Original Article

Comparative Diagnostic Performance of Six Clinical Scoring Systems for Identifying Complicated Appendicitis in School-Aged Children: A Multicenter Prospective Study

Qasem Alyhari[1], Saif Ghabisha[2], Saleh Al-Wageeh[3], Faisal Ahmed[4], Abdulfattah Altam[5]

[1,2,3]Department of General Surgery, School of Medicine, Ibb University, Ibb, Yemen, [4]Department of Urology, School of Medicine, Ibb University, Ibb, Yemen

[5]Department of General Surgery, School of Medicine, 21 September University, Sanaa, Yemen

Abstract

Background: Complicated appendicitis (CA) in children is associated with increased morbidity and prolonged hospitalization. Early identification of patients at higher risk may facilitate timely surgical management and optimize resource utilization. **Aim**: to compare the diagnostic performance of six clinical scoring systems for identifying CA in school-aged children. **Patients and Methods**: A multicenter prospective observational study was conducted among children aged 10–15 years with histopathologically confirmed acute appendicitis between January 2024 and December 2025. The Appendicitis Inflammatory Response (AIR), Modified Alvarado Scoring System (MASS), Paediatric Appendicitis Score (PAS), Modified PAS (MPAS), Lintula, and Tzanakis scores were calculated prospectively before surgery. Complicated appendicitis was defined according to the World Society of Emergency Surgery criteria. Diagnostic performance was assessed using receiver operating characteristic (ROC) analysis, comparison of areas under the curve (AUCs), calibration analysis, and decision curve analysis. **Results**: Among 360 children included, 132 (36.7%) had complicated appendicitis and 228 (63.3%) had uncomplicated disease. The AIR score demonstrated the highest discriminatory performance, with an AUC of 0.78 (95% confidence interval [CI]: 0.74–0.82), sensitivity of 76.8%, and specificity of 73.8% at a cut-off value of ≥9. MASS showed moderate discrimination (AUC 0.72, 95% CI: 0.67–0.77), followed by Lintula (AUC 0.71), PAS (AUC 0.69), Tzanakis (AUC 0.68), and MPAS (AUC 0.65). DeLong comparison demonstrated significantly higher discrimination of AIR compared with other scoring systems. The AIR score showed the best calibration and highest net clinical benefit on decision curve analysis. **Conclusion**: The AIR score demonstrated superior diagnostic performance among six evaluated scoring systems for identifying complicated appendicitis in school-aged children. Its use may improve early risk stratification and clinical decision-making, particularly in settings where advanced imaging resources are limited. External validation in broader paediatric populations is recommended.

Keywords: Acute appendicitis; Appendicitis inflammatory response score; Children; Clinical scoring systems; Complicated appendicitis; Diagnostic accuracy.

Introduction

Acute appendicitis is the most common paediatric surgical emergency and a frequent cause of abdominal pain requiring urgent evaluation and treatment in children and adolescents.[1] Complicated appendicitis (CA), including gangrene, perforation, and abscess formation, is associated with increased postoperative morbidity, longer hospital stay, and greater healthcare burden compared with uncomplicated disease.[2,3] Early identification of children at increased risk of CA may support timely surgical decision-making and appropriate perioperative management.[4] Several clinical scoring systems have been developed to assist in the assessment of acute appendicitis, including the Appendicitis Inflammatory Response (AIR) score, Modified Alvarado Scoring System (MASS), Paediatric Appendicitis Score (PAS), Modified PAS (MPAS), Lintula score, and Tzanakis score.[5,6,7,8,9]

Address for correspondence: Faisal Ahmed
Department of Urology, School of Medicine, Ibb University, Ibb, Yemen
Email: fmaaa2006@yahoo.com

Received: 21-05-2026 **Revised:** 10-08-2026 **Accepted:** 18-08-2026 **Available Online:** 11-09-2026

Access this article online

Quick Response Code:

Website: https://wwwafrjpaedsurg.com

DOI: https://doi.org/10.67481/ajps.v23i1.20

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: editor@afrjpaedsurg.com

How to cite this article: Alyhari Q, Ghabisha S, Al-Wageeh S, Ahmed F Altam A. Comparative Diagnostic Performance of Six Clinical Scoring Systems for Identifying Complicated Appendicitis in School-Aged Children: A Multicenter Prospective Study. Afr J Paediatr Surg 2026;23(1):30-37.

