Acute appendicitis diagnosis using artificial neural networks

Sung Yun Park and Sung Min Kim*
Department of Medical Biotechnology, Dongguk University-Seoul, Seoul, Korea

Abstract.
BACKGROUND: Artificial neural networks is one of pattern analyzer method which are rapidly applied on a bio-medical field.
OBJECTIVE: The aim of this research was to propose an appendicitis diagnosis system using artificial neural networks (ANNs).
METHODS: Data from 801 patients of the university hospital in Dongguk were used to construct artificial neural networks for diagnosing appendicitis and acute appendicitis. A radial basis function neural network structure (RBF), a multilayer neural network structure (MLNN), and a probabilistic neural network structure (PNN) were used for artificial neural network models. The Alvarado clinical scoring system was used for comparison with the ANNs.
RESULTS: The accuracy of the RBF, PNN, MLNN, and Alvarado was 99.80%, 99.41%, 97.84%, and 72.19%, respectively. The area under ROC (receiver operating characteristic) curve of RBF, PNN, MLNN, and Alvarado was 0.998, 0.993, 0.985, and 0.633, respectively.
CONCLUSIONS: The proposed models using ANNs for diagnosing appendicitis showed good performances, and were significantly better than the Alvarado clinical scoring system ($p < 0.001$). With cooperation among facilities, the accuracy for diagnosing this serious health condition can be improved.

Keywords: Alvarado clinical scoring system, acute appendicitis, clinical scoring system, artificial neural network

1. Introduction

Acute appendicitis is one of the most common surgical emergencies involving the abdomen. The incidence of appendicitis is approximately 1.1–1.9 per 1,000 in people of all ages [1–3]. Acute appendicitis is treated by surgical diagnosis, and an incorrect or late diagnosis is very commonly associated with morbidity and mortality. Therefore, diagnostic imprecision can result in a high wound infection rate, high perforation rate, and high negative laparotomy rate [3].

Various methods for the diagnosis of appendicitis and acute appendicitis have been developed to increase the speed and accuracy, and to decrease negative laparotomies. In 1986, Alvarado proposed a clinical scoring system, which is called the Alvarado clinical scoring system, to diagnose acute appendicitis. This system consists of clinical symptoms, signs, and laboratory findings [5], and has shown good performances in suspected appendicitis patients. Recently, computerized tomography (CT), and ultrasound have proven to be accurate imaging techniques for diagnosing acute appendicitis [5–8]. A method that combines the clinical scoring system and imaging technique has shown better performance than when the clinical scoring system is used alone.

---

*Corresponding author: Sung Min Kim, Department of Medical Biotechnology, Dongguk University-Seoul, 26, Pil-dong 3-ga, Jung-gu, Seoul, Korea. E-mail: smkim@dongguk.edu.
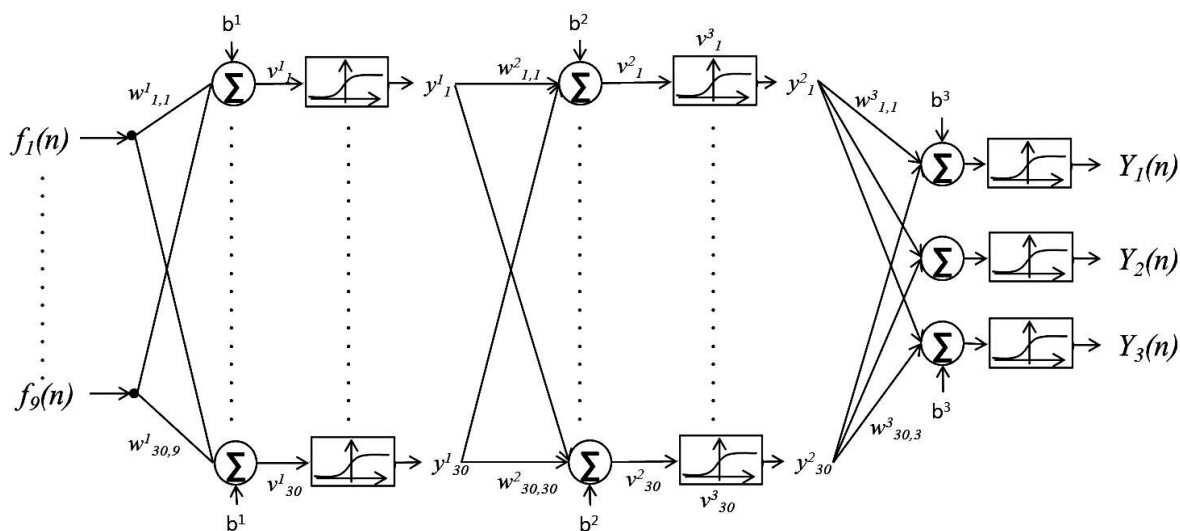
Fig. 1. Multilayer neural network structure with one input layer, two hidden layers, and one output layer.

