Article preview
Abstract
Introduction
Section snippets
References (53)
Cited by (58)

Artificial Intelligence in Medicine
Volume 49, Issue 1, May 2010, Pages 33-42

A machine learning-based approach to prognostic analysis of thoracic transplantations

Dursun Delen [a], Asil Oztekin [b c], Zhenyu (James) Kong [b]

Show more

Add to Mendeley
Share
Cite

https://doi.org/10.1016/j.artmed.2010.01.002
Get rights and content

Abstract

Objective

The prediction of survival time after organ transplantations and prognosis analysis of different risk groups of transplant patients are not only clinically important but also technically challenging. The current studies, which are mostly linear modeling-based statistical analyses, have focused on small sets of disparate predictive factors where many potentially important variables are neglected in their analyses. Data mining methods, such as machine learning-based approaches, are capable of providing an effective way of overcoming these limitations by utilizing sufficiently large data sets with many predictive factors to identify not only linear associations but also highly complex, non-linear relationships. Therefore, this study is aimed at exploring risk groups of thoracic recipients through machine learning-based methods.

Methods and material

A large, feature-rich, nation-wide thoracic transplantation dataset (obtained from the United Network for Organ Sharingâ€"UNOS) is used to develop predictive models for the survival time estimation. The predictive factors that are most relevant to the survival time identified via, (1) conducting sensitivity analysis on models developed by the machine learning methods, (2) extraction of variables from the published literature, and (3) eliciting variables from the medical experts and other domain specific knowledge bases. A unified set of predictors is then used to develop a Cox regression model and the related prognosis indices. A comparison of clustering algorithm-based and conventional risk grouping techniques is conducted based on the outcome of the Cox regression model in order to identify optimal number of risk groups of thoracic recipients. Finally, the Kaplanâ€"Meier survival analysis is performed to validate the discrimination among the identified various risk groups.

Results

The machine learning models performed very effectively in predicting the survival time: the support vector machine model with a radial basis Kernel function produced the best fit with an R^2 value of 0.879, the artificial neural network (multilayer perceptron-MLP-model) came the second with an R^2 value of 0.847, and the M5 algorithm-based regression tree model came last with an R^2 value of 0.785. Following the proposed method, a consolidated set of predictive variables are determined and used to build the Cox survival model. Using the prognosis indices revealed by the Cox survival model along with a k-means clustering algorithm, an optimal number of â€œthreeâ€ risk groups is identified. The significance of differences among these risk groups are also validated using the Kaplanâ€"Meier survival analysis.

Conclusions

This study demonstrated that the integrated machine learning method to select the predictor variables is more effective in developing the Cox survival models than the traditional methods commonly found in the literature. The significant distinction among the risk groups of thoracic patients also validates the effectiveness of the methodology proposed herein. We anticipate that this study (and other AI based analytic studies like this one) will lead to more effective analyses of thoracic transplant procedures to better understand the prognosis of thoracic organ recipients. It would potentially lead to new medical and biological advances and more effective allocation policies in the field of organ transplantation.

Introduction

Thoracic (heart and lung) transplantation has been accepted as a viable treatment for end-stage cardiac and pulmonary failure. The increased experience in cardiac and pulmonary transplantation, improvements in patient selection, organ preservation, and preoperative support have significantly reduced the early threats to patient survival [1]. Over the past decade, the thoracic transplant waiting time for a listed patient has markedly increased, but the number of transplants performed has declined. In addition, the research also found that there is a perceived inequity in access to organs. The organ allocation system needs to be improved since it may become a major factor negatively influencing the survivability of thoracic transplant [2].

