J Med Syst (2007) 31:357–364
DOI 10.1007/s10916-007-9077-9

ORIGINAL PAPER

Comparison of the Levels of Accuracy of an Artificial Neural Network Model and a Logistic Regression Model for the Diagnosis of Acute Appendicitis

Shinya Sakai · Kuriko Kobayashi · Shin-ichi Toyabe · Nozomu Mandai · Tatsuo Kanda · Kohei Akazawa

Received: 27 February 2007 / Accepted: 25 June 2007 / Published online: 17 July 2007

Abstract An accurate diagnosis of acute appendicitis in the early stage is often difficult, and decision support tools to improve such a diagnosis might be required. This study compared the levels of accuracy of artificial neural network models and logistic regression models for the diagnosis of acute appendicitis. Data from 169 patients presenting with acute abdomen were used for the analyses. Nine variables were used for the evaluation of the accuracy of the two models. The constructed models were validated by the ".632+ bootstrap method". The levels of accuracy of the two models for diagnosis were compared by error rate and areas under receiver operating characteristic curves. The artificial neural network models provided more accurate results than did the logistic regression models for both indices, especially when categorical variables or normalized variables were used. The most accurate diagnosis was obtained by the artificial neural network model using normalized variables.

Keywords Acute appendicitis · Artificial neural network · Logistic regression · Bootstrap

The MATLAB program, which we used to construct the ANN models, is available for any interested individuals. Please contact via e-mail: toyabe@med.niigata-u.ac.jp.

S. Sakai · K. Kobayashi · N. Mandai
Division of Information Science and Biostatistics, Niigata University Graduate School of Medical and Dental Sciences,
Asahimachi-Dori 1-754,
Niigata 951-8520, Japan

S.-i. Toyabe · K. Akazawa (✉)
Department of Medical Informatics,
Niigata University Medical and Dental Hospital,
Asahimachi-Dori 1-754,
Niigata 951-8520, Japan
e-mail: akazawa@med.niigata-u.ac.jp

T. Kanda
Division of Digestive and General Surgery, Niigata University Graduate School of Medical and Dental Sciences,
Asahimachi-Dori 1-754,
Niigata 951-8520, Japan

Introduction

An early and accurate diagnosis of acute appendicitis is essential to minimize morbidity. However, making an accurate diagnosis of acute appendicitis in the early stage remains difficult. In many cases of acute appendicitis, the initial conditions, including clinical histories, clinical signs, and results of laboratory tests, are often uncertain. Errors in diagnosis lead not only to perforation but also to a resection of a normal appendix in from 5.2% to 42.2% of cases [1]. Difficulty in making an early diagnosis continues to result in a burden on patients, leading to unnecessary surgical operations, a prolonged length of hospital stay and increased medical costs [2].

Some diagnostic scores and computer-aided diagnostic systems have been proposed to improve the accuracy of diagnosis of acute appendicitis [3–8]. However, some of these scoring methods are not suitable for practical use because of the inclusion of too many variables and out-of-date variables in medical examinations. The sensitivity and specificity of some systems are also not sufficient for making a diagnosis. For example, if the Eskelinen score had been used for diagnosis, normal appendices would have been resected in 21.9% of the patients at this hospital who were suspected to have acute appendicitis.

Recently, artificial neural network (ANN) models have become popular in decision-making and outcome prediction of clinical medicine [9–11]. Pesonen et al. used an ANN model for the diagnosis of acute appendicitis, and they reported a good accuracy using various ANN algorithms [12].

This study compared the levels of accuracy of ANN models and logistic regression models for the diagnosis of acute appendicitis. The clinical characteristics of cases in which normal appendices were resected or in which abdominal operations were performed were also analyzed.

Methods

Patients

Data from patients who suffered from acute abdominal pain and were suspected of having acute appendicitis in Niigata University Medical and Dental Hospital in Japan during the period between 1993 and 2005 were used for analysis purposes in this study. Patients aged less than 4 years of age and patients who had been admitted for other diseases were excluded from this study. There were 184 patients who met the inclusion criteria. Fifteen patients (8.2%) were excluded because of missing values of C-reactive protein (CRP). Consequently, data from 169 patients were used for the analyses. Seventy-seven (45.6%) of the patients were males (aged from 4 to 74 years) and 92 (54.4%) of the patients were females (aged from 4 to 87 years). The diagnosis of acute appendicitis was confirmed by an appendectomy and a pathological examination of the resected specimens.

