

UCSF

UC San Francisco Previously Published Works

Title

A situational awareness Bayesian network approach for accurate and credible personalized adaptive radiotherapy outcomes prediction in lung cancer patients

Permalink

<https://escholarship.org/uc/item/462857w7>

Authors

Luo, Yi
Jolly, Shruti
Palma, David
et al.

Publication Date

2021-07-01

DOI

10.1016/j.ejmp.2021.05.032

Peer reviewed



Published in final edited form as:

Phys Med. 2021 July ; 87: 11–23. doi:10.1016/j.ejmp.2021.05.032.

A Situational Awareness Bayesian Network Approach for Accurate and Credible Personalized Adaptive Radiotherapy Outcomes Prediction in Lung Cancer Patients

Yi Luo^{1,*}, Shruti Jolly¹, David Palma², Theodore S. Lawrence¹, Huan-Hsin Tseng¹, Gilmer Valdes³, Daniel McShan¹, Randall K. Ten Haken¹, Issam El Naqa¹

¹Department of Radiation Oncology, The University of Michigan, Ann Arbor, MI, USA

²London Health Sciences Centre, Western University, London, ON, Canada

³Department of Radiation Oncology, UCSF Medical Center at Mission Bay, San Francisco, CA, USA

Abstract

Purpose: A situational awareness Bayesian network (SA-BN) approach is developed to improve physicians' trust in the prediction of radiation outcomes and evaluate its performance for personalized adaptive radiotherapy (pART).

Methods: 118 non-small-cell lung cancer patients with their biophysical features were employed for discovery ($n=68$) and validation ($n=50$) of radiation outcomes prediction modeling. Patients' important characteristics identified by radiation experts to predict individual's tumor local control (LC) or radiation pneumonitis with grade 2 (RP2) were incorporated as expert knowledge (EK). Besides generating an EK-based naïve BN (EK-NBN), an SA-BN was developed by incorporating the EK features into pure data-driven BN (PD-BN) methods to improve the credibility of LC or / and RP2 prediction. After using area under the free-response receiver operating characteristics curve (AU-FROC) to assess the joint prediction of these outcomes, their prediction performances were compared with a regression approach based on the expert yielded estimates (EYE) penalty and its variants.

Results: In addition to improving the credibility of radiation outcomes prediction, the SA-BN approach outperformed the EYE penalty and its variants in terms of the joint prediction of LC and RP2. The value of AU-FROC improves from 0.70 (95% CI: 0.54–0.76) using EK-NBN, to 0.75 (0.65–0.82) using a variant of EYE penalty, to 0.83 (0.75–0.93) using PD-BN and 0.83 (0.77–0.90) using SA-BN; with similar trends in the validation cohort.

*Corresponding author at: Department of Radiation Oncology, The University of Michigan, 519 West William Street, Ann Arbor, MI, USA. YL1515@gmail.com (Y. Luo).

Declarations of interest: none

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions: The SA-BN approach can provide an accurate and credible human-machine interface to gain physicians' trust in clinical decision-making, which has the potential to be an important component of pART.

Keywords

Situational awareness; Bayesian networks; Accuracy and credibility; Personalized adaptive radiotherapy

1. Introduction

Lung cancer is the leading cause of cancer death in the world. About 80% to 85% of lung cancer cases are non-small-cell lung cancer (NSCLC). Radiotherapy is one of the main treatment modalities for locally advanced NSCLC. While some patients may get tumor local control (LC) from the radiotherapy without any complications, some of them may not get cured, and in the meantime, they may suffer from additional radiation-induced toxicities (RITs). In order to improve NSCLC patients' therapeutic satisfaction, personalized adaptive radiotherapy (pART) was proposed to explore an individual patient's radiation treatment plans based on his / her biophysical characteristics before and during the course of radiotherapy by maximizing the patient's tumor LC and minimizing the probability of receiving RITs simultaneously [1]. However, the mechanisms of tumor response to the radiotherapy are still under-explored. Although trial and error methods cannot be ethically used in clinical practice, the increasing amount of available clinical data before and during radiotherapy has the potential to provide useful biophysical patterns for the realization of pART. Also, the improvement of computing hardware allows more sophisticated machine learning (ML) algorithms to investigate these patterns from the data directly.

While datasets in the field of radiation oncology usually have a small sample size, each patient owns high-dimensional biophysical features datasets, including his / her physical, biological, imaging, genomic and dosimetric information along the course of radiation treatment. As accuracy is typically an essential criterion to evaluate the performance of an outcome prediction model, explainability is another important aspect due to its potential to produce insights into the cause of the algorithmic decisions [2]. Being a generative ML approach, Bayesian networks (BNs) enjoy an explainable network structure to display the dependent / independent relationship among random variables. A pure data-driven BN (PD-BN) approach was developed in our previous research to unravel the biophysical pathways among patients' characteristics, radiation outcomes and treatment plans from radiation oncology datasets for LC or / and radiation pneumonitis (RP) with grade 2 (RP2) prediction [3–5]. However, a purely data-driven radiation outcome prediction model with a good performance may not be recognized or accepted by physicians in clinical practice and still fail to garner their trust because it may fail to reflect their conceived knowledge or known predictors in the literature. For instance, our previously developed PD-BNs included existing but also new findings that have yet to be proven in the lab or to be adopted for clinical practice; however, it didn't explicitly select some other risk factors known by physicians such as smoking or chemotherapy.

Since an outcome prediction model needs to be accepted by physicians before it could be applied to clinical decision-making, *credibility* has been proposed recently to indicate the capability of a model to gain clinicians' trust [6]. This concept originated from studies related to explainability and interpretability [7–9]. Credibility generally represents the ability of an outcome prediction model to provide reasons for its predictions that are, at least in part, in line with the physicians' understanding (or prior knowledge) while its prediction accuracy does no worse (is not inferior) than that of a corresponding (data-driven) model [6]. When the explainable new findings from the data-driven models can barely gain physicians' trust, credibility becomes a new criterion to guide the application of ML approaches in developing outcome prediction models. The purpose of this study is to develop and demonstrate a new BN-based approach for accurate and credible pART outcomes prediction in lung cancer patients.

Including the phases of perception, comprehension and projection from human factors engineering research [10], situational awareness (SA) is recognized as a critical foundation for effective clinical communication and successful decision-making in healthcare [11]. The realization of SA depends on data-driven and goal-driven processing. While the former is a bottom-up way to indicate how changes in the environment can affect a switch in active goal states, the latter is a top-down approach to guide the search for better interpretation of information and plans to achieve those goals [10]. By alternating these two ways to simulate human information processing, SA intends to direct attention and interpret information in the environment. Due to its potential to increase the credibility of an outcome prediction model and gain physicians' trust in medical care settings, the concept of SA is employed to build our new outcome prediction model for pART applications.

Obviously, previous PD-BN approaches only represent SA's data-driven processing for radiation outcome estimation. To gain physicians' trust in the outcome prediction for pART, goal-driven processing needs to come into play. The physicians' trust is related to their accumulative knowledge gained from years of experience, reading articles, training, colleagues, which is named *expert knowledge* (EK) in this study. In addition to bypassing otherwise complex systems and providing parsimonious solutions that focus on key aspects of a given situation [12], the EK can add new information to the models learned from data only [13,14]. Using the incorporation of the EK into the PD-BN method to capture SA's goal-driven processing, we developed an SA-BN approach for radiation outcome prediction in pART. Additionally, we extended a well-known credible concept, the EYE penalty approach [6], to radiation outcomes prediction and evaluated its performance as benchmark to that of our proposed credible model.

The developed SA-BNs offer a human-machine interface to allow potential human involvement in pART. In addition to enabling exploring biophysical pathways starting with EK, the SA-BNs can help physicians to conduct their familiar "what if" counterfactual analysis along the biophysical pathways, which can provide great potentials to gain their trust in clinical practice. The rest of the paper is organized as follows. Section 2 introduces the SA-BN approach to develop credible outcome prediction BN models, and presents EYE penalty and its variant methods to benchmark our new approach. The performance comparisons among EK based naïve BN (EK-NBN), PD-BN, SA-BN, LASSO, EYE penalty

and its variants are introduced in Section 3. Section 4 mainly discusses the credibility and accuracy of the SA-BNs with correct EK compared to the PD-BNs. Conclusions are drawn in Section 5.

2. Methods and Materials

2.1. Study participants and data collection

Our study to explore a SA-BN approach for the clinical decision support of pART has been approved by institutional review board (IRB). The datasets to build and validate SA-BNs contain 118 stage III NSCLC patients including adenocarcinoma, squamous cell carcinoma sub-categories treated by volumetric modulated arc therapy (VMAT) and / or chemotherapy. All stereotactic body radiation therapy (SBRT) patients were excluded from this study due to varying regimens. Planning information followed standard clinical protocols [15,16]. Relevant to this study, the generalized equivalent uniform dose (gEUD) was used to evaluate the effect of radiation dose to treatment outcomes; the clinical target volume margin was developed based on clinical standards at our institution with 6–8 mm isotropic extension; and the addition of the random error margin due to respiratory motion (~1 cm) gave the planning target volume (PTV).

The number of NSCLC patients, radiation outcome events and median follow-ups time in discovery and validation datasets are shown in Table 1. While each patient in these datasets had 297 features, a radiation outcome such as LC or RP2 is not necessarily related to all of them. In our study, positron emission tomography (PET) radiomics before and during radiotherapy are designed to predict LC only. Including all the features for LC or RP2 prediction not only can add noise to the development of an outcome prediction model, but also may mislead it. For example, PET radiomics features for LC prediction may be selected as the part of a RP2 prediction model based on a data-driven approach. Then, features in the whole dataset were allocated into *LC's or RP2's feature datasets* for LC or / and RP2 prediction based on the experience of physicians and medical physicists in our study. Table 2 shows the categories of biophysical features before and during radiation treatment and the number of features in each category of the whole dataset, LC's and RP2's feature datasets. As we can see from the table, 15 common dosimetric parameters were distributed to LC's or RP2's feature dataset, all the pre- and during treatment PET radiomics features were included in LC's feature dataset.

In our study, radiomics investigates the extraction of quantitative, sub-visual image features to create mineable databases from radiological images [17], and it includes widely used gray-level co-occurrence matrix (GLCM), neighborhood gray-tone difference matrix (NGTDM), run-length matrix (RLM), and gray-level size-zone matrix (GLSZM). The slopes (SLP) of cytokines change and the relative changes (RD) of PET tumor imaging / radiomics features with fluorodeoxyglucose as the radiotracer during the courses of radiotherapy were calculated from the patients' responses by the end of weeks 2 and 4 radiation treatment. The details of all the features can be found in our previous studies [3–5].

Patients were considered to have LC if their clinical, radiographic, or biopsy evidence of progression were not observed with a minimum follow-up of six months. While the patients'

LC rate by the end of the follow-up is 71%, their overall survival rate at 3 years after treatment is 26%. RP is a common and explicit complication caused by radiotherapy in lung cancer patients [18]. Five grades based on common terminology criteria for adverse events (CTCAE 3.0) were employed to score the patients' RP based on clinical assessment and imaging findings, and the level of RP was identified by the maximal RP score during follow-up. As a serious complication in radiation treatment practice, RP2 is usually studied in radiation oncology literatures [18–20]. To keep the same caliber, RP2 is also considered as representative of a typical RIT in our study.

