ORIGINAL ARTICLE

Decision Model for Acute Appendicitis Treatment With Decision Tree Technology—A Modification of the Alvarado Scoring System

Hsien-Wei Ting[1,2], Jing-Tang Wu[3], Chien-Lung Chan[2]*, Shoei-Loong Lin[4], Min-Hsiung Chen[5]

[1]*Department of Neurosurgery, Taipei Hospital, Department of Health, Taipei,* [2]*Department of Information Management, Yuan Ze University,* [3]*Department of Surgery, Taoyuan General Hospital, Department of Health, Taoyuan,* [4]*Department of Surgery, Taipei Hospital, Department of Health, and* [5]*Department of Neurosurgery, Taipei Veterans General Hospital, Taipei, Taiwan, R.O.C.*

Background: How to decide the proper time to do laparotomies for acute appendicitis patients is sometimes very difficult, especially in areas with no imaging diagnostic tools. The Alvarado scoring system (ASS) is a convenient and inexpensive decision making tool; however, its accuracy needs to be improved. The decision tree is the most frequently used data mining technology for diagnostic model building. This study used a decision tree to modify the ASS and to prioritize the variables.

Methods: We collected 532 patients who underwent appendectomy. Patients who had undergone incidental appendectomy were excluded from the study. The decision tree algorithm was constructed with the data mining workbench Clementine version 8.1. It is a top-down algorithm designed to generate a decision tree model with entropy. The algorithm chooses the best decision node with which to separate different classes from empirical data. The Wilcoxon signed rank test, Student *t* test and χ^2 test were used for statistical analysis.

Results: Among the 532 patients recruited into the study, 420 had acute appendicitis and 112 had normal appendix. Women with acute appendicitis were older than their male counterparts ($p < 0.001$). All patients had right lower quadrant tenderness. The new model was constructed with decision tree technology, and the accuracy of the diagnostic rate was better than that of ASS ($p < 0.001$). The sensitivity and specificity of the new model were 0.945 and 0.805, respectively.

Conclusion: The new model is more convenient and accurate than ASS. Right lower quadrant tenderness is an inclusion criterion for acute appendicitis diagnosis. Migrating pain and neutrophil count > 75% were significant factors for acute appendicitis diagnosis if ASS score < 6. Although the criteria of nausea/vomiting and white blood cell count > 10,000/dL were significantly different between acute appendicitis and normal appendix, there was no significant contribution of entropy change below the "neutrophil count > 75%" nodes in the model. So they were erased from the decision tree model. Further studies need to be conducted to investigate why older women are at higher risk for acute appendicitis. [*J Chin Med Assoc* 2010;73(8):401–406]

Key Words: acute appendicitis, Alvarado scoring system, decision tree, medical decision making

Introduction

Acute appendicitis is one of the most common surgical emergencies, with a worldwide incidence ranging from 75 to 120 per 100,000 population per year; however, it is often misdiagnosed.[1–5] Normal appendices are frequently found during laparotomy if only a few symptoms/signs of acute appendicitis are observed. Unfortunately, some patients with acute appendicitis are not diagnosed until peritonitis or other severe complications occur because their surgeons were waiting for more evidence of acute appendicitis. These patients have higher mortality and morbidity than patients who are diagnosed in a timely manner.[2,6–8] Deciding on when is the optimal time to perform laparotomy is important but sometimes difficult.

*Correspondence to: Dr Chien-Lung Chan, Department of Information Management, Yuan Ze University, 135, Yuan-Tung Road, Chung-Li 320, Taiwan, R.O.C.
E-mail: clchan@im.yzu.edu.tw • Received: September 30, 2009 • Accepted: June 29, 2010

ELSEVIER

There are several appendicitis scoring systems that help surgeons to make this decision; most are based on symptoms, signs and laboratory data. The Alvarado scoring system (ASS) is a simple scoring system for diagnosing appendicitis.[9] This scoring system includes 8 variables: 3 symptoms, 3 signs and 2 laboratory data, and is used in many countries and hospitals because of its ease of use, especially in areas with no imaging diagnostic tools. Although it is very convenient to use in clinical diagnosis, there are controversial conclusions with regard to its accuracy.[7,10,11] Hallan et al modified the ASS by adding some inflammatory factors,[12] but these laboratory studies would increase the health care costs.

