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

Delen D, Oztekin A, Kong ZJ.

Spears School of Business, Oklahoma State University, T-NCB 378, 700 North Greenwood Avenue, Tulsa, OK, 74106, USA.

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 Sharing-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 Kaplan-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 "three" risk groups is identified. The significance of differences among these risk groups are also validated using the Kaplan-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.

PMID: 20153956 [PubMed - as supplied by publisher]