Critical weak points of the established methods for diagnosing acute appendicitis have been investigated in literature [9–12], and it has been found that the Alvarado clinical scoring system does not always provide stable results (58.5–91.7% for sensitivity, and 78.7–93.3% for accuracy). As such, some researchers have proposed new clinical scoring systems (e.g. the Eskelinen and Lintula scoring systems), or a modified Alvarado scoring system suitable for their individual clinical environments [3,5,10,12]. CT imaging is associated with exposure to radiation, and the examination is costly. Also, the quality of ultrasound imaging is highly dependent on the operator [8,13]. Another thing to consider with both imaging technique methods is that they are not always feasible, especially during off hours.

Artificial neural networks (ANNs) are nonlinear computational mathematics techniques which are developed from a biological neural system [14]. ANNs are useful because of the complex interconnected structure of each input data, and the learning systems by training algorithms based on the input-data. With these properties, ANNs have been widely used and have shown a good diagnostic rate. The purpose of this study was to propose ANN models for diagnosing acute appendicitis. The performance of the ANNs was compared to that of clinical scoring systems, i.e. the Alvarado.

2. Materials and methods

2.1. Patients

Patients who came to the emergency center of Dongguk University Hospital with abdominal pain between August 2011 and February 2012 were enrolled in this research. Experienced surgeons diagnosed the appendicitis and classified the patients. The categories of the diagnostic case report form were set by the WOGRC (World Organization of Gastroenterology Research Committee). The clinical surgeons were trained in uniform clinical protocols to maintain the quality of diagnoses of abdominal pain. Patients suffering from abdominal pain were interviewed and physically examined, followed by the completion of a diagnostic case report form, including the Alvarado clinical scoring system. Afterwards, patients with suspected appendicitis were taken for ultrasound or MRI. Then, patients with

Table 1
Clinical and laboratorial factors and their input for ANNS

Variables	Type
Pain location	Binary
Migration of RLQ	Binary
Tenderness of RLQ	Binary
Rebound tenderness of RLQ	Binary
Bowel sound	Binary
Nausea	Binary
Vomiting	Binary
Body temperature	Continuous
WBC counts	Continuous

suspected acute appendicitis underwent surgery to remove the inflamed appendix. Finally, all of the patients were allocated to one of three categories; Group 1 was the No Appendicitis patients (NA), Group 2 was Normal Appendicitis patients (NorA), and Group 3 was the Acute Appendicitis patients (AA). All protocols of this study were approved by the IRB (Institutional Review Board) of Dongguk University Ilsan Hospital.

2.2. Artificial neural networks

The ANNs consisted of three layers: an input layer, a hidden layer, and an output layer. The input layer had nine features, as shown in Table 1. The output layer consisted of indices of three classes (NA, NorA, and AA). The hidden layer had a different structure, as types of ANN structures. Three types of ANNs structures were used; a multilayer neural network structure (MLNN), probabilistic neural network structure (PNN), and radial basis function neural network structure (RBF) [15–17]. In this study, the MLNN consisted of two hidden layers and had thirty-one neurons for each hidden layer. The number of neurons was optimized by the value of the mean squared errors (MSE). For example, a structure of MLNN with one-input layer, two hidden layers, and one output layer were constructed. Each layer, except the input layer, consisted of a summation part and an activation part. In this study, the summation function was a linear type, as in Eq. (1), and the activation function was a nonlinear sigmoid type, as in Eq. (2) [14,15].

$$v_i^1 = \left(\sum_{j=1}^{30} w_{j,i}^1 \times f_i \right) + b^1 \tag{1}$$

$$y_j^1 = \left(1 + \exp\left(-v_i^1 \right) \right)^{-1} \tag{2}$$

where v_i^1 is the output of summation of the ith neuron, $w_{j,i}^1$ is the weight factor of the first hidden layer, b^1 is a bias of the first hidden layer, and y_j^1 is the output of activation function of jth neuron in the first hidden layer. The hidden layers and output layer used a linear summing junction, and nonlinear sigmoid activation function, respectively. A MLNN was successfully used for the diagnostic prediction of appendicitis. To train the MLNN structure, a back propagation (BP) and Levenberg-Marquardt (LM) algorithm were used, which are widely recognized [18].