Variables used in the analyses

Nine variables, including age, gender, migration of abdominal pain to right lower quadrant (migration), tenderness at right lower quadrant (tenderness), rebound tenderness, muscular guarding, body temperature (BT), white blood cell count (WBC), and the CRP levels, were used to compare the levels of accuracy for the two methods of analysis. These variables, excluding age and gender, were selected on the basis of results of a meta-analysis [1].

All categorical data for clinical symptoms were coded as 0 for absent and 1 for present. The gender of the subjects was also coded as 0 for males and 1 for female. The presence of acute appendicitis was coded as 1 and other

Table 1 Patient characteristics and comparison with the controls

Variables and categories	Acute appendicitis (%; n=86)	No appendicitis (%; n=83)	p value
Age (mean±SD)	24.4±20.3	27.5±17.4	0.289
<16	40 (46.5)	27 (32.5)	0.117
≥16	46 (53.5)	56 (67.5)	
Gender			
Male	42 (48.8)	35 (42.2)	0.441
Female	44 (51.2)	48 (57.8)	
Migration			
Presence	33 (38.4)	23 (27.7)	0.146
None	53 (61.6)	60 (72.3)	
Tenderness			
Presence	69 (80.2)	65 (78.3)	0.850
None	17 (19.8)	18 (21.7)	
Rebound tenderness			
Presence	62 (72.1)	53 (63.9)	0.322
None	24 (27.9)	30 (36.1)	
Muscular guarding			
Presence	44 (51.2)	25 (20.1)	0.008
None	42 (48.8)	58 (69.9)	
BT (°C; mean±SD)	37.9±0.96	37.2±0.83	<0.001
≤37.5	34 (39.5)	61 (73.5)	<0.001
>37.5	52 (60.5)	22 (26.5)	
WBC (10^3/μl; mean±SD)	13.2±4.7	11.4±4.0	0.014
≤8.5	11 (12.8)	26 (31.3)	0.005
>8.5	75 (87.2)	57 (68.7)	
CRP (mg/dl; mean±SD)	8.6±8.4	4.1±6.7	<0.001
≤2.45	24 (27.9)	52 (62.7)	<0.001
>2.45	62 (72.1)	31 (37.3)	

Differences between the patients and the controls were analyzed by the chi-square test, Fisher's exact test for categorical variables and Wilcoxon's rank sum test for continuous variables

BT Body temperature, *WBC* white blood cell count, *CRP* C-reactive protein diagnoses (acute abdominal pain, ileitis, mesenteric lymphadenitis, diverticulitis, etc.) were coded as 0. Continuous variables were analyzed by the original values and by being transformed into categorical values and normalized values. Because children with appendicitis often show an atypical course [13], the age at diagnosis was categorized into two categories: 0 for patients younger than 16 years and 1 for patients aged 16 years or more [14]. The other continuous variables, BT, WBC and CRP, were also categorized into two categories: 0 for less than the cutoff point and 1 for more than the cutoff point. The cutoff points were determined using recursive partitioning analyses [8]. The normalized variables ranged from 0 to 1.

Chi-square tests and Fisher's exact tests were used to detect differences in proportions of patients and controls. Wilcoxon's rank sum tests were used for the univariate comparison of continuous data.

ANN model and logistic regression model

The ANN models used in this study were feed-forward networks, which were trained with a back propagation algorithm. The network has three layers: an input layer with nine neurons, a hidden layer with several neurons and an output layer with one neuron. In order to determine the optimal number of neurons in the hidden layer, experiments were performed by altering the number of neurons in the hidden layer. In the logistic regression models, all of the aforementioned independent variables were included in the regression model. The cutoff point for the model derived predicted probability of acute appendicitis versus other diseases was assigned a value of 0.5. The error rates of the logistic regression models and ANN models were calculated by using the ".632+ bootstrap method," which was designed to be unbiased and have low variance [13, 15]. Briefly, this method is a weighted combination of the apparent error rate and the leave-one-out bootstrap estimator: .632+ error rate=$(1-\omega)\times$(apparent error rate)+$\omega\times$(leave-one-out bootstrap estimator), where ω denotes the weight for the estimated error rates [13]. The apparent error rate is obtained by testing the model on the same training set. The leave-one-out bootstrap estimator was calculated by using 1,000 random samples sampled from original data with replacement (bootstrap sample). This estimator was evaluated by using the samples that were not included in each bootstrap sample (validation sample).