However, their ability to distinguish complicated from uncomplicated appendicitis remains variable among paediatric populations. The PAS, although widely used in children, has shown limited accuracy for predicting CA, with reported moderate discrimination in some studies.[9] Combining clinical scores with inflammatory markers such as C-reactive protein (CRP) and neutrophil-related parameters may improve prediction of complicated disease.[10]

The AIR score incorporates clinical findings and graded inflammatory markers, including white blood cell count, neutrophil percentage, and CRP. Previous studies have suggested that AIR may provide better discrimination for advanced appendicitis than symptom-based scores such as Alvarado and PAS, particularly when inflammatory response is prominent.[11,12] Nevertheless, most available studies have evaluated individual scoring systems or single-centre cohorts, and prospective multicentre comparisons of multiple scoring tools in children remain limited. Reliable risk stratification tools are particularly important in settings where access to advanced imaging and specialist resources may be limited. A validated scoring approach could assist clinicians in identifying children who require closer monitoring, early intervention, or additional diagnostic evaluation.

Therefore, this multicentre prospective study aimed to compare the diagnostic accuracy and clinical utility of six commonly used scoring systems, namely AIR, MASS, PAS, MPAS, Lintula, and Tzanakis scores, for predicting complicated appendicitis in school-aged children.

Patients and Methods

Study Design and Setting

This multicentre prospective observational diagnostic accuracy study was conducted between January 2024 and December 2025 in the surgical and emergency departments of three Ibb university-affiliated hospitals in Ibb governorate, southwestern Yemen: Al-Thora hospital, a government tertiary hospital; Al-Nasr hospital, a private hospital, both located in Ibb city, and Al-Manar hospital, a private hospital approximately 15 km from the city. These hospitals provide paediatric surgical services to urban and semi-rural populations and serve a largely overlapping referral population.

A common study protocol, identical eligibility criteria, standardised outcome definitions, and uniform case-report forms were used across all centres. Surgeons, emergency physicians, radiologists, and research assistants received standardised training before study initiation. Demographic, clinical, laboratory, operative, and histopathological data were collected prospectively.

The primary objective was to compare the diagnostic accuracy of six clinical scoring systems (AIR, MASS, PAS, MPAS, Lintula, and Tzanakis) for predicting CA in school-aged children. The study is reported in accordance with the STROBE and STARD guidelines. Sample Size Estimation Sample size was estimated to provide adequate precision for the primary diagnostic accuracy analysis and sufficient power for pairwise comparisons of receiver operating characteristic (ROC) curve areas. Based on published validation studies, the expected area under the ROC curve (AUC) for the Appendicitis Inflammatory Response (AIR) score was conservatively specified as 0.78 and 0.68 for the next best-performing score (Modified Alvarado Score or Paediatric Appendicitis Score).[13,14,15] Assuming a two-sided α of 0.05, 80% power, a complicated appendicitis prevalence of 35%, and a within-patient correlation of 0.70 between scoring systems, approximately 400 children were required using the method of Hanley and McNeil for comparing correlated ROC curves.[16] Allowing for exclusions, the target sample size was increased to 450 participants. During the study period, 360 consecutive eligible children, including 132 with complicated appendicitis, were enrolled. Although the planned target sample size was not fully achieved, the number of events exceeded the recommended minimum for multivariable logistic regression (approximately 22 events per variable), supporting the planned analyses.

Ethical Considerations

The study protocol was approved by the Institutional Ethics Committee of the Faculty of Medicine and Health Sciences, Ibb University (Reference No. 021). The study was conducted according to the principles of the declaration of Helsinki. Written informed consent was obtained from parents or legal guardians, and age-appropriate assent was obtained from children aged seven years or older.

Study Population

Children aged 10–15 years presenting with suspected acute appendicitis and undergoing appendectomy were prospectively screened. Eligible participants had clinical suspicion of acute appendicitis, complete clinical and laboratory data, and histopathological confirmation. Exclusion criteria were previous or incidental appendectomy, incomplete records, chronic inflammatory disease, non-operative management, or a histologically normal appendix. Consecutive eligible patients were enrolled according to predefined criteria.

Definition of Complicated Appendicitis

The reference standard consisted of operative findings together with histopathological confirmation of appendicitis. According to the World Society of Emergency Surgery (WSES) criteria, complicated appendicitis was defined as gangrene, perforation, or peri-appendiceal abscess;[17] all other cases were classified as uncomplicated appendicitis. Histopathological assessment was performed by specialist pathologists blinded to the clinical scores.