Some acute abdominal pain diagnostic models have also been constructed, including Bayesian statistic models, discrimination rule, logistic regression and neural networks. The sensitivities of these models range from 55% to 99%.[13–15] Although most of the diagnostic models have a high sensitivity and specificity, they require computer aids and are not very convenient to use because of their complexity. Some researchers have combined the scoring systems and imaging studies or used only ultrasound or computed tomography as tools for acute appendicitis diagnosis.[6,16–21] But the diagnostic accuracy of imaging studies remain controversial.

A decision tree is a model of data that encodes the distribution of the class attributes in terms of the predictor attributes.[22,23] The success of a decision tree algorithm can also be measured at the level of meaningfulness. Tanner et al[24] used decision tree for the diagnosis and outcome of dengue fever; the sensitivity and specificity was relatively satisfactory. Frey et al[25] discovered prognostic molecular markers of lung cancer using a decision tree algorithm. Constructed decision trees are simple models for diagnosis based on empirical data,[26,27] and convenient to use.

This study modified the ASS with decision tree technology and constructed a convenient and accurate decision support model for acute appendicitis diagnosis and timing of laparotomy.

Methods

We retrospectively collected patients who underwent appendectomy between January 1, 1999 and June 30, 2004 at a 600-bed general hospital in northern Taiwan. All data were confirmed by 2 gastrointestinal surgeons. Patients in whom laparotomies were performed for reasons other than appendicitis, such as acute cholecystitis or diverticulitis, were excluded. Patients in whom appendectomies were performed during routine cesarean section and patients with incidental findings of acute appendicitis were also excluded. Normal appendectomy was defined as the removal of an appendix that did not have any pathological abnormalities.

There were 532 patients (327 men, 205 women) included in this study. All Alvarado score-related data from patients were collected. The patient data analyzed in this study included age, sex, length of hospital stay, presence or absence of migratory pain, anorexia, nausea/vomiting, right lower quadrant (RLQ) tenderness, and rebound pain. The body temperature at admission, white blood cell count (WBC), and differential count were also analyzed.

C5.0 is a series of decision tree models constructed by Quinlan.[22,23] It is a top-down algorithm designed to generate a decision tree model for diagnosis. The algorithm chooses the best decision node with which to separate different classes from empirical data.[25] The main induction loop of the decision tree is shown below.[22,23]

- Step 1: Assume A as the possible "best" decision attribute for the next node.
- Step 2: Assign A as the decision attribute for the node.
- Step 3: For each value of A, create a new descendent of the node.
- Step 4: Count the entropies of the training examples to leaf nodes.
- Step 5: Stop searching for new leaf nodes if training examples are well classified or continue the new leaf nodes if they are not well classified. (This loop needs to return back to Step 1 for the leaf nodes.)

The decision tree in this research was built using the C5.0 component of the workbench Clementine version 8.1 (SPSS Inc., Chicago, IL, USA). The Wilcoxon signed rank test, Student t test and χ^2 test were used to analyze the data of patients and the accuracies between old and new diagnostic models. Statistical significance was defined as $p < 0.05$.

Results

Among the 532 patients enrolled in this study, 340 (63.91%) had acute appendicitis, 80 (15.04%) had perforated appendicitis and 112 (21.05%) had normal appendix. There was no significant difference in age between the patients with normal appendix (mean age, 29.9 years) and those with acute appendicitis (mean age, 31.9 years; $p = 0.264$). Women with perforated appendicitis (mean age, 47.8 years) were significantly older than men with perforated appendicitis (mean age, 32.5 years; $p < 0.001$). Women with acute appendicitis

Table 1. Demographic characteristics of patients ($n=532$)

	Normal appendix ($n=112$)	Acute appendicitis ($n=340$)	Ruptured appendicitis ($n=80$)
Case number (male/female)	57/55	214/126	56/24
Mean age (yr)	29.9	31.9	37.1*
Male	29.3	29.7	32.5
Female	30.7	35.2†	47.8‡
Mean Alvarado score	5.81§	6.84	6.99
Mean length of hospital stay (d)	4.64	4.66	8.08‖

*$p<0.001$; †female acute appendicitis patients were older than male acute appendicitis patients ($p<0.001$); ‡female patients with ruptured appendicitis were older than other patients ($p<0.001$); §mean Alvarado score of normal appendix patients was lower than for other patients ($p<0.001$); ‖ruptured appendicitis patients had longer length of hospital stay than the other patients ($p<0.001$).

