Patient Classification Model of Appendicitis Severity Based on Bayesian Stepwise Discriminant Analysis

Lijun Liang¹, Xinyi Li¹, Lixia Cui¹ and Huaqiang Wang²,*

¹ Department of Management, Tianjin University of Traditional Chinese Medicine, Tianjin, China
² Economy and Trading Management Academy, Tianjin Sino-German University of Applied Sciences, Tianjin, China
* Corresponding author: whqiang216@126.com

Abstract. To emphasize distinctions in diagnosis and provide valuable guidance for treatment based on patient heterogeneity, it is useful to classify the various forms of appendicitis into three groups. A bayesian stepwise discriminant model was proposed to predict the patient’s severity of Appendicitis. The age, BMI index, diabetes status, perforation status, other complications, WBC and hemoglobin index were discriminant factors for patient classification according to a bayesian stepwise discriminant analysis. The classification accuracies for severe appendicitis and mild appendicitis were 96% and 97% respectively. The results showed that the presented method was verified effective and practically applicable.

1. Introduction

Appendicitis is a common abdominal disease, and the worldwide incidence rate is from 7.5 per million to 12 per million each year [1]. Now, the treatment trend of acute appendicitis is changing. Emergency operation can be deferred, and conservative antibiotic treatment is preferred alternate choice [2]. So, it is very important to diagnosis the severity of appendicitis. For example, some patients can develop peritonitis or other severe complications and have a higher mortality, if they get delayed treatment [3, 4]. Otherwise, some patients with mild appendicitis would be likely to choose conservative antibiotic treatment [2]. However, the clinical diagnosis of appendicitis remains a challenge to physicians and patients. MC Daly et al. [5] proposed the probability of complicated appendicitis was forecasted by classification and regression tree (CART) analysis. M Singh et al. [6] analyzed a medical database with 257 patients thought to have acute appendicitis to identify the risk factors for complications in AA patients by applying Chi-square test and independent t-test. Although some classification methods have been considered, there are few large studies to identify the key factors and classify the severity of appendicitis.

Computer aided diagnosis systems are widely being used to solve problems. Interest in the use of computer aided diagnosis to help in risk prediction of diseases has been increasing ever since 1970s. Horrocks JC et al. [7] introduced the application of computer-aided diagnosis system in operational experience with 2034 cases in 1972. Sarmad Shafique et al. [8] reviewed some computer-aided leukaemia diagnosis methods, such as preprocessing techniques, feature extraction, classification methods. The standard approach for promoting clinical diagnosis has been the development and use of treatment decision aids. Classification analysis entails automatically constructing a model on the basis of key factors between the decision classes, for example outcomes or decision classes [9]. A host of
classification methods have been put forward, such as decision tree [10], neural network [11], fisher classification [12], Bayesian classification [13], k-nearest-neighbor classification [14] and so on. Jingming ning et al. [12] presented an identification model of six tea based on theanine, caffeine and catechins contents by using stepwise identification and Fisher discriminant analysis. Varmaghani et al. [15] put forward a classification method for early-season corn and soybean applying bayesian discriminant model. This paper uses the bayesian stepwise discriminant method to classify the appendicitis patients into three distinct grades: grade I-severe appendicitis, grade II-moderate appendicitis, grade III-mild appendicitis. Based on the patient classification, the treatments would be greatly beneficial for enhancement of treatment effect, improvement of patient satisfaction and the cost control.

The main contents in this paper are structured as follows. In Section 1, a brief overview of patient classification methods is shown. In section 2, we present our approach to classify the patients into three grades in order to provide guidance for better treatment. The results for patient classification of appendicitis are given in section 3. Section 4 contains the discussion. Our conclusions and recommendations are given in section 5.

2. Methodology
Better patient-centered treatment alternatives can be chosen based on various severity of appendicitis. Therefore, a patient classification model is proposed by using Bayesian stepwise discriminant method. The principles of the Bayesian discriminant analysis and stepwise introducing approaches and the model of patient classification are given bellow.