Data Collection and Scoring Systems

Clinical and laboratory variables were recorded prospectively at presentation before surgery, including age, sex, symptom duration, fever (>38°C), abdominal tenderness, rebound tenderness, guarding, white blood cell count, neutrophil percentage, and C-reactive protein. AIR, MASS, PAS, MPAS, Lintula, and Tzanakis scores were calculated according to their original published criteria [18,19,20,21] by trained research assistants before surgery and

were unavailable to the treating surgeons. Surgeons and pathologists were blinded to the calculated scores. Ultrasonography, when clinically indicated, was interpreted by experienced radiologists or trained sonographers according to local practice, and findings required for the Tzanakis score were recorded before surgery (Supplementary Appendix).

Outcome Measures

The primary outcome was the diagnostic accuracy of each scoring system for predicting complicated appendicitis, assessed by the AUC. Secondary outcomes included sensitivity, specificity, positive and negative predictive values, optimal cut-off values, calibration, and clinical utility assessed by decision curve analysis.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA) and MedCalc Statistical Software version 22.013 (MedCalc Software Ltd., Ostend, Belgium). Continuous variables were presented as mean ± standard deviation or median (interquartile range), according to their distribution, and categorical variables were presented as frequencies and percentages. Group comparisons used the independent-samples t-test or Mann–Whitney U test for continuous variables and the chi-square or Fisher's exact test for categorical variables. Diagnostic performance was evaluated using receiver operating characteristic (ROC) curve analysis, with optimal cut-off values determined by the Youden index. Areas under the ROC curves were compared using DeLong's test. Multivariable logistic regression identified independent predictors of complicated appendicitis. Variables with clinical relevance or statistical significance in univariable analysis were considered for inclusion. Multicollinearity was assessed using the variance inflation factor (VIF), with values <5 considered acceptable. Model calibration was evaluated using the Hosmer–Lemeshow goodness-of-fit test, discrimination by the C-statistic, and clinical utility by decision curve analysis. All tests were two-sided, and P<0.05 was considered statistically significant.

Results

Patient Characteristics

A total of 521 children with suspected acute appendicitis were assessed during the study period. After exclusion of 161 patients who did not meet eligibility criteria, 360 children with histopathologically confirmed appendicitis were included in the final analysis (Figure 1). Based on operative and histopathological findings according to WSES criteria, 228 patients (63.3%) had uncomplicated appendicitis and 132 (36.7%) had complicated appendicitis.

Baseline characteristics were presented in Table 1. Children with CA were slightly older than those with uncomplicated disease (13.9 ± 0.8 vs. 13.7 ± 0.9 years, P = 0.042). The proportion of males was similar between groups (73.5% vs. 73.2%, P = 0.950).

The duration of symptoms was longer among children with CA (median 36.0 hours, IQR 24.0–72.0) compared with uncomplicated appendicitis (median 14.0 hours, IQR 10.0–24.0; P <0.001). Fever, rebound tenderness, and guarding were more frequent in the CA group (all P <0.05).

Inflammatory laboratory markers were higher among children with CA. White blood cell count (11.2 ± 3.1 vs. 8.9 ± 2.4 $\times 10^3/\mu L$), neutrophil percentage (78.2 ± 9.5% vs. 68.4 ± 8.2%), and CRP level (58.0 [IQR 42.0–85.0] vs. 32.0 [IQR 20.0–45.0] mg/L) differed significantly between groups (all P <0.001).

Diagnostic Performance of Clinical Scoring Systems

The diagnostic performance of the six scoring systems was shown in Table 2 and Figure 2A. The AIR score demonstrated the highest discrimination, with an AUC of 0.78 (95% CI 0.74–0.82). At a cut-off value of ≥9, AIR showed 76.8% sensitivity, 73.8% specificity, PPV of 62.9%, and NPV of 84.6%.

