Application of an Artificial Intelligence Method for Diagnosing Acute Appendicitis: The Support Vector Machine

Sung Yun Park[1], Jun Seok Seo[2], Seung Chul Lee[2], and Sung Min Kim[1,*]

[1] Department of Medical Bio Engineering, Dongguk University-Seoul, Seoul, Republic of Korea
{sypark,smkim}@dongguk.edu
[2] Department of Emergency Medicine, Dongguk University Ilsan Hospital, Dongguk University-Seoul, Seoul, Republic of Korea
{drsjs,edlee}@dumc.or.kr

Abstract. The aim of this study is to suggest an artificial intelligence model to diagnosis acute appendicitis using a support vector machine (SVM). Acute appendicitis is one of the most common abdominal surgery emergencies. Various methods have been developed to diagnose appendicitis, but they have not performed well in the Middle East, Asia, or the West. A total of 760 patients were used to construct the SVM. Both the Alvarado clinical scoring system (ACSS) and multilayer neural networks (MLNN) were used to compare performance. The accuracies of the ACSS, MLNN, and SVM were 54.87%, 92.89, and 99.61%, respectively. The areas under the curve of ACSS, MLNN, and SVM were 0.621, 0.969, and 0.997 respectively. The performance of the AI model was significantly better than that of the ACSS ($P < 0.001$). We consider that the developed models are a useful method to reduce both negative appendectomies and delayed diagnoses, particularly for junior clinical surgeons.

Keywords: appendicitis, artificial intelligence, support vector machine, clinical scoring system, a receiver operating characteristics graph.

1 Introduction

Acute appendicitis is one of the most common surgical emergencies of the abdomen. The lifetime incidence of acute appendicitis is approximately 7%, and acute appendicitis is clearly treated by a surgical diagnosis [1, 2, 3]. An early diagnosis of suspected appendicitis is important for treating acute cases, as a missed or delayed acute appendicitis diagnosis is associated with high morbidity and mortality. Diagnostic imprecision can result in a high wound infection rate, high perforation rate, and high negative laparotomy rate, which ranges from 20–30% [3, 4].

Several clinical methods for early and correct diagnosis of acute appendicitis have been suggested and developed to increase diagnostic accuracy and to decrease

* Corresponding author.

James J. (Jong Hyuk) Park et al. (eds.), *Future Information Technology*, Lecture Notes in Electrical Engineering 276,
DOI: 10.1007/978-3-642-40861-8_13,

negative laparotomies [4, 5, 6]. In 1986, Alvarado suggested a clinical scoring system consisting of signs, symptoms, and laboratory findings, and several clinical scoring systems have been developed and modified based on Alvarado's clinical scoring system (ACSS) [4]. However, several researchers have shown that the performance of these clinical scoring systems is insufficient for diagnosis. Image analysis methods including computed tomography (CT), and ultrasound (US) have significantly higher performance than other diagnostic methods, but they have some disadvantages [2, 5]. The quality of a CT image is highly related to radiation exposure and the diagnostic performance of US is highly dependent on the operator and cannot be used during off-hours. Moreover, the image analysis method occasionally becomes the cause for a delayed diagnosis of acute appendicitis.

More recently, artificial intelligence (AI) methods have been applied to diagnose or predict disease [7, 8, 9, 10]. Among AI algorithms, the support vector machine (SVM), which is derived from statistical learning theory by Vapnik [7], has been increasingly investigated as an aid for clinical decisions and has shown good diagnostic performance in various clinical fields, particularly cancer prediction including cervical [8], prostate [9], and breast cancers [10]. Because of the properties of SVM its outstanding performance with a small data set, relationship of nonlinear and high dimension in input data [7]. SVM can help with diagnostic guidelines and minimize possible errors in complicated diseases, particularly for inexperienced clinicians. Most importantly, AI methods including SVM can reduce the time for a diagnosis.

In this study, we used the SVM method to diagnose acute appendicitis. We compared the performance of a multilayer neural network (MLNN), and the ACSS. The MLNN method is commonly used in pattern recognition problems and shows good performance in clinical fields. The aim of this study is to propose an AI method for diagnosing acute appendicitis in patients with abdominal pain. The results showed better diagnostic performance for the AI method than that of the ACSS.