The RBF had one hidden layer, called a radial basis layer, and was defined by Gaussian function. The equations are shown in Eqs (3)–(6).

$$x_j = \phi \left(\| \vec{f} - \vec{c}_j \| * b^{ih} \right) \tag{3}$$

$$\phi(x) = \exp(-x^2) \tag{4}$$

$$b^{ih} = 0.833/\text{s} \tag{5}$$

$$Y_i = \sum\nolimits_{j=1}^{h} w_{ji}^{h_0} * x_j \tag{6}$$

where $i = 1, 2, 3, j = 1, 2, 3, \ldots, h, Y_i$ is the ith output, $\vec{f}$ is the 9-dimentional real value input vector, W^{h0} are the output layer weights, $\vec{c_j}$ is the center vector of the jth hidden node, s is the real constant known as spread factor, $b_{i\text{th}}$ is the biasing term of radical basis layer, and $\varphi(.)$ is the nonlinear radial basis function [21].

The PNN also had one hidden layer, also called a radial basis layer, and was defined by Gaussian function; and the output layer of the PNN consisted of a Gaussian function and a competitive activation function instead of a nonlinear sigmoid activation function. The equations are shown as Eqs (7)–(11).

$$x_j = \phi\left(\|\vec{f} - \vec{c_j}\| * b^{ih}\right) \tag{7}$$

$$\phi(x) = \exp(-x^2) \tag{8}$$

$$b^{ih} = 0.833/\text{s} \tag{9}$$

$$s_i = \sum_{j=1}^{h} w_{ji}^{h_0} * x_j \tag{10}$$

$$Y_i = \begin{cases} 1, & \text{if } S_i \text{ is max of } \{S_1, S_2, S_3\} \\ 0, & \text{else} \end{cases} \tag{11}$$

where $i = 1, 2, 3, j = 1, 2, 3, \ldots, h, Y_i$ is the ith output, $\vec{f}$ is the 9-dimentional real valued input vector, W^{h0} are the output layer weights, $\vec{c_j}$ is the center vector of the jth hidden node, s is the real constant known as spread factor, $b^{i\text{th}}$ is the biasing term of radical basis layer, and $\varphi(.)$ is the nonlinear radial basis function [21]. MATLAB (Mathworks, MA) was used to set up the ANN structures.

2.3. Statistical analysis

All analyses were performed using a commercial statistical program (SPSS for Windows 18, SPSS, Chicago, IL, USA). The significance of the difference in parameters was assessed by Wilcoxon's rank-sum test among groups, Kruskal-Wall test for continuous variables, and chi-squared test for categorical variables. $P < 0.05$ was considered a statistically significant difference. The performance of methods was measured by confusion matrix, ROC (receiver operating characteristic), and AUC (area under the ROC curve). The ROC curve was a plot of sensitivity vs. specificity. The ROC curve rose rapidly toward the upper right-hand corner of the graph if the test method indicated a successful model for the diagnosis tool. The AUC ranged from 0 for poor performance to 1 for perfect performance. The classification accuracy was measured using Eq. (12):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

Table 2
Statistical analysis among groups with physical signs

	I. No appendicitis (n = 596)	II. Appendicitis (n = 205)		P§	$P_{I,II}$†	$P_{I,III}$†	$P_{I,IV}$†	$P_{II,IV}$†
		III. Normal appendicitis (n = 143)	IV. Acute appenditis (n = 62)					
Age (years)	29.68 ± 13.63	30.55 ± 13.99	31.53 ± 16.32	0.475	0.868	0.235	0.527	0.313
				P‡	P_{I}‡	P_{II}‡	P_{III}‡	P_{IV}‡
Gender (Male/Female)	290/306	53/90	32/30	0.015	< 0.001	0.414	0.732	0.394

Data was expressed as mean ± standard deviation on age. † Wilcoxon's rank-sum test, § Kruskal-Wallis test, and ‡ χ2-test.