2.1 Bayesian discriminant analysis
Suppose the data consist of $k$ real-valued vectors $x = (x_1, x_2, ..., x_k)^T$, sampled from $g$ populations which follow $k$ multivariate normal distributions. Let $x$ be assigned to one of the $g$ populations, with prior probability $p_i$ which can be calculated as follows.

$$p_i = \frac{n_i}{n}, i = 1, 2, ..., g; \sum_{i=1}^{g} p_i = 1 \quad (1)$$

The probability density function of population $i$ is defined by:

$$f_i(x) = \frac{1}{(2\pi)^{k/2} |V_i|^{1/2}} \exp \left\{-\frac{1}{2} (x - \bar{x})^T V_i^{-1} (x - \bar{x}) \right\} \quad (2)$$

Where $V_i$ is the variance-covariance matrix of population $i$.

Related to the Bayesian criterion (minimize the expected misclassification loss), the discriminant function is given by:

$$y_i(x) = \ln q_i + c_{0i} + c_{1i} X_1 + \cdots + c_{ki} X_k \quad (3)$$

$$k = 1, 2, \cdots, K \quad i = 1, 2, \cdots, g$$

The posterior probability of misclassifying a subject in class $t$ into class $i$ can be calculated by:
\[ p\left( x \in i \mid x \in t \right) = \frac{\exp(y_i(x))}{\sum_{j=1}^{g} \exp(y_j(x))}, i = 1, 2, \ldots, g \]  

2.2 Stepwise introducing method
In a multiple linear regression model, not all of the independent variables have significant relationship with dependent variables, and sometimes the role of some independent variables can be ignored. The critical factors of appendicitis severity are selected by using stepwise introducing method. The stepwise regression introduces the concept nodes as variables one by one, and the introduction of variable conditions can be based on the mean square error (MSE). Each new variable introduced will be eliminated when its adding doesn’t significantly enhance the performance.

2.3 The patient classification model of Appendicitis
A framework is proposed to classify the appendicitis patients into different risk grades. Fig.1 provides a graphical depiction of this framework, mainly including:

a) Prepare data for modeling: Firstly, determine the factors of appendicitis severity based on literature researches and expert consultations. Then collect information of appendicitis patients and judge the risk grade of each patient based on the disease information by experienced physicians.

b) Build the function of bayes discriminant: Based on the above dataset, build the function of bayes discriminant which is used as feature function.

c) Identify the critical features: Identify the critical features one by one using stepwise introducing method.

d) Calculate the correlation coefficient: Calculate the correlation coefficient of discriminant function based on training samples and provide the patient classification model finally.

![Figure 1. The patient classification model of appendicitis](image)

3. Results
This analysis included 206 patients who underwent laparoscopic appendectomy. Each case was consisted of several attributes that covered the patient’s basic characteristics (gender, age, height, weight), diabetes status, hypertension status, perforation status, other complications, temperature, the length of fever, white blood cell count (WBC), hemoglobin index, preoperative hospital stay. According to the histopathological results, patients were classified into three distinct grades: grade I-severe appendicitis, grade II-moderate appendicitis, grade III-mild appendicitis. And there were 25 cases in grade I, 48 cases in grade II and 133 cases in grade III.
Table 1. Description of attributes

| Variable | Attribute | Description |
|----------|-----------|-------------|
| x1       | gender    | female=1, male=2 |
| x2       | age       | (14,18]=1, [18,55]=2, [55,80]=3 |
| x3       | BMI index | Normal[18.5,30]=1, abnormal(<18.5 or >30)=2 |
| x4       | diabetes  | No=0, Yes=1 |
| x5       | hypertension | No=0, Yes=2 |
| x6       | perforation | No=0, Yes=2 |
| x7       | other complications | No=0, Yes=2 |
| x8       | temperature | Normal=1, abnormal=2 |
| x9       | the length of fever | continuous variable |
| x10      | WBC       | Normal=1, abnormal=2 |
| x11      | the length of abnormal WBC | continuous variable |
| x12      | hemoglobin index | Normal=1, abnormal=2 |
| x13      | preoperative hospital stay | continuous variable |

All the categorical variables were treated, as showed in Table 1. Then, the translated data were analyzed by IBM SPSS statistics.

