Expert Systems with Applications 36 (2009) 10663–10672

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer

Chun-Lang Chang *, Ming-Yuan Hsu

Institute of Industrial Engineering and Management, National Formosa University, 64, Wun-Hwa Rd., Huwei, Yunlin 632, Taiwan, ROC

ARTICLE INFO

Keywords:
Artificial intelligence
Artificial neural network
Genetic algorithm
Logistic regression
Pancreatic cancer

ABSTRACT

Recent medical researches indicate that pancreatic cancer is the eighth leading cause of death of the populations in Taiwan. Each year approximately 800 victims die of this cancer, and the number is increasing year by year. Since most early symptoms of pancreatic cancer is non-specific, doctors' diagnostic decisions might differ based on individual experience, knowledge of the disease, and influenced by their mental conditions at examinations. Certain diagnostic errors are inevitable to occur and mislead the following treatment plans and thus result in insignificant and inefficient follow-up tests. This phenomenon not only caused wastes of the medical resources but also severely delay the golden timing of "early detection and early treatment" for patients. This study used artificial neural network and genetic algorithm of artificial intelligence (AI) as well as logistic regression of statistics to construct three types of screening models for pancreatic cancer and acute pancreatitis. Additionally, it adopted the ROC curves to compare and analyze discriminations of the above-mentioned three screening models. It used 234 case patient data as its training samples and 117 cases as test data. The results of pairwise comparisons and analysis indicate that AUC values of the tree models have no significant differences regardless. Except the fact that GALR model is obviously better than SLR model, both the pairwise comparisons between SLR and BPN or BPN and GALR have no significant difference. On the contrary, however, if under the condition of obtaining the optimal threshold of the three models, GALR model has the best performance with 96.7% in sensitivity and 82.5% in specificity, which are both better than SLR model with 96.7% in sensitivity and 73.7% in specificity and BPN model with 88.3% in sensitivity and 84.2% in specificity. Finally, artificial intelligent approaches will have more optimal predictions in the future with larger and more comprehensive data base as well as more accurate computing methods.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The pancreas is an important body organ that has both exocrine and endocrine functions. Since the pancreas is located in the posterior of the abdomen, signs and symptoms of pancreatic cancer strike at a slower pace and often go unnoticed before it spreads. Therefore, comparing with other types of cancers, regardless of European or oriental countries, prognosis of pancreatic cancer is generally regarded as poor, the causes are the hardest to found, and it is the type of abdominal cancer that has the worst survival rate. Clinically when physicians encounter patients with symptoms similar to both acute pancreatitis and pancreatic cancer, diagnosis is made based on individual experiences. First, a doctor may initially suspect the patient with a certain disease. Following the hypothesis, more tests like the endoscopic procedures, imaging tests or blood tumor markers will be prescribed by the doctor to obtain further information of the disease before making final assessment and confirming diagnosis. However, diagnosis results still might differ depending on the mental and emotional conditions of the doctor. The possibility of false diagnosis is still high. From past literatures in the related filed of both international researches and local studies in Taiwan, many cases of pancreatic cancer were initially diagnosed as other abdominal diseases due to its non-specific symptoms at early stage and thus making treatment plans ineffective (NTUCM Clinicopathological Conference, 2004). To date, there is no effective screening method focusing on early stage of pancreatic cancer nor is there a specific diagnostic tool that has sufficient sensitivity (Chen, 1997). Lacking a proper diagnostic tool at the early stage, the following-up diagnostic screening and treatments might not reach satisfactory recovery results, either. Therefore, if a series of intelligent computer supplemental system can be established to assist in diagnosis and providing references and/or suggestions for a doctor to conduct follow-up tests tailored to a specific cause and to make correct diagnostic decisions, human judgmental errors will decrease, patients will receive better treatments, and most importantly, quality care is enhanced. As a result, this study attempts to use AI to establish an auxiliary

* Corresponding author. Tel.: +886 5 6315721.
E-mail address: jcchang@nfu.edu.tw (C.-L. Chang).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.046

diagnostic screening model for pancreatic cancer. Mainly, this model is to help clinical doctors, when facing patients with abdominal diseases, take advantage of the high-performance computing technology of computers and use the huge volume of information stored in the database to build a suitable diagnostic model based on symptom information, physical examination results, and lab tests of patients. Then, as the next step, the doctors can prescribe follow-up endoscopic procedures, imaging tests, or tumor marks as a referential indicator to reduce man-made judgmental mistakes and eliminate wastes of medical resources.