A receiver operating characteristic (ROC) curve was performed to evaluate the sensitivity and specificity of each model by using validation samples [16]. The optimum cutoff points were determined from the ROC curve [17]. Differences in the mean area under the curve (AUC) between the models were analyzed using Kruskal–Wallis tests and Wilcoxon's rank sum tests. Bonferroni's method was used to adjust for multiple comparison. In all statistics, a two-tailed p value of less than 0.05 was considered statistically significant. All analyses were performed using the SPSS software program (SPSS Inc., Chicago, IL), Clementine (SPSS) in addition to the MATLAB Neural Network Toolbox and Statistics Toolbox (MathWorks Inc., Natick, MA).

Results

Patient characteristics

Table 1 shows the characteristics of the patients and controls. An appendectomy was performed in 98 (58.0%) of the 169 subjects. Eighty-six of those patients were diagnosed to have acute appendicitis after the pathological examination of tissue samples. Therefore, normal appendices were removed in 12 (12.2%) of the patients. Final diagnoses in those patients were mesenteric lymphadenitis in three patients, diverticulitis in two patients, ileitis in two patients, and unknown origin in five patients. Thirteen (15.1%) of the 86 patients diagnosed as having acute appendicitis were diagnosed 12 h or more after the first medical examination. Consequently, acute appendicitis was diagnosed in 50.9% (86 of 169) of the patients, and 49.1% (83 of 169) of the patients were diagnosed as having other

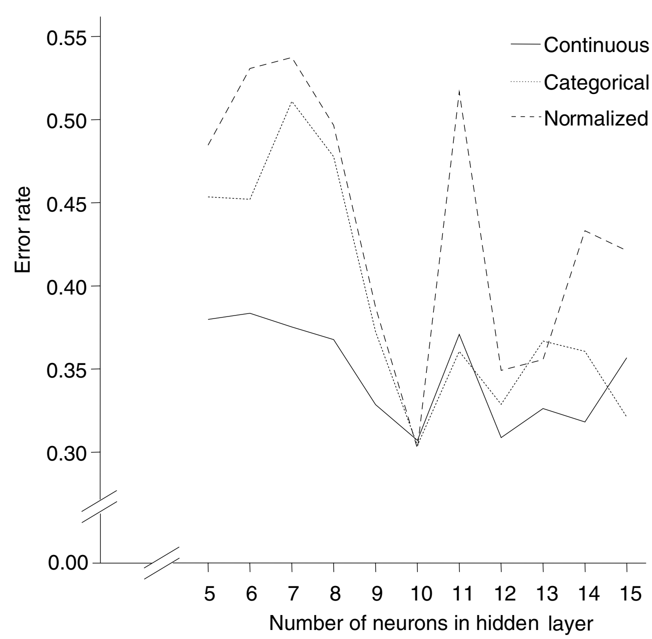

Fig. 1 Optimal number of neurons for the hidden layer. The relationship between the number of neurons for the hidden layer and the error rate of the corresponding artificial neural network (ANN) model is shown. *Continuous*: The model is constructed by using continuous variables, *Categorical*: the model is constructed by using categorical variables, *Normalized*: the model is constructed using normalized variables