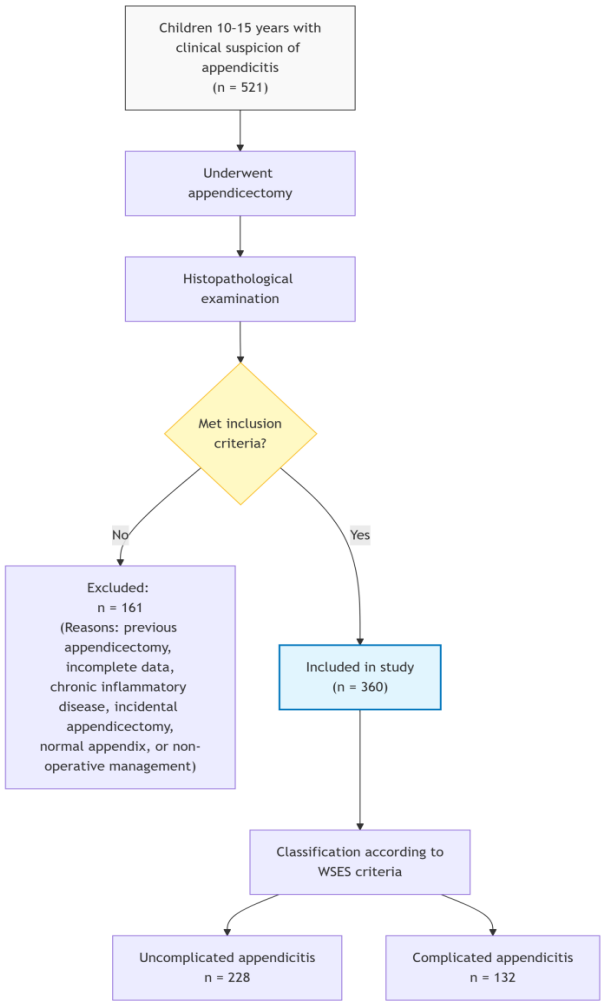


Figure 1. Study flow diagram showing patient enrolment (n = 521), exclusions (n = 161), surgical and histopathological confirmation, and final classification into uncomplicated (n = 228) and complicated (n = 132) appendicitis groups according to WSES criteria.

Table 1: Baseline clinical and laboratory characteristics of school-aged children with acute appendicitis, stratified by disease severity

Variable	Uncomplicated appendicitis (n = 228)	Complicated appendicitis (n = 132)	*P* value
Age (years), mean ± SD	13.7 ± 0.9	13.9 ± 0.8	0.042
Male sex, n (%)	167 (73.2)	97 (73.5)	0.950
Symptom duration (hours), median (IQR)	14.0 (10.0–24.0)	36.0 (24.0–72.0)	< 0.001
Fever >38 °C, n (%)	96 (42.1)	74 (56.1)	0.011
Rebound tenderness, n (%)	172 (75.4)	110 (83.3)	0.048
Guarding, n (%)	160 (70.2)	104 (78.8)	0.035
WBC count (×10³/µL), mean ± SD	8.9 ± 2.4	11.2 ± 3.1	< 0.001
Neutrophil percentage (%), mean ± SD	68.4 ± 8.2	78.2 ± 9.5	< 0.001
CRP (mg/L), median (IQR)	32.0 (20.0–45.0)	58.0 (42.0–85.0)	< 0.001

SD: standard deviation; IQR: interquartile range; WBC: white blood cell; CRP: C reactive protein. Continuous variables were compared using the independent samples t test or the Mann–Whitney U test; categorical variables were compared using the chi-square test or Fisher's exact test. Exact P values were reported.

Table 2: Diagnostic performance of six clinical scoring systems for predicting complicated appendicitis

Scoring system	Optimal cut-off	AUC (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV %	NPV %	*P* value*
AIR score	≥9	0.78 (0.74–0.82)	76.8 (71.2–81.5)	73.8 (68.4–78.6)	62.9	84.6	Reference
MASS	≥7	0.72 (0.67–0.77)	68.2 (62.1–73.8)	65.4 (59.8–70.6)	53.6	78.1	0.003
Lintula	≥21	0.71 (0.66–0.76)	64.5 (58.3–70.3)	67.1 (61.5–72.3)	52.1	77.4	< 0.001
PAS	≥6	0.69 (0.64–0.74)	62.1 (55.8–68.0)	63.5 (57.8–68.9)	50.3	74.2	< 0.001
Tzanakis	≥8	0.68 (0.63–0.73)	59.4 (53.1–65.4)	61.2 (55.5–66.7)	47.6	72.1	< 0.001
MPAS	≥4	0.65 (0.60–0.70)	55.8 (49.5–62.0)	58.4 (52.7–63.9)	44.3	69.2	< 0.001

AUC: area under the receiver operating characteristic curve; CI: confidence interval; PPV: positive predictive value; NPV: negative predictive value; AIR: Appendicitis Inflammatory Response; MASS: Modified Alvarado Scoring System; PAS: Paediatric Appendicitis Score; MPAS: Modified Paediatric Appendicitis Score.
*P value for pairwise AUC comparison versus AIR score (DeLong's test).
PPV and NPV were calculated using the observed prevalence of complicated appendicitis (36.7%).