A radiation outcome's EK dataset was selected from the outcome's feature dataset based on two lung cancer physicians' experience in our study. It is assumed that LC's EK dataset includes the following EK factors for LC prediction before and during the courses of radiation treatment, "*Stage*", "*gross tumor volume (GTV)*", "*PTV*", "*Age*", "*chemotherapy (Chemo)*", "*Tumor gEUD*", "*dose that covers 95% of planning target volume (PTVD95)*", "*dose that covers 95% of the GTV (GTV95)*", "*biologically effective dose (BED)*", "*Dose Per Fraction*", "*Treatment Duration*", "*Total Treatment Time*", and the rest of factors in LC's feature dataset constitute LC's non-EK (NEK) dataset; RP2's EK dataset consists of the following EK variables for RP2 prediction before and during radiation treatment, "*Total Lung Volume*", "*Smoking*", "*Lung gEUD*", "*the volume of normal lung receiving 20 Gy (V20)*", "*the volume of normal lung receiving 5 Gy (V5)*", "*Dose Per Fraction*", "*Chemo*", and RP2's NEK dataset includes the rest of variables in RP2's feature dataset.

Although there may exist other factors that can be explored in the future for better radiation outcomes prediction, these EK factors are so far the most common and well-recognized knowledge based on the experience of lung cancer experts. Since the discovery and validation datasets are collected from patients treated in different time periods and the latter is not considered in the discovery phase of the SA-BNs, our validation strategy for model development and assessment satisfies the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) type 2*b* recommendations. (Note: type 2*b* of prediction model studies covered by the TRIPOD statement indicates that the data are nonrandomly split (e.g., by location or time) into two groups: one to develop the prediction model and another to evaluate its prediction performance [21].)

2.2. Situational awareness BN (SA-BN) approach

Our previous PD-BN approach for LC or / and RP2 prediction mainly includes feature selection and BN structure learning. Markov blanket (MB) developed based on incremental association Markov blanket (IAMB) or its variants [22,23] has the potential to identify a variable's patients, descendances and spouses such that the variable is independent of other variables given its inner family. The first step of the PD-BN approach is to identify the most important features related to LC or RP2 by exploring its extended Markov blankets (EMBs) based on the discovery dataset, which includes not only its inner family and but also each family member's next of kin. The second step is to find the best stable structure of a single or a multi-objective BN in terms of their radiation outcome(s) prediction performance from these important features [3–5]. While the PD-BNs enjoy encoding causal relations in their structure (or topology) and good prediction performance, well-known EK features may not

be selected from the first step to participate in BN structure learning at the second step, resulting in barriers to gain the trust of physicians. Thus, a SA-BN approach is developed to incorporate the EK into these two steps without comprising its prediction performance compared to the PD-BN approach.

The major differences between the SA-BN approach and the PD-BN method is during the process of feature selection as illustrated in Fig. 1. While LC's or RP2's feature dataset in the former approach is separated into two parts including an EK dataset and a NEK dataset, the concept of EMB in the latter method [5] is employed to identify LC's or RP2's EMB from their EK and NEK datasets in the former approach, which are denoted as LC's or RP2's EK-EMB and NEK-EMB respectively. If an EMB of LC or RP2 includes both EK and NEK features, these features together with important EK features identified from an EK-EMB of LC or RP2 are treated as the inputs of SA-BN structure learning as illustrated by the left path in Fig. 1; otherwise, whether the EK features can be incorporated to the structure of a SA-BN depends on the property of the EK-EMB of LC or RP2. If it is not empty, the contained important EK features together with critical NEK features identified from a NEK-EMB of LC or RP2 are considered as inputs to learn BN structure as shown by the right path in the figure; otherwise, only important NEK features are considered as the inputs of BN structure learning as described by the middle path in Fig. 1.

For the joint radiation outcomes prediction of LC and RP2, important EK features (if possible) and NEK features are identified by following the above feature selection process for LC or RP2 separately, and then they are combined as inputs to learn the structure of two-objective SA-BN. Being similar as that of the PD-BN approach, the structure learning of a single or a multi-objective SA-BN is a backward feature elimination process from the selected EK and NEK features by removing so far the most un-important "leaf node" in the network to improve the SA-BN's prediction performance. Here, the importance of each leaf node is assessed by the strength of the arcs that connect them to the outcome(s) and other features in the SA-BN. At the beginning of the network structure learning or after eliminating a "leaf node" from the network, the best SA-BN structure is identified by score-based algorithms such as Tabu search [24]. However, the final SA-BN may have multiple different structures with or without directed edges opposite to known cause-effect mechanisms, which is called Markov equivalent [25] in literature. Then, the most reasonable SA-BN is identified by following timestamps, literature, and expert experience to determine the order of different kinds of features in biophysical pathways.

2.3. Comparison models of the SA-BN for radiation outcomes prediction

EK-NBNs are designed to explore the impact of the EK to radiation outcomes by connecting important EK (if possible) and NEK features selected from the EK-EMBs and NEK-EMBs to LC or / and RP2 directly. Also, a logistic regression prediction model with an EYE regularization term, named EYE penalty approach, is employed for LC or / and RP2 prediction by constraining weights for EK features with L_2 -norm and weights for NEK variables with L_1 -norm. While the L_2 -norm intends to maintain a dense structure among the EK factors, the L_1 -norm is designed to encourage sparsity on the NEK covariates like least absolute shrinkage and selection operator (LASSO) approach. By assuming the EK could be

wrong, the EYE penalty approach is proven to have the desired properties of a linear credible model in a least-square regression setting [6].

It is recognized that the impact of comparably high-dimensional features with less samples in radiation oncology datasets on the performance of the EYE penalty is unknown. Then, a variant of the EYE penalty approach, named LASSO-EYE, is developed in this study to evaluate it by identifying important features related to LC or RP2 before and during treatment from NEK datasets through the LASSO and combining them with all the corresponding EK features as the inputs of the EYE penalty approach for LC or / and RP2 prediction.

2.4. Prediction performance evaluation based on discovery and validation datasets

In a single objective model for LC or RP2 prediction, nested cross-validation to evaluate the entire process of PD-BN approach [3] was employed to tune parameters of EK-NBN, SA-BN, EYE, LASSO and LASSO-EYE models and to evaluate their prediction performance before and during radiotherapy from the NSCLC patients in a discovery dataset ($n=68$). The prediction performance of pre- and during treatment outcome prediction models developed from the whole discovery dataset was further evaluated by other NSCLC patients in a validation dataset ($n=50$), which is named external validation in our study.

For the joint prediction of LC and RP2, single EK-NBN and SA-BN can predict these two objectives simultaneously; however, in contrast a logistic regression model cannot consider two radiation outcomes at the same time. Then, we developed an analytic model for LC or RP2 prediction separately using EYE, LASSO or LASSO-EYE approach and combined these two models to evaluate the joint prediction of LC and RP2 of these methods. Being an extension of the conventional receiver operating characteristic (ROC) used with single endpoints, a free-response ROC (FROC) curve was employed to evaluate these approaches' prediction performance with two endpoints (LC and RP2) for the nested cross-validation and external validation [4,26].

In our study, the development and evaluation of the EK-NBN, SA-BN and LASSO approaches were conducted in R version 4.0.0 designed for statistical computation and graphics, and the development and evaluation of the EYE and LASSO-EYE methods were implemented using Pytorch version 1.5 in the Python environment modified from the code given in [6]. Moreover, Delong test was used to evaluate the difference of the prediction performance between the SA-BN and other approaches with single or two radiation outcomes before or during radiotherapy based on the discovery or validation dataset. In order to compare two FROC curves, each of them was transferred into the corresponding ROC curves [26], and Delong test was employed to compare these two ROC curves [27].

3. Results

3.1. Single-objective SA-BN models for LC or RP2 prediction

Figs. 2(a) and 2(b) illustrate the stable structure of pre- and during treatment SA-BNs to predict LC based on the discovery dataset, respectively. In addition to SNPs, miRNAs, cytokines, PET radiomics features and dosimetric information, EK features, such as "Age",

“GTV” and “Tumor_gEUD”, were selected and displayed as nodes in the SA-BNs through the feature selection and BN structure learning algorithms. In a SA-BN, nodes represent selected important variables, and their dependencies are described by directed edges. The thickness of an edge denotes the strength of a connection, where the thicker edge represents a stronger connection. While a green or red color is employed to denote monotonic positive or negative influences through the edges between the connected nodes, a gray color indicates non-monotonic or mixed influences between them. The performance of the SA-BN approach to predict LC before and during radiation treatment are evaluated by area under ROC curves (AUCs) based on nested cross-validations within the discovery dataset as shown in Table 3. The performance of the SA-BNs in Fig. 2 for LC prediction are evaluated by the AUCs of external validations based on the validation dataset as listed in Table 3.

Based on the discovery dataset, the pre- and during treatment SA-BNs for RP2 prediction were developed as shown in Figs. 3(a) and 3(b) respectively. EK features, such as “Chemo”, “Smoking”, “Total Lung Volume” and “Lung_gEUD”, were selected and displayed as nodes in the SA-BNs via the feature selection and BN structure learning algorithms. The performances of the SA-BN approach and the SA-BNs in Fig. 3 to predict RP2 before or during radiation treatment are evaluated by nested cross-validations and external validations based on the discovery and validation datasets respectively as illustrated in Table 3.

3.2. Multi-objective SA-BN models for joint prediction of LC and RP2

Figs. 4(a) and 4(b) show pre- and during treatment SA-BNs for the joint prediction of LC and RP2 based on the discovery dataset. EK features, such as “Age”, “GTV”, “Chemo”, “Lung_gEUD”, “Tumor_gEUD”, “Smoking” and “Total Lung Volume”, were selected as key variables for the joint prediction of PR2 and LC before and during the radiotherapy. The performance of the SA-BN approach or the SA-BNs in Fig. 4 for the joint prediction of LC and RP2 before and during radiation treatment are evaluated by nested cross-validations or external validations based on the discovery or validation dataset as listed in Table 3.

Especially, the performance of the SA-BN approach for joint LC and RP2 prediction based on the discovery dataset is 0.83 with the 95% CI of 0.77–0.90 based on 2000 bootstrap samples. Figs. 5(a) and 5(b) show the FROC curves of the SA-BN approach and the SA-BN in Fig. 4(b) to predict both LC and RP2 during treatment, respectively.

3.3. Other radiation outcomes prediction methods

Based on target nominal type I error rate 0.1, pre-treatment important EK features related to LC prediction such as “Stage”, “GTV”, “Age”, “Chemo” and “Tumor_gEUD” were selected from the EK-EMB of LC based on its EK dataset, and pre-treatment important EK variables related to RP2 prediction like “Total Lung Volume”, “Smoking”, “Chemo” and “Lung_gEUD” were chosen from the EK-EMB of RP2 based on its EK dataset. The above important EK features related to LC or RP2 were combined for joint LC and RP2 prediction. Since there are no additional EK features identified from during treatment, the EK-NBNs for LC or / and RP2 prediction before and during the treatment based on the discovery or validation dataset are the same, resulting in their similar prediction performances as shown in Table 4. The performance of EK-NBN for joint LC and RP2 prediction based on the discovery dataset is 0.70 with the 95% CI of 0.54–0.76 based on 2000 bootstrap samples.

The performance of the PD-BN approach and its associated PD-BNs in our previous research [3–5] are also listed in the same table for comparison. Especially, the performance of PD-BN for joint LC and RP2 prediction based on the discovery dataset is 0.83 with the 95% CI of 0.75–0.93. The performance of EYE, LASSO and LASSO-EYE approaches or their associated prediction models for radiation outcomes prediction were evaluated from the nested cross-validation or external validation based on the discovery or validation dataset as summarized in Table 4. The values of AU-FROC of EYE, LASSO and LASSO-EYE approaches for joint LC and RP2 prediction based on the discovery dataset are 0.70 (95% CI: 0.60–0.78), 0.53 (0.43–0.64) and 0.75 (0.65–0.82), respectively. In the meantime, Table A1 in *Appendix A* shows the important features and their coefficients in LC or RP2 prediction models before and during treatment developed from LASSO and EYE approaches based on the discovery dataset.