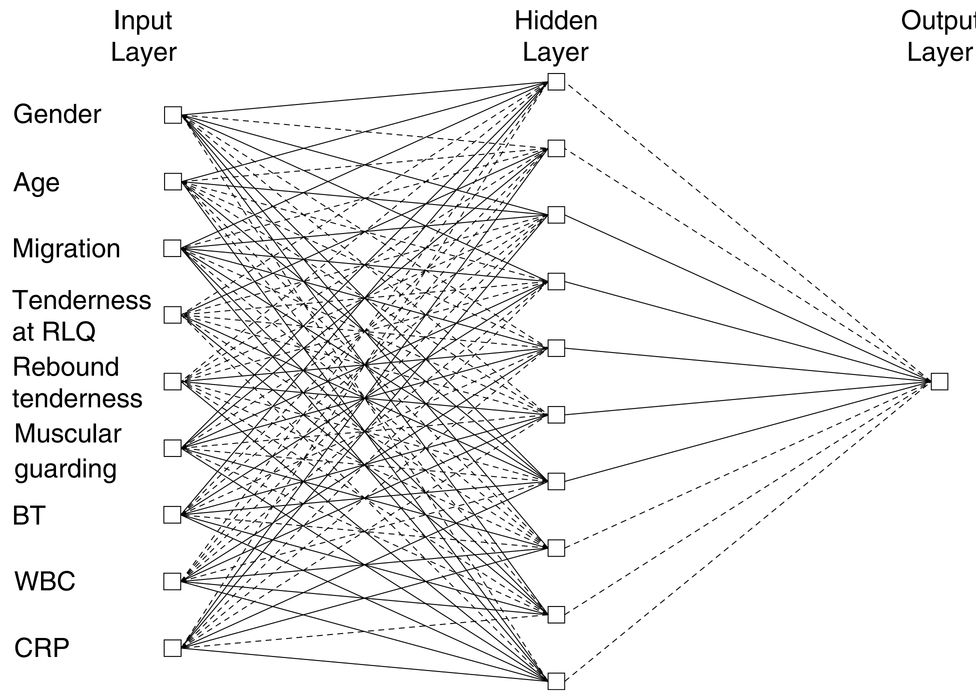

Fig. 2 ANN model using normalized variables. Summary of models including nodes and neurons is shown. The connection weight between two nodes is depicted by two lines. *Solid lines* show higher weight than *dotted lines*. *BT* Body temperature, *WBC* white blood cell count, *CRP* C-reactive protein

diseases. There were significant differences between the patients with appendicitis and those without appendicitis regarding the presence of muscular guarding, body temperature, WBC and level of CRP.

Model construction

The first step in the construction of the model is to try to determine the optimal number of neurons in the hidden layer. The error rate was found to change depending on the number of neurons in the hidden layer (Fig. 1). Since the minimum error rate was obtained at ten neurons, ten neurons were used for the hidden layer. The ANN model constructed by bootstrap sampling is summarized in Fig. 2. The bold lines and dotted lines indicate high connection weights and low connection weights between nodes, respectively. The cutoff point of weight is the median of absolute values of connection weights. Table 2 is a

Table 2 Coefficients of logistic regression analysis

Type of variables	Continuous			Categorical			Normalized		
	Original	Bootstrap		Original	Bootstrap		Original	Bootstrap	
	β	β (mean)	Frequency (%)	β	β (mean)	Frequency (%)	β	β (mean)	Frequency (%)
Intercept	−27.830	−30.921	95.7	−2.157	−2.343	88.2	−2.849	−3.276	95.3
Gender	0.133	0.168	8.5	0.296	0.326	13.4	0.133	0.168	8.5
Age	−0.657	−0.712	41.7	−0.789	−0.849	53.1	−0.657	−0.712	41.7
Migration	0.758	0.826	47.4	0.836	0.905	54.8	0.758	0.826	47.4
Tenderness	0.471	0.531	20.1	0.265	0.302	12.6	0.471	0.531	20.1
Rebound tenderness	0.245	0.264	9.6	0.190	0.206	8.8	0.245	0.264	9.5
Muscular guarding	0.591	0.648	37.0	0.539	0.589	30.6	0.591	0.648	37.0
BT	0.698	0.776	93.3	1.004	1.093	76.9	3.071	3.417	93.3
WBC	0.048	0.051	18.2	0.687	0.745	33.3	1.021	1.096	18.3
CRP	0.068	0.074	64.3	1.248	1.339	89.7	2.496	2.723	64.3