2. Literature review

2.1. Relevant studies on pancreatic cancer

As imaging technology is advancing and knowledge regarding genetic mutation is maturing, early detection and diagnosis of pancreatic cancer is made more possible (Zhan & Lin, 2002). Some commonly used imaging tests include abdominal ultrasound, computerized axial tomography (CAT scan or CT scan), endoscopic retrograde cholangiopancreatography (ERCP), etc.:

1. Abdominal ultrasonography:
   It is the fastest and most convenient tumor screening tool, especially for the internal organs like the thyroid, liver, gallbladder, spleen, pancreas, and kidneys, abdominal ultrasound is the best choice. However, it still remains insufficient in diagnosing pancreatic cancer. Since the pancreas hides behind the stomach and the intestines, images of the pancreas often are affected by the air in the intestines when taking the ultrasound, and the actual conditions of the pancreas are not discovered. For patients with larger sizes or those suffering from bloating and abdominal distention, "psudonegative" images are very likely to exist.
2. Computerized tomography, CT:
   Using CT san to detect pancreatic tumor can avoid the obstruction of regular abdominal ultrasound scan, and obtain a clear visualization of the exterior pancreas appearance and its neighboring organs, including the swelling of the pancreatic duct and common bile duct. CT scan can further confirm lymphatic involvement, if any, to accurately provide surgeons a referential basis for surgical procedures. Since such method is much more costly than others, it is not suitable for use as a normal screening method.
3. Endoscopic retrograde cholangiopancreaticography, ERCP:
   It primarily uses the duodenoscope-assisted cholangiopancreatoscopy to directly observe the fresh and bones or the duodenum to find out if they are oppressed and if the duodenal nipple presents any abnormalities. A dye will then be injected into the endoscope. Once the exam shows narrowing, obstruction, misshaped, and/or blockage of the pancreatic duct and its branches, they are considered symptoms of pancreatic cancer. Unfortunately, ERCP is an invasive exam. During the process, patients usually experience more discomforts and the method cannot compromise early diagnosis and/or detection, either. Clinically, when a doctor examines a patient, usually prognosis is made based on patient symptoms, medical history, and risk factors. However, since pancreatic cancer symptoms are mostly non-specific in addition to the insufficiency of the above-mentioned imaging testing, making clinical judgments for doctors is fairly challenging. If a suitable and effective supplemental indexing tool can be developed to help in diagnosis, not only can it provide the physicians important references but also help patients receive prompt and suitable diagnostic treatments.

2.2. Logistic regression

Logistic regression is generalized from linear regression model. Assume Y a binominal dependent variable ($Y = 0$ or 1), the probability function is

$$P(Y = 1|x) = \pi(X) = \frac{e^{\beta' X}}{1 + e^{\beta' X}}. \tag{2.1}$$

Note that $\beta' X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$.

The traditional multi-variant regression usually uses the classic ordinary least square to minimize the residual value to obtain the optimal estimation of independent variable parameters. However, in logistic regression analysis, maximum likelihood estimation (MLE) is used instead to maximize the actual observed probability of dependant variables to reach optimum of independent variable parameters as a result (Wang, 2004). Comparing with ordinary least square, MLE can be applied to both the linear model and even to the more complex non-linear estimations. Since logistic regression is a non-linear model, MLE is thus the preferred and most commonly used method (Wang & Guo, 2004).

Assume that each observed value is a Bernoulli test, random variable Y_i yields the Bernoulli probability distribution, then both are binominal Bernoulli random variables and the expectation value is $E\{Y_i\} = \pi_i$. Therefore,

$$P(Y_i = 1) = \pi_i,$$
$$P(Y_i = 0) = 1 - \pi_i.$$

We can describe its probability distribution as follows:

$$P(Y_i = y_i) = \pi_i^{y_i} \cdot (1 - \pi_i)^{1-y_i}, \quad i = 1, 2, 3, \ldots, n \; y_i = 0, 1. \tag{2.2}$$

Let Y_i be an independent random variable. Then according to the above equation, probability distribution function can be obtained as follows (Hauck & Donner, 1977):

$$\begin{aligned} L &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) \\ &= P(Y_1 = y_1) \cdot P(Y_2 = y_2) \cdots P(Y_n = y_n) \\ &= \prod_{i=1}^{n} \pi_i^{y_i} \cdot (1 - \pi_i)^{1-y_i} = \frac{\exp(\beta') \sum_{i=1}^{n} X_i Y_i}{\prod_{i=1}^{n} [1 + \exp(\beta' X_i)]}. \end{aligned} \tag{2.3}$$

To calculate MLE conveniently, LOG the probability distribution function ($\log_e$) and transform Eq. (2.3) into log-likelihood function as in Eq. (2.4)

$$\log_e L(\beta) = \sum_{i=1}^{n} Y_i(\beta' X_i) - \sum_{i=1}^{n} \log_e[1 + \exp(\beta' X_i)]. \tag{2.4}$$

In logistic regression, MLE of parameter β is to obtain the maximum of its log-likelihood function. Let $S(\beta)$ one of $\log_e L(\beta)$ derivative function values be 0

$$S(\beta) = \frac{\partial \log_e L(\beta)}{\partial \beta} = -\sum_{i=1}^{n} \frac{\exp(\beta' X_i)}{1 + \exp(\beta' X_i)} + \sum_{i=1}^{n} Y_i X_i' = 0. \tag{2.5}$$

From Eq. (2.5), it is found that $S(\beta)$ is a non-linear function of β, and no closed-form solution obtained. Namely, no simple computing equation can represent it. Thus, we must seek for the computerized data iterator mode to locate the solutions until obtaining the maximal value of the likelihood function of β convergence.

2.2.1. Model estimation and examination

(1) Goodness-of-fit test:
   The overall goodness-of-fit test can adopt −2Log Likelihood (short as −2LL) ratio test. One model is only consisted of intercept items and the other model of all explanatory variables and intercept items entering the logistic regression. Compare the −2LL differences of the two models and use G to represent the difference:

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات