Medicine®

OPEN

Evaluation of the diagnostic performance of a decision tree model in suspected acute appendicitis with equivocal preoperative computed tomography findings compared with Alvarado, Eskelinen, and adult appendicitis scores
A STARD compliant article

Hyo Jung Kang, MD[a], Hyuncheol Kang, PhD[b], Bohyun Kim, MD[c], Min Seok Chae, MD[d], Young Rock Ha, MD[d], Seong Beom Oh, MD[e], Jung Hwan Ahn, MD[a,f,*]

Abstract
This study evaluated the diagnostic performance of a new clinical approach based on decision tree (DT) analysis in adult patients with equivocal computed tomography (CT) findings of acute appendicitis (AA) compared with previous scoring systems.

This retrospective study of 244 adult patients with equivocal CT findings included appendicitis (AG, n = 80) and non-appendicitis (NAG, n = 164) groups. The chi-squared automatic interaction detection algorithm was for AA prediction. A receiver operating characteristic curve analysis and area under the curve (AUC) were used to compare the DT analysis with Alvarado, Eskelinen score, and adult appendicitis scores (AAS).

The following factors were selected for AA prediction: rebound tenderness severity, migration, urinalysis, symptom duration, leukocytosis, neutrophil count, and C-reactive protein levels. The DT comprised 11 final nodes with the following AA probabilities: node 1, 100% (16/16); node 2, 90% (9/10); node 3, 80% (8/10); node 4, 60.9% (14/23); node 5, 50% (3/6); node 6, 43.8% (7/16); node 7, 22.6% (12/53); node 8, 13% (10/77); node 9, 5.6% (1/18); node 10, 0% (0/12); and node 11, 0% (0/3). The AUC of the DT was higher (0.850 [95% confidence interval {CI}; 0.799–0.893]) than the Alvarado score (0.695 [95% CI; 0.633–0.752]), AAS (0.749 [95% CI; 0.690–0.802]), and the Eskelinen score (0.715 [95% CI; 0.654–0.770]). The results were statistically significant when compared with the AUCs of the Alvarado score, Eskelinen score, and AAS ($P < .001$, $P < .001$, $P = .003$, respectively).

The DT-based approach facilitated AA diagnosis and determination of clinical status in patients with equivocal preoperative CT findings and ambiguous results.

Abbreviations: AA = acute appendicitis, AAS = adult appendicitis score, AG = appendicitis group, AIR = appendicitis inflammation score, AUC = area under the curve, CI = confidence interval, CRP = C-reactive protein, CT = computed tomography, DT = decision tree, NAG = non-appendicitis group, RIPASA = Raja Isteri Pengiran Anak Saleha score, RLQ = right lower quadrant, ROC = receiver operating characteristic, RT = rebound tenderness, WBC = white blood cell.

Keywords: acute appendicitis, computed tomography, decision tree, equivocal, scoring system

1. Introduction

Acute appendicitis (AA) is a medical condition that requires a prompt and accurate diagnosis to ensure that morbidity, mortality, and unnecessary surgical intervention rates are as low as possible.[1] In recent years, most of the emergency physicians or general surgeons have used abdominal computed tomography (CT) scans to accurately establish AA diagnosis clinically, before surgical intervention.[2–8] However, the equivocal findings of an abdominal CT scan complicate the diagnosis

and surgical decision-making in AA.[9] A study suggested that the incidences of equivocal CT findings of AA are approximately 5% to 13.1%.[10] Among these patients with equivocal preoperative CT findings, AA is detected in up to 30%.[9] Daly et al[3] suggested that equivocal CT findings involving suspected appendicitis represent a third group. These patients with symptoms carry an approximately 30% chance of appendicitis.[3] In these patients, re-evaluation and reassessment using ultrasound, CT images and magnetic resonance imaging (MRI) have been reported to improve diagnostic accuracy.[11–13] However, in most institutions, these methods are traditionally not frequently used to aid in the diagnosis of AA. Therefore, clinicians may find it challenging to establish appendicitis based on clinical presentation alone.[3,9,10]

Clinical scoring systems have been used to support the diagnosis of AA and reduce the use of abdominal CT scans. The Alvarado score was introduced in 1986 for its simplicity and clinical advantage.[2,14–17] A previous study reported the diagnostic performance of the clinical scoring systems (Alvarado score, Eskelinen score, appendicitis inflammation score [AIR], adult appendicitis score [AAS], and Raja Isteri Pengiran Anak Saleha score [RIPASA]) in adult patients with suspected AA with equivocal preoperative CT findings. In previous studies, the diagnostic power of this clinical scoring system was moderately accurate (Alvardo score; 0.698, AIR; 0.668, RIPASA; 0.653, Eskelinen; 0.710, AAS; 0.726). In the study that evaluated the diagnostic power of these scoring systems on equivocal AA, the scores were found to be less accurate than other studies.[5,14–16,18,19] Therefore, if patients with suspected AA and equivocal CT findings are classified as a third group, the previous scoring systems may not be accurate for AA diagnosis. Therefore, a new scoring system or method is needed.