3.4. Comparison of prediction performance between the SA-BN and other approaches

The differences of prediction performance between the SA-BN model and other prediction models with single or two radiation outcomes before or during treatment based on the discovery or validation dataset were evaluated by Delong test [27,28], which can be described by p values as shown in Table 4. In the table, if the p -values are greater than or equal to 0.05, their original values are kept; otherwise, they are classified into different categories and replaced by corresponding labels “<0.05”, “<0.01” or “<0.001”. The testing results in the table show that:

- The differences between SA-BN and EK-NBN or LASSO for LC or / and RP2 prediction are significant (p -values<0.001 or 0.01 or 0.05).
- The differences between SA-BN and EYE or LASSO-EYE for RP2 prediction and the joint prediction of LC and RP2 are significant (p -values<0.001 or 0.01 or 0.05).
- SA-BN has similar prediction performance as PD-BN for LC or / and RP2 prediction (p -values> 0.05).
- SA-BN has similar prediction performance as EYE or LASSO-EYE for LC prediction (p -values> 0.05).

4. Discussion

4.1. Related studies on Bayesian networks and expert knowledge

Standard literature on BN construction describes in details the valuable contributions of the EK to help identify BN structures and parameters [29]. However, the construction and analysis of BNs and their variant influence diagrams are conducted mainly based on datasets with a limited number of features. In the field of radiation oncology, more and more patients’ information including their physical, clinical, biological, genetic, imaging features are becoming available from different resources for outcome prediction as shown in Table 2. Being a small part of these high-dimensional datasets, the EK features might not be selected by a pure data-driven approach such as the PD-BN for radiation outcomes prediction as shown in our previous study [3–5].

However, this does not mean that the EK expertise is not important anymore, which is to the contrary as shown in this resulting model with tighter confidence intervals. Although big data has the potential to help outcome prediction achieve a better performance, the data-driven approach may not be able to recognize and distinguish complex biophysical relationships from too many variables and limited sample sizes without human being's intervention, which has the potential to mislead physicians. Also, the process of identifying useful patterns to build the prediction model might not be efficient. Hence, by incorporating the EK features into the PD-BN method, we developed a novel SA-BN approach to explore EK involved biophysical pathways and identify the best treatment plans from maximizing the probability of a patient's getting tumor LC and minimizing the probability of his / her receiving RP2 before and during radiotherapy.

The EK has been intensively used to help BN development in the field of radiation oncology. In a study of distributed learning of lung cancer to allow the developments of prediction models on data originating from multiple hospitals while avoiding many of the data sharing barriers, both the EK and data-driven approach were used to determine the structure of the BN from known clinical and physical variables and the performances of these two structures were found to be similar [30,31]. In order to model the radiation therapy process of prostate cancer and prognostic indicators such as distant metastasis, rectal and bladder complications for more outcome-focused decision making, an influence diagram network / BN was developed using expert opinion, results of clinical trials, and published research [32]. Considering the critical outcomes of tumor eradication and normal tissue sparing, the role of PET was evaluated in the treatment of occult disease in head and neck cancer by using influence diagram to get the optimal policy with maximum expected utility. The structure of the influence diagram was developed from expert opinions in this study [33]. Unlike using either EK or pure data-driven approaches to develop the BNs / influence diagrams in these studies, the SA-BN approach incorporates the former into the latter for BN's construction in order to improve the credibility of radiation outcome prediction models.

In another study to evaluate the feasibility of BNs for personalized survival estimates and treatment selection recommendations based on the English Lung Cancer Database, a hybrid approach was developed by incorporating the EK into the learning process of the BNs, and its associated BNs had little effect on the Bayesian score or the predictive performances attained and helped yield structures that look more similar to the expert elicited structure [34]. Our SA-BN approach is different from this hybrid method mainly in the following two aspects. First, it includes a feature selection process, which can efficiently handle high-dimensional datasets obtained from different biophysical resources. Secondly, by assuming that the EK could be wrong, it can choose the right EK features for BN structure learning by using the EMBs or the EK-EMBs through the process of feature selection. Details of credibility and accuracy roles of our SA-BN approach for better informing radiation outcomes prediction in pART are further discussed below.

4.2. The credibility of the SA-BN approach for radiation outcomes prediction

The basic concept of credibility is increasing the explainability of a prediction model without compromising its prediction accuracy. While an EK-NBN is developed only based

on the EK features related to radiotherapy resulting in a good interpretation of patients' status in pART, its prediction accuracy performance is undesired compared to other presented models in Table 4 due to the limited information involved. Although a PD-BN built from all biophysical features has a better prediction accuracy compared to the EK-NBN as listed in Table 4, the EK features might not be selected as part of the PD-BN. It is harder for the physicians to relate or interpret the biophysical pathways in the PD-BN for clinical decision-making in pART. Then, neither the EK-NBN nor the PD-BN can by themselves meet the criterion of credibility.

The SA-BN approach is developed in this study with the intention of improving the credibility of PD-BN models by incorporating the EK features into the BN construction. Compared the SA-BNs with the PD-BNs [4] for the joint prediction of LC and RP2 before and during the radiotherapy, we can find that while the former keeps the most of the latter's important features, there exists some differences between them. Not being in the selected EK features' EMBs, some important features in the PD-BNs could be replaced by other variables in the EMBs. For example, SNPs "atm_Rs664143", "tp53_Rs1042522" appear in the SA-BNs due to the incorporation of the EK, and pre-treatment PET radiomics "pre_MTV" in the PD-BNs is replaced by "GLRLM_LGRE" in the SA-BNs as shown in Fig. 4. Similarly, pre-treatment cytokines "pre_IL_4" and "pre_IL_10" are replaced by "pre_eotaxin", the relative change of during treatment PET radiomics "RD_GLSZM_LZLGE" is replaced by the slope of the change in during treatment cytokine "SLP_GM_CSF".

In comparison to PD-BNs, SA-BNs display more common biophysical pathways related to radiation outcomes. For instance, smoking is the major factor along with other environmental and genetic risk factors; patients are eligible for certain treatments from surgery to radiation to chemotherapy as well as targeted therapy depending on the staging of lung cancer; carriers of TP53 germline sequence variations who also smoke are more than 3 times more likely to develop lung cancer than nonsmokers [35,36]. In addition to radiation treatment, patients' treatment outcomes also depend on their individual characteristics and other therapies. The integration of these important causes of lung cancer into the biophysical pathways for the joint prediction of LC and RP2 in a pre-treatment SA-BN as illustrated in Fig. 4(a) has the potential to help physicians understand the cause-effect of the radiation outcomes before identifying the best treatment plans for pART. For NSCLC patients with three-dimensional conformal radiation therapy, GTV has been identified as a highly-valued prognostic for overall and cause-specific survival and local tumor control compared to gender, race, histology, tumor stage and node stage, and may be important in stratification of patients in prospective therapy trials [37,38]. Our SA-BNs illustrate the importance of GTV in LC prediction and the joint prediction of LC and RP2.

The biophysical relationships explored by the SA-BN approach are supported by cited literatures. Tumor cell growth and migration can be directly regulated by chemokines such as eotaxin, and the level of eotaxin can be used to improve the creditability of tumor local progression before and during treatment [39]. Also, the contribution of IL-5 was examined in an experimental model of lung fibrosis induced by bleomycin. The findings show that IL-5 is a key mediator in the recruitment of lung eosinophils to exacerbate lung fibrosis by

secreting profibrotic mediators [40], which supports the role of IL-5 for RP2 prediction in SA-BNs as shown in Figs. 3(a) and 3(b). Moreover, the connection between pre-IL-8 and SLP-TGF-beta1 for during treatment RP2 prediction as illustrated in Fig. 3(b) is supported by a NSCLC validation study, which concluded that lower pre-treatment IL-8 and higher the change of TGF-beta during the treatment were associated with higher risk of RP2 [41]. The consideration of both LC and RP2 is usually related to overall survival (OS). In a systematic assessment of the clinicopathological prognostic significance of tissue cytokine expression for lung adenocarcinoma (LUAD) based on integrative analysis of TCGA data, it was observed that for LUAD patients diagnosed <45 years old, low expression of CXCL10 (encoding IP-10), CCL11 (encoding eotaxin), IL-15 (encoding IL-15) and CSF3 (encoding G-CSF) was associated with shorter OS [42]. The observation supports the explored biophysical pathway in a during treatment SA-BN for the joint prediction of LC and RP2 as illustrated in Fig. 4(b).

Furthermore, biophysical pathways explored from the SA-BNs can help gain physicians' trust in radiation outcomes prediction and treatment planning before and during the radiotherapy as shown in the following example. By employing Netica (developed by Norsys Software Corp. located in Vancouver, Canada) as an interface, the SA-BN model in Fig. 3(a) to predict RP2 before radiotherapy can also be represented by another format in Fig. 6. In the figure, while yellow rectangular boxes indicate the probability distribution of patients' characteristics in different categories at a population level, full-length bars associated with 100-percentage probability in shadowed rectangular boxes denote a patient's specific properties. According to physicians' knowledge and experience, a smoker with large GTV treated by chemotherapy usually has a relatively high probability to get RP2 under a standard (medium) radiation treatment level. To avoid the serious complication, the physicians may choose a low radiation dose to the patient instead of the medium treatment level.

Without knowing a patient's SNPs, miRNAs, and cytokines information, physicians can predict his / her potential of receiving RP2 based on the probability distribution of these features at the population level and the patient's EK features as described by 100-percentage bars in shadow nodes "total lung volume", "smoking", "chemo" under medium and low radiation treatment levels as illustrated in Figs. 6(a) and 6(b) respectively. The SA-BNs in these figures show that the probability of the patient's getting RP2 decreases from 43.4% to 12.6% after changing the treatment plan as indicated by a red arrow in Fig. 6(b), which is clinically significant. As additional information such as SNPs, miRNAs and cytokines become available as denoted by the 100-percentage bars of these nodes in Fig. 6(c), the physicians can then get more affirmatory RP2 prediction from the SA-BN by considering the low radiation dose as the best treatment plan.

The above is just an example to show how a SA-BN has the potential to help physicians identify the best treatment plan by only considering a patient's RP2. However, their decision-making process is more complicated in clinical practice, since whether the patient's tumor can be locally controlled or not with the reduced dose is another important concern in determining the best treatment plan. Then the SA-BNs for the joint prediction of LC and RP2 before and during the courses of radiation treatment as shown in Figs. 4(a) and 4(b) can

help the physicians identify the best treatment plan by balancing a patient's LC and RP2 based on their EK on the patient's personal characteristic and therapeutic satisfaction. On the other hand, except radiation treatment decision variables such as "Lung_gEUD" and "Tumor_gEUD", the rest of EK features did not show in the PD-BNs based on our previous research work [3–5]. Since the new findings of biophysical pathways among the SNPs, miRNAs, and cytokines in the PD-BNs have not become the physicians' common knowledge in clinical practice yet, the SA-BNs have a higher probability of gaining their trust than the PD-BNs for radiation outcomes prediction in pART.

Overall, the results of our data analysis show that a SA-BN supports or extends well-known biophysical understandings, explores more credible (known) relationships in radiotherapy, provides more confidence in physicians, and then it is more likely to be used as a human-machine interface to help physicians' clinical decision-making in pART compared to other models listed in Table 4. While the single-objective SA-BN in Fig. 2 or 3 specifies an EK involved biophysical pathway related to LC or RP2, the joint SA-BN in Fig. 4 describes a comprehensive one related to both. By exploring the differences among them, physicians can evaluate cross-talks between LC's and RP2's pathways. Moreover, the SA-BN approach has the similar prediction performance as the PD-BN method as shown in Tables 3 and 4. Therefore, the SA-BN is a credible outcome prediction model for pART. In addition, it demonstrated tighter confidence intervals, which is an important added value for future adoption.

