinical Pathways:** Create and implement institutional clinical pathways for right lower quadrant pain that incorporate observation protocols for low-to-medium risk patients, clear imaging algorithms, and surgical referral criteria.

- **Audit Outcomes:** Regularly monitor key performance indicators such as negative appendectomy rate, perforation rate, and time-to-surgery to evaluate and refine local diagnostic practices.

3. For Future Research:

- **Conduct Pragmatic Trials:** Perform large, multi-center trials comparing the clinical outcomes and cost-effectiveness of different integrated diagnostic pathways (e.g., US-first vs. selective CT vs. MRI-second) in well-defined patient subgroups.
- **Validate AI Tools:** Rigorously evaluate artificial intelligence models for diagnosing appendicitis from imaging and clinical data in diverse, real-world settings to determine their clinical utility and integration potential.
- **Standardize Reporting:** Future diagnostic accuracy studies should adhere to STARD guidelines, clearly report operator experience, and transparently handle indeterminate results to allow for more precise meta-analyses.

REFERENCES

- A. Acharya, S. Markar, M. Ni, and G. Hanna. "Biomarkers of Acute Appendicitis: Systematic Review and Cost–Benefit Trade-Off Analysis." *Surgical Endoscopy*, 2016.
- A. Başaran, and M. Basaran. "Diagnosis of Acute Appendicitis During Pregnancy: A Systematic Review." *Obstetrical and Gynecological Survey*, 2009.
- A. Doria, R. Moineddin, C. Kellenberger, M. Epelman, J. Beyene, S. Schuh, P. Babyn, and P. Dick. "US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis." *Radiology*, 2006.

A. van Randen, S. Bipat, A. Zwinderman, D. Ubbink, J. Stoker, and M. Boermeester. “Acute Appendicitis: Meta-Analysis of Diagnostic Performance of CT and Graded Compression US Related to Prevalence of Disease.” *Radiology*, 2008.

A. Weston, T. Jackson, and S. Blamey. “Diagnosis of Appendicitis in Adults by Ultrasonography or Computed Tomography: A Systematic Review and Meta-Analysis.” *International Journal of Technology Assessment in Health Care*, 2005.

Abdul Rafeh Awan, Zaka Ullah Khan, Hira Saleem, Hanniya Iqbal, Waqas Ahmad, Ali Raza Khan, and Mobeen Farooqi. “A Comparison of the Accuracy of Tzanakis and Alvarado Score in the Diagnosis of Acute Appendicitis: A Systematic Review and Meta-Analysis.” *The Surgeon*, 2024.

B. A. Anderson, Leon Salem, and D. Flum. “A Systematic Review of Whether Oral Contrast Is Necessary for the Computed Tomography Diagnosis of Appendicitis in Adults.” *American Journal of Surgery*, 2005.

B. Saba. “Systematic Review of the Accuracy of Magnetic Resonance Imaging in the Diagnosis of Acute Appendicitis in Children : Comparison with Computed Tomography,” 2017.

B. Xiong, Baishu Zhong, Zhenwei Li, F. Zhou, Ruying Hu, Zhan Feng, Shunliang Xu, and Feng Chen. “Diagnostic Accuracy of Noncontrast CT in Detecting Acute Appendicitis: A Meta-Analysis of Prospective Studies.” *The American Surgeon*, 2015.

Bethaney Miller, David J McCreary, and Jon Rees. “How Accurate Is Point-of-Care Ultrasound for Detecting Paediatric Appendicitis? A Systematic Review and Meta-Analysis.” *Archives of Disease in Childhood*, 2025.

C. Douglas, Neil E Macpherson, P. Davidson, and J. S. Gani. “Randomised Controlled Trial of Ultrasonography in Diagnosis of Acute Appendicitis, Incorporating the Alvarado Score.” *British Medical Journal*, 2000.

C. Franke, H. Böhner, Qingmo Yang, C. Ohmann, H. Röher, and Acute Liver Failure Study Group. “Ultrasonography for Diagnosis of Acute Appendicitis: Results of a Prospective Multicenter Trial.” *World Journal of Surgery*, 1999.