The decision tree (DT) analysis is a data mining framework resembling a flowchart for the extraction of predictive models.[20] The DT model facilitates analysis of relationships between different covariates, or creation of algorithms that predict a specific target variable. Recently, this model was used in a clinical setting to predict the diagnosis of various diseases by exploring the relationships between different covariates.[21,22]

Therefore, we hypothesize that a clinical approach based on a DT facilitates the diagnosis of AA with equivocal preoperative CT findings compared with the previous scoring systems. To the best of our knowledge, studies investigating the diagnostic performance of clinical approaches based on DT in adult patients with suspected AA with equivocal preoperative CT findings have yet to be reported. The aims of this study were to evaluate the diagnostic performance of a new clinical approach based on DT and compare the diagnostic performance of DT-based approaches with previous scoring systems such as the Alvarado score, Eskelinen score, and AAS in adult patients with suspected AA and equivocal preoperative CT findings.

2. Methods

2.1. Study design

This study was designed as a single-center, observational, retrospective study conducted between April 2011 and November 2015 after obtaining the approval (IRB no. EMC16-03) of the Institutional Review Board of Bundang Jesaeng General Hospital, Seongnam, South Korea. Informed consent was waived for the current study.

2.2. Enrolled patients

During the study period, intravenously enhanced abdominal CT scans were performed on a total of 4690 patients (15 years or older) with suspected AA to evaluate the appendix. Preoperative diagnosis was confirmed in 4427 of these patients as follows: definitive appendicitis (n=1577), probable appendicitis (n=346), probably not appendicitis (n=355), and normal appendix (n=2149). Nineteen patients with incomplete medical records could not be compared with the previous clinical scoring systems, and therefore, were not included in the study. A total of 244 adult patients (≥15 years) with suspected AA and equivocal CT findings were finally enrolled (Fig. 1).

2.3. Clinical scores and data collection

The clinical scoring systems based on the Alvarado, Eskelinen, and AAS were compared with the new clinical approach based on DT. The scoring systems were selected based on their performance in predicting AA in patients with equivocal CT findings in a previous study.[19] The parameters used to calculate the Alvarado score, Eskelinen score, and AAS were obtained from patient medical records and laboratory report systems. Patient demographics, history, and urinalysis results were also collected. All data were collected comprehensively by a single emergency medicine physician. We defined fever as a body temperature greater than 37.3 °C. The laboratory reference values were based on the high reference values at the hospital. Leukocytosis was defined as a white blood cell (WBC) count of >10.0 (×10^9/L). An elevated neutrophil count was defined as higher than 80% of the total WBC count, and an elevated C-reactive protein (CRP) was defined by levels >0.8 mg/L. The Alvarado score was calculated based on the cumulative scores for nausea/vomiting (score 1), anorexia (score 1), migratory pain to the right lower quadrant (RLQ) (score 1), RLQ direct tenderness (score 2), rebound tenderness (score 1), elevated body temperature (≥37.3 °C) (score 1), elevated WBC (≥10.0 × 10^9/L) (score 2), and elevated neutrophil count (>75%) (score 1).[14] The Eskelinen score was calculated as a sum of the total scores in the following categories: RLQ pain (3.51), migratory pain to RLQ (3.51), duration of pain (2.13), RLQ direct tenderness (11.41), rebound tenderness (4.25), guarding (6.62), and elevated WBC count (≥10.0 × 10^9/L) (5.88).[23] The AAS was also calculated using the same method but under the following categories: RLQ pain (2), migratory pain to the RLQ (2), RLQ direct tenderness (3; men and women older than 50 years, score 1; women aged 16–49 years), rebound tenderness (2, mild; and 4, moderate to strong), elevated WBC (score 1, 7.2 ≤ WBC < 10.9 [×10^9/L]; score 2, 10.9 ≤ WBC < 14.0 [×10^9/L]; score 3, ≥14.0 [×10^9/L]); elevated neutrophil count (score 2, 62% ≤ neutrophil < 75%; score 3, 75% ≤ neutrophil < 83%; and score 4, ≥83%), and CRP (mg/L) (<24-h symptoms [score 2, 4 ≤ CRP < 11; score 3, 11 ≤ CRP < 25; score 5, 25 ≤ CRP < 83; score 1, CRP ≥ 83], and ≥24-h symptoms [score 2, 12 ≤ CRP < 152; score 1, CRP ≥ 152]).[16]