4.3. The accuracy of the SA-BN approach for radiation outcomes prediction

Besides credibility, the accuracy of an outcome prediction model is another important aspect to gain physicians' trust for clinical decision-making in pART. While the SA-BN approach has similar accuracy performance as the EYE penalty method for LC prediction before and during radiotherapy, its AUC values for RP2 prediction are better than those of the linear credible approach as described in Tables 3 and 4. The impact of high-dimensional dataset on the performance of the EYE penalty approach was evaluated in this study. A LASSO-EYE method was developed by selecting important features from the NEK dataset based on the LASSO and considering them and EK features as inputs for EYE penalty analysis. Since the performance of the LASSO-EYE method for radiation outcomes prediction is similar as that of the EYE penalty approach as shown in Table 4, high-dimensional dataset is not the reason for the EYE penalty to have poor RP2 prediction.

In general, radiation oncology datasets are associated with a large proportion of LC (70% in our study) and a small proportion of radiation-induced toxicities such as RP2 (17% in our study) after radiation treatment. Then the performance of the EYE and its variants for RP2 prediction may have been affected by imbalanced class distribution in the radiation oncology datasets. Also, Table 4 shows that the difference between the SA-BN and EYE penalty approaches for the joint prediction of LC and RP2 is significant. In addition to low RP2 event rate in the discovery and validation datasets, another reason of the poor joint prediction is that LC and RP2 cannot be predicted simultaneously in a logistic regression model. Unlike a SA-BN with two objectives, the performance of the joint prediction can be

compromised by combining two logistic regression models from the EYE penalty approach for LC or RP2 prediction.

Without distinguishing EK and NEK features, the feature selection of the PD-BN approach depends on the EMB that only includes a radiation outcome's parents, descendants and spouses and their next-of-kin in a BN. Even though some EK features are strongly relevant to the outcome, they may not be selected if they are not within its EMB. To avoid the impact of NEK features on important EK features identification, the SA-BN's feature selection process directly explores a radiation outcome's EK-EMB from its EK dataset based on the targeted nominal type I error rate. After identifying important EK features from the EK-EMB (if possible), they are combined with critical EK features from the EMB (if possible) and important NEK features from the NEK-EMB as the inputs of SA-BN structure learning. In other words, even if no or few important EK features have been selected by the EMB in the PD-BN approach, the EK-EMB offers another chance to explore the radiation outcome's EK dataset and tries to incorporate as many EK features as possible into SA-BN structure learning for the credibility improvement.

However, if no EK feature is identified from both EMB and EK-EMB, it turns out that all the EK features are not critical to the radiation outcome prediction, and the structure of the SA-BN would be the same as that of the PD-BN. Therefore, the SA-BN's prediction accuracy is not likely to be worse than that of the PD-BN approach as shown in Tables 3 and 4. Moreover, while the important features selected from an EMB based on the constraint-based algorithm may vary in terms of different training folds during the cross-validation of the PD-BN approach, the integration of important EK and NEK features at the stage of feature selection has the potential to stabilize the prediction performance of the SA-BN approach. This could be the reason why its 95% CI is narrower compared to that of the PD-BN approach.

4.4. Limitations of our study

The SA-BN approach has the potential to improve the credibility and accuracy of radiation treatment outcome prediction compared to other methods. However, the number of NSCLC patients in the discovery dataset to build the SA-BNs for LC or / and RP2 prediction is limited and there are few events in the validation dataset for RP2 prediction performance evaluation. Then our approach is still at its infancy for gaining the physician's trust and providing better decision support for pART in clinical practice. While the highest RP value was adopted from the clinical evaluation and diagnosis imaging to classify RP according to CTCAE 3.0, this grade determination seems to be general but not quantitative, which is a limitation of present outcome studies. However, this is currently the clinical standard for evaluation. Future use of pulmonary function or advances in quantitative imaging may provide more objective measures to improve the performance of these models. Moreover, since PET is not the worldwide accepted approach and is not always available in clinical practice, CT scan may be used in our future research to expand the usability of the SA-BN approach.

The limitations of this study related to the credibility can be described as follows. The accumulated effect of the EK along the biophysical pathway is still unknown, and the

causal-effect relationships among the EK and NEK features also need to be further investigated. Since the EK in the current study is provided by two main lung physicians' experiences, involving more physicians' expertise, incorporating more EK such as causal-effect relation, monotonicity constraints into BN structure learning can further improve the SA-BNs' credibility for the realization of pART. Furthermore, the calibration of probabilities obtained from our SA-BN approach to reflect the likelihood of true events still needs to be evaluated in the next step of our research.

5. Conclusions

We developed a new SA-BN approach to improve credibility and enhance physicians' trust and help the clinical decision-making for pART. In addition to exploring biophysical pathways among patients' physical, clinical, biological, genomic and PET radiomics features, the SA-BN approach has the potential to help physicians understand why, when, and how to conduct radiation treatment for the improvement of patients' therapeutic satisfaction by incorporating EK into these biophysical pathways. Based on the nested cross-validation and external validation, the SA-BN approach outperforms other credible models such as the EYE penalty and the LASSO-EYE in terms of the joint prediction of LC and RP2. While its prediction accuracy is not significantly better than that of the PD-BN approach, it has smaller 95% CIs on the performance, and more trustable ways to assess patients' best treatment plans before and during the radiotherapy. As an accurate and credible model for radiation outcomes prediction, the SA-BN approach has the potential to be an important component of future pART. However, it still needs to be validated with external independent datasets via multi-institutional collaborations.

Acknowledgments:

This work was supported in part by the National Institutes of Health P01 CA059827, R37-CA222215 and R01-CA233487.

Appendix A.

Table A1

The selected features and their coefficients in outcome prediction models developed from LASSO and EYE approaches

Pre-treatment LC (215 features) (20 dosimetric / clinical features, 62 miRNAs, 60 SNPs, 43 PET radiomics, 30 Cytokines)		
Methods	Selected EK features	Selected NEK features
LASSO	Tumor gEUD (3.821 [*])	Rs2234671 (0.043), Rs1799796 (-0.595), Rs1040363 (-1.707), pre-ifn-gama (-0.033), pre-il10 (-0.040), pre-tgfa (-0.020), GLSZM-SZLGE (-0.018)
EYE	Chemo (0.05), Stage (-0.17), Age (-0.06), GTV (-0.11), BED (0.09), PTVD95 (-0.12), GTVD95 (-0.05), Tumor_gEUD (0.12)	Rs2234671 (0.06), Rs3857979 (-0.06), Rs1800795 (-0.05), Rs180925 (0.05), Rs689470 (-0.05), Rs7333607 (0.05), Rs1800469 (0.05), Rs1061622 (0.05), Rs609261 (0.05), Rs238406 (0.05), Rs1042522 (0.05), pre_il13 (0.06), pre_il1b (-0.05), pre_il5 (-0.06), pre_il8 (0.05), pre_tnfa (-0.05), miR_100_5p (-0.05), miR_143_3p (-0.06), miR_192_5p (0.06), miR_19b_3p (-0.05), miR_25_3p (0.06), miR_29a_3p (-0.05), miR_145_5p (0.06), miR_15a_5p (-0.05), miR_30e_5p (-0.06), miR_15b_5p (-0.07), miR_191_5p (-0.05), miR_20a_5p (0.06), miR_93_5p (0.05), pre_MTV (0.07),

		GLCM_Correlation (0.05), GLRLM_SRLGE (-0.06), GLRLM_LRHGE (-0.06), GLRLM_GLV (0.05)
During-treatment LC (288 features) (20 dosimetric / clinical features, 62 miRNAs, 60 SNPs, 86 PET radiomics, 60 Cytokines)		
<i>Methods</i>	<i>Selected EK features</i>	<i>Selected NEK features</i>
LASSO	Tumor gEUD (3.962), Treatment Duration (-0.018)	Rs2234671 (2.577), Rs25487 (-0.85), Rs1799796 (-3.208), Rs1040363 (-4.902), pre-ifn-gama (-0.055), pre-il10 (-0.049), pre-tgfa (-0.039), GLSZM-SZLGE (-0.016), GLCM-IDM (-0.508), SLP-tgfa (-0.067), RD-GLRLM-GLV (-0.126), RD-GLRLM-RP (0.238)
EYE	Chemo (0.11), Stage (-0.16), Age (-0.07), GTV (-0.22), Treatment Duration (-0.19), BED (0.18), PTVD95 (-0.15), GTVD95 (-0.07), Tumor_gEUD (0.19)	Rs2234671 (0.07), Rs235756 (-0.05), Rs2070874 (0.05), Rs180925 (0.05), Rs4760259 (-0.05), Rs4776342 (-0.05), miR_23a_3p (-0.06), miR_24_3p (0.05), GLCM_Homogeneity (-0.05), GLRLM_RLN (0.05), GLSZM_GLN (-0.05), RD_GLRLM_SRE (0.05), SLP_ip10 (0.05)
Pre-treatment RP2 (175 features) (23 dosimetric / clinical features, 62 miRNAs, 60 SNPs, 30 Cytokines)		
<i>Methods</i>	<i>Selected EK features</i>	<i>Selected NEK features</i>
LASSO	Lung gEUD (3.736)	Rs1040363 (-0.761), Rs1799796 (-0.418), Rs1800468 (0.345), miR-92a-3p (-0.229), miR-124-3p (0.156), pre-ifn-gama (-0.027), pre-il10 (-0.038), pre-tgfa (-0.010)
EYE	Chemo (0.09), Smoking (0.21), Age (0.11), Total Lung Volume (0.06), Lung gEUD (1.7)	Rs3857979 (-0.07), Rs235756 (-0.24), Rs12906898 (0.05), Rs1800057 (-0.05), Rs11615 (-0.05), Rs17655 (-0.05), Rs9293329 (-0.05), Rs1478486 (0.05), Rs2228001 (-0.05), miR_100_5p (-0.05), miR_125b_5p (0.05), miR_143_3p (-0.05), miR_17_3p (-0.07), miR_21_5p (0.06), miR_221_3p (-0.05), miR_23a_3p (-0.07), miR_25_3p (0.07), miR_296_5p (0.07), miR_423_5p (0.06), miR_193a_5p (-0.06), miR_7_5p (0.07), pre_il10 (-0.08), pre_il2 (-0.08), pre_il5 (0.07), pre_il7 (0.08), pre_mcp1 (0.06), pre_mip1a (-0.06), pre_tnfa (0.07)
During-treatment RP2 (205 features) (23 dosimetric / clinical features, 62 miRNAs, 60 SNPs, 60 Cytokines)		
<i>Methods</i>	<i>Selected EK features</i>	<i>Selected NEK features</i>
LASSO	Smoking (1.159), Treatment Duration (-0.269), Lung gEUD (4.092)	pre-TGF-beta1 (-0.610), Rs235756 (0.977), Rs11724777 (-3.143), Rs1800468 (4.680), Rs13181 (-3.733), Rs17655 (1.437), Rs1047768 (2.395), Rs25487 (-3.454), Rs6464268 (-1.377), Rs1799796 (-4.245), Rs1040363 (-8.268), miR-885-5p (0.276), pre-gmcsf (-0.005), pre-il10 (-0.093), pre-il12-p70 (0.056), pre-il6 (0.097), pre-tgfa (-0.1891), pre-tnfa (0.174), SLP-fractal (-0.019), SLP-il12-p40 (-0.007), SLP-mcp1 (0.009), SLP-mip1a (0.016), SLP-tgfa (-0.367),
EYE	Chemo (0.07), Smoking (0.14), Age (0.12), Treatment Duration (-0.14), Lung gEUD (1.2)	Rs1801275 (0.05), Rs4760259 (0.05), Rs689470 (0.05), Rs1800469 (-0.06), Rs1047768 (-0.07), Rs12917 (0.06), Rs1805794 (-0.05), Rs1625895 (0.06), Rs1042522 (-0.06), Rs2075685 (-0.06), pre_eotaxin (0.05), pre_il13 (0.07), pre_il2 (0.05), pre_il4 (-0.05), pre_il7 (-0.05), pre_mcp1 (0.14), pre_tgfa (-0.06), pre_tnfa (-0.05), miR_122_5p (0.05), miR_125b_5p (-0.07), miR_143_3p (0.05), miR_223_3p (-0.06), miR_224_5p (0.06), miR_27a_3p (-0.06), miR_885_5p (-0.06), miR_92a_3p (0.05), miR_193a_5p (0.07), miR_103a_3p (-0.05), miR_93_5p (0.05), miR_16_5p (0.06), SLP_eotaxin (-0.05), SLP_ifn_gamma (-0.06), SLP_il1a (0.05), SLP_il6 (0.05), SLP_tnfa (0.05), SLP_TGFbeta1 (-0.12)

* the coefficient of a feature based on a prediction model.