The logistic regression models were constructed for original data (Original) and each bootstrap sample (Bootstrap). A summary of the results is shown

Frequency Frequency that each correlation coefficient was statistically significant ($p<0.05$), *BT* body temperature, *WBC* white blood cell count, *CRP* C-reactive protein, *Continuous* model constructed by using continuous variables, *Categorical* model constructed by using categorical variables, *Normalized* model constructed by using normalized variables

Table 3 Comparison of error rate and AUC between logistic regression analysis and ANN

	Error rate		AUC		Cutoffs at 95% specificity		Cutoffs at 95% sensitivity	
	Logistic	ANN	Logistic (mean±SD)	ANN[a] (mean±SD)	Logistic (mean±SD)	ANN (mean±SD)	Logistic (mean±SD)	ANN (mean±SD)
Continuous	0.317	0.307	0.719±0.054	0.715±0.056	0.804±0.015	0.803±0.014	0.214±0.013	0.215±0.014
Categorical	0.337	0.304	0.718±0.056	0.736±0.051*	0.813±0.010	0.876±0.008*	0.197±0.014	0.104±0.008*
Normalized	0.317	0.302	0.719±0.054	0.741±0.052*	0.804±0.015	0.785±0.013	0.214±0.013	0.199±0.009

The logistic regression models (Logistic) and the artificial network models (ANN) were applied to samples that were not used in the model construction. The error rate, the mean area under the receiver operating characteristics curve (AUC) and the cutoff points at 95% specificity and sensitivity for each model are shown. Variables were used in both analyses in three ways of transformation, i.e., original variables (Continuous), variables categorized by cutoff points (Categorical) and variables normalized in the range 0 to 1 (Normalized)
*There were significant ($p<0.05$) differences between logistic regression model and ANN model
[a] There were significant differences among models

summary of the constructed logistic regression models. BT and CRP were frequently obtained as significant variables in the constructed models.

Comparison of the levels of accuracy of the logistic regression models and ANN models

The levels of accuracy in the logistic regression models and the ANN models were compared for diagnosis by using the error rate (Table 3). In each type of variable, the ANN models tended to yield better results than the results of the logistic regression models. Among the three types of variables that were used, the normalized variable model showed the lowest error rate.

The accuracy of the models was evaluated by accurate diagnosis rates (Fig. 3). The accurate diagnosis rates were calculated by dividing the number of accurately diagnosed subjects by the number of all subjects only in validation samples. As in the case of the error rates, the normalized variable model showed the most favorable results.

On the other hand, the values of AUC were significantly higher in the ANN models than in the logistic regression models when categorical variables or normalized variables were used for analyses (Table 3). There was no significant difference between the ANN model and the logistic regression model when continuous variables were used. The cutoff points at 95% specificity and 95% sensitivity were also shown in Table 3. The range between the cutoff points at 95% specificity and that at 95% sensitivity were wide in ANN model using normalized variables. The level of accuracy of the ANN model using normalized variables was better than the levels of accuracy of other models.

The actual data on the 169 patients was then entered into the constructed models (Table 4). Although an accurate diagnosis was made for 86/169 (50.9%) patients on a clinical basis, the neural network model using normalized variables gave the accurate diagnosis in 127/169 (75.1%) patients. As for the operated patients, 90/98 (91.8%) patients were diagnosed correctly by the neural network model using normalized variables, and this finding was superior to the results of the clinical diagnosis (86/98, 87.8%).

Clinical characteristics of the patients in whom normal appendices were resected

There were no significant differences between the clinical characteristics of the patients who did not undergo abdominal operations and the patients in whom normal appendices were resected (data not shown).