Three classification function equations were established based on the Bayesian discriminant coefficients:

\[ f_I = -91.701 + 18.528x_1 + 27.955x_2 + 5.732x_4 + 64.349x_6 + 44.135x_8 + 11.398x_{10} + 6.840x_{12} \]  

(5)

\[ f_{II} = -64.564 + 19.955x_1 + 27.955x_2 + 5.683x_4 + 6.392x_6 - 4.494x_8 + 9.572x_{10} + 7.351x_{12} \]  

(6)

\[ f_{III} = -49.067 + 18.391x_1 + 23.661x_2 - 0.85x_4 + 0.768x_6 - 8.6x_8 + 7.462x_{10} + 6.425x_{12} \]  

(7)

The results showed that age, BMI index, diabetes status, perforation status, other complications, WBC and hemoglobin index were critical factors that could affect the severity of appendicitis. According to the Bayesian stepwise discriminant model, associations of the above attributes and patients’ risk grades were observed. And the other attributes including gender, hypertension status, preoperative hospital stay, temperature, the length of fever, the length of abnormal WBC were eliminated by stepwise introducing method.

The classification capability of the Bayesian stepwise discriminant model was tested by cross validation. The total accuracy of this model was 84%. The results were given in Table 2. From the table 2, we found that the classification accuracy was 96% for grade I- severe appendicitis and the classification accuracy was 97% for grade III- mild appendicitis, which were two extremely important grades for selecting an optimal treatment for most patients. The results showed that the presented method was verified effective.

Table 2. Classification results

| Prediction classes | I | II | III | Summary | Accuracy |
|--------------------|---|----|-----|---------|----------|
| Expert             |   |    |     |         |          |
4. Discussions
Service quality and cost containment are two key challenges in healthcare management. As a common illness, appendicitis is important and requires timely accurate diagnosis. However, its diagnosis is sometimes difficult to make, even for experienced physicians. Bayesian discriminant analysis is a well method for disease classification in medicine [16]. But not all of the independent variables have significant relationship with dependent variables, and sometimes the role of some independent variables can be ignored in a multiple linear regression model. For example, only the factors of hypertension status, preoperative hospital stay and the length of fever were no significant in the tests of equality. But there might be the existence of multi-collinearity among the other factors. Therefore, we proposed a Bayesian stepwise discriminant model to classify the appendicitis patients into three different grades. Stepwise introducing method was used to select the discriminant factors of appendicitis severity, and Bayesian discriminant analysis approach was used to establish the patient classification model.

In our opinion, the discriminant model was an effective method to classify the patients. To illustrate the contribution of the etiological attributes, the tests of equality of discriminant factors were statistically different ($P < 0.05$), so this method could be carried out. The total accuracy of this model was 84%. Especially, the classification accuracies for severe appendicitis and mild appendicitis came up to 96% and 97% respectively.

While the model is effective in patient classification, there are some issues that can be considered in future research.

Firstly, selecting illness severity-related attributes is very important that may reflect the discriminant model. More attributes can be considered, such as abdominal pain, nausea, migratory pain which need to be observed in clinic, and other indicators which are discovered until operation.

Secondly, invite more experienced physicians to judge the risk grade of each patient combined with ultrasound and operation. Sometimes, it is difficult to diagnose appendicitis, even for experienced physicians. From Table 2, we find that the classification accuracy is only 41.7% for grade II, and it is hard to separate from grade I and grade III. However, from Fig.2, we find that the physicians can misclassify some cases in grade I into grade II which has closely correlative with the accuracy of model.

![Figure 2. The chart of classification results](image)

Thirdly, patients represent different preferences in treatment decisions [17], so the preferences of patients can be elicited and considered in the patient classification model.

