



OPEN ACCESS

EDITED BY

Tonghe Wang,
Emory University, United States

REVIEWED BY

Kevin Camphausen,
National Cancer Institute (NIH),
United States
Jung Hun Oh,
Memorial Sloan Kettering Cancer
Center, United States

*CORRESPONDENCE

Yi Luo
Yi.Luo@moffitt.org

SPECIALTY SECTION

This article was submitted to
Radiation Oncology,
a section of the journal
Frontiers in Oncology

RECEIVED 04 October 2022

ACCEPTED 01 November 2022

PUBLISHED 09 December 2022

CITATION

Luo Y, Cuneo KC, Lawrence TS,
Matuszak MM, Dawson LA, Niraula D,
Ten Haken RK and El Naqa I (2022) A
human-in-the-loop based Bayesian
network approach to improve
imbalanced radiation outcomes
prediction for hepatocellular cancer
patients with stereotactic
body radiotherapy.
Front. Oncol. 12:1061024.
doi: 10.3389/fonc.2022.1061024

COPYRIGHT

© 2022 Luo, Cuneo, Lawrence,
Matuszak, Dawson, Niraula, Ten Haken
and El Naqa. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

A human-in-the-loop based Bayesian network approach to improve imbalanced radiation outcomes prediction for hepatocellular cancer patients with stereotactic body radiotherapy

Yi Luo^{1*}, Kyle C. Cuneo², Theodore S. Lawrence²,
Martha M. Matuszak², Laura A. Dawson³, Dipesh Niraula¹,
Randall K. Ten Haken² and Issam El Naqa¹

¹Department of Machine Learning, Moffitt Cancer Center, Tampa, FL, United States, ²Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States, ³Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada

Background: Imbalanced outcome is one of common characteristics of oncology datasets. Current machine learning approaches have limitation in learning from such datasets. Here, we propose to resolve this problem by utilizing a human-in-the-loop (HITL) approach, which we hypothesize will also lead to more accurate and explainable outcome prediction models.

Methods: A total of 119 HCC patients with 163 tumors were used in the study. 81 patients with 104 tumors from the University of Michigan Hospital treated with SBRT were considered as a discovery dataset for radiation outcomes model building. The external testing dataset included 59 tumors from 38 patients with SBRT from Princess Margaret Hospital. In the discovery dataset, 100 tumors from 77 patients had local control (LC) (96% of 104 tumors) and 23 patients had at least one grade increment of ALBI (I-ALBI) during six-month follow up (28% of 81 patients). Each patient had a total of 110 features, where 15 or 20 features were identified by physicians as expert knowledge features (EKFs) for LC or I-ALBI prediction. We proposed a HITL based Bayesian network (HITL-BN) approach to enhance the capability of selecting important features from imbalanced data in terms of accuracy and explainability through humans' participation by integrating feature importance ranking and Markov blanket algorithms. A pure data-driven Bayesian network (PD-BN) method was applied to the same discovery dataset of HCC patients as a benchmark.

Results: In the training and testing phases, the areas under receiver operating characteristic curves of the HITL-BN models for LC or I-ALBI prediction during SBRT are 0.85 (95% confidence interval: 0.75-0.95) or 0.89 (0.81-0.95) and 0.77

or 0.78, respectively. They significantly outperformed the during-treatment PD-BN model in predicting LC or I-ALBI based on the discovery cross-validation and testing datasets from the Delong tests.

Conclusion: By allowing the human expert to be part of the model building process, the HITL-BN approach yielded significantly improved accuracy as well as better explainability when dealing with imbalanced outcomes in the prediction of post-SBRT treatment response of HCC patients when compared to the PD-BN method.

KEYWORDS

accuracy and explainability, Bayesian networks, human-in-the-loop, hepatocellular cancer, outcome prediction, stereotactic body radiotherapy

1 Introduction

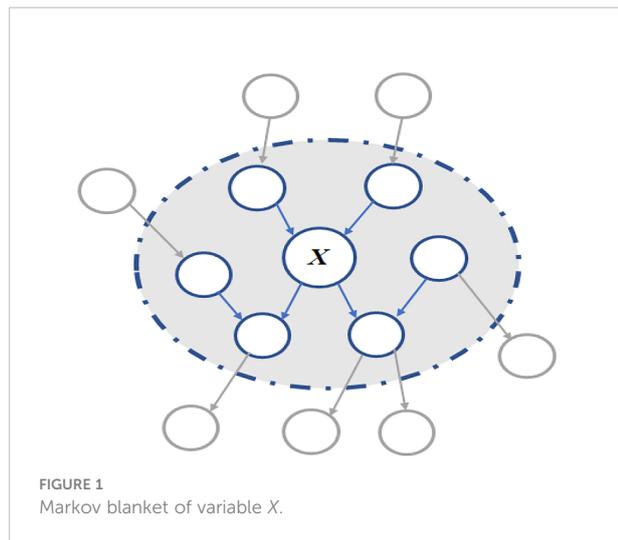
Hepatocellular cancer (HCC) is the third leading cause of cancer death worldwide. In 2020, the American Society of Clinical Oncology (ASCO) estimated that 830,180 people around the world died from the disease. While radiotherapy is designed to achieve tumor local control (LC) in HCC patients, it may also lead to radiation-induced toxicities (RITs). As a relatively newer radiation treatment technique, stereotactic body radiation therapy (SBRT) uses focused beams of radiation aimed at the tumor from many different angles given in one to five treatment fractions. Thus, the aim of SBRT is to cure tumors in the meanwhile decreasing the radiation to nearby healthy tissues. While it is more effective for tumor LC and RITs reduction compared to conventional approaches, stringent dose volume constraints of SBRT require the treatment planning to be highly personalized to meet its intended goals (1).

In HCC SBRT, LC can be evaluated radiologically from a lesion that is no longer arterially enhancing and has not spread to neighboring lymph nodes without any failures within the irradiated area over long-term follow-up (2). The impact of RITs to baseline liver function of HCC patients before and after SBRT can be evaluated by albumin-bilirubin (ALBI) grades for personalized standard or adaptive implementation (3–5). Specifically, physicians are concerned whether patients' ALBI grades will increase at least by one grade or not during 6-month follow-up, which is denoted as I-ALBI. Thus, we considered I-ALBI as another relevant SBRT outcome in addition to tumor LC in this study. The literature on outcomes prediction models for HCC patients with SBRT and their explainability capability remains limited and challenging (6, 7). The purpose of this study is to fill these gaps by developing accurate and explainable LC or I-ALBI prediction models for HCC patients with SBRT.

In clinical practice, oncology datasets usually have high dimensional features with limited sample size making

susceptible to spurious correlations including the Simpson paradox (8). The dataset of HCC patients with SBRT in this study is not an exception. Machine learning (ML) is defined as the task of extracting information from possibly high-dimensional and noisy data to give some guarantees of performance on unseen data. However, extracting the structure based on the proximity between empirical and population densities becomes challenging in the higher dimensions, since the distance between objects may be heavily dominated by noise, and the associated optimization process has an exponential dependency on these dimensions (9). Then, feature selection is designed to help conventional ML approaches handle high-dimensional datasets. For example, in our previous study on personalized adaptive radiotherapy for non-small-cell lung cancer patients, a pure data-driven Bayesian network (PD-BN) approach is developed including feature selection and BN structure building two steps. While Markov blanket (MB) algorithms were employed in the first step to identify the most important features from high-dimensional oncology datasets, Tabu Search was used in the second step to learn network structure based on the selected features. In addition to unraveling the biophysical relationships among lung cancer patients' personal characteristics, radiation treatment, and outcomes, the PD-BNs can predict lung tumor LC or/and RITs and identify the best treatment strategies before and during the radiotherapy to improve patients' therapeutic satisfaction (10–12).

Initially proposed by Pearl (13), the concept of variable X 's MB is to identify its optimal feature subset containing strongly relevant and non-redundant features, such as the variable's parents, children, and spouses as shown in the shadow area of Figure 1. Given these features in the subset, the variable is independent to other features outside it. Due to its capability of fully explaining a target variable, the MB has the potential of selecting the features that have strong relevance to an outcome



for building its prediction models. Then the MB algorithms such as incremental association MB (14) and its variants (15) were successfully employed in the feature selection process of our previous PD-BN approach to develop accurate and interpretable outcome prediction models.

However, in addition to high dimensional features with limited sample size, oncology datasets usually have imbalanced outcomes, such as HCC patients' LC or I-ALBI in this study. The prediction of these treatment outcomes can be modeled as a binary classification problem under supervised machine learning. Class imbalance occurs when the minority group, such as non-LC or I-ALBI, contains significantly fewer events samples than the majority group, such as LC or non-I-ALBI. Learning these imbalanced outcomes from high-dimensional datasets can be very difficult (15, 16), and non-standard machine learning methods are often guaranteed to achieve desirable results (14). Moreover, features selected from the above theoretically sound MB algorithms to have a strong relevance with an outcome may not be able to build the BN-based outcome prediction model with high accuracy, since accuracy and explainability are two different criteria for feature selection (17). Then, the PD-BN based outcome prediction models have a limited prediction performance in this case. Furthermore, the developed PD-BNs are not necessarily following physicians' common practice knowledge, and unconfirmed biophysical interactions explored from the PD-BN approach can barely gain physicians' trust for application in routine clinical decision making. Therefore, the goal of this study is to develop a new ML approach in handling imbalanced oncology data to improve the accuracy and explainability of predicting HCC patients' outcomes with SBRT.