LC means “local control”; RP2 means “radiation pneumonitis with grade 2”; EK-NBN means “expert knowledge based naïve Bayesian network”; PD-BN means “pure data-driven Bayesian network”; SA-BN means “situational awareness Bayesian network”; EYE means “expert yielded estimates”; LASSO means “least absolute shrinkage and selection operator”; gEUD means “generalized equivalent uniform dose”; EK means “expert knowledge”; RD means “relative difference”; SLP means “slope of changes”; GTV means “gross tumor volume”; BED means “biologically effective dose”; PTV means “planning target volume”; GLSZM means “gray level size zone matrix”; SZLGE means “small zone low gray-level”; GLCM means “grey level co-occurrence matrix”; GLRLM means “grey level run length matrix”; GLV means “gray level variance”; RP means “run percentage”; IDM means “inverse difference moment”; RLN means “run-length

nonuniformity”; **GLN** means “gray-level nonuniformity”; **SRE** means “short run emphasis”; **SNP** means “single nucleotide polymorphism”; **Rs** means “RefSNPs”; **miR** means “microRNA”.

References

- [1]. Tseng HH, Luo Y, Ten Haken RK, El Naqa I. The Role of Machine Learning in Knowledge-Based Response-Adapted Radiotherapy. *Front Oncol* 2018;8:266. 10.3389/fonc.2018.00266. [PubMed: 30101124]
- [2]. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. 2018 Ieee 5th Int Conf Data Sci Adv Anal 2018:80–9. 10.1109/Dsaa.2018.00018.
- [3]. Luo Y, McShan D, Ray D, Matuszak M, Jolly S, Lawrence T, et al. Development of a Fully Cross-Validated Bayesian Network Approach for Local Control Prediction in Lung Cancer. *IEEE Trans Radiat Plasma Med Sci* 2018. 10.1109/TRPMS.2018.2832609.
- [4]. Luo Y, McShan DL, Matuszak MM, Ray D, Lawrence TS, Jolly S, et al. A multiobjective Bayesian networks approach for joint prediction of tumor local control and radiation pneumonitis in nonsmall-cell lung cancer (NSCLC) for response-adapted radiotherapy. *Med Phys* 2018;45:3980–95. 10.1002/mp.13029.
- [5]. Luo Y, El Naqa I, McShan DL, Ray D, Lohse I, Matuszak MM, et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. *Radiother Oncol* 2017;123:85–92. 10.1016/j.radonc.2017.02.004. [PubMed: 28237401]
- [6]. Wang JX, Oh J, Wang HZ, Wiens J. Learning Credible Models. *Kdd'18 Proc 24th Acn Sigkdd Int Conf Knowl Discov Data Min* 2018:2417–26. 10.1145/3219819.3220070.
- [7]. Ben-David A Monotonicity Maintenance in Information-Theoretic Machine Learning Algorithms. *Mach Learn* 1995;19:29–43. 10.1023/A:1022655006810.
- [8]. Pazzani MJ, Mani S, Shankle WR. Acceptance of rules generated by machine learning among medical experts. *Methods Inf Med* 2001;40:380–5. [PubMed: 11776735]
- [9]. Martens D, Vanthienen J, Verbeke W, Baesens B. Performance of classification models from a user perspective. *Decis Support Syst* 2011;51:782–93. 10.1016/j.dss.2011.01.013.
- [10]. Endsley RM. Situation Awareness Misconceptions and Misunderstandings. *J Cogn Eng Decis Mak* 2015;9:4–32. 10.1177/1555343415572631.
- [11]. Wright MC, Taekman JM, Endsley MR. Objective measures of situation awareness in a simulated medical environment. *Qual Saf Heal Care* 2004;13 Suppl 1:i65–71. 10.1136/qhc.13.suppl_1.i65.
- [12]. Perera HA, Drew AC, Johnson JC, editors. *Expert Knowledge and Its Application in Landscape Ecology*. Springer Science & Business Media; 2012. 10.1007/978-1-4614-1034-8.
- [13]. Gennatas ED, Friedman JH, Ungar LH, Pirracchio R, Eaton E, Reichmann LG, et al. Expert-augmented machine learning. *Proc Natl Acad Sci* 2020;117:4571 LP – 4577. 10.1073/pnas.1906831117. [PubMed: 32071251]
- [14]. Sun J, Hu J, Luo D, Markatou M, Wang F, Edabollahi S, et al. Combining knowledge and data driven insights for identifying risk factors using electronic health records. *AMIA . Annu Symp Proceedings AMIA Symp* 2012;2012:901–10.
- [15]. Kong F-MS, Frey KA, Quint LE, Ten Haken RK, Hayman JA, Kessler M, et al. A pilot study of [18F]fluorodeoxyglucose positron emission tomography scans during and after radiation-based therapy in patients with non small-cell lung cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 2007;25:3116–23. 10.1200/JCO.2006.10.3747.
- [16]. Zhao L, West BT, Hayman JA, Lyons S, Cease K, Kong F-M. High radiation dose may reduce the negative effect of large gross tumor volume in patients with medically inoperable early-stage non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 2007;68:103–10. 10.1016/j.ijrobp.2006.11.051. [PubMed: 17363189]
- [17]. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6. 10.1016/j.ejca.2011.11.036. [PubMed: 22257792]
- [18]. Kouloulia V, Zygogianni A, Efstathopoulos E, Victoria O, Christos A, Pantelis K, et al. Suggestion for a new grading scale for radiation induced pneumonitis based on radiological

- findings of computerized tomography: correlation with clinical and radiotherapeutic parameters in lung cancer patients. *Asian Pac J Cancer Prev* 2013;14:2717–22. 10.7314/apjcp.2013.14.5.2717. [PubMed: 23803021]
- [19]. Yu H, Wu H, Wang W, Jolly S, Jin J-Y, Hu C, et al. Machine Learning to Build and Validate a Model for Radiation Pneumonitis Prediction in Patients with Non–Small Cell Lung Cancer. *Clin Cancer Res* 2019;25:4343 LP – 4350. 10.1158/1078-0432.CCR-18-1084. [PubMed: 30992302]
- [20]. Wang W, Xu Y, Schipper M, Matuszak MM, Ritter T, Cao Y, et al. Effect of normal lung definition on lung dosimetry and lung toxicity prediction in radiation therapy treatment planning. *Int J Radiat Oncol Biol Phys* 2013;86:956–63. 10.1016/j.ijrobp.2013.05.003. [PubMed: 23845844]
- [21]. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;162:55–U103. 10.7326/M14-0697. [PubMed: 25560714]
- [22]. Aliferis CF, Tsamardinos I, Statnikov A. HITON: a novel Markov Blanket algorithm for optimal variable selection. *AMIA. Annu Symp Proceedings AMIA Symp* 2003;2003:21–5.
- [23]. Tsamardinos I, Aliferis C, Statnikov A. Algorithms for Large Scale Markov Blanket Discovery. *Proc. Sixt. Int. Florida Artif. Intell. Res. Soc. Conf*, 2003, p. 376–81.
- [24]. Glover F Tabu Search: A Tutorial. *INFORMS J Appl Anal* 1990;20:74–94. 10.1287/inte.20.4.74.
- [25]. F I, L PJ. Markov Equivalence in Bayesian Networks. In: L P, A J, G AS, editors. *Adv. Probabilistic Graph. Model. Stud. Fuzziness Soft Comput*, vol. 213, Berlin, Heidelberg: Springer; 2007. 10.1007/978-3-540-68996-6_1.
- [26]. Bandos AI, Rockette HE, Song T, Gur D. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics* 2009;65:247–56. 10.1111/j.1541-0420.2008.01049.x. [PubMed: 18479482]
- [27]. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* 1988;44:837–45. 10.2307/2531595. [PubMed: 3203132]
- [28]. Sun X, Xu W. Fast Implementation of DeLong’s Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves. *IEEE Signal Process Lett* 2014;21:1389–93. 10.1109/LSP.2014.2337313.
- [29]. Kjærulff BU, Madsen LA. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. New York, NY: Springer; 2013. 10.1007/978-1-4614-5104-4.
- [30]. Jochems A, Deist TM, van Soest J, Eble M, Bulens P, Coucke P, et al. Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital - A real life proof of concept. *Radiother Oncol* 2016;121:459–67. 10.1016/j.radonc.2016.10.002. [PubMed: 28029405]
- [31]. Jochems A, Deist TM, El Naqa I, Kessler M, Mayo C, Reeves J, et al. Developing and Validating a Survival Prediction Model for NSCLC Patients Through Distributed Learning Across 3 Countries. *Int J Radiat Oncol Biol Phys* 2017;99:344–52. 10.1016/j.ijrobp.2017.04.021. [PubMed: 28871984]
- [32]. Smith WP, Kim M, Holdsworth C, Liao J, Phillips MH. Personalized treatment planning with a model of radiation therapy outcomes for use in multiobjective optimization of IMRT plans for prostate cancer. *Radiat Oncol* 2016;11. [PubMed: 26822015]
- [33]. Phillips MH, Smith WP, Parvathaneni U, Laramore GE. Role of positron emission tomography in the treatment of occult disease in head-and-neck cancer: a modeling approach. *Int J Radiat Oncol Biol Phys* 2011;79:1089–95. 10.1016/j.ijrobp.2009.12.037. [PubMed: 20510538]
- [34]. Sesen MB, Nicholson AE, Banares-Alcantara R, Kadir T, Brady M. Bayesian networks for clinical decision support in lung cancer care. *PLoS One* 2013;8:e82349. 10.1371/journal.pone.0082349. [PubMed: 24324773]
- [35]. Hwang SJ, Cheng LSC, Lozano G, Amos CI, Gu XJ, Strong LC. Lung cancer risk in germline p53 mutation carriers: association between an inherited cancer predisposition, cigarette smoking, and cancer risk. *Hum Genet* 2003;113:238–43. 10.1007/s00439-003-0968-7. [PubMed: 12802680]