Table 3
Comparison of performances between clinical scoring system and artificial neural networks

Methods of diagnosis		Specificity [%]	Sensitivity [%]	Positive predictive value [%]	Negative predictive value [%]	Accuracy [%]
Clinical scoring system	Alvarado*	87.42	23.16	43.16	77.88	72.19
ANNs	MLNN†	96.62	99.46	94.82	99.68	97.84
	RBF†	99.69	100	99.46	100	99.80
	PNN†	99.08	100	98.40	100	99.41

*$P < 0.001$, Alvarado vs. ANNs; † No significant difference ($P = 0.697$), among ANNs.

where FN, FP, TN, and TP are false negatives, false positives, true negatives, and true positives, respectively. The sensitivity and specificity were measured using Eqs (13) and (14).

$$Sensitivity = \frac{TP}{TP + FN}(\%) \tag{13}$$

$$Specificity = \frac{TN}{FP + TN}(\%) \tag{14}$$

3. Experimental results

A total of 801 patients aged 30.27 ± 18.58 years (mean ± SD (standard deviation)) were enrolled from August 2011 to February 2012, as seen in Table 2. The NA group consisted of 596 patients (290 males and 306 females), aged 29.68 ± 13.63 years. The total of the appendicitis groups, which was made up of the normal appendicitis (NorA) group and acute appendicitis (AA) group, consisted of 205 patients, including 143 patients (53 males and 90 females) in the NorA group (aged 30.55 ± 13.99 years) and 62 patients (32 males and 30 females) in the AA group (aged 31.53 ± 16.32 years). There was no age significant difference among the NA group, NorA group, and AA group.

The sensitivity, specificity, negative predictive value, positive predictive value, and accuracy of the Alvarado clinical scoring system were 28.16%, 87.42%, 77.88%, 43.61%, and 72.19%, respectively, as can be seen in Table 3. All the values of the ANN methods were over 90%, and the sensitivity and negative predictive values, in particular, were 100% for both the RBF and PNN. The accuracy of RBF (99.80%) was a higher value than for the other methods (97.84% for MLNN, and 99.41% for PNN). All the parameters of the MLNN had lower values than the other methods (RBF and PNN), but the difference was not significant between them ($P = 0.697$). The AUC of the Alvarado clinical scoring system, MLNN, RBF, and PNN was 0.633, 0.985, 0.998, and 0.993, respectively, as shown in Table 4.

Table 4
AUC values of the clinical scoring system and artificial neural networks

Method of diagnosis		AUC
Clinical scoring system	Alvarado	0.633
ANNs	MLNN	0.985
	RBF	0.998
	PNN	0.993

4. Discussion

Acute appendicitis is one of the most difficult diseases to diagnose clinically, even for experienced surgeons. Several methods, such as the clinical scoring system and laboratory tests, have been used to increase the accuracy of appendicitis diagnoses. However, these methods have not shown stable performances, including accuracy, sensitivity, and specificity, due to complex clinical protocols. The Alvarado clinical scoring system has shown good results in previous studies, especially in the West, whereas the results of this study showed poor values (28.16%, 87.42%, 77.88%, 43.61%, and 72.19% for sensitivity, specificity, negative predictive value, positive predictive value, and accuracy, respectively). The purpose of this study, therefore, was to develop a simple diagnosis method for appendicitis using artificial neural networks (ANNs).

Saki et al. [19] in Japan created an ANN construct (0.741 for AUC) for appendicitis diagnosis, with 169 patients (86 patients with appendicitis and 83 patients without appendicitis). Hsieh et al. [11] in Taiwan used an ANN construct with 180 patients (116 patients with appendicitis and 65 patients without appendicitis), which performed well (94% for sensitivity and 85% for specificity). Some researchers in western countries have reported new clinical diagnosis methods by using ANNs, instead of the Alvarado clinical scoring system, and image analysis [20]. Prabhudesai et al. [20] in the United Kingdom used an MLNN structure of ANNs with 60 patients (24 patients with appendicitis and 36 patients without appendicitis), and had high-performance results (100% for sensitivity, and 97.2% for specificity).

In this study, various structures (MLNN, RBF, and PNN) of ANNs were used with 801 patients (205 patients with appendicitis, 596 patients without appendicitis). All structures of the ANNs showed significantly higher performances compared to the Alvarado clinical scoring system ($P < 0.001$). However, Prabhudesai et al. [20] and Hsieh et al. [11] did not show significant differences ($P = 0.063$, and $P = 0.223$, respectively), even though they performed their studies in Asia. Prabhudesai et al. [20] explained this result as having small subject groups, and Hsieh et al. [18] also suggested that it might be the number of patients or incomplete ANNs. The number of patients in this study was considerably greater than in previous studies. ANN is still one of pattern analysis, and its performance in the application phase is highly dependent on how much experience is gained in the training phases. This implies that developed ANNs can obtain stable performances in various environments.