Building accurate and explainable outcome prediction models from high-dimensional imbalanced data is a complex process that requires nontrivial understanding of complex ML algorithms (18). Humans are typically involved in unstructured manner at various points in the processes of the model development, model training,

and testing of the underlying ML algorithm implementation. Human-in-the-loop ML (HITL-ML) approaches are proposed to rather define a new type of structured interactions between humans and machine learning algorithms. Being developed initially from reinforcement learning, preference learning, and active learning, the HITL-ML is a hybrid of data-driven and knowledge-driven approach that integrates *a priori* expert knowledge (EK) into ML frameworks to overcome issues related to model bias and uncertainty (19). In addition to making ML more accurate or to obtain the desired accuracy faster, the HITL-ML approach makes humans more effective and efficient (18). Especially, it is useful in handling imbalanced data (20, 21). Due to the transparency of the BN for potential clinical causal inferences, in this study we develop an HITL-BN approach to build HCC SBRT outcome prediction models from imbalanced oncology data by incorporating EK features and allowing human agents to participate in the BN feature selection process. The accuracy and explainability of HITL-BN based outcome prediction models are evaluated and compared to the PD-BN based models that do not involve human agents.

The rest of paper is organized as follows. Section 2 introduces the properties of our dataset and the details of the HITL-BN approach. Section 3 shows and compares outcome prediction models developed from the PD-BN and HITL-BN approaches. Section 4 discusses the accuracy and explainability of our new approach and verifies the relationships among biophysical features in developed HITL-BNs based on related literatures. Section 5 concludes our paper.

2 Material and methods

2.1 Participation and data collection

Our study uses 81 HCC patients with SBRT on prospective protocols under institutional review board (IRB) approval from University of Michigan Hospital (Michigan Medicine). Since each patient may have one or more tumors, there are totally 104 tumors in our discovery dataset. In this study, two or more tumors in an HCC patient are assumed to be independent from each other for the sake of simplicity. There are 23 patients with I-ALBI during six-month follow up, and 100 tumors from 77 patients achieved LC. Each patient has 97 features, including dosimetric information, clinical factors, pre- and during-treatment labs and cytokines as summarized in Table 1. The change of a lab or cytokine value during treatment was calculated from the difference between its post treatment (or three months after treatment) and pre-treatment (or baseline) values, and it is formulated by adding prefix "D_" to its name in our study. To avoid confusion in outcome prediction, biophysical features related to LC or I-ALBI were specified and manually designated by a human expert. For example, "gross tumor volume (GTV)" is considered for predicting LC

instead of I-ALBI. The number of features for LC or I-ALBI prediction before and during SBRT are listed in Table 2. For independent external validation, we tested our developed models on a dataset from the Princess Margaret Hospital, which included 59 tumors from 38 HCC patients.

Physicians' acquired knowledge and experience from treating HCC SBRT patients were collected and treated as expert knowledge (EK) for SBRT outcomes prediction in this study. The EK features (EKFs) related to I-ALBI prediction comprise "LIVER_GTV_Mean_Dose", "pre_ICGR15", "D_ICGR15", "Child_Pugh", "Cirrhosis", "Liver_GTV_Volume", "pre_HGF", "D_HGF", "Age", "Sex", "pre_CD40_L", "D_CD40_L". The EKFs for LC prediction include "Child_Pugh", "GTV", "Total_BED", "LIVER_GTV_Mean_Dose", "Prior_Liver_Occurrences", "pre_MELD", "pre_Bilirubin", "pre_Albumin", "Tumor_Size", "GTV_Mean_Dose_LQ", "pre_ICGR15", "pre_HGF", and "pre_TGF_Beta". Except the above EKFs related to two different outcomes, the rest of features in Table 1 were denoted as non-EK features (non-EKFs) in this study.

2.2 Human-in-the-loop to handle imbalanced data in feature selection

Gained from many years of experience, reading articles, training, peers' interaction, EK has a potential of bypassing

otherwise complex systems and providing parsimonious solutions that focus on key aspects of a given situation. By incorporating EKFs into the feature selection of the PD-BN approach, we previously developed a situational awareness BN (SA-BN) method to predict the radiation outcomes of lung cancer patients (29). With the SA-BNs, the physicians not only are able to know patients' situation and predict LC and potential RITs starting from their acquired knowledge, but also can evaluate the best treatment strategies to maximize the LC and minimize the RITs before and during the course of radiotherapy. Focusing on improving the PD-BN based models' explainability, the SA-BN method has limitations in alleviating the impact of high-dimensional imbalanced data on the PD-BN approach's feature selection process to improve the accuracy of its associated outcome prediction models (29). However, the tighter confidence intervals of prediction performance and well-known biophysical relationships in the SA-BNs indicated that the EK has potential to improve the accuracy and explainability of outcome prediction models. Then, the EK methodology was employed in the HITL-BN approach to guide feature selection from imbalanced HCC SBRT data.

Selecting an ML approach for outcome prediction often involves a trade-off between prediction accuracy and explainability (30). While some ML approaches may lead to relatively more accurate outcome prediction models, other ML methods can result in more explainable ones. According to the explainability of their associated outcome prediction models, the

TABLE 1 Features of HCC patients with SBRT in the discovery dataset.

Categories	Names
Clinical Factors (25 features)	Sex, Age, pre-treatment_Cirrhosis (pre_Cirrhosis), Portal_Vein_Thrombosis, pre_SBRT, pre_SBRT_Liver, Active_Extrahepatic_Disease, Prior_Liver_Occurrences, Previously_Treated, Active_Liver_Lesions, Tumor_Size, gross tumor volume (GTV), planning target volume (PTV), Fiducials, Initial_Fraction, treatment break (Tx_Break) (22), Adapted, Total_Time, Break_Period, Number_of_Initial_Fractions (N_Initial_Fractions), N_Final_Fractions, Total_N_Fractions, Liver_GTV_Volume, pre-treatment eastern cooperative oncology group performance status (pre_ECOG_PS), D_ECOG_PS
Dosimetric Information (35 features)	Initial_Dose, biologically effective dose (BED)_Initial_Dose, equivalent dose in 2 Gy fractions (EQD2)_Initial_Dose, Total_SBRT_Dose, Total_BED, Total_EQD2, BED_Manual (23), dose that covers 98% of GTV (GTV_D98), generalized equivalent uniform dose of GTV (GTV_gEUD), GTV_Min_Dose, dose that covers 98% of GTV using linear-quadratic (LQ) model (GTV_D98_LQ), GTV_gEUD_LQ, GTV_Mean_Dose_LQ, GTV_Min_Dose_LQ, dose that covers 98% of GTV using linear-quadratic-linear (LQL) model with threshold dose 20Gy (GTV_D98_LQL_DT20), gEUD of GTV with $a = -20$ using LQL model with $\alpha/\beta = 10$ (GTV_gEUD_N20_LQL_10), GTV_Mean_Dose_LQL_DT20, GTV_Min_Dose_LQL_DT20, PTV_Mean_Dose, dose that covers 98% of PTV (PTV_D98), PTV_gEUD, PTV_Min_Dose, PTV_D98_LQ, PTV_gEUD_LQ, PTV_Mean_Dose_LQ, PTV_Min_Dose_LQ, PTV_D98_LQL_DT20, PTV_gEUD_LQL_DT20, PTV_Mean_Dose_LQL_DT20, PTV_Min_Dose_LQL_DT20, mean dose of total liver excluding gross tumor volume (LIVER_GTV_Mean_Dose), LIVER_GTV_Mean_Dose_LQ, LIVER_GTV_Mean_Dose_LQL, the 700 cm ³ subvolume EQD2 of LIVER_GTV using LQ model with $\alpha/\beta = 2.5$ Gy (LIVER_GTV_DC_LQ_EQD2) (24), the ≤ 15 Gy cold volume EQD2 of LIVER_GTV using LQ model with $\alpha/\beta = 2.5$ Gy (LIVER_GTV_CV_EQD2_LQ) (24)
Pre- and During-Treatment Labs (27 features)	pre_Na, D_Na*, pre_Creatinine, D_Creatinine, pre_Albumin, D_Albumin, pre-treatment aspartate aminotransferase (pre_AST), D_AST, pre-treatment alanine transaminase (pre_ALT), D_ALT, pre_Alphos, D_Alphos, pre_Bilirubin, D_Bilirubin, pre-treatment international normalized ratio (pre_INR), D_INR, pre_Protime_INR, D_Protime_INR, pre-treatment alpha fetoprotein (pre_AFP), D_AFP, baseline model for end-stage liver disease (MELD_baseline), D_MELD, MELD_Na_baseline, D_MELD_Na, Child_Pugh, D_Child_Pugh, Barcelona_Score
Pre- and During-Treatment Cytokines (10 features)	pre-treatment indocyanine green after 15 minutes (pre_ICGR15) (25, 26), D_ICGR15, pre-treatment transforming growth factor beta (pre_TGF_Beta), D_TGF_Beta, pre-treatment CD40 ligand (pre_CD40_L) (27, 28), D_CD40_L, pre-treatment hepatocyte growth factor (pre_HGF) (28), D_HGF, pre_Eotaxin, D_Eotaxin

*"D_" is a prefix to indicate the change of during-treatment labs or cytokines, which was calculated from the difference of their values between post-treatment (or three months after treatment) and pre-treatment (or baseline).

TABLE 2 The number of features associated with each of 104 tumors before and during treatment for I-ALBI or LC prediction.

Outcome	I-ALBI		LC	
	Pre- Treatment	During Treatment	Pre- Treatment	During Treatment
# of features associated with each tumor	45	69	68	93

ML approaches can be generally classified into explainable ML (EML) and unexplainable ML (UML) methods. The former includes Decision Trees, Logistic Regression and its variants, Naïve BNs, BNs, etc., and the latter comprises Random Forests (RFs), Support Vector Machines (SVMs), Gradient Boosting Machines (GBMs), Deep Learning (DL), etc. Although the EML-based outcome prediction models generally have relatively lower prediction accuracy compared to the UML-based models, they can be used to identify the most relevant features in explaining an outcome. On the other hand, while the UML-based outcome prediction models have difficulties in interpreting the relationships between specific features and the outcome, a list of ranked features can be generated from each of them based on features' importance in terms of outcome prediction (31). However, the ranking lists generated from different UML approaches may not be the same, resulting in different important features selected from the top rank of these lists for outcome prediction. An integrated feature ranking list is developed in this study by combining these lists based on the performance of its associated UML-based outcome prediction models to achieve robust feature selection as introduced in the next section.