Table 2. Percentage of patients with positive symptoms/signs and laboratory data of Alvarado score ($n=532$)

Percentage of positive symptoms/signs	Normal appendix ($n=112$)	Acute appendicitis* ($n=420$)
Migrating pain†	22.32%	43.81%
Anorexia ($p=0.239$)	23.21%	28.81%
Nausea/vomiting†	27.68%	51.43%
RLQ tenderness‡	100%	100%
Rebound pain ($p=0.170$)	70.54%	63.57%
Temperature > 37.5°C ($p=0.069$)	25.00%	34.05%
WBC > 10,000/dL†	72.32%	88.10%
Neutrophil count > 75%†	67.86%	88.57%

*Including patients with ruptured appendicitis; †$p<0.001$; ‡all patients had RLQ tenderness as the presenting symptom. RLQ = right lower quadrant; WBC = white blood cell count.

(mean age, 35.2 years) were also significantly older than men with acute appendicitis (mean age, 29.7 years; $p<0.001$). However, there was no significant difference in age between men and women with normal appendix ($p=0.617$; Table 1).

Patients with normal appendix had lower Alvarado scores (mean score, 5.81) than patients with acute appendicitis (mean score, 6.84) and those with perforated appendicitis (mean score, 6.99; $p<0.001$); however, there was no significant difference in scores between patients with acute appendicitis and those with perforated appendicitis ($p=0.348$). The duration of hospital stay was significantly longer for patients with perforated appendix (8.08 days) than for the other groups of patients ($p<0.001$; Table 1).

All patients presented with RLQ tenderness. Migrating pain, nausea/vomiting, WBC > 10,000/dL (leukocytosis) and neutrophil count > 75% were all significantly more frequent in acute appendicitis patients than in patients with normal appendix ($p<0.001$). However, anorexia ($p=0.239$), rebound abdominal pain ($p=0.170$), and body temperature > 37.5°C ($p=0.069$) on admission were less frequently encountered in patients with acute appendicitis than in those with normal appendix (Table 2).

Decision tree is a classification structure in which the leaf nodes are the class attributes and the branches represent conjunctions of features that lead to those classifications. These class attributes are the attributes that need to be predicted.[18,19] Decision trees can be compared and evaluated on the basis of predictive accuracy, speed, robustness and flexibility. Predictive ability is the correct prediction of class attribute for unseen data. Speed deals with the computational cost involved in generating and predicting test data. Robustness is the ability to make correct decisions if noisy data or data with missing attributes are provided. Flexibility deals with the creation of an efficient tree even if large amounts of data are given. In this study, 3 decision levels and 6 leaf nodes were found. The results are shown in Figure 1.

Although the ASS could be used to make good decisions for laparotomies when ASS score > 6, it was not good enough when ASS score ≤ 6. Also, the ASS treated all diagnostic attributes equally. For example, it gives the same score if neutrophil count > 75% or body