- [36]. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* 2016;5:288–300. 10.21037/tlcr.2016.06.07. [PubMed: 27413711]
- [37]. Bradley JD, Ieumwananonthachai N, Purdy JA, Wasserman TH, Lockett MA, Graham MV, et al. Gross tumor volume, critical prognostic factor in patients treated with three-dimensional conformal radiation therapy for non-small-cell lung carcinoma. *Int J Radiat Oncol Biol Phys* 2002;52:49–57. 10.1016/S0360-3016(01)01772-2. [PubMed: 11777621]
- [38]. Pierson C, Grinchak T, Sokolovic C, Holland B, Parent T, Bowling M, et al. Response criteria in solid tumors (PERCIST/RECIST) and SUVmax in early-stage non-small cell lung cancer patients treated with stereotactic body radiotherapy. *Radiat Oncol* 2018;13. [PubMed: 29378610]
- [39]. Levina V, Nolen BM, Marrangoni AM, Cheng P, Marks JR, Szczepanski MJ, et al. Role of Eotaxin-1 Signaling in Ovarian Cancer. *Clin Cancer Res* 2009;15:2647–56. 10.1158/1078-0432.Ccr-08-2024. [PubMed: 19351767]
- [40]. Huaux F, Gharaee-Kermani M, Liu TJ, Morel V, McGarry B, Ullenbruch M, et al. Role of eotaxin-1 (CCL11) and CC chemokine receptor 3 (CCR3) in bleomycin-induced lung injury and fibrosis. *Am J Pathol* 2005;167:1485–96. 10.1016/S0002-9440(10)61235-7. [PubMed: 16314464]
- [41]. Wang SL, Campbell J, Stenmark MH, Zhao J, Stanton P, Matuszak MM, et al. Plasma Levels of IL-8 and TGF-beta 1 Predict Radiation-Induced Lung Toxicity in Non-Small Cell Lung Cancer: A Validation Study. *Int J Radiat Oncol Biol Phys* 2017;98:615–21. 10.1016/j.ijrobp.2017.03.011. [PubMed: 28581403]
- [42]. Dong YM, Liu Y, Bai H, Jiao SC. Systematic assessment of the clinicopathological prognostic significance of tissue cytokine expression for lung adenocarcinoma based on integrative analysis of TCGA data. *Sci Rep* 2019;9. [PubMed: 30626887]

- A situational awareness Bayesian network is developed based on expert knowledge
- It enables exploring biophysical pathways starting with the expert knowledge
- It allows physicians to conduct their familiar “what if” counterfactual inference
- It outperforms other credible models for the joint prediction of treatment outcomes
- It has the potential to be a key component of personalized adaptive radiotherapy

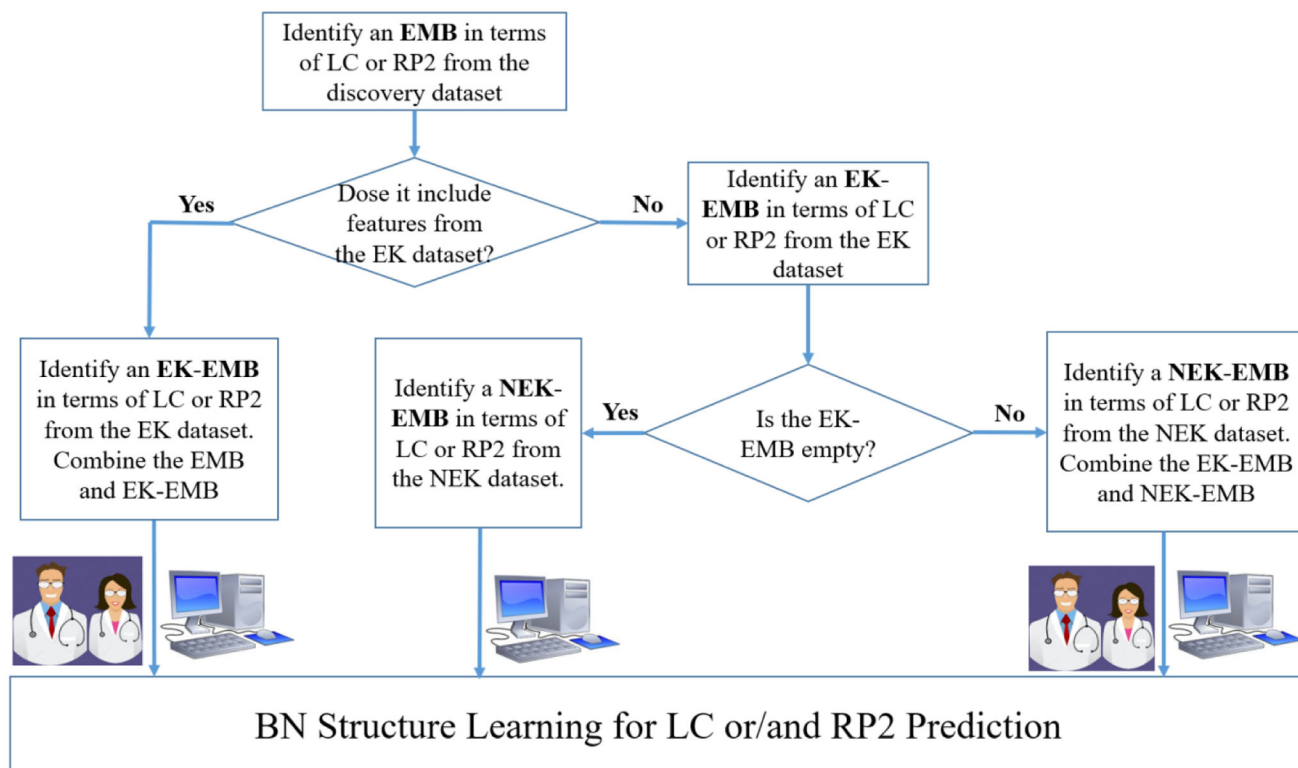


Fig. 1. SA-BN approach with focus on feature selection. EK dataset means “dataset with features from expert knowledge”; NEK dataset means “dataset without features from expert knowledge”; EK-EMB means “extended Markov blanket based on the EK dataset”; NEK-EMB means “extended Markov blanket based on the NEK dataset”.

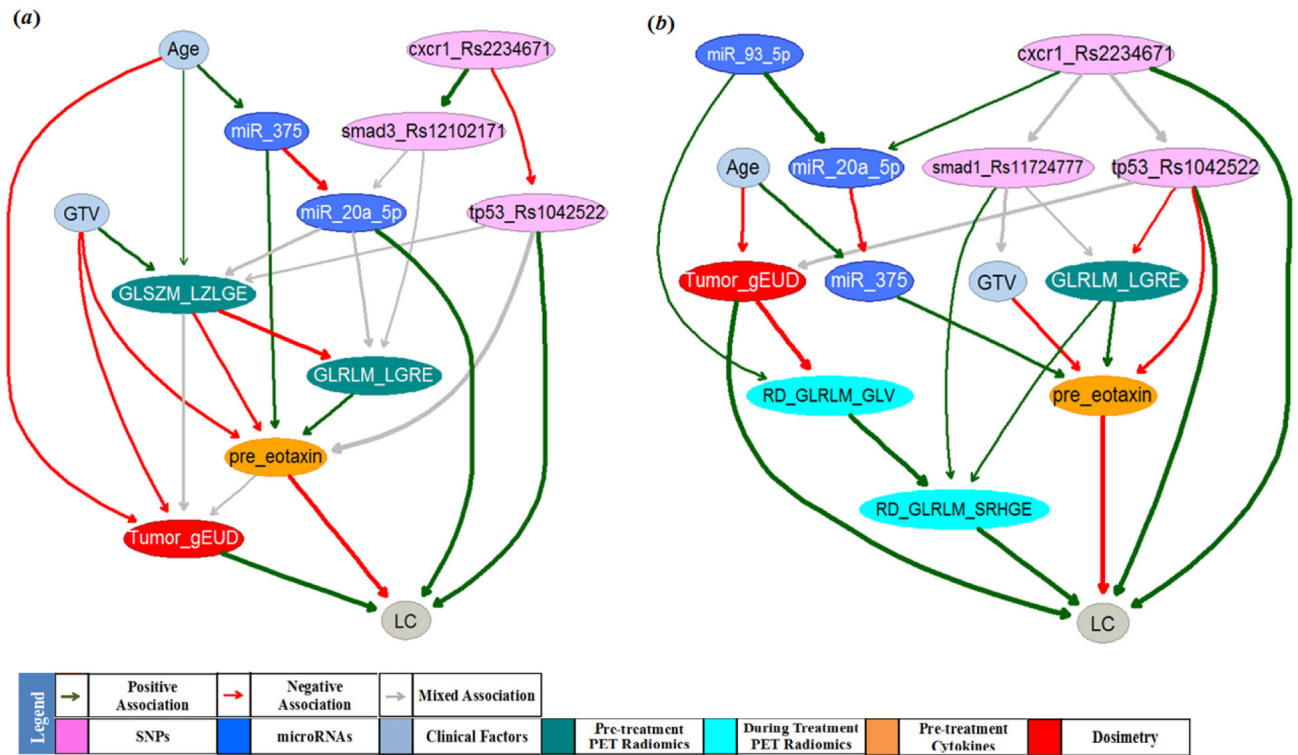


Fig. 2. Pre (a) and during (b) treatment SA-BNs for LC prediction based on the discovery dataset.

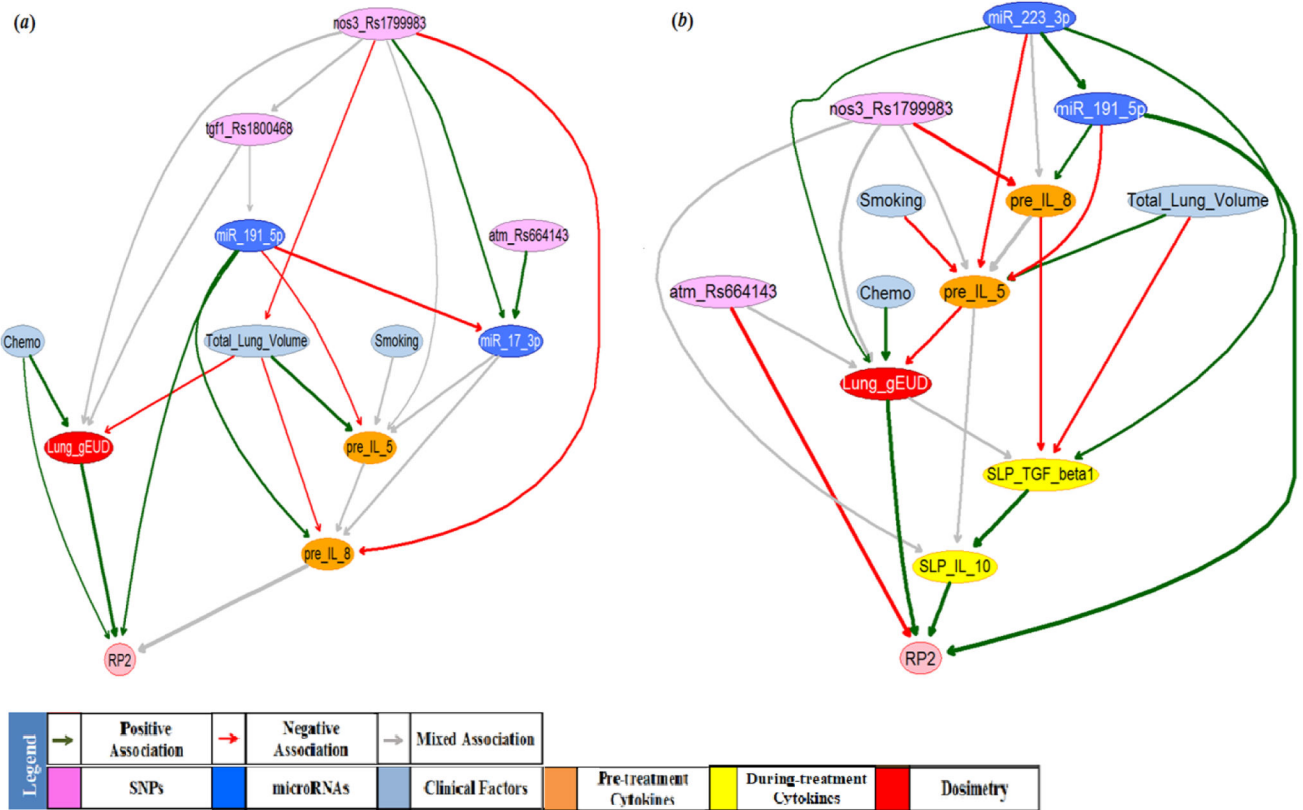


Fig. 3. Pre (a) and during (b) treatment SA-BNs for RP2 prediction based on the discovery dataset.