The selected features from the EML and UML approaches are generally different, even though they are evaluated from one single dataset. The former and latter have the potential of improving an outcome prediction model's explainability and accuracy respectively. While the MB algorithm and network structure learning were considered as a *computational agent* to improve the prediction model's explainability by exploring EKF and non-EKF that have strong relevance to an outcome, the integrated feature ranking list was treated as another *computational agent* to enhance the prediction model's accuracy by investigating each feature's importance in terms of outcome prediction. Then, the HITL-BN approach can improve its capability of learning from the imbalanced HCC SBRT data by allowing human agents to interact these two computational agents during the process of feature selection.

2.3 The human-in-the-loop BN approach

As stated previously, the UML approaches include RF, SVM, GBM, DL, etc., and each of them can generate a feature ranking

list in terms of importance in outcome prediction from all features including EKFs and non-EKFs in a dataset. Let K be the total kinds of these UML approaches, k be the index of these approaches ($k = 1, 2, 3, \dots, K$), L^k be a feature ranking list obtained from UML approach k ($k = 1, 2, 3, \dots, K$) with the most important feature for outcome prediction at the top of the list, and AUC^k be the performance of an outcome prediction model developed from UML approach k based on cross validation in the discovery dataset ($k = 1, 2, 3, \dots, K$). Let J be the total number of features in the discovery dataset, j be the index randomly assigned to them ($j = 1, 2, 3, \dots, J$), $N_j(L^k)$ be the rank of feature j in ranking list L^k ($j = 1, 2, 3, \dots, J, k = 1, 2, 3, \dots, K$). The rankings $N_j(L^k)$ of the feature in different lists L^k may not be the same, and the performance AUC^k of UML approaches for outcome prediction could be different. It is assumed that a robust feature ranking list can be developed by integrating all these ranking lists based on their corresponding UML approaches' prediction performance. Let L^* be an integrated feature ranking list based on K UML approaches, be the weighted ranking score (WRS) of feature j to determine its ranking in list L^* , and its value can be evaluated from the following equation by integrating its ranks $N_j(L^k)$ in different ranking lists L^k

$$WRS_j = \sum_{k=1}^K \frac{N_j(L^k) \cdot \sum_{k=1}^K AUC^k}{AUC^k} \quad j = 1, 2, 3, \dots, J \quad (1)$$

Then, the ranking list L^* in terms of features' importance in outcome prediction can be obtained from sorting all the features based on their WRS_j , where the feature with the minimal score value is ranked at the top of the list.

Including feature selection and BN structure learning processes, HITL-BN based outcome prediction models are mainly developed based on the integrated ranking list. Let I be the total number of EKFs in list L^* with $I < J$, i represent the order of an EKF within all EKFs ($i = 1, 2, 3, \dots, I$). An initial HITL-BN is developed from the top n percent of features in the list. The value of n depends on the total number N of features in a dataset and appropriate feature dimension D to satisfy the MB algorithms' faithfulness assumption, and we assumed $n = 100^* \frac{D}{N}$. Suppose the top n percent of features in list includes i EKFs ($i \leq D$), an initial HITL-BN based outcome prediction model can be denoted as HITL-BN(i). Since some EKFs in the top rank of list L^* may be redundant or less relevance to an outcome compared to other ones, the most relevant EKFs can be identified from the outcome's MB. Given the selected EKFs,

important non-EKFs in the top rank of the list to improve the outcome prediction should also be strongly related to these EKFs, which can be identified from each of their MBs. Thus, important EKFs and non-EKFs can be selected from the top rank of the list with balanced accuracy and explainability for HITL-BN(i) development. Note that the structure learning of HITL-BN(i) is the same as that of PD-BNs, where less important EKFs and non-EKFs are eliminated from the network to maximize its prediction performance. The rest of our HITL-BN approach is to repeatedly evaluate whether a next EKF and additional non-EKFs before it in list L^* can improve the accuracy of previous outcome prediction models or not.

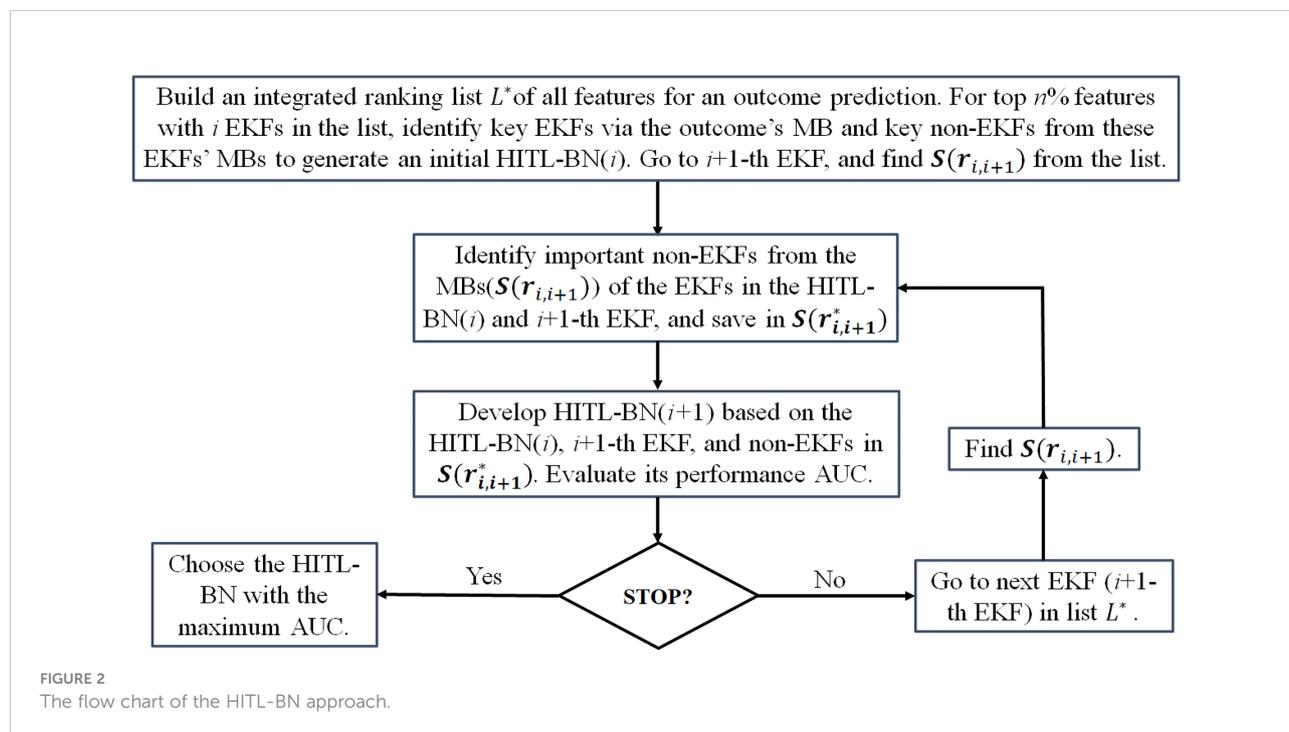
Let r_i be the rank of i -th EKF in list L^* ($i = 1, 2, 3, \dots, I$). As the evaluation moves from i -th EKF to $i+1$ -th EKF in the integrated feature ranking list, the set of additional indices between them is denoted as $r_{i, i+1}$. Let $S(r_{i, i+1})$ represents the set of non-EKFs associated with $r_{i, i+1}$, and the number of non-EKFs in the set could be zero when two EKFs are consecutive in the list. If set $S(r_{i, i+1})$ is not empty, the importance of these non-EKFs for the outcome prediction depends on whether they have strong relevance with selected EKFs, including the EKFs in HITL-BN(i) and $i+1$ -th EKF. Let MBs ($S(r_{i, i+1})$) be these EKFs' MBs based on non-EKFs in $S(r_{i, i+1})$, and the set of selected non-EKFs from these MBs is indicated as $S(r_{i, i+1}^*)$. Then HITL-BN($i+1$) can be developed based on $i+1$ -th EKF, non-EKFs in $S(r_{i, i+1}^*)$ together with all the features in HITL-BN(i) by employing PD-BN's structure learning process. The process continues along list until the performance of prediction model cannot be improved or meet a target

prediction performance. The details of the HITL-BN approach to generate an accurate and explainable outcome prediction model are described in Figure 2.

3 Results

3.1 PD-BN models for I-ALBI or LC prediction

As a comparison of the HITL-BN approach, PD-BN models for I-ALBI or LC prediction were developed based on our HCC SBRT patients as shown in Figures 3 or 4. Numerical experiments in this study were conducted in an R environment, where function "inter-MB" in R package "bnlearn" was employed as the MB algorithm for feature selection and function "boot.strength" in the same R package was used for BN structure learning. Figures 3A or 3D shows pre- or during-treatment PD-BN model for I-ALBI prediction developed from the discovery dataset. While the PD-BN method selected biophysical features "pre_Bilirubin", "pre_Cirrhosis", "Portal_Vein_Thrombosis", "pre_Creatinine", "pre_CD40_L", "pre_HGF", and "Liver_GTV_DC_LQ_EQD2" for pre-treatment I-ALBI prediction, additional variables "D_Protime_INR", "D_Bilirubin", and "D_ICGR15" were chosen for during-treatment I-ALBI prediction. The prediction performances of the former and the latter based on the discovery dataset are 0.78 (95%CI: 0.67-0.83) and 0.82 (95%CI: 0.74-0.88) as described in Figures 3B and 3E respectively. The prediction



performance of the former or the latter based on the testing dataset is 0.68 or 0.73 as illustrated by Figures 3C or 3F.