C. Ohmann, C. Franke, and Qin Yang. “Clinical Benefit of a Diagnostic Score for Appendicitis: Results of a Prospective Interventional Study. German Study Group of Acute Abdominal Pain.” *Archives of Surgery*, 1999.

Chnoor Shirwan, Erica Hawrami, Kirubel Tesfaye Ricci, Korlos Hailu, Salib, Sanath Savithri, Alousious Nandeesha, Pousette Kasagga, and Hamid. “Accuracy of Ultrasound and CT Scans in the Diagnosis of Acute Appendicitis in Children: A Systematic Review.” *Journal For International Medical Graduates*, 2024.

D. Diercks, E. Adkins, N. Harrison, P. Sokolove, H. Kwok, S. Wolf, John D. Anderson, et al. “Clinical Policy: Critical Issues in the Evaluation and Management of Emergency Department Patients With Suspected Appendicitis: Approved by ACEP Board of Directors February 1, 2023.” *Annals of Emergency Medicine*, 2023.

D. Fawkner-Corbett, G. Hayward, Mohammed Alkhmees, A. Van den Bruel, JM Ordóñez-Mena, and G. Holtman. “Diagnostic Accuracy of Blood Tests of Inflammation in Paediatric Appendicitis: A Systematic Review and Meta-Analysis.” *BMJ Open*, 2022.

D. Kulik, E. Uleryk, and J. Maguire. “Does This Child Have Appendicitis? A Systematic Review of Clinical Prediction Rules for Children with Acute Abdominal Pain.” *Journal of Clinical Epidemiology*, 2013.

D. Schriger, and T. Reynolds. “Annals of Emergency Medicine Journal Club. Journal Club: The Conduct and Reporting of Meta-Analyses of Studies of Diagnostic Tests, and a Consideration of ROC Curves.” *Annals of Emergency Medicine*, 2010.

Diana Isabel Castro-Luna, Juan D Porras-Hernandez, J. A. Flores-García, Pilar Dies-Suarez, M. F. Servin-Martinez, and Mauricio Pierdant-Pérez. “Contemporary Ultrasound, Computed Tomography, or Magnetic Resonance Imaging for Acute Appendicitis Diagnosis in Children and Adolescents: Systematic Review and Meta-Analysis.” *Pediatric Radiology*, 2025.

Dong Wook Kim, H. Yoon, Jeong-Yong Lee, Jung Heon Kim, A. Jung, Jin Seong Lee, and Y. Cho. “Diagnostic Performance of CT for Pediatric Patients with Suspected Appendicitis in Various Clinical Settings: A Systematic Review and Meta-Analysis.” *Emergency Radiology*, 2018.

Dongchan Kim, B. Woodham, Kathryn T. Chen, Vinushan Kuganathan, and M. Edye. “Rapid MRI Abdomen for Assessment of Clinically Suspected Acute Appendicitis in the General Adult Population: A Systematic Review.” *Journal of Gastrointestinal Surgery*, 2023.

E. Duke, B. Kalb, Hina Arif-Tiwari, Z. J. Daye, Dorothy Gilbertson-Dahdal, S. Keim, and Diego R. Martín. “A Systematic Review and Meta-Analysis of Diagnostic Performance of MRI for Evaluation of Acute Appendicitis.” *AJR. American Journal of Roentgenology*, 2016.

Edwin Kinesya, Elen Putri Cintya, Maria Juliana Dorothy, Nur Nadhirah Ennaidi, Hafidh Fahreza Rusti, Yusuf Mannagalli, and E. Pasaribu. “Diagnostic Accuracy of Alvarado Score Components in Patients with Appendicitis: Systematic Review and Meta-Analysis Approach.” *Health Sciences Review*, 2022.

Elina Lietzén, I. Ilves, P. Salminen, H. Paajanen, T. Rautio, P. Nordström, M. Aarnio, et al. “Clinical and Laboratory Findings in the Diagnosis of Right Lower Quadrant Abdominal Pain: Outcome Analysis of the APPAC Trial.” *Clinical Chemistry and Laboratory Medicine*, 2016.

Elio Stefan Arruzza, S. Milanese, L. Li, and J. Dizon. “Diagnostic Accuracy of Computed Tomography and Ultrasound for the Diagnosis of Acute Appendicitis: A Systematic Review and Meta-Analysis.” *Radiography*, 2022.