The survivability prediction is becoming increasingly more important in medicine. When a resource is scarce, the need for accurate prediction becomes acute [3]. Especially *prediction of survival time* and *prognosis prediction of medical treatments* are clinically important and challenging problems [4]. Scarceness of organs necessitates the development of effective and efficient procedures to select the most optimal organ receiver since demand for organs of all patients might not be satisfied. To achieve this, one critical step is to reveal the knowledge underlying huge amount of data collected and stored from organ transplantation procedures performed in the past. The objectives are (1) to maximize the patientsâ€™ survival time after the organ transplantation surgery, and (2) to optimize the prognosis for the organ recipients. These can be potentially achieved by discovering the knowledge that may be contained in large dataset consisting of more than hundreds of determinative variables regarding the donors, the potential recipients, and transplantation procedures. Therefore, in this study a data mining method is proposed to process large amount of transplantation data obtained from UNOS to identify the important factors as well as their relationships to the survival of the graft and the patient. Thereafter, a prognostic index [5], [6] is developed to classify the patients into different risk groups for better understanding of the transplantation phenomenon. In short, this study will address the following questions: (1) what are the most important variables to be included in an effective prognostic index related to thoracic organ transplantations? (2) what are the most coherent risk groups that can be formed based on the prognostic index? Predicting the thoracic survivability and classifying the patients (potential thoracic organ receivers) into different classes of risks would help decision makers in determining patientsâ€™ priority for transplantation source assignment.

In the recent past, a number of studies were conducted using data-driven analytics on various organ transplantation datasets. Closely related to the study reported herein, Hariharan et al. [7] focused on the analysis of improved graft survival rate using cyclosporine after renal transplantation in both short-term (less than 1 year) and long-term (more than 1 year). A regression analysis was used to predict the probability of the graft failure after kidney transplantation in both short-term and long-term period in the light of demographic characteristics, transplant-related variables, and post-transplantation variables. The study performed by Herrero et al. [8] included 116 patients who received a liver transplant between the years 1994 and 2000. Statistical tests are used to compare the demographic and characteristic variables, pretransplant, and intra-operative variables between the two groups, namely younger and older than 60. The results indicate that there is a clear trend showing that older patients have lower survival after liver transplantation. Hong et al. [9] presented a survival analysis of liver transplant patients in Canada by considering some factors such as age, blood type, donor type (cadaveric or alive), race, and gender of recipient and donors. However, having limited the variables with this scope, they also admitted that the clinical information lacks of many potential details.

Taking a data mining approach, Kusiak et al. [10] compared two rule-based data mining techniques, i.e. decision trees and rough sets, to predict survival time of kidney dialysis patients. This study achieved satisfactorily high prediction accuracy. The main limitation of the study was the utilization of a small dataset with only 188 patients in total and also many patient-related parameters were neglected in the problem formulation. Using more traditional methods, and specifically having focused on thoracic transplantation, Jenkins et al. [11] and Fernandez-Yanez et al. [12] had a rich pool of independent variables for survivability prediction. Their studies used popular statistical techniques such as Kaplanâ€"Meier method of survival analysis with Mantelâ€"Haenszel log-rank test. However, both of these techniques have been criticized with two major limitations: (1) linear relationships are assumed, which hence cannot capture the nonlinearity among the variables, and (2) the independent variables were selected solely based on the experiences and intuitions of the analysts who conducted these studies. Thus, many potentially significant variables might be left outside the scope of this study. Tjang et al. [13] added more explanatory variables to determine the survivability in heart transplantation, such as body mass index, waiting time on the list, and previous cardiac surgery, their study also ignored the non-linear relationships among the pool of survivability-related variables. Similar limitations exist in some other studies focused directly or indirectly on thoracic transplantation [14], [15], [16].

The existing studies implicitly assume that the relationships among the predictive variables and output variable are linear and the predictor variables are independent of each other, which may not be valid in reality. Moreover, the abovementioned studies focus on small datasets with limited number of predictors for survivability of patients after transplantation. This limitation may cause incomprehensive modeling due to the insufficient information contents (i.e., omission of a number of potentially important predictor variables).

Prognostic index (PI) provides compact prognosis information regarding a specific patient based on the results of a Cox proportional hazards model [5]. Cox proportional hazards model helps identify variables of prognostic importance and hence prognostic index can be used to define groups of individuals at different risk categories. Even though prognostic index is a convenient tool to measure how well the patients are doing after the transplantation, its use in the organ transplantation area has been limited mostly due to the lack of follow-up data. Some existing studies related to devising a PI in transplant area are summarized as follows.