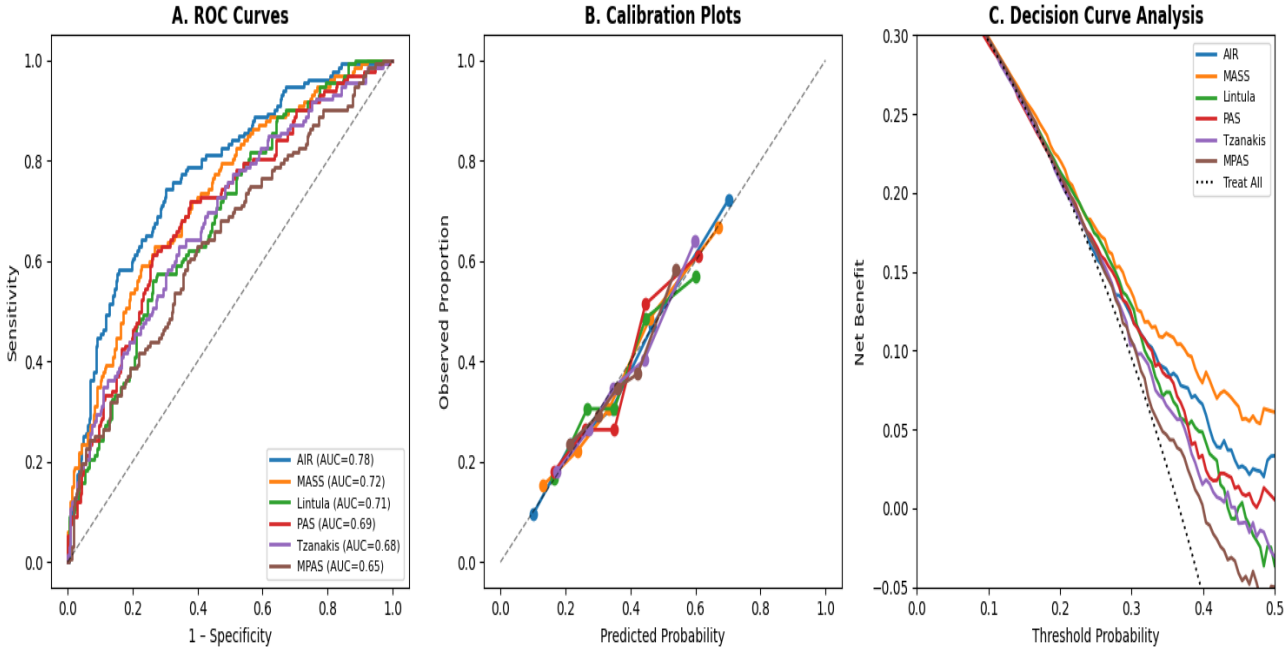


Figure 2: Diagnostic performance and clinical utility of six scoring systems for predicting complicated appendicitis. (A) Receiver operating characteristic (ROC) curves for the Appendicitis Inflammatory Response (AIR), Modified Alvarado Scoring System (MASS), Paediatric Appendicitis Score (PAS), Modified PAS (MPAS), Lintula, and Tzanakis scores; the AIR score showed the highest area under the curve (AUC 0.78, 95% CI 0.74–0.82). (B) Calibration plots comparing predicted probabilities against observed frequencies; the AIR score demonstrated the closest agreement. (C) Decision curve analysis of net clinical benefit across threshold probabilities; the AIR score provided the highest net benefit over the clinically relevant range.

Table 3: Multivariable logistic regression analysis for predictors of complicated appendicitis

Variable	Adjusted OR	95% CI	*P* value	VIF
AIR score (per 1-point increase)	1.42	1.25–1.61	< 0.001	1.82
Symptom duration >48 hours	2.15	1.42–3.26	< 0.001	1.45
CRP ≥50 mg/L	1.84	1.15–2.94	0.011	2.10
Neutrophil percentage >85%	1.62	1.04–2.52	0.032	1.95
Age (per year increase)	1.08	1.02–1.15	0.015	1.12
Male sex	1.32	1.05–1.66	0.018	1.08

OR: odds ratio; CI: confidence interval; VIF: variance inflation factor. Model performance: Hosmer–Lemeshow goodness-of-fit test, $P = 0.42$; C-statistic (AUC) = 0.81 (95% CI 0.77–0.85); Nagelkerke $R^2 = 0.38$.