Eszter Mán, Z. Simonka, Ákos Varga, F. Rárosi, and G. Lázár. “Impact of the Alvarado Score on the Diagnosis of Acute Appendicitis: Comparing Clinical Judgment, Alvarado Score, and a New Modified Score in Suspected Appendicitis: A Prospective, Randomized Clinical Trial.” *Surgical Endoscopy*, 2014.

G. Favara, A. Maugeri, M. Barchitta, Andrea Ventura, G. Basile, and A. Agodi. “Comparison of RIPASA and ALVARADO Scores for Risk Assessment of Acute Appendicitis: A Systematic Review and Meta-Analysis.” *PLoS ONE*, 2022.

Guixin Shen, Jing Wang, Faming Fei, Minghang Mao, and Z. Mei. “Bedside Ultrasonography for Acute Appendicitis: An Updated Diagnostic Meta-Analysis.” *International Journal of Surgery*, 2019.

H. Yoon, C. Suh, Y. Cho, Jeong Rye Kim, Jin Seong Lee, A. Jung, Jung Heon Kim, Jeong-Yong Lee, and S. Kim. “The Diagnostic Performance of Reduced-Dose CT for Suspected Appendicitis in Paediatric and Adult Patients: A Systematic Review and Diagnostic Meta-Analysis.” *European Radiology*, 2018.

Hae Young Kim, Ji Hoon Park, Yoon Jin Lee, S. Lee, Jong-June Jeon, and K. Lee. “Systematic Review and Meta-Analysis of CT Features for Differentiating Complicated and Uncomplicated Appendicitis.” *Radiology*, 2017.

Hae Young Kim, Y. Ko, Ji Hoon Park, and K. Lee. “Detection and False-Referral Rates of 2-mSv CT Relative to Standard-Dose CT for Appendiceal Perforation: Pragmatic Multicenter Randomized Controlled Trial.” *AJR. American Journal of Roentgenology*, 2020.

Hanfei Zhang, M. Liao, Jie Chen, Dongyong Zhu, and Sama Byanju. “Ultrasound, Computed Tomography or Magnetic Resonance Imaging - Which Is Preferred for Acute Appendicitis in Children? A Meta-Analysis.” *Pediatric Radiology*, 2017.

Ibrahim Hadadi. “Is Ultrasonography or Computed Tomography More Accurate in Diagnosing Adults with Suspected Appendicitis? A Systematic Review,” 2017.

J. Arredondo Montero, Blanca Paola Pérez Riveros, and N. Martín-Calvo. “Diagnostic Performance of Total Platelet Count, Platelet-to-Lymphocyte Ratio, and Monocyte-to-Lymphocyte Ratio for Overall and Complicated Pediatric Acute Appendicitis: A Systematic Review and Meta-Analysis.” *Surgical Infections*, 2023.

J. Matthew Fields, Joshua J Davis, C. Alsup, A. Bates, A. Au, S. Adhikari, and I. Farrell. “Accuracy of Point-of-care Ultrasonography for Diagnosing Acute Appendicitis: A Systematic Review and Meta-analysis.” *Academic Emergency Medicine*, 2017.

Jian Fu, Xumeng Zhou, Liang Chen, and Sheng Lu. “Abdominal Ultrasound and Its Diagnostic Accuracy in Diagnosing Acute Appendicitis: A Meta-Analysis.” *Frontiers in Surgery*, 2021.

John J. Hong, Stephen M. Cohn, A. P. Ekeh, M. Newman, M. Salama, and Suzanne D. LeBlang. “A Prospective Randomized Study of Clinical Assessment Versus Computed Tomography for the Diagnosis of Acute Appendicitis.” *Surgical Infections*, 2003.

Jungheum Cho, Hae Young Kim, Seungjae Lee, Ji Hoon Park, and Kyoung Ho Lee. “Radiology Residents’ Independent Diagnosis of Appendicitis Using 2-mSv Computed Tomography: A Secondary Analysis of a Large Pragmatic Randomized Trial.” *Korean Journal of Radiology*, 2023.

K. Al-Khayal, and M. Al-Omran. “Computed Tomography and Ultrasonography in the Diagnosis of Equivocal Acute Appendicitis. A Meta-Analysis.” *Saudi Medical Journal*, 2007.

