Computers in Biology and Medicine 63 (2015) 124–132

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques

Guanjin Wang [a,b,*], Kin-Man Lam [c], Zhaohong Deng [d], Kup-Sze Choi [a,b,e]

[a] School of Nursing, Hong Kong Polytechnic University, Hong Kong, China
[b] Centre for Smart Health, School of Nursing, Hong Kong Polytechnic University, Hong Kong, China
[c] Department of Surgery, Tseung Kwan O Hospital, Hong Kong, China
[d] School of Digital Media, Jiangnan University, Wuxi, Jiangsu, China
[e] Interdisciplinary Division of Biomedical Engineering, Hong Kong Polytechnic University, Hong Kong, China

A R T I C L E I N F O

Article history:
Received 2 February 2015
Accepted 17 May 2015

Keywords:
Bladder cancer
Radical cystectomy
Mortality
Prediction
Prognosis
Machine learning

A B S T R A C T

Bladder cancer is a common cancer in genitourinary malignancy. For muscle invasive bladder cancer, surgical removal of the bladder, i.e. radical cystectomy, is in general the definitive treatment which, unfortunately, carries significant morbidities and mortalities. Accurate prediction of the mortality of radical cystectomy is therefore needed. Statistical methods have conventionally been used for this purpose, despite the complex interactions of high-dimensional medical data. Machine learning has emerged as a promising technique for handling high-dimensional data, with increasing application in clinical decision support, e.g. cancer prediction and prognosis. Its ability to reveal the hidden nonlinear interactions and interpretable rules between dependent and independent variables is favorable for constructing models of effective generalization performance. In this paper, seven machine learning methods are utilized to predict the 5-year mortality of radical cystectomy, including back-propagation neural network (BPN), radial basis function (RBFN), extreme learning machine (ELM), regularized ELM (RELM), support vector machine (SVM), naive Bayes (NB) classifier and k-nearest neighbour (KNN), on a clinicopathological dataset of 117 patients of the urology unit of a hospital in Hong Kong. The experimental results indicate that RELM achieved the highest average prediction accuracy of 0.8 at a fast learning speed. The research findings demonstrate the potential of applying machine learning techniques to support clinical decision making.

1. Introduction

Bladder cancer is a common cancer in genitourinary tract [1], affecting mainly the elderly population. In Hong Kong, a survey reported in 2011 that bladder cancer constituted 1.4% of all new cases of cancers [2]. Among various types of bladder cancer, muscle invasive bladder cancer has poor prognosis, with high tendency of metastasis and mortality that necessitate prompt treatment [3]. The most effective treatment approach is the surgical excision of bladder and the surrounding lymphatic tissue, which is known as radical cystectomy. Radical cystectomy is a major surgery that has significant morbidity and mortality [2]. The early postoperative complication rate is between 25% and 57% [4,5], and the early mortality rate is around 3%. The rates are higher for elderly patients [6–8]. Efforts have been made to identify the risk factors in order to maximize the operative outcomes, particularly the long-term survival after surgery. In a retrospective review, the association between clinicopathological factors and mortality for 117 patients treated with radical cystectomy for bladder cancer was investigated from statistical inference perspective [2]. In this study, instead of statistical inference, seven machine learning methods – back-propagation neural network (BPN), radial basis function network (RBFN), extreme learning machine (ELM), regularized ELM (RELM), support vector machine (SVM), naive Bayes (NB) classifier and k-nearest neighbour (KNN) algorithm-are exploited as alternative approaches to predict the overall 5-year survival based on the same clinicopathological dataset of 117 patients treated with radical cystectomy.

1.1. Bladder cancer

The indications for radical cystectomy include treatment failures in non-muscle invasive bladder cancer, and T2-T4aN0M0

* Corresponding author at: School of Nursing, Hong Kong Polytechnic University, Hong Kong, China. Tel.: +852 2766 3913; fax: +852 2364 9663.
E-mail address: guanjin.br.wang@connect.polyu.hk (G. Wang).

http://dx.doi.org/10.1016/j.compbiomed.2015.05.015
0010-4825/

muscle invasive bladder cancer [9,10]. Radical systectomy includes the excision of bladder, prostate (in case of men), urethra, part of distal ureter, and lymphatic tissue of pelvis. After the removal of urinary bladder, urinary diversion is performed to divert the urine produced from the kidneys outside of body. Urinary diversion can either be continence diversion by a neo-bladder or continence pouch, or urinary conduit and ureterocutaneostomy. Radical cystectomy and urinary diversion can be performed by conventional open surgery or minimally invasive surgery.