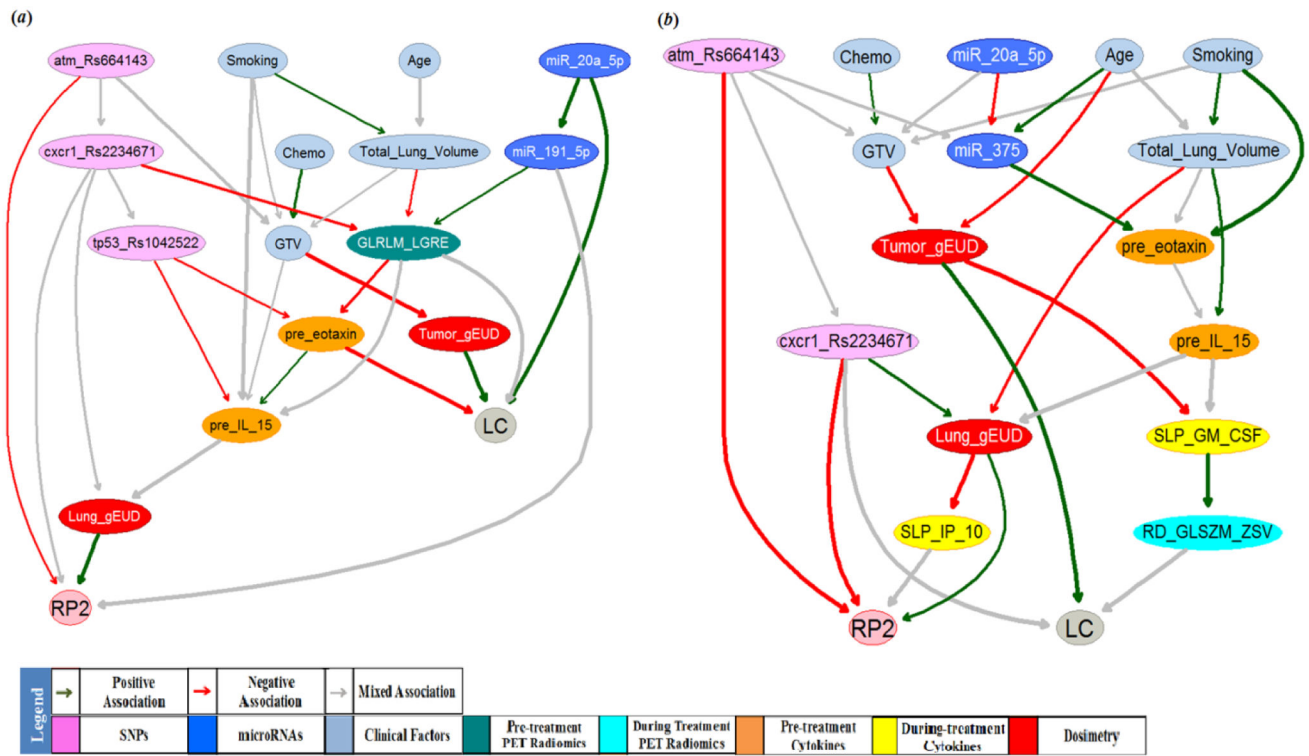


Fig. 4. Pre (a) and during (b) treatment SA-BNs for the joint prediction of LC and RP2 based on the discovery dataset.

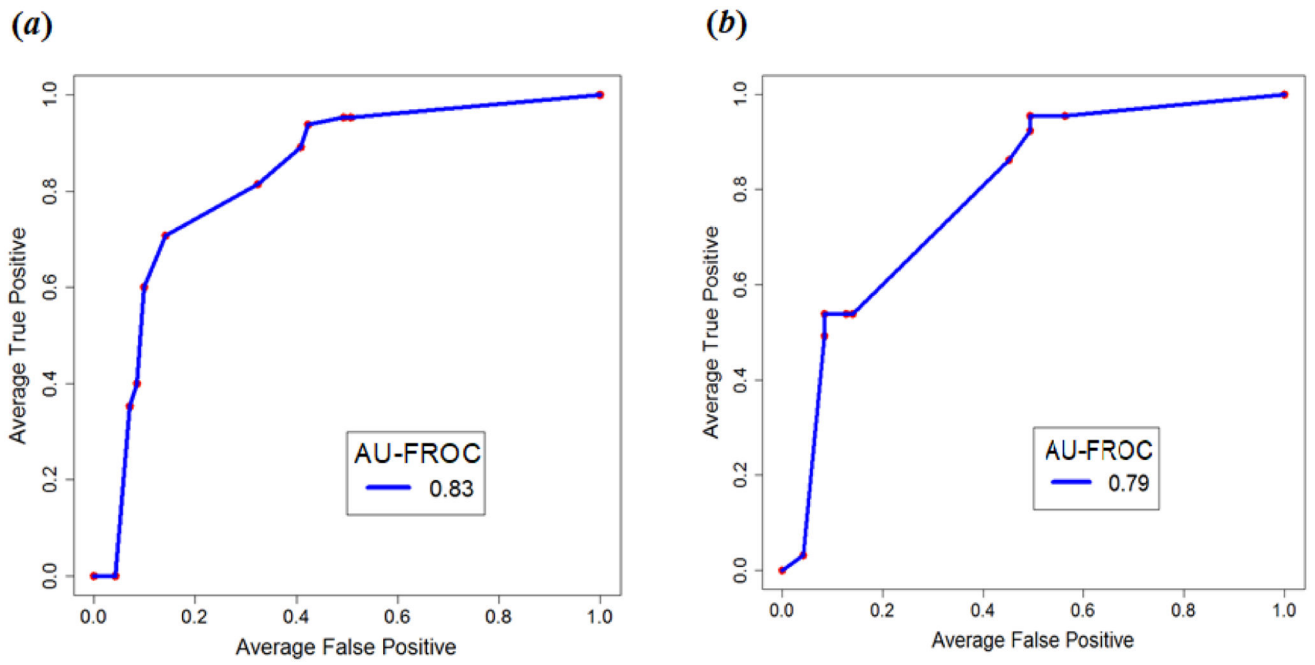
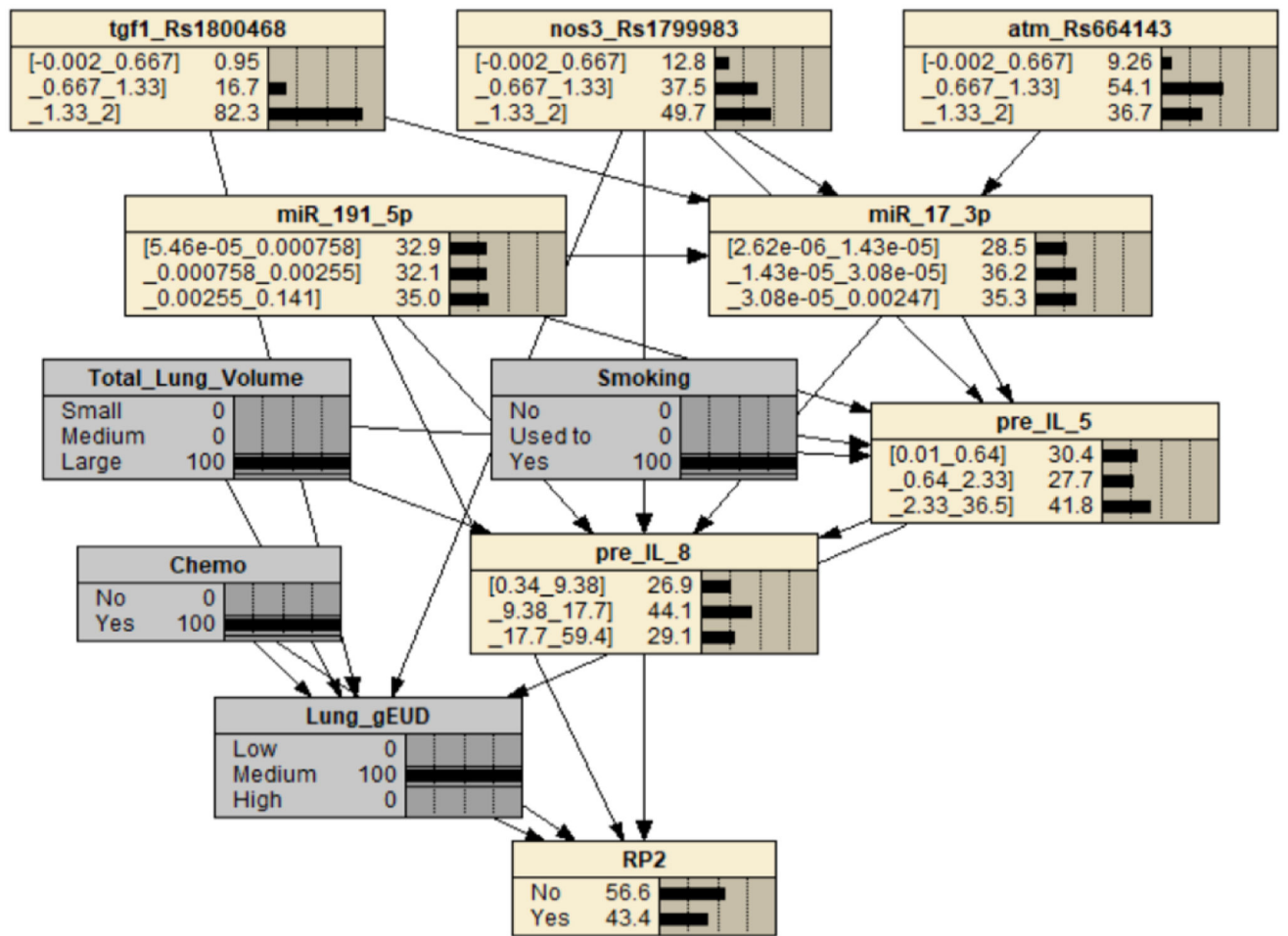
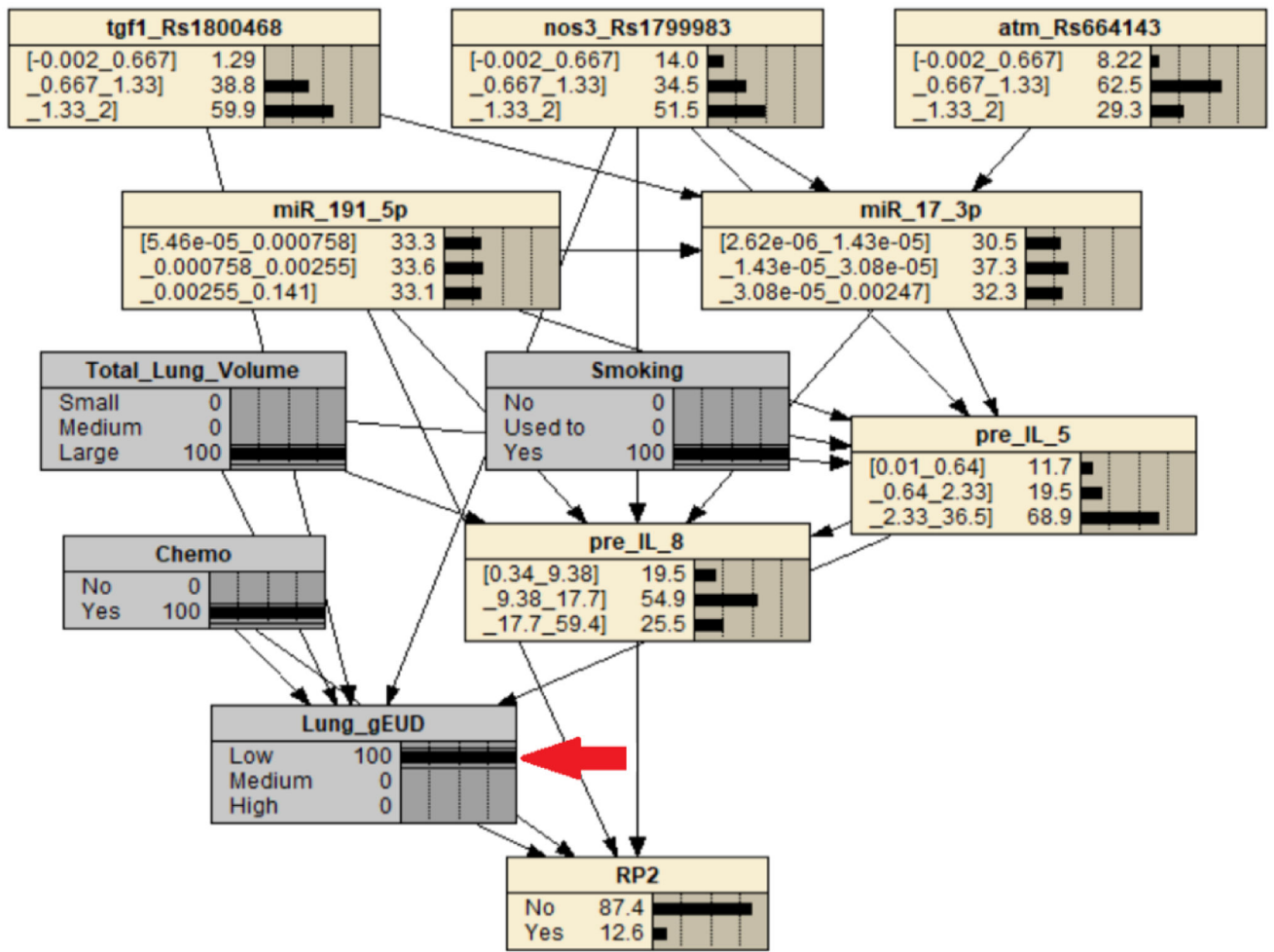


