Bayesian Network framework for biophysical radiation pneumonitis modeling

Statement of Impact

This study represents the first attempt, to date, to apply a Bayesian Network framework to incorporating biomarker information along with dosimetric variables into a radiation pneumonitis (RP) risk model, which demonstrates the potential to enhance patient-specific prediction of RP risk.

Introduction

The role of biomarkers in RP has been emphasized in response to limited prediction accuracy of dosimetric RP models (Rodrigues et al., 2004), yet there is a lack of an approach to incorporate biological knowledge into the current models. A Bayesian network (BN) not only facilitates classification of a high RP risk group based on a biological profile and a radiotherapy plan but also provides visual interpretation of interactions among variables of interests. However, the high-dimensional nature of biological information remains a challenge for BN structure learning. This pilot study presents a computational scheme to identify relevant biomarkers and their interactions that lead to RP onset using the BN framework.

Materials and Methods

Data collection: 22 inoperable NSCLC patients who received radiotherapy (RT) between 2006 to 2009 were selected for this study. Five of them were diagnosed with RP (CTCAE v3.0 grade ≥2). Blood serum was collected from each patient before and during the RT. Enzyme-linked immunoassay (ELISA) was performed on each blood sample to measure the concentration of the following four candidate biomarkers: alpha-2-macroglobulin (α2M), angiotensin converting enzyme (ACE), transforming growth factor β (TGF-β), and interleukin-6 (IL-6). Two known dosimetric variables, mean lung dose (MLD) and GTV location in the superior-inferior direction (GTVCOMSI) were extracted from the electronic treatment plans (Bradley et al., 2007).

Variable selection and pre-processing: Dimensionality of biomarker measurement data was reduced to a subset of variables which linear combination via a logit link function showed the highest Spearman’s correlation with RP. For this purpose, the model order selection feature implemented in the Dose Response Explorer (DREES) (El Naqa et al., 2006) was used. The selected biomarker variables along with the known two dosimetric variables were discretized into 2 bins and used as features for BN with the RP occurrence set as a class variable. Missing data were filled with the median of the remaining values.

Bayesian network learning: The optimal BN structure and maximum likelihood parameters were learned by cross-validation on the original dataset using the method described by Oh et al. (Oh et al., 2011a). The constructed BN was used as a classifier which predicted the RP occurrence when the probability of RP given a set of input variables was inferred to be larger than 0.5.

Model validation and comparison: Due to limited sample size, the constructed BN-based classifier was validated on a simulated testing dataset. The testing dataset was generated from random numbers and then transformed non-linearly in a way that the probability distribution of each variables was equal to the original dataset (van der Schaaf et al. 2012). The predictive power of the model was evaluated in terms of its classification accuracy, positive predictive value (PPV), and negative predictive value (NPV). The results were compared to those of a logistic regression model built using DREES upon the same set of variables and a simpler Naïve Bayes classifier where all the selected variables are assumed to influence RP independently.

Results and Discussion

The variable selection method revealed the following 5 biomarker variables were relevant to predicting RP: 1) pre-RT concentration level of α2M (a2Mpre), 2) ratio of pre- to intra-RT levels of α2M (a2Mratio), 3) intra-RT IL6 level (IL6intra), 4) ratio of pre- to intra-RT levels of TGF-β (TGFratio), and 5) pre-RT level of ACE (ACEpre). The learned BN structure identified the probabilistic relations amongst the 7 features (5 biomarker variables + 2 dosimetric variables) and a class variable (RP) (Figure 1). The illustrated structure hints into the important role of α2M as a mediator of acute radiation response as has been previously reported (Oh et al., 2011b). In the simulated testing dataset, Performance of the proposed BN model was estimated to be 0.873, 0.958, and
0.676 for overall accuracy, NPV, and PPV respectively. It outperformed a logistic regression and a Naïve Bayes counterparts in terms of prediction accuracy, and quite remarkably, PPV (Figure 2). It has to be noted though that positive prediction power was inferior to negative prediction regardless of the chosen model, possibly due to a low event rate in the dataset.

Figure 1. A Bayesian network structure learned from the dataset. RP was fixed as a root variable prior to the learning. Note that the variable MLD was disconnected from the network and thus took no part in the classification process.

Figure 2. A bar graph comparing the performance of the proposed BN model (blue) against a logistic regression model (red) and a Naïve Bayes model (green). Error bars represent 1 standard deviation from 10 validation trials.

Conclusion
The presented Bayesian Network approach has a potential to enhance the prediction power of RP by explicitly modeling interaction between input physical and biological variables, which can also be used to validate or guide further biomarker RP research.

Reference