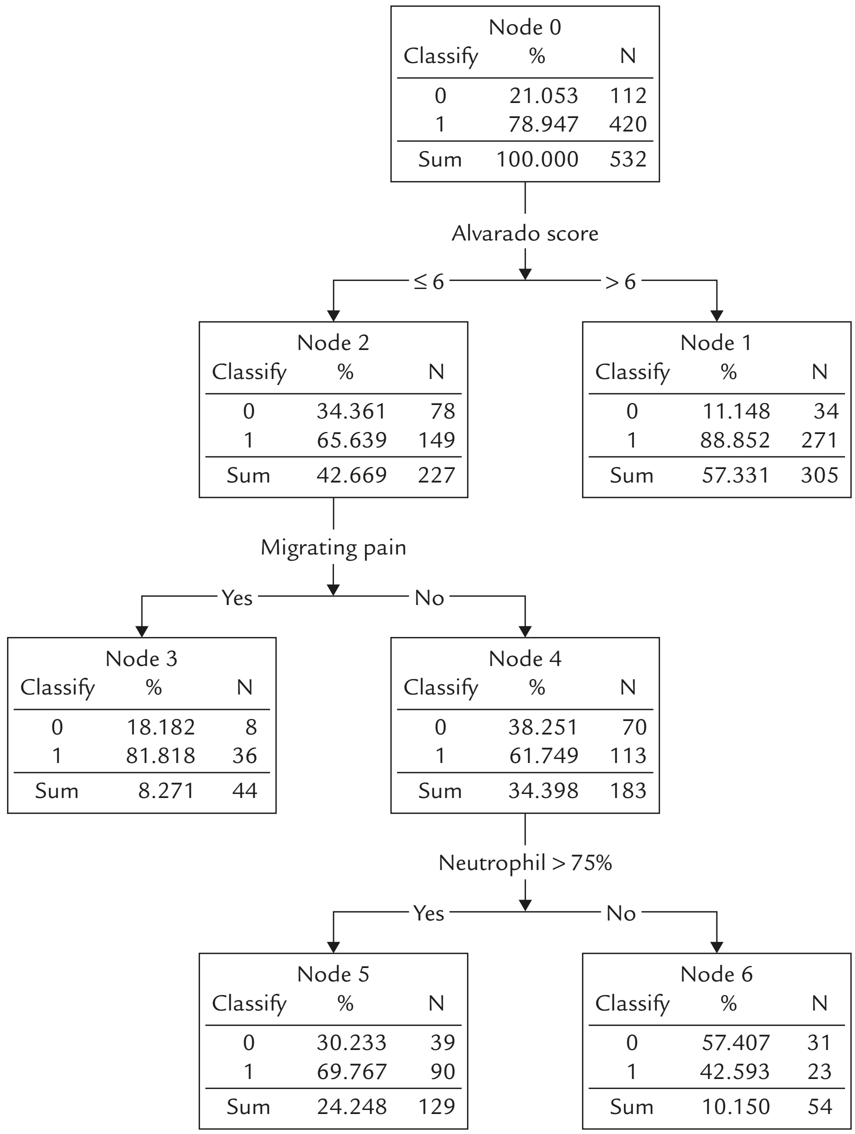

Figure 1. Decision tree process and nodes of acute appendicitis. 0 = normal appendix, 1 = acute appendicitis, including ruptured appendicitis. The decision tree and nodes were built using the C5.0 component of the workbench Clementine version 8.1.

temperature > 37.5°C. But in our study, neutrophil count > 75% was found to be more important than body temperature > 37.5°C for the diagnosis of acute appendicitis. The results of our decision tree were re-adjusted by the priority of attributes. Because 100% of acute appendicitis patients will have RLQ tenderness, the new decision model for the diagnosis of acute appendicitis treated RLQ tenderness as an inclusion criterion. Patients can be managed with observation if they do not have RLQ tenderness. Laparotomy is indicated in patients with Alvarado score > 6 and Alvarado score ≤ 6 with the presentation of migratory pain. Laparotomy is also indicated in patients with both Alvarado score ≤ 6 and no migrating pain if laboratory data reveal a neutrophil count > 75% (Figure 2). The sensitivity and the specificity of the new flow chart were 0.945 and 0.805, respectively.

Discussion

Kalan et al[28] modified the ASS by removing the neutrophil count from the model; however, we found that neutrophil count is a very important factor in evaluating patients with acute appendicitis ($p < 0.001$). It should not be excluded from the ASS. The new decision model modified the ASS and used patients' symptoms/signs to aid the decision for laparotomy. RLQ tenderness became an inclusion criterion in this flow chart. Migrating tenderness and neutrophil count > 75% are decision