The final diagnoses were established as follows. In patients who underwent surgery, the diagnosis of AA was based on histological findings showing transmural infiltration of neutrophils in the appendix. In the case of patients who did not undergo surgery, the 2-week follow-up medical records were reviewed. Patients who were not followed up at our medical center were contacted directly via telephone, and for patients who were transferred to another medical center, the medical center was
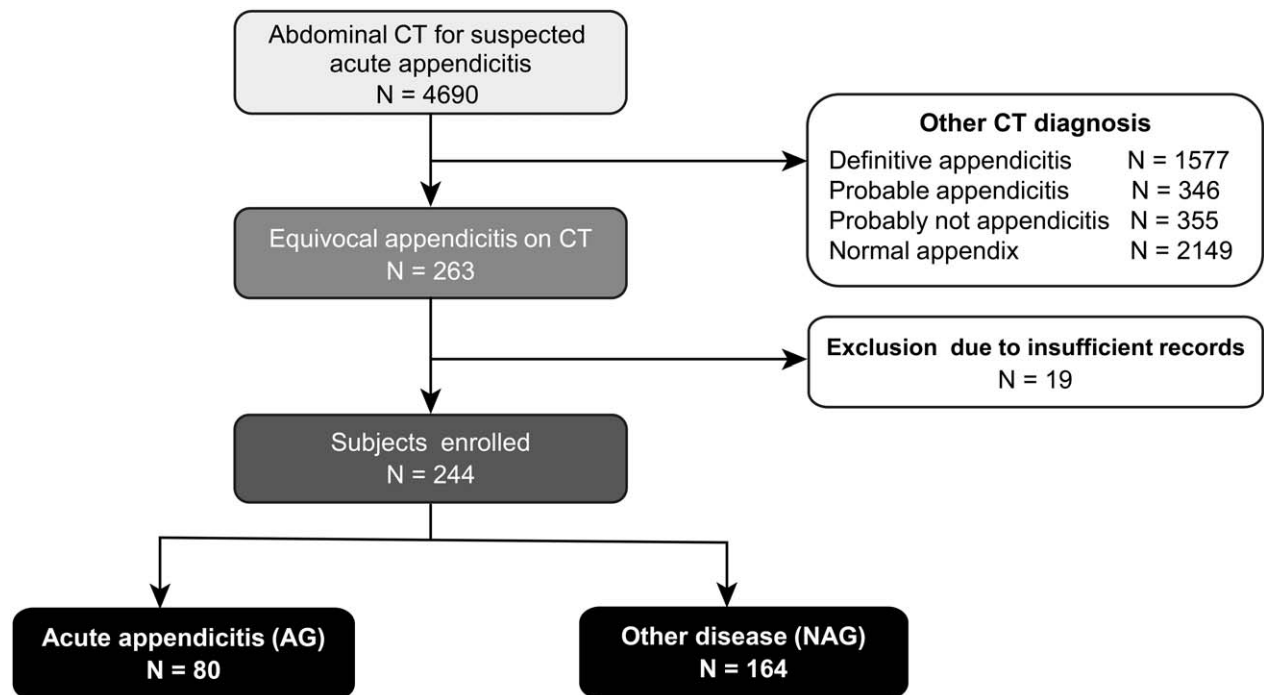
Figure 1. Study flowchart. AG=appendicitis group, CT=computed tomography, NAG=non-appendicitis group.

contacted directly via telephone. Enrolled patients were classified into 2 groups according to the final diagnosis: appendicitis group (AG, n=80) and non-appendicitis group (NAG, n=164).

2.4. Imaging methods and interpretation

A 16-slice multidetector CT scanner (Brilliance 16, Philips Healthcare, Eindhoven, the Netherlands) with intravenous administration of contrast material was used for all CT scans. Oral contrast medium was not administered in this study. A whole abdominal scan was performed from the diaphragm to the symphysis pubis. Other parameters included collimation, 1.5 mm; rotation time, 0.75 seconds; and pitch, 1.188. The reconstructed images consisted of axial and coronal sections measuring 3 to 5 mm. The tube voltage and tube current was 120 kVp and 150 to 300 mA, respectively. All CT scans were contrast-enhanced, with a 60-second delay after a starting dose of 2 mL/kg body weight of iohexol (Omnipaque 350, GE Healthcare, Princeton, NJ), and iopamidol (Pamiray 370, Dongkook Pharmaceutical, Seoul, Korea). The contrast medium was administered as an infusion through an antecubital vein at 4 mL/s.

A single abdominal radiologist with >10 years of experience in interpreting abdominopelvic CT scans retrospectively reviewed equivocal cases to identify the following CT signs of appendicitis: outer-to-outer diameter ≥6 mm (without intraluminal air), peri-appendiceal fat stranding or fluid, enhancement of the appendiceal wall, and concentric wall thickening. The presence of any one of these findings was defined as equivocal appendicitis. The diameter of the appendix was measured based on the axially enhanced sections and the largest portion of the appendix that was visible. If the appendix was not visualized or if the radiologist failed to trace the entire appendix, the patient was considered negative for appendicitis.[17]

2.5. Statistical analysis

All data analysis was performed using SPSS 23 statistics software (IBM, Inc., Armonk, NY) and MedCalc ver. 7.4 (MedCalc Software, Mariakerke, Belgium). Continuous data were expressed as means ± standard deviation or median (interquartile range) based on the results of normal distribution analysis. Categorical data are presented as absolute values, together with frequency distribution where appropriate.

The Mann–Whitney U test and independent t test were used to compare age and each of the clinical scores between AG and NAG according to the normal distribution analysis. The Chi-square test was used to compare the sexes of the 2 groups. A P value of <.05 was considered statistically significant.

The DT analysis was conducted using SPSS Decision Trees Version 23 (IBM, Inc., Armonk, NY). The variables used in this study were established in previous studies of AA diagnosis. The variables included sex, age, symptoms (anorexia, nausea or vomiting, migrated pain to RLQ), signs (RLQ direct tenderness, guarding, rebound tenderness, guarding, elevated body temperature), laboratory findings (elevated neutrophils and leukocystosis), and symptom duration (defined as the time between symptom onset and the CT scan). The chi-squared automatic interaction detection (CHAID) algorithm was used to create the tree models for AA prediction. The analysis was run automatically with the nodes defined as follows: parent node, 10 subjects, child nodes, 1 subject and the depth of tree in 3. The significance of the merged nodes and node splitting was set at a P value of <.05 and <.2, respectively.