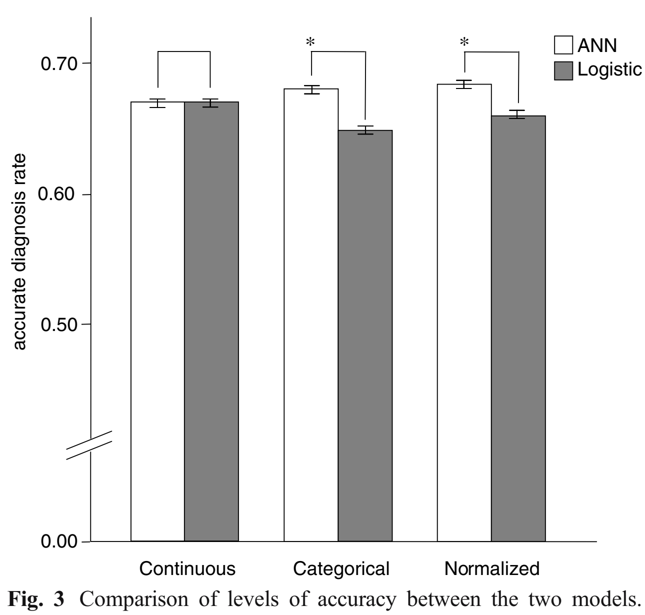

Fig. 3 Comparison of levels of accuracy between the two models. Samples for validation were analyzed by the artificial neural network model and the logistic regression model. *ANN* Artificial neural network model, *Logistic* logistic regression model, *Continuous* model constructed by using continuous variables, *Categorical* model constructed by using categorical variables, *Normalized* model constructed by using normalized variables. *$P<0.001$

Table 4 Performance of the constructed models

Subjects	Constructed models	Variables	Specificity	Sensitivity	Positive predictive value	Negative predictive value	AUC
Operated subjects	Clinical diagnosis		0/12 (0.0%)	86/86 (100.0%)	86/98 (87.8%)	0/0	
	Logistic regression	Continuous	7/12 (58.3%)	59/86 (68.6%)	59/64 (92.2%)	7/34 (20.6%)	0.735
		Categorical	6/12 (50.0%)	62/86 (72.1%)	62/68 (91.2%)	6/30 (20.0%)	0.700
		Normalized	12/12 (100.0%)	43/86 (50.0%)	43/43 (100.0%)	12/55 (21.8%)	0.720
	Neural network	Continuous	10/12 (83.3%)	54/86 (62.8%)	54/56 (96.4%)	10/42 (23.8%)	0.684
		Categorical	3/12 (25.0%)	80/86 (93.0%)	80/89 (89.9%)	3/9 (33.3%)	0.710
		Normalized	9/12 (75.0%)	81/86 (94.2%)	81/84 (96.4%)	9/14 (64.3%)	0.754
All subjects	Clinical diagnosis		0/83 (0.0%)	86/86 (100%)	86/169 (50.9%)	0/0	
	Logistic regression	Continuous	64/83 (77.1%)	59/86 (68.6%)	59/78 (75.6%)	64/91 (70.3%)	0.788
		Categorical	56/83 (67.5%)	62/86 (72.1%)	62/89 (69.7%)	56/80 (70.0%)	0.788
		Normalized	77/83 (92.8%)	43/86 (50.0%)	43/49 (87.8%)	77/120 (64.2%)	0.774
	Neural network	Continuous	74/83 (89.2%)	51/86 (59.3%)	51/60 (85.0%)	74/109 (67.9%)	0.781
		Categorical	68/83 (81.9%)	54/86 (62.8%)	54/69 (78.3%)	68/100 (68.0%)	0.775
		Normalized	61/83 (73.5%)	66/86 (76.7%)	66/88 (75.0%)	61/81 (75.3%)	0.801

The actual data on 98 operated patients (upper part of the table) and all 169 patients (lower) were entered into the models, and the analyzed results are shown

Discussion

There results showed that the level of accuracy of the ANN models was better than that of the logistic regression models for the diagnosis of acute appendicitis. The ANN model provided a more accurate diagnosis when normalized variables were used than when continuous variables or categorical variables were used. The performance of the ANN model and logistic regression model is almost equal in the analysis of a large sample size, but the performance of the models is often different between models when only a smaller number of samples is analyzed [18]. Since the sample size in the present study was relatively small, the data structure in this study might thus be more suited for a ANN model than a logistic regression model. In addition the level of CRP was another important predictor for an accurate diagnosis of acute appendicitis.