Kevin A Eng, Aryan Abadeh, C. Ligocki, Yvonne K Lee, R. Moineddin, T. Adams-Webber, S. Schuh, and A. Doria. “Acute Appendicitis: A Meta-Analysis of the Diagnostic Accuracy of US, CT, and MRI as Second-Line Imaging Tests After an Initial US.” *Radiology*, 2018.

L. Neumayer, and A. Kennedy. “Imaging in Appendicitis: A Review With Special Emphasis on the Treatment of Women.” *Obstetrics and Gynecology*, 2003.

M. Favot, and R. Ehrman. “What Elements of the History, Examination, Laboratory Testing, or Point-of-Care Ultrasonography Are Most Useful in the Diagnosis of Acute Appendicitis in Children?” *Annals of Emergency Medicine*, 2017.

M. Frountzas, K. Stergios, Dimitra Kopsini, Dimitrios Schizas, K. Kontzoglou, and K. Toutouzas. “Alvarado or RIPASA Score for Diagnosis of Acute Appendicitis? A Meta-Analysis of Randomized Trials.” *International Journal of Surgery*, 2018.

M. Moore, A. Kulaylat, C. Hollenbeak, B. Engbrecht, J. Dillman, and S. Methratta. “Magnetic Resonance Imaging in Pediatric Appendicitis: A Systematic Review.” *Pediatric Radiology*, 2016.

Mahbod Issaiy, Diana Zarei, and A. Saghaezadeh. “Artificial Intelligence and Acute Appendicitis: A Systematic Review of Diagnostic and Prognostic Models.” *World Journal of Emergency Surgery*, 2023.

Malsha Kularatna, M. Lauti, C. Haran, Wiremu S. MacFater, L. Sheikh, Ying Huang, J. McCall, and A. MacCormick. “Clinical Prediction Rules for Appendicitis in Adults: Which Is Best?” *World Journal of Surgery*, 2017.

Mania Kave, F. Parooie, and M. Salarzaei. “Pregnancy and Appendicitis: A Systematic Review and Meta-Analysis on the Clinical Use of MRI in Diagnosis of Appendicitis in Pregnant Women.” *World Journal of Emergency Surgery*, 2019.

Marc Horton, Steven F Counter, M. Florence, and Michael J. Hart. “A Prospective Trial of Computed Tomography and Ultrasonography for Diagnosing Appendicitis in the Atypical Patient.” *American Journal of Surgery*, 2000.

Michael D. Repplinger, Joseph F. Levy, E. Peethumnongsin, Megan E. Gussick, J. Svenson, S. Golden, W. Ehlenbach, R. Westergaard, S. Reeder, and D. Vanness. “Systematic Review and Meta-analysis of the Accuracy of MRI to Diagnose Appendicitis in the General Population.” *Journal of Magnetic Resonance Imaging*, 2016.

Mohamad Badlis, Kamil Amari, Maya Alkheshi, Karim Alolaby, and B. Alsaid. “Ultrasound and Computed Tomography in Differentiating Between Simple and Complicated Appendicitis in Pediatric Patients.” *Pediatric Surgery International (Print)*, 2024.

Mohsen Motavaselian, Fateme Bayati, Reza Amani-beni, Amirreza Khalaji, Sara Haghverdi, Zeynab Abdollahi, Arash Sarrafzadeh, et al. “Diagnostic Performance of Magnetic Resonance Imaging for Detection of Acute Appendicitis in Pregnant Women; a Systematic Review and Meta-Analysis.” *Archives of Academic Emergency Medicine*, 2022.

N. Shah, M. Islam, I. Sabir, Tahira Mehreen, and Mumtaz Khan. “DIAGNOSTIC ACCURACY OF CLINICAL EXAMINATION VERSUS COMBINATION OF ABDOMINAL ULTRASOUND AND ALVARADO SCORE, IN PATIENTS WITH ACUTE APPENDICITIS,” 2011.

Nigel D’Souza, G. Hicks, R. Beable, A. Higginson, and B. Rud. “Magnetic Resonance Imaging (MRI) for Diagnosis of Acute Appendicitis.” *Cochrane Database of Systematic Reviews*, 2021.