Figures 4A or 4D shows pre- or during-treatment PD-BN for LC prediction generated from the discovery dataset. While the PD-BN method selected features “*pre_Albumin*”, “*Active_Liver_Lesions*”, “*Portal_Vein_Thrombosis*”, “*pre_ECOG_PS*”, “*pre_TGF_Beta*”, “*pre_HGF*”, and “*GTV_gEUD_N20_LQL_10*” for pre-treatment LC prediction, additional variables “*pre_Cirrhosis*”, “*Adapted*”, “*D_CD40_L*”, and “*D_TGF_Beta*” were chosen for during-treatment LC prediction. The prediction performances of the former and the latter based on the discovery dataset are 0.75 (95%CI: 0.60-0.86) and 0.79 (95%CI: 0.69-0.89) as shown in Figures 4B and 4E respectively. The prediction performance of the former or the latter based on the testing dataset is 0.66 or 0.72 as illustrated by Figures 4C and 4F.

3.2 HITL-BN models for HCC SBRT patients’ outcomes prediction

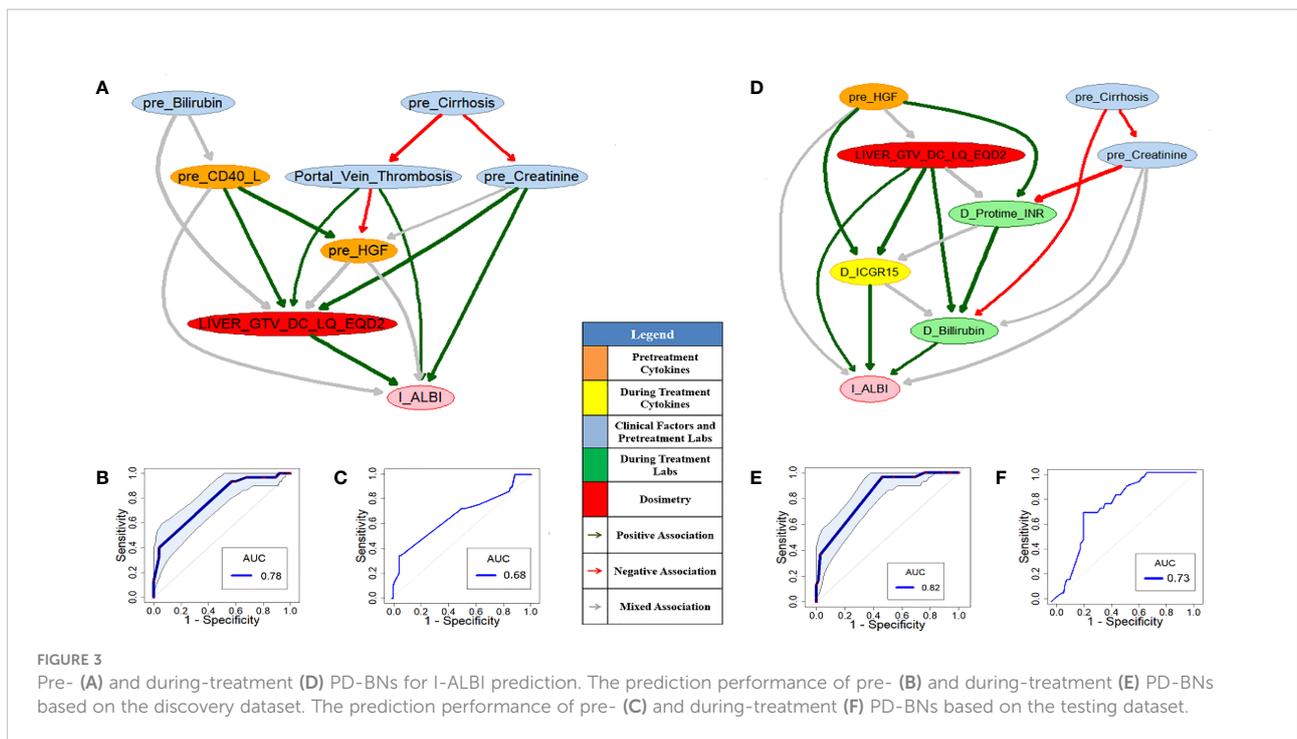
We conducted numerical experiments to develop or test HITL-BN models for I-ALBI or LC prediction based on the discovery and testing datasets in the same R environment as that of developing or testing the PD-BN models. Two UML approaches, the RF and GBM (K=2), were employed in this study to generate an integrated feature ranking list for a HITL-BN based outcome prediction model development before or during treatment. RF and GBM share similar tree/graph

structure learning to BN. Packages ‘randomForestSRC’ and ‘gbm’ were used to identify feature ranking lists from the former and latter approaches based on the discovery dataset respectively. After evaluating the two UML approaches’ prediction performances, each feature’s WRS was computed based on its ranks in two different ranking lists and the corresponding UML approaches’ prediction performances from Equation (1). Then, an integrated feature ranking list to rank all the features in the discovery dataset for I-ALBI or LC prediction before or during SBRT can be generated from their WRSs.

3.2.1 HCC SBRT patients’ I-ALBI prediction

Figures 5A or 5D shows pre- or during-treatment HITL-BN for I-ALBI prediction developed from the discovery dataset. While the HITL-BN approach selected features “*Sex*”, “*Age*”, “*pre_Na*”, “*pre_Cirrhosis*”, “*pre_Alkphos*”, “*pre_Bilirubin*”, “*pre_ICGR15*”, and “*LIVER_GTV_DC_LQ_EQD2*” for pre-treatment I-ALBI prediction, additional variables “*D_MELD*”, “*D_Albumin*”, and “*D_ICGR15*” were chosen for during-treatment I-ALBI prediction. Tables 3 and 4 show the integrated feature ranking lists of all features according to their WRSs for I-ALBI prediction before and during SBRT respectively. The features in PD-BNs as shown in Figure 3 are highlighted with *italic* font in these tables, and the features in HITL-BNs as illustrated in Figure 5 are emphasized with **bold** font in them. Especially, the features marked with italic and bold fonts come from both the PD-BN and HITL-BN.

The performances AUCs of pre- and during-treatment HITL-BNs for I-ALBI prediction based on the discovery



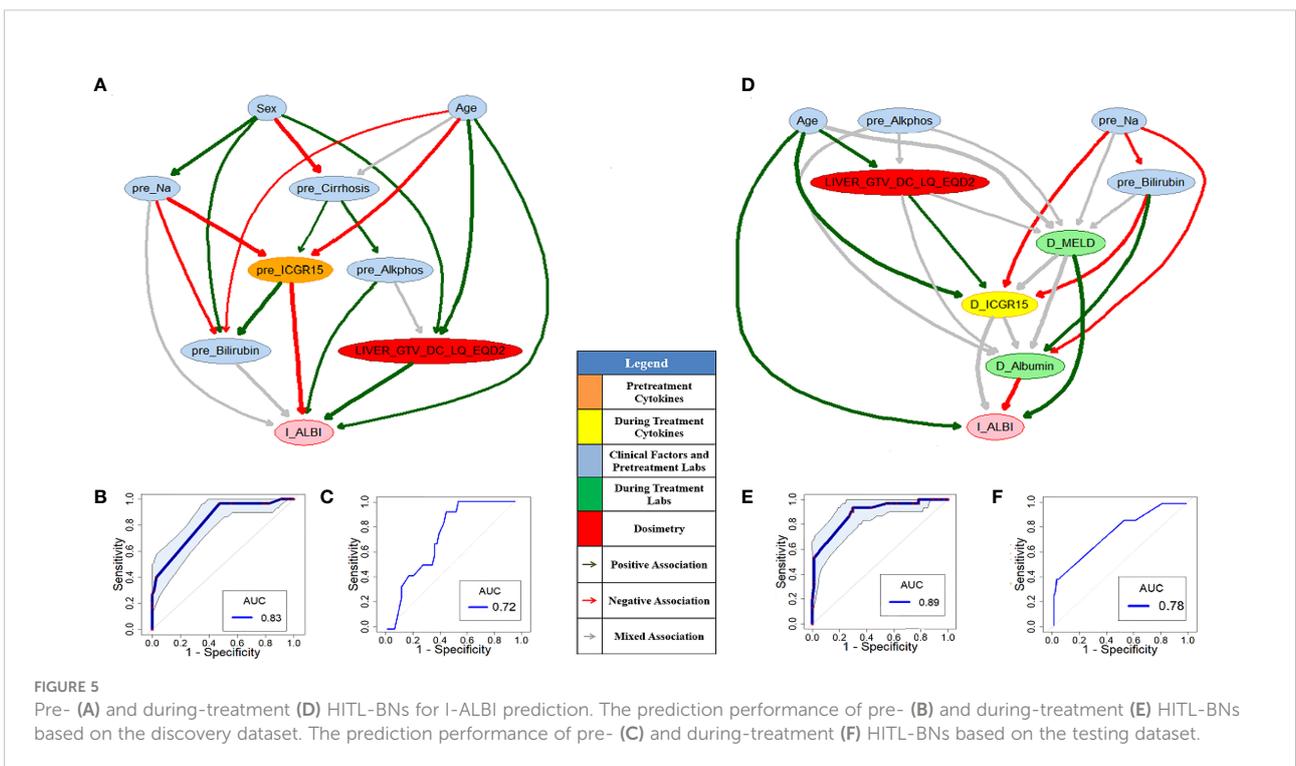
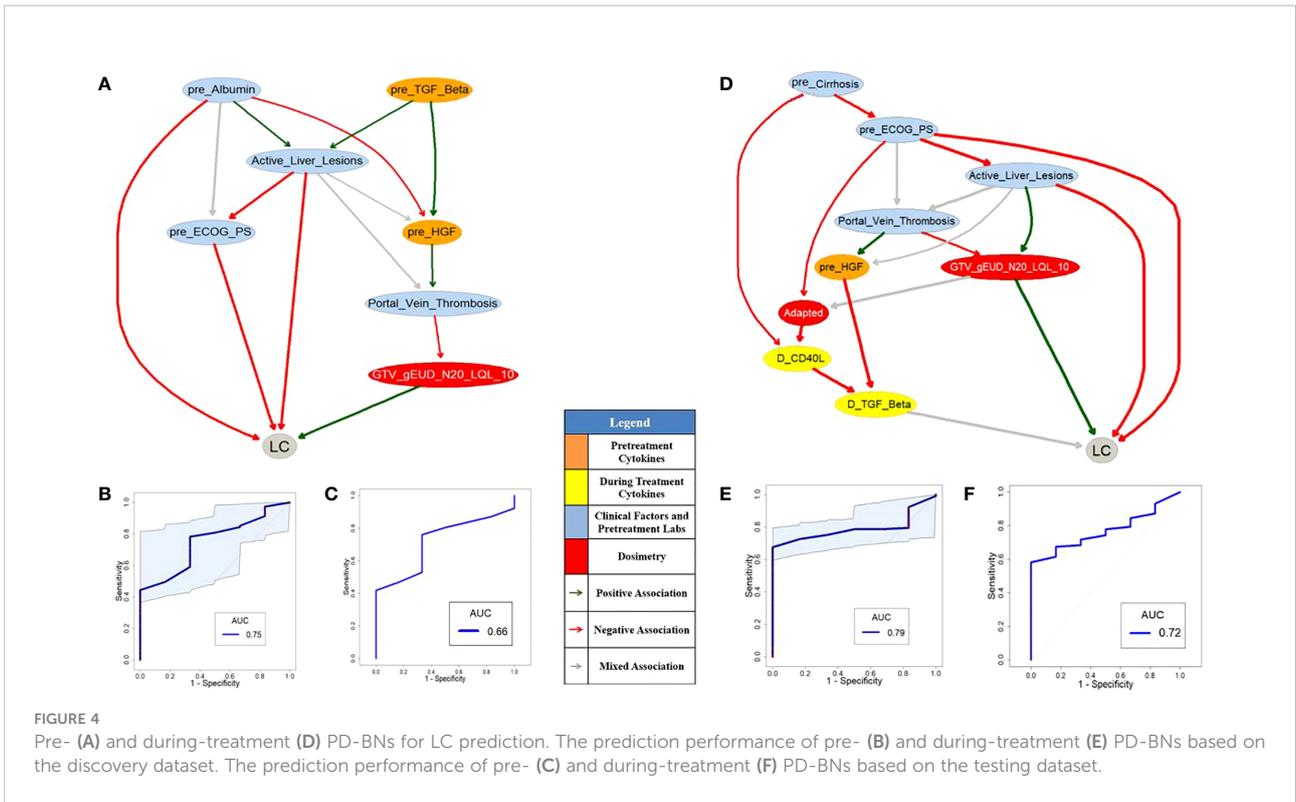