The level of CRP was found to be a predictive variable, although previous studies have shown that WBC was the best laboratory test, especially early in the course of acute appendicitis [19, 20]. In other studies, CRP may not have been used in the diagnosis of acute appendicitis [3–8]. However, the present results show that CRP was a significant predictor in the diagnosis of acute appendicitis, especially when the logistic regression model was used. One possible reason for this discrepancy is the timing of the laboratory investigation. CRP rises 12 to 24 h after the onset of acute appendicitis when other symptoms have appeared [21]. On the other hand, WBC rises early in the course of acute appendicitis [19, 20]. Since this hospital is a secondary or tertiary care hospital, the timing of the laboratory investigation tends to be late compared with that in a primary care hospital. In fact, 91 (53.1%) of the patients in this study were admitted to our hospital from other clinics or hospitals. Therefore, CRP seems to be an effective variable when the timing of laboratory investigation is late enough. To use these models in a primary care setting, an improvement of this model is therefore required, especially regarding how the CRP values are used as a predictive variable.

When the number of parameters is large in relation to the available sample size in ANN models, close correspondence between a dependent variable and independent variables does not necessarily imply that the independent variables would be a good predictor for the dependent variable [22]. To resolve this problem, which is known as overfitting, the methodology that was used was based on a meta-analysis [1].

In the ANN model, normalization of data ranging from 0 to 1 is essential to prevent the premature saturation of hidden nodes, which impedes the learning process [23]. In this study, the ANN model using normalized variables gave the best accuracy of all models (Table 3). This result suggests that the ANN model using normalized variables accurately reflected the influence of each variable on diagnosis.

When continuous values were divided into categories in the logistic regression model, the type I error (false negative) rate increased [24] and necessary information might be discarded [25]. In accordance with the results of those studies, the error rate increased in the logistic regression model when categorical variables were used, in the present study. However, the ANN model using categorical variables showed a low error rate near using normalized variables (Table 3). This result suggests that the ANN model could minimize the bias by categorization.

There are two limitations in this study. First, it was performed in a single institution. These findings must be validated in a multi-center study. Second, the proportion of subjects with tenderness at the right lower quadrant was smaller than the proportions in other studies. In other studies, this symptom was a significant variable for diagnosis of acute appendicitis [3–7]. However, it was not a significant variable in the present study ($p=0.850$). Since this hospital is a secondary or tertiary care hospital, the symptoms of the patients might be more diverse and ambiguous in comparison to the symptoms of the patients in a primary care hospital.

Finally, the diagnostic accuracy of the ANN model in the present study was 75.1%. The ability is far from a level that would permit its reliable use in a clinical setting, although it was much better than the initial diagnosis which was only based on the clinical and laboratory findings. The above findings therefore suggest that a precise diagnosis of acute appendicitis therefore cannot be made without the use of imaging examinations. Lee et al. [26] stressed the clinical benefit of computed tomography (CT) of the abdomen based on their clinical trial. In fact, in the present series, 96 of 169 patients (56.8%) underwent abdominal CT scans. The efficacy of CT scans for the diagnosis of acute appendicitis has been widely recognized although the increased medical cost is now also becoming a major issue that needs to be resolved. The next issue that must be addressed is whether ANN models can efficiently select patients who truly need refined imaging examinations, thereby helping to reduce the medical costs involved in the management of patients suspected of having acute appendicitis.

Acknowledgements We thank Ms. Akane Inaizumi, Ms. Atsuko Sugiyama and Mr. Toshikazu Abe for valuable assistance in the acquisition of the data and references. We also thank Dr. Ralph Grams for his kind suggestions.