Image analyses of ultrasound and computer tomography (CT) play a very useful role in the diagnoses of acute appendicitis, and show excellent performances [7,8,13]. However, the use of CT has to be considered in regard to radiation exposure, high cost, and operating time. The quality of analytic ultrasound images is highly dependent upon the operator, but without the disadvantages of CT. Despite these disadvantages, image analysis is currently one of the primary tools for the diagnosis of acute appendicitis.

5. Conclusions

ANN structures were developed in this study, and showed impressive performances for the diagnoses of appendicitis compared to the Alvarado clinical scoring system. The developed models have the poten-

tial to help junior clinical surgeons diagnose appendicitis. And since nearly all of the input data of ANNs use objective values, developed ANNs have the potential to show good performances in other regions. However, further studies are needed to develop the ANNs using data from other districts or countries.

Acknowledgment

References

[1]   Chong CF, Adi MI, Thien A, Suyoi A, Mackie AJ, Tin AS, et al. Development of the RIPASA score: a new appendicitis scoring system for the diagnosis of acute appendicitis. Singapore Med J. 2010; 51(3): 220.
[2]   Lintula H, Kokki H, Kettunen R, Eskelinen M. Appendicitis score for children with suspected appendicitis. A randomized clinical trial. Langenbecks Arch Surg. 2009; 394(6): 999.
[3]   Lintula H, Kokki H, Pulkkinen J, Kettunen R, Gröhn O, Eskelinen M. Diagnostic score in acute appendicitis. Validation of a diagnostic score (Lintula score) for adults with suspected appendicitis. Langenbecks Arch Surg. 2010; 395(5): 495.
[4]   Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. Arch Surg. 2002; 137(7): 799.
[5]   Alvarado A. A practical score for the early diagnosis of acute appendicitis. Ann Emerg Med. 1986; 15(5): 557.
[6]   Paulson EK, Coursey CA. CT protocols for acute appendicitis: time for change. AJR. 2009; 193(5): 1268.
[7]   Poortman P, Lohle PN, Schoemarker CM, Cuesta MA, Oostvogel HJ, de Lange-de Klerk ES, et al. Improving the false-negative rate of CT in acute appendicitis-reassessment of CT images by body imaging radiologists: a blinded prospective study. Eur J Radiol. 2010; 74(1): 67.
[8]   Israel GM, Malguria N, McCarthy S, Copel J, Weinreb J. MRI vs. ultrasound for suspected appendicitis during pregnancy. J Magn Reson Imaging. 2008; 28(2): 428.
[9]   Hawkins JD. Thirlby RC. The accuracy and role of cross-sectional imaging in the diagnosis of acute appendicitis. Adv Surg. 2009; 43: 13.
[10] Kalan M, Talbot D, Cunliffe WJ, Rich AJ. Evaluation of the modified Alvarado score in the diagnosis of acute appendicitis: a prospective study. Ann R Coll Surg Engl. 1994; 76(6): 418.
[11] Hsieh CH, Lu RH, Lee NH, Chiu WT, Hsu MH, Li YC. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery. 2011; 149(1): 87.
[12] Konan A, Hayran M, Kilic YA, Karakoc D, Kaynaroglu V. Scoring systems in the diagnosis of acute appendicitis in the elderly. Ulus Travma Acil Cerrahi Derg. 2011; 17(5): 396.
[13] John H, Neff U, Kelemen M. Appendicitis diagnosis today: clinical and ultrasonic deductions. World J Surg. 1993; 17(2): 243.
[14] Zhou ZH, Jiang Y, Yang YB, Chen SF. Lung cancer cell identification based on artificial neural network ensembles. Artif Intell Med. 2002; 24(1): 25.
[15] Temurtas F, A comparative study on thyroid disease diagnosis using neural networks. Expert Sys. Appl. 2009; 36(1): 944.
[16] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artif Intell Med. 2005; 34(2): 113.
[17] Wu D, Warwick K, Ma Z, Gasson MN, Burgess JG, Pan S, Aziz TZ. Prediction of Parkinson's disease tremor onset using a radial basis function neural network based particle swarm optimization. Int J Neural Sys. 2010; 20(2): 109.
[18] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. Massachusetts: MIT Press Cambridge; 1986. p. 318.
[19] 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(6): 723.
[20] 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(2): 310.
[21] Er O, Yumusak N, Temurtas F. Chest diseases diagnosis using artificial neural networks. Expert Sys. Appl. 2010; 37(12): 7648.
[22] Teicher I, Landa B, Cohen N, Kabnick LS, Wise L. Scoring system to aid in diagnoses of appendicitis. Ann Surg. 1983; 198(6): 753.