The diagnostic accuracy of the new clinical approach based on DT and the previous scoring system was compared using the receiver operating characteristic (ROC) curves and the area under the curve (AUC) values plotted and calculated with MedCalc ver. 7.4 (MedCalc Software, Mariakerke, Belgium). A P value of <.05 was considered statistically significant.

The sample size based on our previous study yielded a sensitivity of 82.0%, a specificity of 53.9%, and a 24.8% incidence of appendicitis with a cut-off value of AAS 8. Based on the sample size calculation, 230 subjects were needed to achieve a precision of .10 for sensitivity, or specificity with an alpha error of .05, and a power of 80%, according to a previous study.[24] Considering the study duration, a total sample size of 244 subjects was included in this study.

3. Results

3.1. General characteristics of enrolled patients

The patient demographics, different scoring systems, and the clinical symptoms and signs are summarized in Table 1. No significant difference was observed in sex ratio, age, nausea or vomiting, anorexia, elevated body temperature, or elevated CRP between the groups (P=.388, P=.429, P=.463, P=.880, P=.936, P=.398, respectively). Among the 102 operations performed, 21.6% (n=22 in NAG) tested negative with appendectomy. Alternative diagnoses in patients who underwent surgery but had negative appendectomy results were as follows: acute gastroenteritis (n=6), pelvic inflammatory disease (n=2), diverticulitis (n=2), appendiceal mucocele (n=2), mesenteric lymphadenopathy (n=2), and non-specific findings (n=8). The diagnosis of 142 patients who did not undergo surgery was as follows: acute gastroenteritis (n=55, 38.7%), non-specific findings (n=32, 26.0%), pelvic inflammatory disease (n=27, 19.0%), mesenteric lymphadenopathy (n=8, 5.6%), diverticulitis (n=6, 4.2%), ovarian cystic rupture (n=4, 2.8%), acute pyelonephritis (n=3, 2.1%), and ureteric stones (n=2, 1.4%).

3.2. DT analysis

The factors selected for DT analysis to predict AA were as follows: rebound tenderness severity, migration, urinalysis, symptom duration, leukocytosis, elevated/normal neutrophil count, and elevated/normal CRP levels. The DT comprises 11 final nodes, as shown in Fig. 2. The severity of rebound tenderness was selected in the parent node. Node 11 (n=3, 1.2% of the total number of patients) showed absence of appendicitis in patients with grade 2 rebound tenderness severity and symptom duration >3288 minutes (54.8 hours). Patients were diagnosed with appendicitis if the node 1 showed grade 2 rebound tenderness severity and the symptom duration was <3288 minutes (54.8 hours), and normal CRP levels (n=16, 6.6% of the total number of patients). Node 10 (n=12, 4.9% of the total number of patients) showed that none of the patients manifesting rebound tenderness severity (grade 1), normal urinalysis results, and normal neutrophil counts had appendicitis. Node 2 (n=10, 4.1% of the total number of patients) showed that 90% of patients with grade 1 severity of rebound tenderness, abnormal urinalysis results, and migratory pain had appendicitis. In node 9 (n=18, 7.4% of the total number of patients), 94.4% of the patients without rebound tenderness but exhibiting migratory pain, and a lapse of >1638 minutes (27.3 hours) from the symptom onset until CT scans (17/18), showed no appendicitis. Finally, in nodes 8, 7, 6, 5, 4, and 3, the likelihood of testing negative for appendicitis was 87.0% (67/77), 77.4% (41/53), 56.2% (9/16), 50% (3/6), 39.1% (9/23), and 20% (2/10), respectively. The results for each final node are summarized in Table 2.

3.3. ROC analysis comparing the diagnostic performance of DT-based clinical approach and clinical scores

The results of the calculated ROC curve and AUC are presented in Fig. 3. Compared with the AUC in the previous scoring systems, the AUC of the DT in this study scored higher (0.850

Table 1

Patient and clinical demographics under each scoring system in different groups.

	Appendicitis group (n=80)	Non-appendicitis group (n=164)	P value
Gender (male:female)	31:49	54:109	.388
Age, y	35 [23, 42]	31 [23, 41]	.429
The duration of symptoms, min	747 [428, 1590]	1617 [569, 3259]	.001
Nausea/Vomiting (presence/absence)	36:44	82:82	.463
Anorexia (presence/absence)	14:66	30:134	.880
Migration to RLQ (presence/absence)	39:41	36:128	<.001
RLQ direct tenderness (presence/absence)	79:1	152:12	.048
Guarding (presence/absence)	71:9	163:1	<.001
The severity of rebound tenderness			<.001
Absence	37	134	
1	19	25	
2	24	5	
Fever (≥37.3°C) (presence/absence)	5:75	10:154	.963
Leukocytosis (presence/absence)	49:31	64:100	.001
Elevated neutrophil (presence/absence)	26:54	35:129	.059
Elevated CRP (presence/absence)	23:57	56:108	.398
Urinalysis (normal/abnormal)	27:53	85:79	.008
Alvarado score	5 [4, 7]	4 [3, 5]	<.001
Eskelinen score	20.8 [19.2, 25.1]	17.1 [14.9, 20.8]	<.001
AAS	12 [10, 15]	9 [7, 11]	<.001

