Abstract

Omneya Amr Attallah

Feature Selection through validation and un-censoring of endovascular repair survival data for predicting the risk of re-intervention

Background Feature Selection (FS) process is essential in the medical area as it reduces the effort and time needed for physicians to measure unnecessary features. Choosing useful variables is a difficult task with the presence of censoring which is the unique characteristic in survival analysis. Most survival FS methods depend on Cox's proportional hazard model however, machine learning techniques (MLT) are preferred but not commonly used due to censoring. Techniques that have been proposed to adopt MLT to perform FS with survival data cannot be used with the high level of censoring. The researcher's previous publications proposed a technique to deal with the high level of censoring. It also used existing FS techniques to reduce dataset dimension. However, in this paper a new FS technique was proposed and combined with feature transformation and the proposed uncensoring approaches to Select a reduced set of features and produce a stable predictive model. Methods In this paper, a FS technique based on artificial neural network (ANN) MLT is proposed to deal with highly censored Endovascular Aortic Repair (EVAR). Survival data EVAR datasets were collected during 2004 to 2010 from two vascular centers in order to produce a final stable model. They contain almost 91% of censored patients. The proposed approach used a wrapper FS method with ANN to Select a reduced subset of features that predict the risk of EVAR re-intervention after 5 years to patients from two different centers located in the United Kingdom, to allow it to be potentially applied to cross-centers predictions. The proposed model is compared with the two popular FS techniques Akaike and Bayesian information criteria (AIC, BIC) that are used with Cox's model. Results The final model outperforms other methods in distinguishing the high and low risk groups as they both have concordance index and estimated AUC better than the Cox's model based on AIC, BIC, Lasso, and SCAD approaches. These models have p-values lower than 0.05, meaning that patients with different risk groups can be separated significantly and those who would need re-intervention can be correctly predicted. Conclusion The proposed approach will save time and effort made by physicians to collect unnecessary variables. The final reduced model was able to predict the long-term risk of aortic complications after EVAR. This predictive model can help clinicians decide patients' future observation plan.