MASS showed an AUC of 0.72 (95% CI 0.67–0.77), followed by Lintula (AUC 0.71), PAS (AUC 0.69), Tzanakis (AUC 0.68), and MPAS (AUC 0.65).

DeLong comparison demonstrated higher discrimination for AIR compared with the other scoring systems. The difference between AIR and MASS was significant (P = 0.003), while comparisons with the remaining scores showed P <0.001.

Factors Associated with Complicated Appendicitis

Factors associated with CA were presented in Table 3.

Longer symptom duration (>48 hours), CRP ≥50 mg/L, and neutrophil percentage >85% were independently associated with CA. Symptom duration >48 hours showed an adjusted OR of 2.15 (95% CI 1.42–3.26, P <0.001), CRP ≥50 mg/L showed an adjusted OR of 1.84 (95% CI 1.15–2.94, P =0.011), and neutrophil percentage >85% showed an adjusted OR of 1.62 (95% CI 1.04–2.52, P =0.032).

Age and male sex also showed statistically significant associations with CA.

The final model demonstrated acceptable calibration

(Hosmer–Lemeshow P =0.42) and good discrimination with a C-statistic of 0.81 (95% CI 0.77–0.85).

Calibration and Clinical Utility Analysis

Calibration analysis showed closer agreement between predicted and observed probabilities for AIR compared with other scoring systems (Figure 2B).

Decision curve analysis demonstrated higher net clinical benefit for AIR across clinically relevant threshold probabilities (Figure 2C).

Discussion

This multicenter prospective study compared six clinical scoring systems for identifying CA in school-aged children. The AIR score demonstrated the highest discriminatory performance among the evaluated tools, with better calibration and greater net clinical benefit compared with the MASS, PAS, Modified PAS (MPAS), Lintula, and Tzanakis scores. These findings indicate that inflammatory marker-based scoring may provide useful risk stratification for complicated disease in children with acute appendicitis.

Early recognition of CA remains clinically important because complicated disease is associated with perforation, abscess formation, longer hospitalization, and increased postoperative morbidity.[2,3] However, distinguishing complicated from uncomplicated appendicitis based only on clinical assessment can be challenging, particularly in children. Objective scoring systems may support clinical decision-making by combining relevant symptoms, examination findings, and laboratory marker.[9]

In this study, the AIR score achieved an AUC of 0.78, with 76.8% sensitivity and 73.8% specificity at a cut-off value of ≥9. AIR showed significantly better discrimination than the other evaluated scoring systems. These findings are consistent with previous studies demonstrating favourable performance of AIR in assessing advanced appendicitis. Andersson and Styrud reported high diagnostic accuracy of AIR in a systematic review and meta-analysis, with better performance than traditional scores such as Alvarado.[12] The lower AUC observed in our cohort compared with pooled estimates may be related to differences in patient age, disease prevalence, healthcare setting, and study methodology. Our analysis included only school-aged children with histopathologically confirmed appendicitis, whereas previous studies included broader populations and different outcome definitions.[11,12]

Similar findings have been reported in paediatric cohorts. Lee *et al*., found that AIR achieved an AUC of 0.78 for differentiating complicated from uncomplicated appendicitis in children younger than 10 years, while Alvarado and PAS showed lower discrimination.[11] The performance observed across different paediatric age groups supports the potential usefulness of AIR in severity assessment. The inclusion of inflammatory parameters such as white blood cell count, neutrophil percentage, and CRP may contribute to its performance, as these variables reflect systemic inflammatory response in addition to clinical findings.[22]

The other scoring systems showed moderate or limited discrimination. MASS, PAS, and MPAS depend mainly on symptoms and physical examination findings, which may vary according to patient age, communication ability, and examiner experience.[20,23] Although PAS remains widely used in paediatric appendicitis assessment, previous studies have reported limited accuracy for predicting complicated disease when used alone.[24] Ding *et al*., showed that combining PAS with inflammatory markers improved prediction of complicated appendicitis in children.[9] Similarly, Zachos *et al*., demonstrated improved risk classification when PAS was combined with neutrophil percentage and CRP.[9]