References

1. Andersson, R. E., Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br. J. Surg.* 91:28–37, 2004.
2. Eldar, S., Nash, E., Sabo, E., Matter, I., Kunin, J., Mogilner, J. G., and Abrahamson, J., Delay of surgery in acute appendicitis. *Am. J. Surg.* 173:194–198, 1997.
3. Arnbjornsson, E., Scoring system for computer-aided diagnosis of acute appendicitis. The value of prospective versus retrospective studies. *Ann. Chir. Gynaecol.* 74:159–166, 1985.
4. Eskelinen, M., Ikonen, J., and Lipponen, P., A computer-based diagnostic score to aid in diagnosis of acute appendicitis. A prospective study of 1333 patients with acute abdominal pain. *Theor. Surg.* 7:86–90, 1992.
5. Alvarado, A., A practical score for the early diagnosis of acute appendicitis. *Ann. Emerg. Med.* 15:557–564, 1986.
6. Ohmann, C., Franke, C., and Yang, Q., Clinical benefit of a diagnostic score for appendicitis: results of a prospective interventional study. German Study Group of Acute Abdominal Pain. *Arch. Surg.* 134:993–996, 1999.
7. Tzanakis, N. E., Efstathiou, S. P., Danulidis, K., et al., A new approach to accurate diagnosis of acute appendicitis. *World J. Surg.* 29:1151–1156, 2005.
8. Kharbanda, A. B., Taylor, G. A., Fishman, S. J., and Bachur, R. G., A clinical decision rule to identify children at low risk for appendicitis. *Pediatrics* 116:709–716, 2005.
9. Hosseini, H. G., Luo, D., and Reynolds, K. J., The comparison of different feed forward ANN architectures for ECG signal diagnosis. *Med. Eng. Phys.* 28:372–378, 2006.
10. Ottenbacher, K. J., Linn, R. T., Smith, P. M., Illig, S. B., Mancuso, M., and Granger, C. V., Comparison of logistic regression and ANN analysis applied to predicting living setting after hip fracture. *Ann. Epidemiol.* 14:551–559, 2004.
11. Baxt, W. G., Shofer, F. S., Sites, F. D., and Hollander, J. E., A neural computational aid to the diagnosis of acute myocardial infarction. *Ann. Emerg. Med.* 39:366–373, 2002.
12. Pesonen, E., Eskelinen, M., and Juhola, M., Comparison of different neural network algorithms in the diagnosis of acute appendicitis. *Int. J. Biomed. Comput.* 40:227–233, 1996.
13. Efron, B., and Tibshirani, R., Improvement on cross-validation: The .632+ bootstrap method. *J. Am. Stat. Assoc.* 92:548–560, 1997.
14. Sivit, C. J., Siegel, M. J., Applegate, K. E., and Newman, K. D., When appendicitis is suspected in children. *Radiographics* 21:247–262, 2001.
15. Wehberg, S., and Schumacher, M., A comparison of nonparametric error rate estimation methods in classification problems. *Biom. J.* 46:35–47, 2004.
16. Hanley, J. A., and McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36, 1982.

17. Zweig, M. H., and Campbell, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39(4):561–577, 1993.
18. Bartfay, E., Mackillop, W. J., and Pater, J. L., Comparing the predictive value of neural network models to logistic regression models on the risk of death for small-cell lung cancer patients. *Eur. J. Cancer Care* 15:115–124, 2006.
19. Marchand, A., Van Lente, F., and Galen, R. S. The assessment of laboratory tests in the diagnosis of acute appendicitis. *Am. J. Clin. Pathol.* 80:369–374, 1983.
20. Gronroos, J. M., Forsstrom, J. J., Irjala, K., and Nevalainen, T. J., Phospholipase A2, C-reactive protein, and white blood cell count in the diagnosis of acute appendicitis. *Clin. Chem.* 40:1757–1760, 1994.
21. Clyne, B., and Olshaker, J. S., The C-reactive protein. *J. Emerg. Med.* 17:1019–1025, 1999.
22. Hua, J., Lowey, J., Xiong, Z., and Dougherty, E. R., Noise-injected neural networks show promise for use on small-sample expression data. *BMC Bioinformatics* 7:274, 2006.
23. Basheer, I. A., and Hajmeer, M., Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* 43:3–31, 2000.
24. Austin, P. C., and Brunner, L. J., Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat. Med.* 23:1159–1178, 2004.
25. Altman, D. G., Lausen, B., Sauerbrei, W., and Schumachar, M., Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J. Natl. Cancer Inst.* 86:829–835, 1994.
26. Lee, C. C., Golub, R., Singer, A. J., Cantu, R., Jr., and Levinson, H., Routine versus selective abdominal computed tomography scan in the evaluation of right lower quadrant pain: a randomized controlled trial. *Acad. Emerg. Med.* 14:117–122, 2007.