2 Methods

2.1 Patient Data

We recruited patients who presented to the emergency department of Dongguk University Hospital with abdominal pain between August 2011 and July 2012. This trial was approved by the Institutional Review Board of Dongguk University Hospital. The clinical protocol including history, physical examination, and laboratory tests was designed using the standardized terminology of the World Federation of Gastroenterology and the ACSS. Patients were allocated into three categories of no appendicitis (NA), normal appendicitis (NorA), and acute appendicitis (AA).

2.2 Alvarado Clinical Scoring System

The ACSS consists of nine factors (1 point for migration of pain to the right lower quadrant, anorexia, nausea/vomiting, rebound tenderness, elevated temperature ≥ 37.5°C, and neutrophil shift to the left > 75% and 2 points for tenderness in the right lower quadrant and leukocytosis (white blood cells > 10,000/µl). The ACSS has a range of 0–10 points and is used to predict the presence or absence of acute appendicitis. The patients were allocated into three groups; ≤ 5 points for NA, ≥ 6 points and ≤ 7 points for NorA, and ≥ 8 points for AA.

2.3 Artificial Intelligence Method

We designed the structure of the SVM in two steps and each step consisted of a SWM method as shown in Fig. 1. The patients were first classified into NA and appendicitis groups. Patients in the appendicitis group were classified into the NorA and AA groups. Each SVM consisted of three spaces (input space, feature space, and output space). The input space had 10 features, including male/female, age, migration of pain to the right lower quadrant, anorexia, nausea/vomiting, rebound tenderness, tenderness in the right lower quadrant, body temperature, neutrophil percentage, and leukocyte count. The features for the input layer were binary (i.e., 0 for no rebound tenderness, and 1 for rebound tenderness) except the continuous data (i.e., body temperature, neutrophil percentage, and leukocyte count). The radial basis function network was used in the feature space, which is commonly used and shows excellent performance, for the mercer kernel as shown (1)

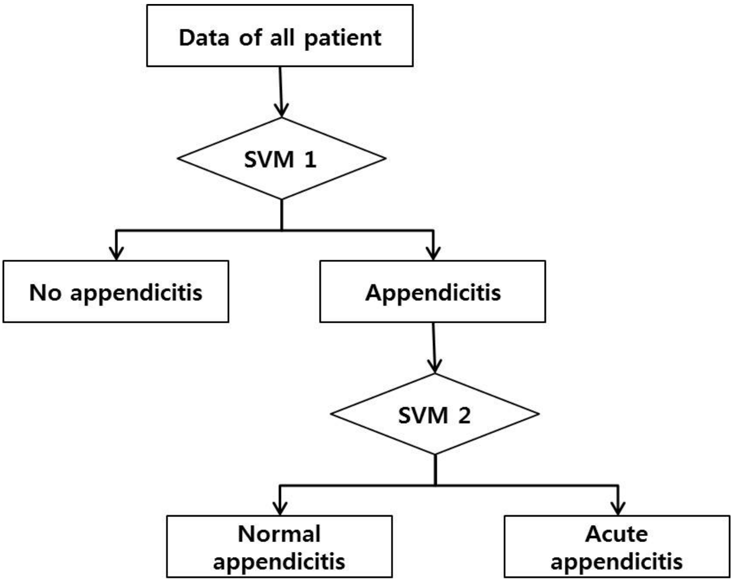

Fig. 1. Diagnosis model of acute appendicitis based on support vector machine (SVM) classifiers

$$k(x^i, x_j^i) = exp\left(-\frac{1}{2\sigma^2}\left\|x^i - x_j^i\right\|^2\right) \quad (1)$$

where x^i_j is the feature of each step (i = 1, 2 for first step, and second step, respectively, and j = 1, 2,…, 10 for features). The output space consisted of two groups for each step (NA and appendicitis groups for the first step, and the NorA and AA groups for the second step). The MLNN also used one input layer with 10 features, two hidden layers, and one output layer consisting of three categories; NA, NorA, and AA. The activation and net functions in MLNN were sigmoidal and linear, respectively.