P. Lopez, S. Cohn, Charlie Popkin, J. Jackowski, and J. Michalek. “The Use of a Computed Tomography Scan to Rule Out Appendicitis in Women of Childbearing Age Is as Accurate as Clinical Examination: A Prospective Randomized Trial.” *The American Surgeon*, 2007.

Raminta Lukšaitė-Lukštė, Igne Gecaite, Kristina Marcinkevičiūtė, Eimantas Dumskis, A. Samuilis, Tadas Žvirblis, E. Jasiunas, et al. “Observation Safely Reduces the Use of the Computerized Tomography in Medium-to-Low-Risk Patients with Suspected Acute Appendicitis: Results of a Randomized Controlled Trial.” *Journal of Clinical Medicine*, 2024.

Richard K. Orr, Deborah Porter, and Daniel Hartman. “Ultrasonography to Evaluate Adults for Appendicitis: Decision Making Based on Meta-Analysis and Probabilistic Reasoning.” *Academic Emergency Medicine*, 1995.

Richard L. Barger, and K. Nandalur. “Diagnostic Performance of Magnetic Resonance Imaging in the Detection of Appendicitis in Adults: A Meta-Analysis.” *Academic Radiology*, 2010.

Roland E Andersson, and J. Stark. “Diagnostic Value of the Appendicitis Inflammatory Response (AIR) Score. A Systematic Review and Meta-Analysis.” *World Journal of Emergency Surgery*, 2025.

Roshanak Benabbas, M. Hanna, Jay Shah, and R. Sinert. “Diagnostic Accuracy of History, Physical Examination, Laboratory Tests, and Point-of-care Ultrasound for Pediatric Acute Appendicitis in the Emergency Department: A Systematic Review and Meta-analysis.” *Academic Emergency Medicine*, 2017.

S. Anand, Nellai Krishnan, M. Jukić, Zvonimir Križanac, Carlos Martin Llorente Muñoz, and Z. Pogorelić. “Utility of Red Cell Distribution Width (RDW) as a Noninvasive Biomarker for the Diagnosis of Acute Appendicitis: A Systematic Review and Meta-Analysis of 5222 Cases.” *Diagnostics*, 2022.

S. Anderson, J. Soto, B. Lucey, A. Ozonoff, J. Jordan, Jirair Ratevosian, A. Ulrich, et al. “Abdominal 64-MDCT for Suspected Appendicitis: The Use of Oral and IV Contrast Material Versus IV Contrast Material Only.” *AJR. American Journal of Roentgenology*, 2009.

S. Cho, and S. Oh. “Accuracy of Ultrasound for the Diagnosis of Acute Appendicitis in the Emergency Department: A Systematic Review.” *Medicine*, 2023.

S. Eun, HO I.G., G. E. Bae, H. Kim, C. M. Koo, M. K. Kim, S. Yoon, Seo Hee, and MD Ph.D Yoon. “Neutrophil-to-Lymphocyte Ratio for the Diagnosis of Pediatric Acute Appendicitis: A Systematic Review and Meta-Analysis.” *European Review for Medical and Pharmacological Sciences*, 2021.

S. Hajibandeh, S. Hajibandeh, N. Hobbs, and Moustafa A. Mansour. “Neutrophil-to-Lymphocyte Ratio Predicts Acute Appendicitis and Distinguishes Between Complicated and Uncomplicated Appendicitis: A Systematic Review and Meta-Analysis.” *American Journal of Surgery*, 2019.