TABLE 3 The rank of features in an integrated feature ranking list for pre-treatment I-ALBI prediction.

Rank	Feature Names	Rank	Feature Names	Rank	Feature Names
1	pre_Albumin	16	Total_SBRT_Dose	31	Prior_Liver_Occurrences
2	pre_ICGR15	17	Total_EQD2	32	EQD2_Initial_Dose
3	pre_Alkphos	18	LIVER_GTV_Mean_Dose_LQ	33	MELD_baseline
4	pre_Na	19	BED_Manual	34	Initial_Dose
5	Age	20	pre_AST	35	Sex
6	LIVER_GTV_DC_LQ_EQD2	21	pre_Eotaxin	36	Active_Liver_Lesions
7	LIVER_GTV_CV_EQD2_LQ	22	Liver_GTV_Volumn	37	Barcelona_Score
8	pre_Bilirubin	23	pre_Protime_INR	38	Child_Pugh
9	pre_AFP	24	LIVER_GTV_Mean_Dose	39	Fiducials
10	pre_Creatinine	25	BED_Initial_Dose	40	pre_INR
11	Total_BED	26	LIVER_GTV_Mean_Dose_LQL	41	pre_ECOG_PS
12	pre_CD40_L	27	pre_Cirrhosis	42	Active_Extrahepatic_Disease
13	pre_TGF_Beta	28	Initial_Fraction	43	pre_SBRT
14	pre_HGF	29	Portal_Vein_Thrombosis	44	Previously_Treated
15	MELD_Na_baseline	30	pre_ALT	45	pre_SBRT_Liver

dataset are 0.83 (95%CI: 0.75-0.89) and 0.89 (95%CI: 0.81-0.95) as shown in Figures 5B, E respectively. While the performance of the former is not significantly better than that of pre-treatment PD-BN as illustrated in Figure 3A, the latter significantly outperforms during-treatment PD-BN as shown in Figure 3D

based on the DeLong test with p -value=0.0253. For the testing dataset, the performance of pre- or during-treatment HITL-BN for I-ALBI prediction is 0.72 or 0.78 as illustrated by Figures 5C or 5F, and the latter significantly outperforms during-treatment PD-BN from the Delong test with p -value=0.0318.

TABLE 4 The rank of features in an integrated feature ranking list for during-treatment I-ALBI prediction.

Rank	Feature Name	Rank	Feature Name	Rank	Feature Name
1	pre_Alkphos	24	pre_Creatinine	47	pre_ALT
2	pre_ICGR15	25	pre_HGF	48	Prior_Liver_Occurrences
3	pre_Albumin	26	pre_TGF_Beta	49	Break_Period
4	D_MELD	27	pre_AST	50	LIVER_GTV_Mean_Dose_LQL
5	D_ICGR15	28	D_INR	51	Child_Pugh
6	pre_Na	29	D_ALT	52	Barcelona_Score
7	D_Creatinine	30	MELD_Na_baseline	53	Portal_Vein_Thrombosis
8	pre_Bilirubin	31	D_TGF_Beta	54	Total_Time
9	D_HGF	32	D_Bilirubin	55	Total_EQD2
10	D_Eotaxin	33	MELD_baseline	56	D_ECOG_PS
11	LIVER_GTV_DC_LQ_EQD2	34	Total_SBRT_Dose	57	Total_N_Fractions
12	pre_Protime_INR	35	Liver_GTV_Volumn	58	D_Child_Pugh
13	D_Alkphos	36	pre_Eotaxin	59	N_Final_Fractions
14	D_CD40_L	37	D_Na	60	Tx_Break
15	D_AFP	38	Initial_Fraction	61	Sex
16	D_MELD_Na	39	Total_BED	62	Active_Extrahepatic_Disease
17	pre_AFP	40	Initial_Dose	63	Fiducials
18	D_Albumin	41	BED_Manual	64	pre_SBRT
19	D_Protime_INR	42	LIVER_GTV_Mean_Dose_LQ	65	pre_ECOG_PS
20	Age	43	BED_Initial_Dose	66	Active_Liver_Lesions
21	LIVER_GTV_CV_EQD2_LQ	44	LIVER_GTV_Mean_Dose	67	pre_INR
22	pre_CD40_L	45	EQD2_Initial_Dose	68	Previously_Treated
23	D_AST	46	pre_Cirrhosis	69	pre_SBRT_Liver

3.2.2 HCC SBRT patients' LC prediction

Figures 6A or 6D shows pre- or during-treatment HITL-BN for LC prediction developed from the discovery dataset. While the HITL-BN for LC prediction approach selected features “Prior_Liver_Occurrences”, “GTV”, “MELD_baseline”, “pre_TGF_Beta”, “pre_HGF”, “GTV_gEUD_LQ”, and “LIVER_GTV_Mean_Dose” for pre-treatment LC prediction, additional variables “MELD_Na_baseline”, “pre_Bilirubin”, “pre_ICGR15”, “GTV_Mean_Dose_LQ”, “D_Protime_INR”, and “D_TGF_Beta” were chosen for during-treatment LC prediction. Tables 5 and 6 show the ranking lists of all features according to their WRSS for LC prediction before and during SBRT respectively. The features from the PD-BNs as shown in Figure 4 are highlighted with *italic* font in these tables, and the features from the HITL-BNs are emphasized with **bold** font in them. Especially, the features marked with italic and bold fonts come from both the PD-BN and HITL-BN.

The performances of pre- and during-treatment HITL-BNs for LC prediction based on the discovery dataset are 0.82 (95% CI: 0.67-0.93) and 0.85 (95%CI: 0.75-0.95) as shown in Figures 6B and 6E and 6E respectively. While the performance of the former is not significantly better than that of pre-treatment PD-BN as illustrated in Figure 4A, the latter significantly outperforms the during-treatment PD-BN as shown in Figure 4D based on the DeLong test with *p*-value=0.0367. For the testing dataset, the performance of pre- or during-treatment HITL-BN for LC prediction is 0.71 or 0.77 as illustrated by Figure 6C or 6F, and the latter significantly

outperforms the during-treatment PD-BN from the Delong test with *p*-value=0.0406. The results of our numerical experiments are summarized in Table 7.