The SVM and MLNN models were developed in three phases of training, validation, and testing. The patient cases were randomly assigned to one of three phases (60%, 20%, and 20% for training, validation, and testing, respectively). The structures of both the SVM and MLNN were constructed using the MATLAB (MathWorks Inc., Ver. 2012b.) program. Detailed information relating the SVM and MLNN can be found in the neural network toolbox section of the MATLAB documentation.

2.4 Statistical Analysis

We used two methods to measure the performances of the SVM, MLNN, and ACSS. The first algorithm was related to a confusion matrix, including sensitivity, specificity, positive predictive value, negative predictive value, and accuracy. The second algorithm for the evaluation used a receiver operating characteristics (ROC) graph, and the area under the ROC curve (AUC). The AUC value indicated the performance of the diagnostic method in a range of 0–1 (excellent, > 0.9; good, 0.8–0.9; moderate, 0.7–0.8; poor, < 0.7). Differences between variables including performance were assessed by Wilcoxon's rank-sum test, the Kruskal–Wallis test, and the $\chi 2$ test for continuous variables and categorical variables respectively. A $P < 0.05$ was considered a significant difference.

3 Results

A total of 760 patients were enrolled from August 2011 to July 2012 in the emergency department of Dongguk University Hospital. In total, 429 (56.45%) patients were in the NA group and 331 (43.55%) were in the appendicitis group including 237 (31.18%) in the NorA group and 94 (12.37%) in the AA group (Table 1). Mean age was 29.57 years, 30.59 years, and 31.31 years for the NA, NorA, and AA groups, respectively (P = 0.427). The number of female patients (294, 122, and 49 for NA, NorA, and AA, respectively) was significantly higher than that of male patients (135, 122, and 45 for NA, NorA, and AA) in the NA group ($P < 0.001$).

Table 1. Results of 760 patients for suspected appendicitis

	No appendicitis	Appendicitis		P value
		Normal appendicitis	Acute appendicitis	
No. of subjects	429	237	94	<0.001†
Male : Female	135:294	115:122	23:71	<0.001‡
Age-mean(years) (min.-max.)	29.57 (0-62)	30.59 (10-69)	31.31 (14-72)	0.427§
Leucocyte-mean±SD ($\times 10^6$/㎣)	3.58±5.58	7.63±5.48	8.62±7.21	<0.001†
Neutrophil-mean±SD (%)	66.95±17.22	72.21±18.08	76.42±16.50	<0.001†

†Kruskal-Wallis test, § Wilconxon's rank-sum test,‡χ^2-test, and SD: Standard deviation.

The laboratory test values, expressed as mean ± standard deviation, are shown in Table 1. The leukocyte counts were 3.58 ± 5.58, 7.63 ± 5.48, and 8.62 ± 7.21 for the NA, NorA, and AA groups, respectively, and the neutrophil percentages were 66.95 ± 17.22, 72.21 ± 18.08, and 76.42 ± 16.50, respectively (both $P < 0.001$).

The ACSS performance was the lowest (77.86%, 25.08%, 46.63%, 57.39%, and 54.87% for specificity, sensitivity, positive predictive value, negative predictive value, and accuracy, respectively) on all parameters, whereas the diagnostic method using the SVM had the highest values (99.53%, 99.70%, 99.40%, 99.77%, and 99.61% for specificity, sensitivity, positive predictive value, negative predictive value, and accuracy, respectively) (Table 2). Although the MLNN method showed lower performance compared with that of the SVM, the performance of MLNN was significantly higher than that of the ACSS.

Table 2. Performance of diagnosis methods

	Specificity (%)	Sensitivity (%)	Positive predictive value (%)	Negative predictive value (%)	Accuracy (%)
ACSS	77.86	25.08	46.63	57.39	54.87
MLNN	95.10	90.03	93.42	92.52	92.89
SVM	99.53	99.70	99.40	99.77	99.61

ACSS: Alvarado clinical scoring system, MLNN: Multilayer neural network, SVM: Support vector machine.