The Tzanakis score and MPAS showed the lowest AUCs. The relatively modest performance of the Tzanakis score may be explained in part by operator-dependent variability in ultrasonography and by incorporation bias, because ultrasound findings both contribute to the score and inform routine decision-making. In a prospective study of 141 children with suspected appendicitis, Sasğ *et al*., found that the Lintula score had the lowest negative appendectomy rate (6.38%), while the Tzanakis score had the highest overall diagnostic accuracy (85.1%); however, the study did not specifically evaluate complicated appendicitis.[25] Lee *et al*., observed that imaging features such as peri-appendiceal free fluid, although helpful, are best used together with clinical and laboratory markers; imaging alone has limited sensitivity for complicated disease.[11] In our setting, differences in sonographic expertise across centres may have further influenced the reproducibility of the Tzanakis score.

Multivariable analysis identified prolonged symptom duration, elevated CRP, and increased neutrophil percentage as factors associated with complicated appendicitis. These findings are consistent with previous reports emphasising the role of disease duration and inflammatory markers in severity assessment. Fujii *et al*., reported associations between symptom duration, CRP elevation, and CA in children.[26] Puputti *et al.,* also demonstrated the value of combining CRP with imaging findings to identify children at lower risk of complicated outcomes.[4]

The AIR score demonstrated better calibration and the highest net clinical benefit in decision curve analysis. These findings suggest that AIR may provide additional clinical information beyond discrimination alone.[27] In healthcare settings where access to advanced imaging is limited, a structured bedside assessment tool may assist clinicians in prioritizing further evaluation and management.

Study Limitations

This study has several limitations. First, only children aged

10–15 years were included, and the findings may not apply to younger children who often have different clinical presentations and higher perforation rates.[28] Second, although the study included three hospitals, all were located in the same geographic region, which may limit generalisability. Differences in surgical practice and ultrasound expertise may influence score performance. Third, only children undergoing appendectomy with histopathological confirmation were included. Patients managed non-operatively and those with normal appendices were excluded, which may introduce verification bias. Finally, external validation in independent paediatric populations was not performed.

Future studies should evaluate the performance of these scoring systems in broader paediatric populations and different healthcare environments. Integration of clinical scores with standardized imaging protocols, biomarkers, or predictive models may further improve risk assessment.

Conclusion

The AIR score demonstrated the highest diagnostic performance among six evaluated clinical scoring systems for predicting complicated appendicitis in school-aged children. It showed better discrimination, calibration, and clinical utility compared with MASS, PAS, MPAS, Lintula, and Tzanakis scores. Prolonged symptom duration, elevated CRP, and increased neutrophil percentage were also associated with complicated disease. AIR may serve as a useful bedside risk stratification tool for paediatric appendicitis, particularly in settings where diagnostic resources are limited. Further external validation across diverse paediatric populations is required before broader clinical implementation.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Waseem M, Wang CF. Paediatric appendicitis. StatPearls. Treasure Island (FL): StatPearls Publishing; 2026. Available from: ncbi.nlm.nih.gov
2. Aoda FS, Abood SS, Noori IF. Factors associated with increased risk of perforation of acute appendicitis. Ann Med Surg (Lond). 2026;88(1):95-9.
3. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World Journal of Emergency Surgery. 2020;15(1):27.
4. Puputti M, Pakarinen MP, Suominen JS. Differentiating uncomplicated from complicated acute appendicitis in paediatric patients with C-reactive protein and imaging findings-a retrospective single centre study in 1149 Patients. Acta Paediatr. 2026;115(2):399-407.
5. Naeem MS, Sadiq Z, Awais M, Rafi M, Javeed S, Ahmed I, *et al.* Accuracy of RIPASA and Lintula scores in diagnosing acute appendicitis using surgical findings as the gold standard. Cureus. 2022;14(11):e31297.
6. Balakrishnan P, Munisamy P, Vijayakumar S, Sinha P. Clinical scoring systems to diagnose complicated acute appendicitis in a rural hospital: are they good enough? Cureus. 2024;16(7):e64927.
7. Bharati YP, Mehra B, Ghoshal S, Dubhashi SP. Diagnostic efficacy study comparing Tzanakis scoring system with Alvarado scoring system in effective diagnosis of acute appendicitis. Cureus. 2024;16(4):e58018.
8. Juma IM, Angolile CM, Salim M, Byomuganyizi M. A comparative analysis of the tzanakis and modified alvarado scoring systems in diagnosing acute appendicitis at Muhimbili National Hospital, Tanzania: a retrospective cross-sectional study. Int J Surg Open. 2024;63:100682.
9. Ding Z, Shi X, Sun H, Chen H, Cheng J, Zhuang W. Diagnostic value of paediatric appendicitis score for complicated appendicitis in preschool children. BMC Pediatr. 2026;26(1):125.
10. Zachos K, Fouzas S, Kolonitsiou F, Skiadopoulos S, Gkentzi D, Karatza A, *et al*. Prediction of complicated appendicitis risk in children. Eur Rev Med Pharmacol Sci. 2021;25:(7)346-53.
11. Lee Y, Cho H, Gwak G, Bae B, Yang K. Scoring system for differentiation of complicated appendicitis in paediatric patients: Appendicitis scoring system in children. Glob Pediatr Health. 2021;8:2333794x211022268.
12. Andersson RE, Styrud J. Diagnostic value of the appendicitis inflammatory response (AIR) score. A systematic review and meta-analysis. World J Emerg Surg. 2025;20(1):12.
13. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg. 2008;32(8):1843-9.
14. Jose T, Rajesh PS. Appendicitis inflammatory response score in comparison to Alvarado score in acute appendicitis. Surg J (N Y). 2021;7(3):e127-e31.
15. Chisthi MM, Surendran A, Narayanan JT. RIPASA and air scoring systems are superior to alvarado scoring in acute appendicitis: diagnostic accuracy study. Ann Med Surg (Lond). 2020;59:138-42.
16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
17. Di Saverio S, Birindelli A, Kelly MD, Catna F, Weber DG, Sartelli M, *et al.* WSES Jerusalem guidelines for diagnosis and treatment of acute appendicitis. World J Emerg Surg. 2016;11:34.
18. Xu T, Zhang Q, Zhao H, Meng Y, Wang F, Li Y, *et al.* A risk score system for predicting complicated appendicitis and aid decision-making for antibiotic therapy in acute appendicitis. Ann Palliat Med. 2021;10(6):6133-44.
19. Korkut M, Bedel C, Karancı Y, Avcı A, Duyan M. Accuracy of Alvarado, Eskelinen, Ohmann, RIPASA and Tzanakis Scores in Diagnosis of Acute Appendicitis; a Cross-sectional Study. Arch Acad Emerg Med. 2020;8(1):e20.
20. Khanafer I, Martin DA, Mitra TP, Eccles R, Brindle ME, Nettel-Aguirre A, *et al.* Test characteristics of common appendicitis scores with and without laboratory investigations: a prospective observational study. BMC Pediatr. 2016;16(1):147.
21. Sencan A, Aksoy N, Yıldız M, Okur Ö, Demircan Y, Karaca I. The evaluation of the validity of Alvarado, Eskelinen, Lintula and Ohmann scoring systems in diagnosing acute appendicitis in children. Pediatr Surg Int. 2014;30(3):317-21.
22. Andersson M, Kolodziej B, Andersson RE. Validation of the Appendicitis Inflammatory Response (AIR) Score. World J Surg. 2021;45(7):2081-91.
23. Bekiaridou K, Kambouri K, Giatromanolaki A, Foutzitzi S, Kouroupi M, Aggelidou M, *et al.* Predicting complicated appendicitis in children: pros and cons of a new score combining clinical signs, laboratory values, and ultrasound images (CLU Score). Diagnostics (Basel). 2023;13(13):2213.
24. Malla BR, Batajoo H. Comparison of Tzanakis score vs Alvarado score in the effective diagnosis of acute appendicitis. Kathmandu Univ Med J (KUMJ). 2014;12(45):48-50.

25. Sağ S, Basar D, Yurdadoğan F, Pehlivan Y, Elemen L. Comparison of appendicitis scoring systems in childhood appendicitis. Turk Arch Pediatr. 2022;57(5):532-7.
26. Fujii T, Tanaka A, Katami H, Shimono R. Usefulness of the paediatric appendicitis score for assessing the severity of acute appendicitis in children. Pediatr Int. 2020;62(1):70-3.
27. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. Spine J. 2021;21(10):1643-8.
28. Li D, Zhang Z, Wan J, Zhao J, Wei G, Pan W, *et al.* Clinical features and perforation predictors of appendicitis in infants and toddlers under 3: A retrospective two-center study. Sci Rep. 2025;15(1):14338.