4 Discussion

4.1 Comparison of the PD-BN and the HITL-BN approaches for class imbalance

Developed from our previous PD-BN method, the HITL-BN approach also includes feature selection and BN structure learning. To handle imbalanced data, the HITL-BN approach allows human agents to integrate the EML-based and UML-based feature selections in identifying important EKFs and non-EKFs in terms of outcome prediction. Tables 3 and 4 show that EKFs and non-EKFs obtained from the HITL-BNs (highlighted by bold font) for I-ALBI prediction are generally ranking higher than those from the PD-BNs (emphasized by italic font) before and during SBRT respectively. Also, a similar situation can be found from Tables 5 and 6 for LC prediction. These findings not only echo that the HITL-BNs outperform the PD-BNs for I-ALBI or LC prediction before and during SBRT as shown in Figures 3 and 5 or Figures 4 and 6, but also indicate that the HITL-BN approach can increase the capability of feature selection from imbalanced data. Since the properties of

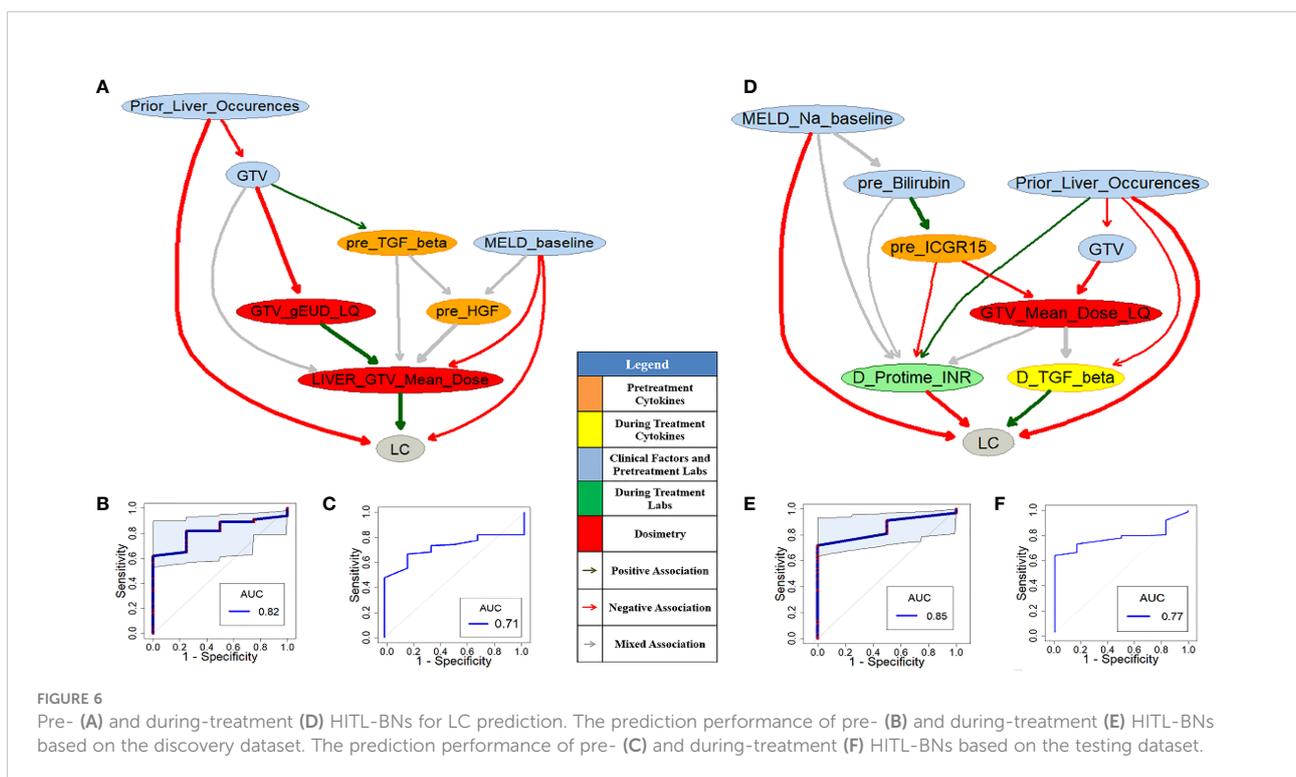


TABLE 5 The rank of features in an integrated feature ranking list for pre-treatment LC prediction.

Rank	Feature Names	Rank	Feature Names	Rank	Feature Names
1	pre_AST	24	GTV_D98_LQ	47	GTV_Mean_Dose_LQ
2	PTV	25	pre_ICGR15	48	GTV_gEUD
3	pre_Protime_INR	26	PTV_D98	49	pre_SBRT
4	MELD_Na_baseline	27	GTV_Min_Dose_LQ	50	EQD2_Initial_Dose
5	pre_TGF_Beta	28	GTV_Mean_Dose	51	BED_Initial_Dose
6	PTV_gEUD_LQ	29	GTV_Min_Dose	52	Sex
7	MELD_baseline	30	pre_Eotaxin	53	Initial_Fraction
8	PTV_Min_Dose	31	GTV_gEUD_LQ	54	Age
9	LIVER_GTV_Mean_Dose	32	GTV_Mean_Dose_LQL_DT20	55	Barcelona_Score
10	GTV	33	<i>pre_Albumin</i>	56	pre_ALT
11	pre_INR	34	Prior_Liver_Occurrences	57	<i>pre_ECOG_PS</i>
12	pre_HGF	35	PTV_D98_LQL_DT20	58	Fiducials
13	pre_AFP	36	<i>GTV_gEUD_N20_LQL_10</i>	59	Previously_Treated
14	PTV_Min_Dose_LQL_DT20	37	PTV_D98_LQ	60	Active_Extrahepatic_Disease
15	PTV_gEUD	38	PTV_Mean_Dose_LQL_DT20	61	GTV_D98
16	PTV_Min_Dose_LQ	39	pre_Na	62	<i>Portal_Vein_Thrombosis</i>
17	PTV_gEUD_LQL_DT20	40	pre_Bilirubin	63	pre_Cirrhosis
18	BED_Manual	41	Child_Pugh	64	Initial_Dose
19	pre_CD40_L	42	<i>Active_Liver_Lesions</i>	65	pre_SBRT_Liver
20	LIVER_GTV_CV_EQD2_LQ	43	PTV_Mean_Dose_LQ	66	Tumor_Size
21	Total_EQD2	44	PTV_Mean_Dose	67	pre_Alkphos
22	Total_BED	45	GTV_Min_Dose_LQL_DT20	68	pre_Creatinine
23	Total_SBRT_Dose	46	GTV_D98_LQL_DT20		

imbalanced outcomes in the testing dataset are not the same as those of the training dataset, the prediction performance of the HITL-BN based outcome prediction models of the former is expectedly less than that of the latter.

The reasons for the improvement of accuracy and explainability of HITL-BN based outcome prediction models in handling the imbalanced proportion of tumors with and without LC or I-ALBI in our HCC SBRT patient dataset could be twofold. First, since traditional ML approaches for crowdsourcing labeled training examples are not effective at locating the scarce minority class examples (32), they have difficulties in handling the high-skewed domain in the real-world, and their associated outcome prediction models may have low accuracy. Active learning is designed to select representative subsets of unlabeled datasets for manual labeling, and an ML algorithm can achieve accuracy with fewer training labels if it is allowed to choose the data from which it learns (14, 33). Originating from active learning, our HITL-BN approach intends to manually label important EKFs and non-EKFs based on their strong relevance to an outcome or/and their importance in the outcome prediction, which is intended to improve the prediction of the imbalanced LC classes and I-ALBI classes. Secondly, while EKFs play an important role in the HITL-BN approach due to its explainability to gain physicians' trust in clinical decision making, not all of them are ranked at the

top of an integrated feature ranking list. They are evenly distributed into the ranking list as shown in Tables 3-6. Only the top-ranked EKFs that are strongly related to the outcome were selected to build an initial HITL-BN. In the meantime, top-ranked non-EKFs have potential to improve the accuracy of the initial HITL-BN model as well. However, given the selected EKFs in the initial model, only the non-EKFs with strong relevance to these EKFs can improve its prediction performance. Our HITL-BN approach is designed to determine the important EKFs or/and non-EKFs from integrating the EML-based and UML-based feature selection methods and maximizing the prediction performance of the developed BNs through feedback. The focused, interactive, incremental process to improve the accuracy and explainability of an outcome prediction model can be considered as an extension of cost-sensitive learning, which is one of traditional methods for class imbalance (14, 34, 35).

As some EKFs may be missing or not available in clinical practice, the HITL-BN approach can skip these EKFs or investigate the EKFs that physicians are most interested in along the integrated ranking list for the outcome prediction model development. The purpose of this study is to verify whether the HITL-BN approach can significantly improve the performance of HCC SBRT patients' outcome prediction models or not based on imbalanced data compared to the PD-BN

TABLE 6 The rank of features in an integrated feature ranking list for during-treatment LC prediction.

Rank	Feature Names	Rank	Feature Names	Rank	Feature Names
1	pre_AST	32	Total_EQD2	63	D_Albumin
2	PTV	33	GTV_Min_Dose_LQL_DT20	64	D_ECOG_PS
3	MELD_Na_baseline	34	PTV_D98_LQL_DT20	65	Initial_Dose
4	pre_Protime_INR	35	D_Bilirubin	66	Barcelona_Score
5	D_TGF_Beta	36	D_Eotaxin	67	Child_Pugh
6	LIVER_GTV_Mean_Dose	37	PTV_D98	68	D_ALT
7	PTV_Min_Dose	38	Total_BED	69	pre_Na
8	MELD_baseline	39	GTV_Min_Dose_LQ	70	BED_Initial_Dose
9	pre_TGF_Beta	40	D_AST	71	pre_ECOG_PS
10	GTV	41	GTV_Mean_Dose	72	Fiducials
11	D_HGF	42	D_Creatinine	73	N_Final_Fractions
12	PTV_gEUD	43	D_MELD	74	Total_N_Fractions
13	pre_AFP	44	GTV_Mean_Dose_LQL_DT20	75	N_Initial_Fractions
14	Total_Time	45	pre_Albumin	76	Active_Liver_Lesions
15	D_ICGR15	46	pre_Alkphos	77	Prior_Liver_Occurences
16	D_AFP	47	pre_ALT	78	Previously_Treated
17	PTV_Min_Dose_LQL_DT20	48	D_Na	79	Active_Extrahepatic_Disease
18	D_Protime_INR	49	BED_Manual	80	Adapted
19	PTV_gEUD_LQ	50	pre_Creatinine	81	Break_Period
20	D_MELD_Na	51	PTV_Mean_Dose_LQL_DT20	82	Age
21	pre_HGF	52	PTV_D98_LQ	83	pre_SBRT
22	pre_CD40_L	53	PTV_Mean_Dose	84	Portal_Vein_Thrombosis
23	pre_INR	54	GTV_gEUD_N20_LQL_10	85	pre_Cirrhosis
24	LIVER_GTV_CV_EQD2_LQ	55	GTV_D98_LQL_DT20	86	Sex
25	GTV_Min_Dose	56	pre_Bilirubin	87	D_CD40_L
26	PTV_Min_Dose_LQ	57	GTV_gEUD_LQ	88	pre_SBRT_Liver
27	pre_ICGR15	58	GTV_D98_LQ	89	Initial_Fraction
28	D_INR	59	GTV_gEUD	90	D_Child_Pugh
29	PTV_gEUD_LQL_DT20	60	GTV_D98	91	D_Alkphos
30	GTV_Mean_Dose_LQ	61	Tumor_Size	92	PTV_Mean_Dose_LQ
31	pre_Eotaxin	62	Tx_Break	93	

TABLE 7 The results of numerical experiments.