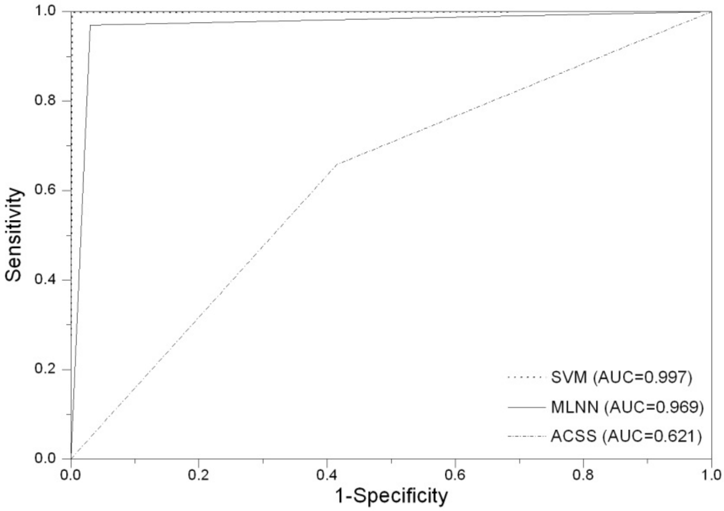

Fig. 2. A receiver operating characteristics (ROC) graph and the area under an ROC curve (AUC) for support vector machine (SVM), Multilayer neural network (MLNN), and Alvarado clinical scoring system (ACSS)

The ROC graphs for the three methods for diagnosing AA are shown in Fig. 2. The method accuracies in decreasing order were SVM (AUC, 0.997), MLNN (AUC, 0.969), and ACSS (AUC, 0.621). Taken together, these results confirm that the AI method had significantly better performance than that of the ACSS.

4 Discussion

Appendicitis is a common abdominal disease in the emergency department. Acute appendicitis, which is considered advanced appendicitis, can lead to death. Although various diagnostic methods have been suggested and have shown good performance, problems have recently come to the fore for the main diagnostic methods such as unstable performance of ACSS, un-usability of ultrasound, and poor safety of CT. We suggested a novel solution using AI methods such as SWM and MLNN.

We enrolled 760 patients with abdominal pain, and the total rate of female patients was significantly higher than that of males (38.82% vs. 61.18%, $P < 0.001$) within the NA (31.47% vs. 68.53%, $P < 0.001$), NorA (40.51% vs. 59.49%, $P < 0.05$), and AA groups (41.69% vs. 58.31%, $P < 0.05$). Hale et al. reported appendectomies in 4,950 patients that were collected over a 12-month period. They noticed that the number of normal appendicitis cases in female patients was significantly higher than that of male patients (19% vs. 9%) [11]. This is because ectopic pregnancy and mittelschmerz in women mimic appendicitis. In the present study, the number of females with appendicitis was also significantly higher than that of males. We thought that some of the women in the appendicitis group may be confused with dysmenorrhea, and many female patients in this study actually were dysmenorrhic (57.14%, 58.87%, and 80.28% for NA, NorA, and AA groups, respectively).

The performance of the ACSS in this study was the lowest of the three methods. This is because of low value of leukocyte count, which plays an important role in the diagnosis of appendicitis, particularly in women and children. Although leukocyte counts were significantly different ($P < 0.001$) among the three groups, the mean leukocyte count ($8.62 \times 106/mm^3$) was lower than $10 \times 10^6/mm^3$, which is a threshold value to receive 2 points in the ACSS. Previous studies have reported that mean leukocyte counts in appendicitis groups are $> 10 \times 10^6/mm^3$ and that the ACSS performed well [1, 7, 12]. We cannot explain why our leukocyte counts were lower compared with those of previous studies. This phenomenon should be investigated in a future study.

However, the performance of the AI method was higher compared with that of the ACSS ($P < 0.001$). de Dombal et al. reported in 1972 that the performance of a computer-aid diagnostic system was significantly higher than that of clinicians [13]. Many researchers have used the AI method to diagnose disease and have shown good performance [8, 9, 10]. The weakness of the AI method is that it is highly dependent on the database (i.e., number of patients), but AI remains the best approach to solve nonlinear problems such as disease diagnosis. To overcome this weak point, the SVM is commonly used to solve nonlinear problems due to kernel function, which converts simple feature dimensions (or input data) into high dimensions [14, 15, 16, 17]. In this study, SVM had better performance on all measurements than that of the MLNN.