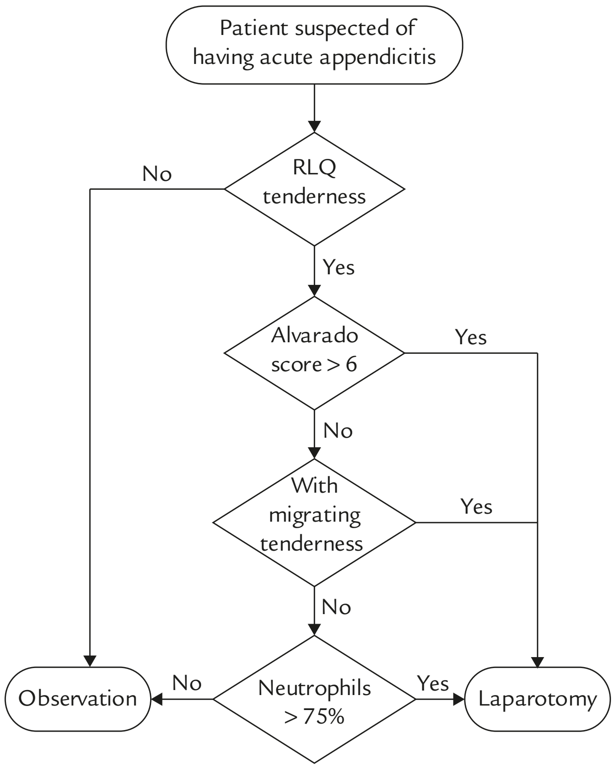

Figure 2. Diagnostic flow chart for acute appendicitis. This flow chart represents a modified Alvarado scoring system for diagnosing acute appendicitis. This new flow chart includes right lower quadrant (RLQ) tenderness. Patients can be managed with observation if there is no symptom of RLQ tenderness. Laparotomy is indicated in patients with an Alvarado score > 6 and also Alvarado score ≤ 6 if patients present with migrating pain. Laparotomy is indicated in patients with an Alvarado score ≤ 6 and no migrating pain if laboratory data reveal a neutrophil count > 75%.

nodes following the total Alvarado score result. The new decision model is more accurate than the ASS alone ($p < 0.001$).

Different from CHAID (χ^2 automatic interaction detector), the methodology of C5.0 is based on the entropy change to leaf nodes. It will choose the best nodes top-to-bottom.[22,23] Although the criteria of nausea/vomiting and WBC > 10,000/dL were significant between acute appendicitis and normal appendix, there was no significant contribution of entropy change below the "neutrophil count > 75%" nodes in the model in C5.0. So they were erased from the decision tree model. But they still play a very important role in this new model.

This study found that female patients with acute appendicitis (including perforated appendicitis) were older than male patients with acute appendicitis ($p < 0.001$); however, there was no significant sex predominance among patients with normal appendices. We also found that age was a risk factor for perforated appendicitis. This finding might be due to the delayed presentation of symptoms and signs in older patients. Some researchers have found that age and sex are risk factors for acute appendicitis and perforated appendix.[29–32] Further studies need to be conducted to investigate why older women are at higher risk for acute appendicitis.

Perforated appendicitis is associated with increased treatment cost. The duration of hospitalization was significantly longer for patients with perforated appendicitis than for those with non-perforated appendicitis ($p < 0.001$). The mean length of hospital stay for ruptured appendicitis (8.08 days) was significant longer than that for normal appendix (4.64 days) and acute appendicitis (4.66 days; $p < 0.001$). Unfortunately, there was no significant ASS score difference between patients with acute appendicitis and perforated appendicitis ($p = 0.348$). Further research for improving the accuracy rate of diagnosing ruptured appendicitis at an early stage is important as it would reduce hospital costs for appendicitis patients.

References

1. Loh DL, Munro FD, Wilson-Storey D, Orr JD. Early appendicitis: a safe diagnosis? *Ann Acad Med Singapore* 2004;33:530–1.
2. Al-Omran M, Mamdani M, McLeod RS. Epidemiologic features of acute appendicitis in Ontario, Canada. *Can J Surg* 2003;46:263–8.
3. Addis DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in United States. *Am J Epidemiol* 1990;132:910–25.
4. Anderson R, Hugander A, Thulin A, Nystrom PO, Olaison G. Indications for operation in suspected appendicitis and incidence of perforation. *BMJ* 1994;308:107–10.
5. Luckmann R. Incidence and case fatality rates for acute appendicitis in California: a population-based study of the effects of age. *Am J Epidemiol* 1989;129:905–18.
6. Jones K, Peña AA, Dunn EL, Nadalo L, Mangram AJ. Are negative appendectomies still acceptable? *Am J Surg* 2004;188:748–54.
7. Al-Hashemy AM, Seleem MI. Appraisal of the modified Alvarado score for acute appendicitis in adults. *Saudi Med J* 2004;25:1229–31.
8. Teicher I, Landa B, Cohen M, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. *Ann Surg* 1983;198:753–9.
9. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557–64.
10. Horzić M, Salamon A, Kopljar M, Skupnjak M, Cupurdija K, Vanjak D. Analysis of scores in diagnosis of acute appendicitis in women. *Coll Antropol* 2005;29:133–8.
11. Bhattacharjee PK, Chowdhury T, Roy D. Prospective evaluation of modified Alvarado score for diagnosis of acute appendicitis. *J Indian Med Assoc* 2002;100:310–1.
12. Hallan S, Asberg A, Edna TH. Additional value of biochemical tests in suspected acute appendicitis. *Eur J Surg* 1997;163:533–8.
13. Liu JL, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, Ohmann C, et al. Systematic reviews of clinical decision tools for acute abdominal pain. *Health Technol Assess* 2006;10:11–67.