AAS = adult appendicitis score, CRP = C-reactive protein, RLQ = right lower quadrant.
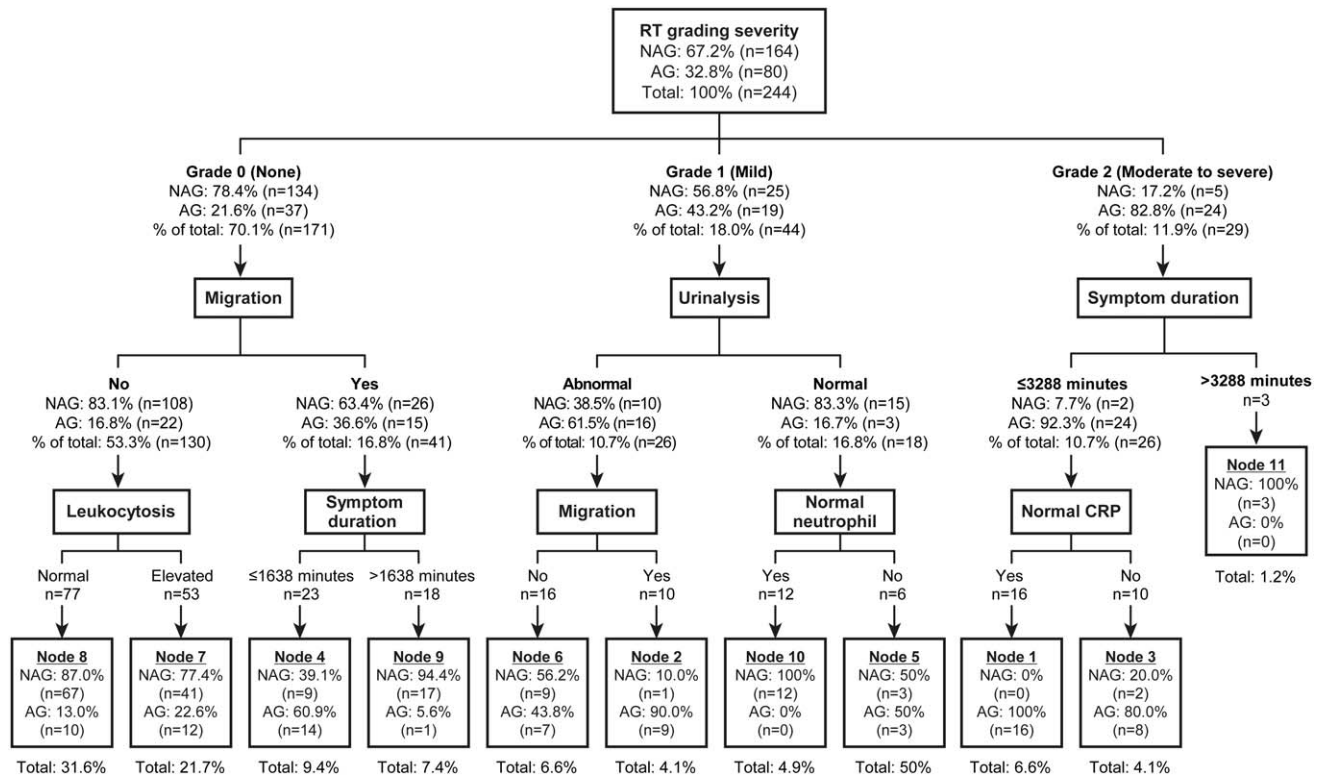
Figure 2. The new clinical approach using decision tree analysis for predicting acute appendicitis in patients with equivocal CT findings. AG = appendicitis group, CRP = C-reactive protein, CT = computed tomography, NAG = non-appendicitis group, RT = rebound tenderness, WBC = white blood cell count.

[95% confidence interval {CI}; 0.799–0.893]) than the Alvarado score (0.695 [95% CI; 0.633–0.752]), AAS (0.749 [95% CI; 0.690–0.802]), and Eskelinen score (0.715 [95% CI; 0.654–0.770]). The DT-based model showed a significant statistical difference compared with the AUCs of the Alvarado score, Eskelinen score, and AAS (P < .001, P < .001, P = .003, respectively).

4. Discussion

In this study, we utilized a DT analysis combining different parameters for the accurate diagnosis of AA in adult patients with

equivocal preoperative CT findings. The new clinical approach based on DT yielded high power (AUC 0.850) for the diagnosis of adult patients with suspected AA and equivocal preoperative CT findings. The DT model is simple to use in a clinical setting. Physicians can use the model by allocating patients to specific subgroups based on the probability of AA by simply following a flowchart-like tree framework.

In a previous study published by Daly et al,[3] it was suggested that patients suspected of AA with equivocal CT results should be classified as a third group, and radiologists should inform the referring physician that symptomatic patients with equivocal CT

Table 2

Summary of each final node, in the descending order of probability of appendicitis.