5 Conclusion

The AI model showed excellent performance to diagnose acute appendicitis without the need for an expert surgeon. This model may help reduce both negative appendectomies and a delayed diagnosis, particularly for junior surgeons.

Acknowledgements. This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry of Health & Welfare, Republic of Korea. (A102058).

References

1. Kim, E., Subhas, G., Mittal, V.K., Golladay, E.S.: C-reactive protein estimation does not improve accuracy in the diagnosis of acute appendicitis in pediatric patients. Int. J. Surg. 7(1), 74–77 (2009)
2. Fergusson, J., Hitos, K., Simpson, E.: Utility of white cell count and ultrasound in the diagnosis of acute appendicitis. ANZ. J. Surg. 72(11), 781–785 (2002)
3. Petroianu, A.: Diagnosis of acute appendicitis. Int. J. Surg. 10(3), 115–119 (2012)
4. Alvarado, A.: A practical score for the early diagnosis of acute appendicitis. Ann. Emerg. Med. 15(5), 557–564 (1986)

5. Pritchett, C., Levinsky, N., Ha, Y., Dembe, A., Steinberg, S.: Management of acute appendicitis: the impact of CT scanning on the bottom line. J. Am. Coll. Surg. 210(5), 699–705 (2010)
6. Yang, H., Wang, Y., Chung, P., Chen, W., Jeng, L., Chen, R.: Laboratory tests in patients with acute appendicitis. ANZ. J. Surg. 76(1-2), 71–74 (2006)
7. Vapnik, V.: The Nature of Statistical Learning Theory. Springer (2010)
8. Hu, X., Cammann, H., Meyer, H., Miller, K., Jung, K., Stephan, C.: Artificial neural networks and prostate cancer-tools for diagnosis and management. Nat. Rev. Urol. 10, 1–9 (2013)
9. Mat-Isa, N., Mashor, M., Othman, N.: An automated cervical pre-cancerous diagnostic system. Artif. Intell. Med. 42(1), 1–11 (2008)
10. Shi, H., Tsai, J., Chen, Y., Culbertson, R., Chang, H., Hou, M.: Predicting two-year quality of life after breast cancer surgery using artificial neural network and linear regression models. Breast Cancer Res. Treat. 135(1), 221–229 (2012)
11. Hale, D., Molloy, M., Pearl, R., Schutt, D., Jaques, D.: Appendectomy: a contemporary appraisal. Ann. Surg. 225(3), 252–261 (1997)
12. Pouget-Baudry, Y., Mucci, S., Eyssartier, E., Guesdon-Portes, A., Lada, P., Casa, C., et al.: The use of the Alvarado score in the management of right lower quadrant abdominal pain in the adult. J. Visc. Surg. 147(2), e40–e44 (2010)
13. de Dombal, F., Leaper, D., Staniland, J., McCann, A., Horrocks, J.: Computer-aided diagnosis of acute abdominal pain. Br. Med. J. 2(5804), 9–13 (1972)
14. Chen, L., Xuan, J., Riggins, R., Clarke, R., Wang, Y.: Identifying cancer biomarkers by network-constrained support vector machines. BMC Syst. Biol. 5, 161 (2011)
15. Mourao-Miranda, J., Reinders, A., Rocha-Rego, V., Lappin, J., Rondina, J., Morgan, C., et al.: Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. Psychol. Med. 42(5), 1037–1047 (2012)
16. Lancashire, L., Roberts, D., Dive, C., Renehan, A.: The development of composite circulating biomarker models for use in anticancer drug clinical development. Int. Cancer 128(8), 1843–1851 (2011)
17. Lv, G., Cheng, H., Zhai, H., Dong, L.: Fault diagnosis of power transformer based on multi-layer SVM classifier. Electr. Power Syst. Res. 75(1), 9–15 (2005)