Dataset	Treatment	Performance	PD-BN		HITL-BN		Delong Test	
			I-ALBI	LC	I-ALBI	LC	I-ALBI	LC
Training	Pre	AUC	0.78	0.75	0.83	0.82	0.0654	0.0875
		95% CI	0.67-0.83	0.60-0.86	0.75-0.89	0.67-0.93	NA	NA
	During	AUC	0.82	0.79	0.89	0.85	0.0253	0.0367
		95% CI	0.74-0.88	0.69-0.89	0.81-0.95	0.75-0.95	NA	NA
Testing	Pre	AUC	0.68	0.66	0.72	0.71	0.0921	0.1063
	During	AUC	0.73	0.72	0.78	0.77	0.0318	0.0406

method. The HITL-BN approach based on two UML approaches with RF and GBM had been implemented in our numerical experiments to test the hypothesis. Our choices of these two because they resemble BN in terms of graph/tree structures. However, if the number of UML algorithms increases, whether the predictive power of the HITL-BN based outcome prediction models could be improved or not and how much it can be improved are interesting research topics that beyond our current scope and we would like to explore in the next step.

Our numerical experiments on developing the HITL-BN based outcome prediction models for HCC SBRT patients have shown that human intelligence can positively augment machine intelligence, and the assistance of human agents involved in the learning phase can enhance the capability of learning from imbalanced data. However, our study still has limitations in terms of small sample size and the assumption of two or more independent lesions in an HCC patient. In the next steps, in addition to developing more robust HITL-BN approaches by removing the within patient tumor independence assumption and conducting further external independent validations, we plan to explore an interactive human-computer interface *via* the HITL-BN approach to conduct prospective personalized SBRT trials for improving HCC patients' radiation treatment outcomes.

4.2 The explainability of the HITL-BNs for HCC SBRT patient outcomes prediction

In addition to outperforming the PD-BN based outcome prediction models in terms of accuracy, the HITL-BN based outcome prediction models also have a better explainability due to the incorporation of the EKFs in their model buildings. The biophysical pathways displayed in our HITL-BNs for I-ALBI prediction before SBRT are supported by cited literatures. Since a longitudinal increase in the ALBI score is closely associated with non-malignancy-related mortality and quality of life (36), the incorporation of mid-treatment change in ALBI in addition to baseline ALBI improves the ability to predict treatment-related toxicity in patients with HCC receiving SBRT (13). Then *change* in albumin–bilirubin score (ALBI) score at three months after SBRT were used in many studies to capture acute toxicity occurring <90 days after SBRT (37). Studies showed that repeated SBRT in patients with advanced liver *cirrhosis* seems to exhibit higher hepatic toxicity (38), and the severity of hepatic *cirrhosis* is a major prognostic factor for radiation induced liver disease (39). Also, researchers found out that direct total bilirubin and total bilirubin are not related to delivery dose, and *age* is a significant predictive factor for radiation-induced liver injury based on univariate analysis of clinical factors (39). Moreover, an elevation in alkaline phosphatase (*alkphos*) of at least 5-fold and/or that of

bilirubin of at least 3-fold compared to either the upper normal limit or the pretreatment level corresponding to grade 3 or higher hepatic toxicity without disease progression within 3 months after SBRT is one of the conditions to define radiation-induced liver disease (40).

The following findings from literatures support the biophysical pathways displayed in the HITL-BN for I-ALBI prediction during SBRT. Increasing *mean liver dose* was associated with larger increases in toxicities (41). As the percentage of retained ICG at 15 minutes, *ICGR15*'s normal value would be in the range of 4–10% (42). While baseline values of *ICGR15* may be associated with the development of radiation induced liver disease, *the change of ICGR15* after radiation therapy appears to be most indication of the toxicity (43, 44). There may exist prognostic significance of baseline serum sodium value (*pre_Na*) in HCC patients complicating with liver *cirrhosis*, and lower serum sodium concentration is a useful predictor for these patients (45). The time course of changes of the liver function after SBRT was analyzed in patients treated for non-resectable HCC. Albumin was the only blood test that changed systematically during a three-month period, and it stabilized thereafter, which indicates the *decrease in albumin* reflects a minor radiation-induced liver disease (46). Model for end-stage liver disease (MELD) is a scoring system used to predict three-month mortality in patients with advanced liver disease (47). An increase in MELD score is associated with a decrease in residual liver function or deterioration in liver function (48).

Moreover, our HITL-BNs to predict LC before and during SBRT are endorsed by the following recorded observations. Higher *treatment dose* was associated with improved freedom from *local progression* (41) (49). Larger *GTV* volume was significantly associated with a higher risk of death (39). While increased TGF- β signaling has demonstrated radiation resistance (50), study shows that inhibition of TGF- β stops disease progression in liver metastases from colon cancer (51) (52). Incorporation of *ICGR15* variables significantly improves the prediction of post-SBRT liver function. The use of *ICGR15* can facilitate the delivery of the maximum safe dose of radiation for patients with hepatocellular carcinoma and has the potential to improve uncomplicated tumor control and survival (43). Prolonged prothrombin time (*Protime*) is the most important score when determining the incidence of radiation-induced liver disease during SBRT in patients with CP-A score 6 (53). International Normalized Ratio (*INR*) is derived from *Protime* which is calculated as a ratio of the patient's *Protime* to a control *Protime* standardized for the potency of the thromboplastin reagent developed by the World Health Organization. The *MELD* is used to prioritize patients for liver transplantation and includes results for creatinine, *bilirubin*, and *Protime* expressed as international normalized ratio (*Protime- INR*) (54). Evidence was provided that the *Protime- INR* was identified as the most important methodologies may influence the *MELD* (54). While lower *MELD* scores were associated with improved survival following SBRT (55), a mathematical equation based on *MELD* and

sodium, named the *MELD-Na* score, is a feasible and independent prognostic predictor for both short- and long-term outcome predictions in patients with hepatocellular carcinoma (56). It turns out that some features related to radiation induced liver disease such as TGF- β , MELD-Na, Bilirubin, etc. appeared in the HITL-BN for LC prediction, and its reason may be related to the fact that liver SBRT was conducted by limiting the toxicity from therapy and not compromising the primary objective of local control.

5 Conclusions

In this study, we have developed a new HITL-BN approach for HCC patients' I-ALBI or LC prediction before and during SBRT based on previous PD-BN method. In addition to incorporating EK into its feature selection process, the HITL-BN approach allows humans to participate in an outcome prediction model building process for better handling of imbalanced HCC SBRT data. Especially, we created a novel feature selection mechanism for the HITL-BN approach by integrating the prediction strength of multiple UML methods and the explainable capability of the theoretically sound MB algorithms. Numerical experiments show that the HITL-BN based outcome prediction models significantly outperform the PD-BN based models during SBRT in terms of accuracy and explainability. In addition to gaining physicians' trust in clinical decision making, the HITL-BN approach has the potential of becoming an important component of future human-computer interface to bridge physicians and advanced ML techniques in improving HCC patients' treatment outcomes. Our approach can be applied to the outcome prediction of treating other types of cancer, but it still needs to be validated in external further independent datasets.

Data availability statement

The raw data supporting the conclusions of this article will be made available upon request from authors and per institutional guidelines.

References

- Cai J, Wijesooriya K, Benedict S. SBRT I: Overview of simulation, planning, and delivery. *Med Phys* (2014) 41(6):412–3. doi: 10.1118/1.4889122
- Matoba M, Tsuchiya H, Kondo T, Ota K. Stereotactic body radiotherapy delivered with IMRT for oligometastatic regional lymph node metastases in hepatocellular carcinoma: a single-institutional study. *J Radiat Res* (2020) 61(5):776–83. doi: 10.1093/jrr/rraa067
- Demirtas CO, Alessio AD, Rimassa L, Sharma R, Pinato DJ. ALBI grade: Evidence for an improved model for liver functional estimation in patients with hepatocellular carcinoma. *Hep Rep* (2021) 3(5):100347. doi: 10.1016/j.jhepr.2021.100347
- Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: A new evidence-based approach—the ALBI grade. *J Clin Oncol* (2015) 33(6):550–U45. doi: 10.1200/JCO.2014.57.9151
- Jackson WC, Hartman HE, Gharzai LA, Maurino C, Karnak D, Mendiratta-Lala M, et al. A mid-treatment increase in albi score is strongly associated with treatment related toxicity following liver radiation therapy. *Int J Radiat Oncol* (2019) 105(1):S206–S7. doi: 10.1016/j.ijrobp.2019.06.277
- El Naqa I, Johansson A, Owen D, Cuneo K, Cao Y, Matuszak M, et al. Modeling of normal tissue complications using imaging and biomarkers after radiation therapy for hepatocellular carcinoma. *Int J Radiat Oncol* (2018) 100(2):335–43. doi: 10.1016/j.ijrobp.2017.10.005
- Pursley J, El Naqa I, Sanford NN, Noe B, Wo JY, Eyer CE, et al. Dosimetric analysis and normal-tissue complication probability modeling of child-pugh score and albumin-bilirubin grade increase after hepatic irradiation. *Int J Radiat Oncol* (2020) 107(5):986–95. doi: 10.1016/j.ijrobp.2020.04.027

Author contributions

KC, TL, MM, RT, IE, and YL conceived of the presented idea. YL developed the theory and performed the computations. IE and DN verified the analytical methods. LD provided the testing dataset. All authors discussed the results and contributed to the final manuscript.