Node	n	Appendicitis % (n)	RT severity	Migratory pain	Symptom duration, min	Urinalysis	WBC	Neutrophil	CRP
1	16	100 (16)	2		≤3288				Normal
2	10	90 (9)	1	+		Abnormal			
3	10	80 (8)	2		≤3288				Elevation
4	23	60.9 (14)	0	+	≤1638				
5	6	50 (3)	1			Normal		Elevation	
6	16	43.8 (7)	1	−		Abnormal			
7	53	22.6 (12)	0	−			Elevation		
8	77	13.0 (10)	0	−			Normal		
9	18	5.6 (1)	0	+	>1638				
10	12	0 (0)	1			Normal		Normal	
11	3	0 (0)	2		>3288				
Total	244	28.9 (80)							

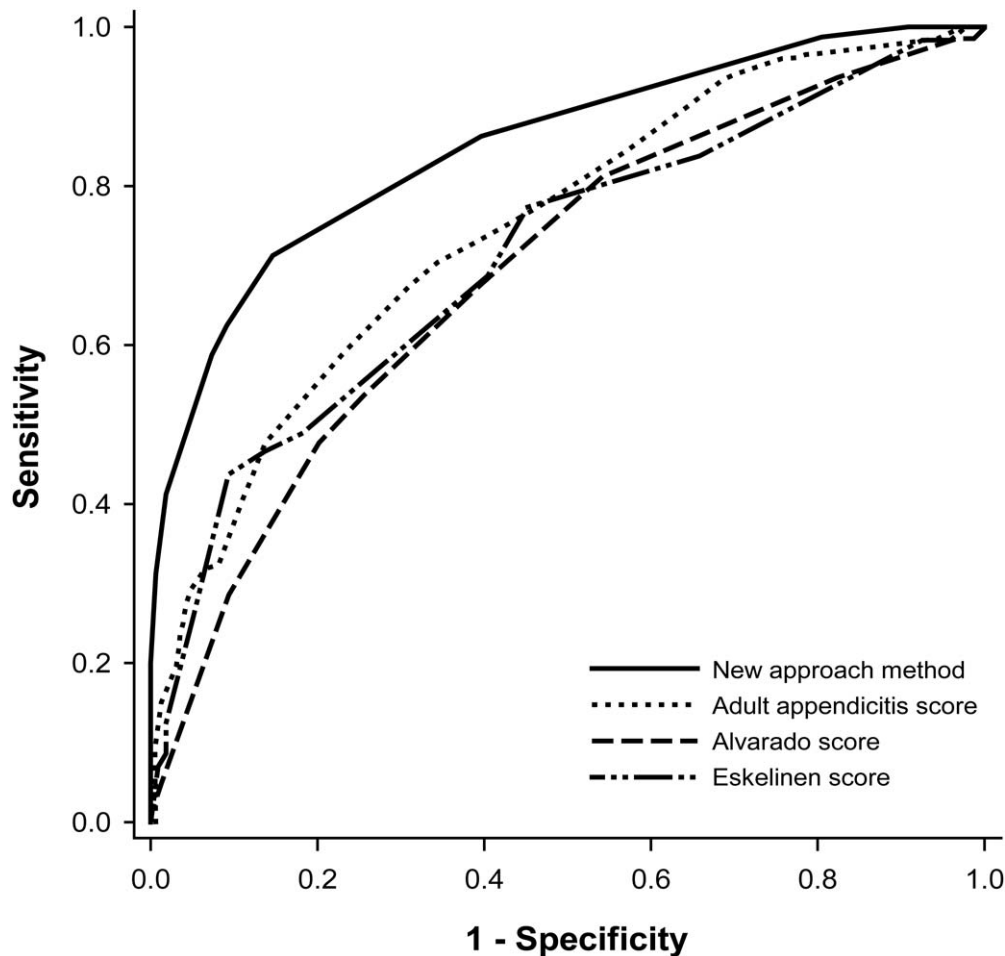CRP = C-reactive protein, RT = rebound tenderness, WBC = white blood cell.

Figure 3. The ROC curve and the AUC of each clinical score, and the new clinical approach based on decision tree analysis. The solid line represents the new clinical approach using the decision tree analysis, with an AUC of 0.850 (95% confidence interval; 0.799–0.893). The Alvarado score is represented by a dashed line, with an AUC of 0.695 (95% confidence interval; 0.633–0.752). The Eskelinen score is represented by the dash-dot-dot line, with an AUC of 0.715 (95% confidence interval; 0.654–0.770). The adult appendicitis score is represented by the dotted line, with an AUC of 0.749 (95% confidence interval; 0.690–0.802). The AUC in the decision tree analysis was statistically significant compared with other scoring systems ($P > .05$). AUC = area under the curve, ROC = receiver operating characteristic.

scans carry an approximately 30% risk of testing positive for appendicitis.[9] Under these conditions, a clinical diagnosis of appendicitis is a challenge.[3,9–11,13] Unfortunately, the diagnostic power of the previous clinical scoring systems were less accurate ($0.5 < AUC \leq 0.7$, Alvardo score; 0.698, AIR; 0.668, RIPASA; 0.653) to moderately accurate ($0.7 < AUC \leq 0.9$, Eskelinen; 0.710, AAS; 0.726) clinically.[19] Furthermore, 74.1% to 94.1% of patients with suspected AA were categorized under an intermediate probability group and therefore, the scoring systems were of limited use in establishing accurate diagnosis.[19] However, under the new model, the results provided diagnostic and treatment decisions (surgery or discharge) in 59 (24.2%) out of 244 patients. Seventy-seven patients (31.6%) may be recommended a discharge plan with short-term follow-up in the outpatient department. While a definitive diagnosis could not be made in another 108 patients (44.3%) using the new model, it may still facilitate prognostic evaluation of these patients based on admission for observation. These results show that DT model is a useful clinical tool in deciding the disposition of patients with equivocal CT findings.