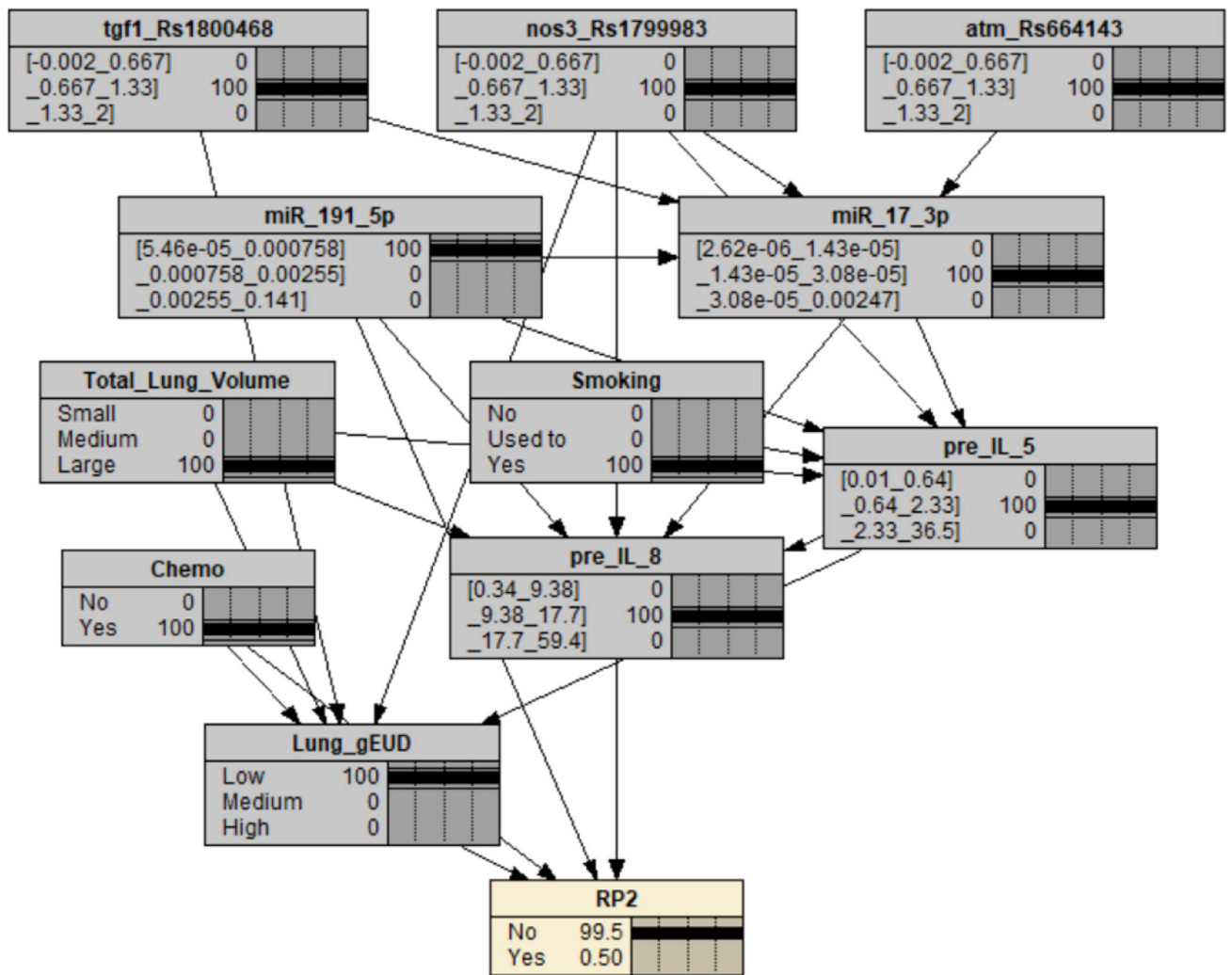
Fig. 5. (a) The performance of the SA-BN approach based on the discovery dataset and (b) the performance of the SA-BN illustrated in Fig. 4(b) based on the validation dataset for the joint prediction of LC and RP2 during radiation treatment.



(a)



(b)



(c)

Fig. 6. RP2 prediction with the EK under medium (a) and low (b) radiation dose levels; (c) RP2 prediction with a patient’s full information under a low radiation dose level.

Table 1

NSCLC patients' information in datasets.

Datasets	# of patients	# of patients with LC	# of patients with RP2	Median follow-ups of surviving patients (months)
Discovery	68	48	17	61
Validation	50	38	3	65

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Number of biophysical features in the whole dataset, LC's, and RP's feature datasets.

Categories	# of features in the whole dataset	# of features in LC's feature dataset	# of features in RP2's feature dataset
Common dosimetric information	15	6 [*]	9 ^{**}
Clinical factors	14	12 [*]	10 ^{**}
MicroRNAs (miRNAs)	62	62	62
Single nucleotide polymorphisms (SNPs)	60	60	60
Pre-treatment positron emission tomography (PET) radiomics	43	43 ^{***}	0
Relative difference (RD) of PET radiomics during treatment	43	43 ^{***}	0
Pre-treatment cytokines	30	30	30
Slopes (SLP) of cytokines change during treatment	30	30	30
Total	297	286	205

* Dosimetric information and clinical factors that are related to LC prediction.

** Dosimetric information and clinical factors that are related to RP2 prediction.

*** PET radiomics features before and during radiotherapy that are only related to LC prediction.

Table 3

Performance evaluation (AUC / AU-FROC) of the SA-BN for LC or / and RP2 prediction before and during treatment based on the discovery and validation datasets.

Time Points	Performance Evaluation					
	LC Prediction (AUC)		RP2 Prediction (AUC)		Joint LC and RP2 Prediction (AU-FROC)	
	Discovery*	Validation**	Discovery	Validation	Discovery	Validation
Pre-Treatment	0.76	0.72	0.77	0.75	0.79	0.76
During Treatment	0.81	0.78	0.82	0.79	0.83	0.79

* performance is evaluated via nested cross-validations based on the discovery dataset.

** performance is evaluated via external validations based on the validation dataset.

AUC means “area under receiver operating characteristics (ROC) curve”; **AU-FROC** means “area under free-response ROC curve”; **LC** means “local control”; **RP2** means “radiation pneumonitis with grade 2”.

Table 4

The performance of prediction models and their differences compared to the SA-BNs

Scenarios	Datasets	Time Points	LC Only		RP2 Only		Joint LC and RP2	
			AUC	<i>p</i> value *	AUC	<i>p</i> value	AU-FROC	<i>p</i> value
EK-NBN	Discovery **	Pre	0.61	<0.001	0.62	<0.001	0.70	<0.01
		During	0.61	<0.001	0.62	<0.001	0.70	<0.05
	Validation ***	Pre	0.54	<0.001	0.56	<0.001	0.63	<0.001
		During	0.54	<0.001	0.56	<0.001	0.63	<0.01
PD-BN	Discovery	Pre	0.75	0.51	0.76	0.56	0.78	0.55
		During	0.80	0.57	0.82	0.60	0.83	0.60
	Validation	Pre	0.76	0.29	0.78	0.34	0.77	0.35
		During	0.79	0.31	0.82	0.37	0.79	0.36
EYE	Discovery	Pre	0.74	0.31	0.60	<0.001	0.66	<0.01
		During	0.76	0.37	0.63	<0.001	0.70	<0.01
	Validation	Pre	0.63	0.16	0.53	<0.001	0.60	<0.001
		During	0.66	0.24	0.57	<0.001	0.65	<0.001
LASSO	Discovery	Pre	0.53	<0.001	0.58	<0.001	0.57	<0.001
		During	0.50	<0.001	0.51	<0.001	0.53	<0.001
	Validation	Pre	0.51	<0.001	0.52	<0.001	0.54	<0.001
		During	0.50	<0.001	0.50	<0.001	0.52	<0.001
LASSO-EYE	Discovery	Pre	0.76	0.32	0.61	<0.001	0.71	<0.05
		During	0.78	0.39	0.66	<0.01	0.75	<0.05
	Validation	Pre	0.64	0.18	0.55	<0.001	0.61	<0.001
		During	0.68	0.25	0.58	<0.001	0.65	<0.01

* The difference of prediction performance compared to the corresponding SA-BN.

** The dataset to evaluate an approach's prediction performance via nested cross-validation.

*** The dataset to evaluate an outcome prediction model's performance via external validation.

LC means "local control"; RP2 means "radiation pneumonitis with grade 2"; CI means "confidential interval"; EK-NBN means "expert knowledge based naïve Bayesian network"; PD-BN means "pure data-driven Bayesian network"; SA-BN means "situational awareness Bayesian network"; EYE means "expert yielded estimates"; LASSO means "least absolute shrinkage and selection operator".