More researches are necessary in order to get the more effective methods for the classification of patient with appendicitis.
5. Conclusions
A classification model for patients with appendicitis was proposed by using Bayesian stepwise discriminant analysis. The contribution of this discriminant model was to classify the appendicitis patients into different risk grades with a view to their early diagnosis and early an optimal treatment alternative selection. The discriminant factors of appendicitis severity were selected by stepwise introducing method, and the patient classification model was established based on Bayesian discriminant analysis. The results showed that the proposed method could reach more reliable and accurate for early predicting of severity in appendicitis patients. This model could provide an effective approach for patient classification.

Acknowledgement
This research is financially supported by Tianjin Philosophy and Social Sciences Program (TJGL16-013Q). The authors thank the editor and anonymous reviewers for their insight comments and suggestions.

References
[1] HW Ting, JT Wu, CL Chan, SL Lin, MH Chen, “Decision model for acute appendicitis treatment with decision tree technology—a modification of the alvarado scoring system,” Journal of the Chinese Medical Association, 73, 401-406 (2010).
[2] N Murugesan Ms, “Appendicitis different treatment options,” IOSR Journal of Dental and Medical Sciences, 14, 36-37 (2015).
[3] Al-Omran M, Mamdani M, McLeod RS, “Epidemiologic features of acute appendicitis in Ontario, Canada,” Canadian Journal of Surgery Journal Canadien De Chirurgie, 46, 263-268 (2003).
[4] Al-Hashemy AM, Seleem MI, “Appraisal of the modified alvarado score for acute appendicitis in adults,” NJ.Saudi Med J, 25, 1229–1231 (2004).
[5] MC Daly, DA Von, HR Wong, “Biomarkers to estimate the probability of complicated appendicitis,” Journal of Pediatric Surgery, 53, 437-440 (2017).
[6] M Singh, YS Kadian, KN Rattan, B Jangra, “Complicated appendicitis: analysis of risk factors in children,” African Journal of Paediatric Surgery, 11, 109-113 (2014).
[7] Horrocks J C, Mccann A P, Staniland J R, Leaper DJ, Dombal FTD, “Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases”, British Medical Journal, 2, 5-9 (1972).
[8] Shafique S, Tehsin S, "Computer-aided diagnosis of acute lymphoblastic leukaemia," Comput Math Methods Med, 11, 1-13 (2018).
[9] YF Li, S Bai, W Carroll, D Y Dan, J C Dort, K Heller, et al. “Validation of the risk model: high-risk classification and tumor pattern of invasion predict outcome for patients with low-stage oral cavity squamous cell carcinoma,” Head & Neck Pathology, 7, 211-223 (2013).
[10] Komal A K, Lalita B, “A review on classification using decision tree,” International Journal of Computing and Technology, 2, 42-46 (2015).
[11] Dreiseitl S, Ohnomachado L, “Logistic regression and artificial neural network classification models: a methodology review,” Journal of Biomedical Informatics, 35, 352-359 (2002).
[12] JM Ning, DX Li, XJ Luo, D Ding, YS Song, ZZ Zhang, et al, “Stepwise identification of six tea (Camellia sinensis, (L.)) categories based on Catechins, Caffeine, and Theanine Contents combined with fisher discriminant analysis,” Food Analytical Methods, 9, 1-9 (2016).
[13] J Zhang, S Wang, L Chen, P Gallinari, “Multiple bayesian discriminant functions for high-dimensional massive data classification,” Data Mining & Knowledge Discovery, 31, 1-37 (2017).
[14] Horton P, Nakai K, “Better prediction of protein cellular localization sites with the k nearest neighbors classifier,” PubMed, 147-152 (1997).
[15] Ali Varmaghani, Varmaghani A, W.E. Eichinger, “Early-season classification of corn and soybean using bayesian discriminant analysis on satellite images,” Agronomy Journal, 108, 1-10 (2016).

[16] Rajaguru H, Prabhakar S K, "Bayesian linear discriminant analysis with hybrid ABC-PSO classifier for classifying epilepsy from EEG signals," International Conference on Computing Methodologies and Communication, 977-980 (2017).

[17] LJ Liang, ZX Liu, HQ Wang, “A shared decision-making model based on DCE and desirability function,” Industrial engineering and management, 18, 79-83 (2013).