14. Prabhudesai SG, Gould S, Rekhraj S, Tekkis PP, Glazer G, Ziprin P. Artificial neural networks: useful aid in diagnosing acute appendicitis. *World J Surg* 2008;32:305–9.
15. Fenyo G, Lindberg G, Blind P, Enochsson L, Oberg A. Diagnostic decision support in suspected acute appendicitis: validation of a simplified scoring system. *Eur J Surg* 1997;163: 831–8.
16. Tzanakis NE, Efstathiou SP, Danulidis K, Rallis GE, Tsioulos DI, Chatzivasiliou A, Peros G, et al. A new approach to accurate diagnosis of acute appendicitis. *World J Surg* 2005;29: 1151–6.
17. Bixby SD, Lucey BC, Soto JA, Theysohn JM, Ozonoff A, Varghese JC. Perforated versus nonperforated acute appendicitis: accuracy of multidetector CT detection. *Radiology* 2006;241: 780–6.
18. McKay R, Shepherd J. The use of the clinical scoring system by Alvarado in the decision to perform computed tomography for acute appendicitis in the ED. *Am J Emerg Med* 2007;25: 489–93.
19. Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. *Arch Surg* 2002; 137:799–804.
20. Flum DR, Morris A, Koepsell T, Dellinger EP. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA* 2001;286:1748–53.
21. Sakai S, Kobayashi K, Nakamura J, Toyabe S, Akazawa K. Accuracy in the diagnostic prediction of acute appendicitis based on the Bayesian network model. *Methods Inf Med* 2007; 46:723–6.
22. Quinlan JR. *C4.5*: *Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.
23. Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1: 81–106.
24. Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai YL, Ng LC, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLOS Neglect Trop D* 2008;12;2:e196.
25. Frey L, Edgerton ME, Fisher DH, Tang L, Chen Z. Using prior knowledge and rule induction methods to discover molecular markers of prognosis in lung cancer. *AMIA Annu Symp Proc* 2005:256–60.
26. Berson A, Smith S, Thearing K. *Building Data Mining Applications for CRM*, 2nd version. New York: McGraw-Hill, 1999.
27. Breiman L, Freidman JH, Olshen RA, Stone CJ. *Classification and Regression Tree*. California: Wadsworth International Group, 1984.
28. Kalan M, Rich AJ, Talbot D, Cunliffe WJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. *Ann R Coll Surg Engl* 1994;76:418–9.
29. Sun JS, Noh HW, Min YG, Lee JH, Kim JK, Park KJ, Kang DK, et al. Receiver operating characteristic analysis of the diagnostic performance of a computed tomographic examination and the Alvarado score for diagnosing acute appendicitis: emphasis on age and sex of the patients. *J Comput Assist Tomogr* 2008; 32:386–91.
30. Mohebbi HA, Mehrvarz S, Kashani MT, Kabir A, Moharamzad Y. Predicting negative appendectomy by using demographic, clinical, and laboratory parameters: a cross-sectional study. *Int J Surg* 2008;6:115–8.
31. Lyon C, Clark DC. Diagnosis of acute abdominal pain in older patients. *Am Fam Physician* 2006;74:1537–44.
32. Sheu BF, Chiu TF, Chen JC, Tung MS, Chang MW, Young YR. Risk factors associated with perforated appendicitis in elderly patients presenting with signs and symptoms of acute appendicitis. *ANZ J Surg* 2007;77:662–6.