In the study conducted by Christensen et al. [17], it is mentioned that primary biliary cirrhosis requires a liver transplantation operation at the end stage. Based on the prognosis analysis with as well as without transplantation, it is decided whether or not the transplantation is required, if so when. To achieve this goal, corresponding PIs and probabilities of surviving are computed for transplantation and non-transplantation cases. Yoo et al. [18] developed a similar index and revealed that socioeconomic status does not influence patient or graft survival that undergoes liver transplantation at the institute where they performed their study. Deng et al. [19] conducted a study with a national dataset in Germany, which discovers the effect of receiving a heart transplant for the patients in a waiting list. The results indicate that cardiac transplant is associated with survival benefit only for patients with a predicted high risk of dying on the waiting list. Ghobrial et al. [20] performed a study to determine prognostic factors for overall survival in 107 adult patients with post-transplantation lymphoproliferative disorders (PTLDs). It is validated that in discriminating the low and high scored patients the proposed prognostic scoring significantly performs better than the International Prognostic Index for the subset of the patients (56 out of 107) with lactate dehydrogenase.

The common limitation in all of these studies is similar to the limitations of the studies summarized in Section 1.2.1. Namely, they directly devise a prognostic index without determining if the variables used in prognostic index devising phase are necessary and sufficient. This motivates a machine learning-based initial step of variable selection procedure. Because, if the critical predictive factors are not captured effectively due to the intuition- and experience-based selection, the resulting prognostic indices developed based on the selected variables would be inaccurate and, in turn, related risk groups of patients would be deviated from the real classes. This may cause mistakes for decision maker in making organ transplantation policies.

Section snippets

Proposed method

Section 1.2 shows that the most of the existing studies for organ transplantation procedures utilize conventional statistical approaches such as Kaplanâ€"Meier function and log-rank test along with expert-selected variables to predict the survivability. However, organ transplantation procedures consist of a large number of variables (several hundred) that may have nontrivial impact on modeling the prognosis of the grafts/patients. Using a somewhat comprehensive variable list may help discriminate â€¦

The case study and discussion

In order to demonstrate and validate the proposed methodology in Section 2, two most popular data mining toolkit are used, namely SPSS PASW ModelerÂ® [48] and SAS 9.1.3Â® [49] statistical software package. Using the UNOS data set, Sections 3.1 Predictive model results, 3.2 Determination of the candidate covariates for Cox regression model, 3.3 Deployment of Cox regression model and devising the prognostic indices, 3.4 Clustering the prognostic indices, 3.5 Validation of risk groups byâ€¦

Conclusions and future research directions

This study demonstrates that machine learning-based methodology for selecting predictor variables in survivability and prognostic modeling of thoracic organ transplantation is superior to the approaches adopting only expert-selected variables. The study showed that of the comprehensive list of predictors, some have been included in the previous studies (such as gender and age of the recipient, his/her medical condition at registration) while some others (which are found to be critical) haveâ€¦

References (53)

R.N. Pierson *et al.*
Thoracic organ transplantation
American Journal of Transplantation (2004)

D. Sheppard *et al.*
Predicting cytomegalovirus disease after renal transplantation: an artificial neural network approach
International Journal of Medical Informatics (1999)

R.S. Lin *et al.*
Single and multiple time-point prediction models in kidney transplant outcomes
Journal of Biomedical Informatics (2008)

J.I. Herrero *et al.*
Liver transplant recipients older than 60 years have lower survival and higher incidence of malignancy
American Journal of Transplantation (2003)

Z. Hong *et al.*
Survival analysis of liver transplant patients in Canada
Transplantation Proceedings (2006)

A. Kusiak *et al.*
Predicting survival time for kidney dialysis patients: a data mining approach
Computers in Biology and Medicine (2005)

P.C. Jenkins *et al.*
Survival analysis and risk factors for mortality in transplantation and staged surgery for hypoplastic left heart syndrome
Journal of the American College of Cardiology (2000)

J. Fernandez-Yanez *et al.*
Prognosis of heart transplant candidates stabilized on medical therapy
Revista Espanola de Cardiologia (2005)

J.T. Cope *et al.*
A cost comparison of heart transplantation versus alternative operations for cardiomyopathy
Annual thoracic Surgery (2001)

J. Aguero *et al.*
Differences in clinical profile and survival after heart transplantation according to prior heart disease
Transplantation Proceedings (2007)

View more references

Cited by (58)