1.2. Prediction and decision support

In the fields of medicine, clinical data are essential for marking diagnosis, formulating treatment and predicting prognosis. Clinicians make use of the knowledge in different specialties to analyze the histological (cell-based), clinical (patient-based), and demographic (population-based) information [11], where statistical methods such as Cox regression, logistic regression and Kaplan–Meier estimator are conventional employed in the analyses. For example, Kaplan–Meier method and Cox proportional hazards model were used to evaluate the prognostic factors of recurrence, progression and disease mortality in patients with bladder cancer [12]. Logistic regression based on 12 variables was used to identify the predictors of overall 5-year survival of patients who had undergone radical cystectomy for bladder cancer [13]. However, with rapid development of health technologies and informatics, medical data of high dimensionality are made available in both volume and variety. The accuracy of outcome prediction depends heavily on effective information integration of data acquired from various sources, clinically or pathologically, making the conventional statistical analyses that rely on clinicians' knowledge and experience a difficult task. The weakness of statistical methods is more apparent when handling medical data with high variability, nonlinear interactions among the variables, and heterogeneous distributions. For example, regression model, a common statistical technique, often requires some explicit assumptions on the relationships among the data that may be practically invalid [13]. Hence, researchers have begun to investigate alternative techniques for clinical outcome prediction, where computational approaches are a main focus. In particular, machine learning has been introduced into the medical domain to overcome the problems with statistical methods and uncover the knowledge hidden in the complex clinical data.

1.3. Machine learning in medicine

Machine learning is a field in computer science leveraging knowledge from artificial intelligence, optimization and statistics to develop algorithms based on the available data. The approach is to build a model by learning from experience (i.e. the existing data, or the known samples acquired) and use the model to make predictions for the new samples [14]. While the quality and size of the samples can affect the prediction performance, machine learning methods are able to handle large, noisy and complex datasets, rendering it a promising technique for broad application in various areas. They have been explored as a more powerful alternative to statistical methods for classifying patterns and making predictions using techniques such as unconventional optimization strategies, conditional probabilities or absolute conditionality [15].

In medicine and healthcare, machine learning has been applied for personalized and predictive medicine [16], cancer diagnosis and detection [15], and for the study of prevention and treatment policy [11]. For bladder cancer, robust outcome predictions of patients undergoing radical cystectomy was achieved using artificial neural work (ANN) prediction model, with configuration optimized by genetic algorithm (GA) [17]. The system was user-friendly and had potential for widespread use for medical decision support. Histology type and bilharziasis datasets were employed to construct a model using ANN and radial basis function network to predict the survival of bladder cancer patients after diagnosis [18]. Besides, clinicopathological and molecular markers were also used to create an ANN model for predicting one-year survival of patients with muscle invasive bladder cancer [19].

As there is never a best algorithm for all problems, it is necessary to test the performance of different algorithms on a specific problem and identify the optimal one [20]. In this paper, seven machine learning methods are investigated to evaluate their performance in predicting bladder cancer mortality after radical cystectomy for the purpose of prognosis. The seven methods – BPN, RBFN, ELM, RELM, SVM, NB classifier and KNN – were selected because they are representatives among the algorithms in their respective domains. Details of these methods will be provided in Section 3. The implementation of these methods included two major processes. In the training process, the learning methods made use of the training dataset, e.g. the supervised input-output pairs, to identify the relationship directly from the clinicopathological data and built the corresponding model. In the testing process, the classification ability of the model is evaluated using the testing dataset. The predicted outputs were compared with the actual outputs of the testing dataset to measure the performance in terms of accuracy, sensitivity, specificity and precision.

The reminder of this paper is organized as follows. Section 2 describes the clinicopathological data adopted in the study. Section 3 briefly reviews the principles of the seven machine learning methods. In Section 4, the 10-fold cross validation strategy and performance indices used in the experiment are introduced. In Sections 6 and 7, the prediction results are presented and discussed. A conclusion is given in Section 7.

2. Clinical data

The dataset employed in this study originated from a retrospective review on the 5-year survival of patients treated with radical cystectomy for bladder cancer [2]. The data were retrieved from computerized clinical records of 117 patients who had undergone radical cystectomy within the period from 2003 to 2011 in a urology unit in Hong Kong. The purpose in the retrospective study was to examine whether age, tumor stage and preoperative serum albumin level are independent predictors of survival after radical cystectomy. Ninety-nine of the patients were male. The mean age was 68 years old (SD = 10 years). There was no loss of follow-up. The mean follow-up time was 31 months (SD = 29 months). The 30-day mortality, 5-year cancer-specific mortality, other-cause mortality, and the overall mortality rates were 3%, 33%, 22% and 55% respectively. Open radical cystectomy was performed in 71 cases and laparoscopic/robotic-assisted radical cystectomy was conducted for the rest. 96 patients had ileal conduit and 21 patients had continent diversion. Other data includes hospital stay duration, preoperative serum albumin level, Charlson comorbidity index, tumor grade, tumor stage and pathological lymph node status. Further details about the dataset can be found in [2].

For those attributes covering wide numerical range when compared with the others, i.e., age, preoperative serum albumin level and follow-up time, pre-processing was performed to normalize them into the range of [0, 1]. With the advice from physicians, irrelevant data were ignored for the study (e.g. date of operation, date of death) and the final dataset used in the study contained 10 attributes (input) and 1 class attribute (output), as shown in Table 1.