Nodes 11, 10, 9, 2, and 1 were useful in diagnosing AA among 59 patients (24.2% of the total number of patients) in our study. As seen in nodes 11, 10, and 9, 32 out of 33 patients did not manifest AA, and nodes 2 and 1, 25 out of 26 patients showed AA. The results can be categorized according to the following variables: in the case of patients without AA, the rebound tenderness severity was grade 2; however, the symptom duration was >3288 minutes (54.8 hours) (node 11 [n = 3, 1.2% of the total number of patients]), the rebound tenderness severity was grade 1, and the urinalysis and neutrophil counts were normal (node 10 [n = 12, 4.9% of the total number of patients]), and presence of rebound tenderness along with migratory pain, and a symptom duration >1638 minutes (27.3 hours) (node 9 [94.4%, 17/18]). On the other hand, patients with AA showed a rebound tenderness severity of grade 2, the symptom duration was <3288 minutes, without CRP elevation (node 1 [n = 16, 6.6% of the total number of patients]), and rebound tenderness severity grade 1, with a normal urinalysis results, and migratory pain (node 2 [90%, 9/10]). Studies show a negative AA rate of 14% to 20% in the purely clinical setting.[25] As our study is based on equivocal

CT findings to establish a definitive clinical diagnosis, we assumed a negative appendectomy rate similar to previous studies in purely clinical setting. Therefore, based on this negative rate in purely clinical setting, we suggest discussing surgical intervention with the patients that fall under the NAG group in node 2. The results in node 8 (n=77, 31.6% of the total number of patients) showed a P-value of .149, which suggested limited statistical significance. However, it may still be useful in the clinical setting as it suggested that only 13% of patients without rebound tenderness and normal leukocytes had a final diagnosis of appendicitis. Therefore, patients in this group may be recommended a discharge plan with short-term follow-up in the outpatient department. On the other hand, 108 (44.3%) patients were categorized into nodes 7, 6, 5, 4, and 3 with a negative probability for appendicitis calculated as 77.4% (41/53), 56.2% (9/16), 50% (3/6), 39.1% (9/23), and 20% (2/10), respectively. Therefore, although none of these cases may require immediate surgery, it may be safe to admit patients for further observation and studies.

Interestingly, symptom duration was a useful factor in evaluation of patients with suspected AA, as seen in nodes 11, 9, and 4. The probability of appendicitis was low if the symptom duration was greater than the time selected by the individual nodes. The cut-off values in the nodes were 54.8 hours in node 11, and 27.3 hours in nodes 9 and 4. Other studies suggest that the duration of symptoms may aid in AA diagnosis.[16,26] The criteria for the CRP levels in AAS varied according to symptom duration and the cut-off value was 24 hours after the onset of symptoms.[16] Another study suggested that appendicitis may be difficult to diagnose within 12 hours from the onset of symptoms, and after 36 hours the likelihood of perforation was high, and therefore, if the symptoms lasted longer than 36 hours, the chances of appendicitis decreased.[26] These results are similar to our study results, wherein a symptom duration of <27 hours and >56 hours yielded a low probability of appendicitis in adult patients with equivocal preoperative CT findings. Therefore, we suggest that symptom duration should be considered when diagnosing AA, without any generalizations based on these few studies. Further studies with larger samples are needed to support this claim.

The present study has several limitations. First, the study was designed retrospectively. Therefore, the diagnostic performance of the new clinical approach based on DT was restricted compared with the previous clinical scoring systems. Second, the study is limited by the small size of the patient population in some final nodes because the DT model was created by setting the child node to one to maximize the probability of AA diagnosis. A further large, randomized controlled study may be required to confirm the results of this study. Third, the CT findings were interpreted by only a single radiologist and therefore, had no interobserver variation, which may have led to greater selection bias in the diagnosis based on equivocal CT findings.

In conclusion, the new clinical approach based on DT presented in this study facilitated the diagnosis of AA in adult patients with suspected AA and equivocal preoperative CT findings. It is also useful in deciding the disposition of patients with equivocal results. The parameters used in the new clinical approach based on DT include rebound tenderness severity, migration, urinalysis, symptom duration, leukocytosis, elevated or normal neutrophil levels, and elevated or normal CRP levels.

Author contributions

Conceptualization: Hyo Jung Kang, Hyuncheol Kang, Seong Beom Oh, Jung Hwan Ahn.
Data curation: Hyo Jung Kang, Hyuncheol Kang, Bohyun Kim, Min Seok Chae, Young Rock Ha, Jung Hwan Ahn.
Formal analysis: Hyo Jung Kang, Hyuncheol Kang, Young Rock Ha, Jung Hwan Ahn.
Investigation: Bohyun Kim, Min Seok Chae, Jung Hwan Ahn.
Methodology: Hyo Jung Kang, Hyuncheol Kang, Bohyun Kim, Min Seok Chae, Seong Beom Oh, Jung Hwan Ahn.
Supervision: Hyuncheol Kang, Bohyun Kim, Young Rock Ha, Seong Beom Oh, Jung Hwan Ahn.
Validation: Hyuncheol Kang.
Writing – original draft: Hyo Jung Kang, Jung Hwan Ahn.

References

[1] Seetahal SA, Bolorunduro OB, Sookdeo TC, et al. Negative appendectomy: a 10-year review of a nationally representative sample. Am J Surg 2011;201:433–7.
[2] Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. Br J Surg 2004;91:28–37.
[3] Daly CP, Cohan RH, Francis IR, et al. Incidence of acute appendicitis in patients with equivocal CT findings. AJR Am J Roentgenol 2005;184:1813–20.
[4] Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. Radiology 2006;241:83–94.
[5] Impellizzeri P, Centonze A, Antonuccio P, et al. Utility of a scoring system in the diagnosis of acute appendicitis in pediatric age. A retrospective study. Minerva Chir 2002;57:341–6.
[6] Jones K, Pena AA, Dunn EL, et al. Are negative appendectomies still acceptable? Am J Surg 2004;188:748–54.
[7] Pickhardt PJ, Lawrence EM, Pooler BD, et al. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. Ann Intern Med 2011;154:789–96. W-291.
[8] Vadeboncoeur TF, Heister RR, Behling CA, et al. Impact of helical computed tomography on the rate of negative appendicitis. Am J Emerg Med 2006;24:43–7.
[9] Levine CD, Aizenstein O, Lehavi O, et al. Why we miss the diagnosis of appendicitis on abdominal CT: evaluation of imaging features of appendicitis incorrectly diagnosed on CT. AJR Am J Roentgenol 2005;184:855–9.
[10] Webb EM, Wang ZJ, Coakley FV, et al. The equivocal appendix at CT: prevalence in a control population. Emerg Radiol 2010;17:57–61.
[11] Kim HC, Yang DM, Kim SW, et al. Reassessment of CT images to improve diagnostic accuracy in patients with suspected acute appendicitis and an equivocal preoperative CT interpretation. Eur Radiol 2012;22:1178–85.
[12] Repplinger MD, Levy JF, Peethumnongsin E, et al. Systematic review and meta-analysis of the accuracy of MRI to diagnose appendicitis in the general population. J Magn Reson Imaging 2016;43:1346–54.
[13] Sim JY, Kim HJ, Yeon JW, et al. Added value of ultrasound re-evaluation for patients with equivocal CT findings of acute appendicitis: a preliminary study. Eur Radiol 2013;23:1882–90.
[14] Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med 1986;15:557–64.
[15] 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;32:1843–9.
[16] Sammalkorpi HE, Mentula P, Leppaniemi A. A new adult appendicitis score improves diagnostic accuracy of acute appendicitis—a prospective study. BMC Gastroenterol 2014;14:114.
[17] Apisarnthanarak P, Suvannarerg V, Pattaranutaporn P, et al. Alvarado score: can it reduce unnecessary CT scans for evaluation of acute appendicitis? Am J Emerg Med 2015;33:266–70.
[18] Erdem H, Cetinkunar S, Das K, et al. Alvarado, Eskelinen, Ohhmann and Raja Isteri Pengiran Anak Saleha Appendicitis scores for diagnosis of acute appendicitis. World J Gastroenterol 2013;19:9057–62.

[19] Chae MS, Hong CK, Ha YR, et al. Can clinical scoring systems improve the diagnostic accuracy in patients with suspected adult appendicitis and equivocal preoperative computed tomography findings? Clin Exp Emerg Med 2017;4:214–21.

[20] Linden A, Yarnold PR. Using data mining techniques to characterize participation in observational studies. J Eval Clin Pract 2016;22:835–43.

[21] Imai S, Yamada T, Kasashi K, et al. Usefulness of a decision tree model for the analysis of adverse drug reactions: evaluation of a risk prediction model of vancomycin-associated nephrotoxicity constructed using a data mining procedure. J Eval Clin Pract 2017;23:1240–6.

[22] Xin Z, Hua L, Wang XH, et al. Reanalysis and external validation of a decision tree model for detecting unrecognized diabetes in rural Chinese individuals. Int J Endocrinol 2017;2017:3894870.

[23] Sitter H, Hoffmann S, Hassan I, et al. Diagnostic score in appendicitis. Validation of a diagnostic score (Eskelinen score) in patients in whom acute appendicitis is suspected. Langenbecks Arch Surg 2004;389:213–8.

[24] Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med 1996;3:895–900.

[25] Wagner PL, Eachempati SR, Soe K, et al. Defining the current negative appendectomy rate: for whom is preoperative computed tomography making an impact? Surgery 2008;144:276–82.

[26] Bickell NA, Aufses AHJr, Rojas M, et al. How time affects the risk of rupture in appendicitis. J Am Coll Surg 2006;202:401–6.