- S. J. Yun, C. Ryu, Na Young Choi, Hyun Cheol Kim, Jiyoung Oh, and D. Yang. "Comparison of Low- and Standard-Dose CT for the Diagnosis of Acute Appendicitis: A Meta-Analysis." *AJR. American Journal of Roentgenology*, 2017.
- S. Kaiser, B. Frenckner, and H. Jorulf. "Suspected Appendicitis in Children: US and CT--a Prospective Randomized Study." *Radiology*, 2002.
- S. Krajewski, Jacqueline A. Brown, P. Phang, M. Raval, and C. Brown. "Impact of Computed Tomography of the Abdomen on Clinical Outcomes in Patients with Acute Right Lower Quadrant Pain: A Meta-Analysis." *Canadian Journal of Surgery. Journal Canadien de Chirurgie*, 2011.
- S. Lee, and S. J. Yun. "Diagnostic Performance of Emergency Physician-performed Point-of-care Ultrasonography for Acute Appendicitis: A Meta-analysis." *American Journal of Emergency Medicine*, 2019.
- S. Sippola, J. Virtanen, Ville Tammilehto, J. Grönroos, S. Hurme, H. Niiniviita, Elina Lietzén, and P. Salminen. "The Accuracy of Low-Dose Computed Tomography Protocol in Patients With Suspected Acute Appendicitis: The OPTICAP Study." *Annals of Surgery*, 2020.
- Seung-Hum Yu, Chun-Bae Kim, J. Park, M. Kim, and D. Radosevich. "Ultrasonography in the Diagnosis of Appendicitis: Evaluation by Meta-Analysis." *Korean Journal of Radiology*, 2005.
- Siyuan Bai, Siyuan Hu, Yi-xing Zhang, Sheng-Xuan Guo, R. Zhu, and Jing Zeng. "The Value of the Alvarado Score for the Diagnosis of Acute Appendicitis in Children: A Systematic Review and Meta-Analysis." *Journal of Pediatric Surgery*, 2023.
- Susan Walker, William Haun, Jeffrey Clark, Kim McMillin, Francine Zeren, and Terry M. Gilliland. "The Value of Limited Computed Tomography with Rectal Contrast in the Diagnosis of Acute Appendicitis." *American Journal of Surgery*, 2000.

T. Terasawa, C. Blackmore, S. Bent, and R. Kohlwes. “Systematic Review: Computed Tomography and Ultrasonography To Detect Acute Appendicitis in Adults and Adolescents.” *Annals of Internal Medicine*, 2004.

V. Giljaca, T. Nadarević, G. Poropat, V. Nadarević, and D. Štimac. “Diagnostic Accuracy of Abdominal Ultrasound for Diagnosis of Acute Appendicitis: Systematic Review and Meta-Analysis.” *World Journal of Surgery*, 2017.

———. “PWE-030 Diagnostic Accuracy of Abdominal Ultrasound for Diagnosis of Acute Appendicitis: Systematic Review and Meta-Analysis.” *Gut*, 2017.

V. Hlibczuk, Judith A. Dattaro, Zhezhen Jin, L. Falzon, and Michael D. Brown. “Diagnostic Accuracy of Noncontrast Computed Tomography for Appendicitis in Adults: A Systematic Review.” *Annals of Emergency Medicine*, 2009.

V. Mittal, J. Goliath, Mubashir Sabir, Rakesh Patel, B. Richards, Idan Alkalay, S. Remine, and M. Edwards. “Advantages of Focused Helical Computed Tomographic Scanning with Rectal Contrast Only Vs Triple Contrast in the Diagnosis of Clinically Uncertain Acute Appendicitis: A Prospective Randomized Study.” *Archives of Surgery*, 2004.

Vinícius Diniz Cedro Araújo, Bianca Cardoso Lopes, Andy Petroianu, and Iure Kalinine Ferraz Souza. “Complementary Computed Tomography to Inconclusive Ultrasonography in Children with Suspected Acute Appendicitis: A Systematic Review and Meta-Analysis.” *The American Surgeon*, 2025.

W. Bom, M. Bolmers, S. Gans, C. V. van Rossem, A. V. van Geloven, P. Bossuyt, J. Stoker, and M. Boermeester. “Discriminating Complicated from Uncomplicated Appendicitis by Ultrasound Imaging, Computed Tomography or Magnetic Resonance Imaging: Systematic Review and Meta-Analysis of Diagnostic Accuracy.” *BJS Open*, 2020.

Yuhao Chen, Jushun Yang, Ruochen Cong, Meirong Wang, Jing Xiao, Jianfeng Zhu, Ying He, and Bosheng He. “Comparison of Contrast-Enhanced Versus Non-Enhanced Helical Computed Tomography in the Diagnosis of Acute Appendicitis: A Meta-Analysis.” *Journal of the College of Physicians and Surgeons--Pakistan : JCPSP*, 2023.