Funding

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

Acknowledgments

The authors would like to thank Princess Margaret Hospital for providing the testing dataset to independently evaluate our models.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

8. Naqa I. Perspectives on making big data analytics work for oncology. *Methods* (2016) 111:32–44. doi: 10.1016/j.ymeth.2016.08.010
9. Bach F. Learning theory from first principles. In: *Draft of a book*, vol. 6. (2021). p. 2021. The MIT Press.
10. 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(1):85–92. doi: 10.1016/j.radonc.2017.02.004
11. 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* (2019) 3(2):232–41. doi: 10.1109/TRPMS.2018.2832609
12. 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(8):3980–95. doi: 10.1002/mp.13029
13. Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. (1988). *Morgan Kaufmann*. doi: 10.1016/C2009-0-27609-4
14. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data* (2019) 6(1):27. doi: 10.1186/s40537-019-0192-5
15. Bauder RA, Khoshgoftaar TM. The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. *Health Inf Syst Syst* (2018) 6(1):9. doi: 10.1007/s13755-018-0051-3
16. Bauder RA, Khoshgoftaar TM, Hasanin T eds. *An empirical study on class rarity in big data*. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (2018). Orlando, FL, USA.
17. Yu K, Wu XD, Ding W, Mu Y, Wang H. Markov Blanket feature selection using representative sets. *IEEE Trans Neural Networks Learn Syst* (2017) 28(11):2775–88. doi: 10.1109/TNNLS.2016.2602365
18. Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: a state of the art. *Artif Intell Rev* (2022). doi: 10.1007/s10462-022-10246-w
19. Wu XJ, Xiao LW, Sun YX, Zhang JH, Ma TL, He L. A survey of human-in-the-loop for machine learning. *Future Generation Comput Systems-the Int J Esience* (2022) 135:364–81. doi: 10.1016/j.future.2022.05.014
20. Chen J, Lim CP, Tan KH, Govindan K, Kumar A. Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments. *Ann operations Res* (2021) 306:1–24. doi: 10.1007/s10479-021-04373-w
21. Wang Y, Gan W, Yang J, Wu W, Yan J eds. *Dynamic curriculum learning for imbalanced data classification*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019). Seoul, Korea (South).
22. Jackson WC, Suresh K, Maurino C, Feng M, Cuneo KC, Ten Haken RK, et al. A mid-treatment break and reassessment maintains tumor control and reduces toxicity in patients with hepatocellular carcinoma treated with stereotactic body radiation therapy. *Radiother Oncol* (2019) 141:101–7. doi: 10.1016/j.radonc.2019.07.027
23. Millar WT, Hopewell JW, Paddick I, Lindquist C, Nordström H, Lidberg P, et al. The role of the concept of biologically effective dose (BED) in treatment planning in radiosurgery. *Physica Medica-European J Med Phys* (2015) 31(6):627–33. doi: 10.1016/j.ejmp.2015.04.008
24. Jackson WC, Tang M, Maurino C, Mendiratta-Lala M, Parikh ND, Matuszak MM, et al. Individualized adaptive radiation therapy allows for safe treatment of hepatocellular carcinoma in patients with child-Turcotte-Pugh b liver disease. *Int J Radiat Oncol* (2021) 109(1):212–9. doi: 10.1016/j.ijrobp.2020.08.046
25. Feng M, Suresh K, Schipper MJ, Bazzi L, Ben-Josef E, Matuszak MM, et al. Individualized adaptive stereotactic body radiotherapy for liver tumors in patients at high risk for liver damage a phase 2 clinical trial. *JAMA Oncol* (2018) 4(1):40–7. doi: 10.1001/jamaoncol.2017.2303
26. Stenmark MH, Cao Y, Wang HS, Jackson A, Ben-Josef E, Ten Haken RK, et al. Estimating functional liver reserve following hepatic irradiation: Adaptive normal tissue response models. *Radiother Oncol* (2014) 111(3):418–23. doi: 10.1016/j.radonc.2014.04.007
27. Cousins MM, Morris E, Maurino C, Devasia TP, Karnak D, Ray D, et al. TNFR1 and the TNF alpha axis as a targetable mediator of liver injury from stereotactic body radiation therapy. *Trans Oncol* (2021) 14(1):100950. doi: 10.1016/j.tranon.2020.100950
28. Cuneo KC, Devasia T, Sun YL, Schipper MJ, Karnak D, Davis MA, et al. Serum levels of hepatocyte growth factor and CD40 ligand predict radiation-induced liver injury. *Trans Oncol* (2019) 12(7):889–94. doi: 10.1016/j.tranon.2019.04.003
29. Luo Y, Jolly S, Palma D, Lawrence TS, Tseng HH, Valdes G, et al. A situational awareness Bayesian network approach for accurate and credible personalized adaptive radiotherapy outcomes prediction in lung cancer patients. *Phys Med* (2021) 87:11–23. doi: 10.1016/j.ejmp.2021.05.032
30. Wu L, Huang R, Tetko IV, Xia Z, Xu J, Tong W. Trade-off predictivity and explainability for machine-learning powered predictive toxicology: An in-depth investigation with Tox21 data sets. *Chem Res Toxicol* (2021) 34(2):541–9. doi: 10.1021/acs.chemrestox.0c00373
31. Wojtas MA, Chen K. Feature importance ranking for deep learning. In: *Proceedings of the 34th international conference on neural information processing systems*. Vancouver, BC, Canada: Curran Associates Inc (2020). p. 429.
32. Lin CH, Weld DS eds. *Active learning with unbalanced classes and example-generation queries*. HCOMP (2018). Zürich, Switzerland.
33. Settles B. Active learning literature survey. In: *Computer sciences technical report*. Madison: University of Wisconsin (2009). p. 1648.
34. Li Z, Sharma P, Lu XH, Cheung J, Reddy S. *Using interactive feedback to improve the accuracy and explainability of question answering systems post-deployment*. Findings of the Association for Computational Linguistics: ACL (2022), p. 926–37. Dublin, Ireland. arXiv:2204.03025.
35. Kumar P, Bhatnagar R, Gaur K, Bhatnagar A. Classification of imbalanced Data: Review of methods and applications. *IOP Conf Series: Materials Sci Eng* (2021) 1099(1):012077. doi: 10.1088/1757-899X/1099/1/012077
36. Sakamaki A, Takamura M, Sakai N, Watanabe Y, Arao Y, Kimura N, et al. Longitudinal increase in albumin-bilirubin score is associated with non-malignancy-related mortality and quality of life in patients with liver cirrhosis. *PLoS One* (2022) 17(2):e0263464. doi: 10.1371/journal.pone.0263464
37. Mathew AS, Atenafu EG, Owen D, Maurino C, Brade A, Brierley J, et al. Long term outcomes of stereotactic body radiation therapy for hepatocellular carcinoma without macrovascular invasion. *Eur J Cancer* (2020) 134:41–51. doi: 10.1016/j.ejca.2020.04.024
38. Huang Y, Chen SW, Fan CC, Ting LL, Kuo CC, Chiou JF. Clinical parameters for predicting radiation-induced liver disease after intrahepatic reirradiation for hepatocellular carcinoma. *Radiat Oncol* (2016) 11(1):89. doi: 10.1186/s13014-016-0663-1
39. Liang SX, Zhu XD, Xu ZY, Zhu J, Zhao JD, Lu HJ, et al. Radiation-induced liver disease in three-dimensional conformal radiation therapy for primary liver carcinoma: the risk factors and hepatic radiation tolerance. *Int J Radiat Oncol Biol Phys* (2006) 65(2):426–34. doi: 10.1016/j.ijrobp.2005.12.031
40. Lee S, Kim H, Ji Y, Cho B, Kim SS, Jung J, et al. Evaluation of hepatic toxicity after repeated stereotactic body radiation therapy for recurrent hepatocellular carcinoma using deformable image registration. *Sci Rep* (2018) 8(1):16224. doi: 10.1038/s41598-018-34676-1
41. Jackson WC, Hartman HE, Gharzai LA, Maurino C, Karnak DM, Mendiratta-Lala M, et al. The potential for midtreatment albumin-bilirubin (ALBI) score to individualize liver stereotactic body radiation therapy. *Int J Radiat Oncol* (2021) 111(1):127–34. doi: 10.1016/j.ijrobp.2021.04.012
42. Koay EJ, Owen D, Das P. Radiation-induced liver disease and modern radiotherapy. *Semin Radiat Oncol* (2018) 28(4):321–31. doi: 10.1016/j.semradi.2018.06.007
43. Suresh K, Owen D, Bazzi L, Jackson W, Ten Haken RK, Cuneo K, et al. Using indocyanine green extraction to predict liver function after stereotactic body radiation therapy for hepatocellular carcinoma. *Int J Radiat Oncol* (2018) 100(1):131–7. doi: 10.1016/j.ijrobp.2017.09.032
44. Yoon HI, Koom WS, Lee JJ, Jeong K, Chung Y, Kim JK, et al. The significance of ICG-R15 in predicting hepatic toxicity in patients receiving radiotherapy for hepatocellular carcinoma. *Liver Int* (2012) 32(7):1165–71. doi: 10.1111/j.1478-3231.2012.02784.x
45. Nishikawa H, Kita R, Kimura T, Ohara Y, Sakamoto A, Saito S, et al. Hyponatremia in hepatocellular carcinoma complicating with cirrhosis. *J Cancer* (2015) 6(5):482–9. doi: 10.7150/jca.11665
46. Dreher C, Hoyer KI, Fode MM, Habermehl D, Combs SE, Hoyer M. Metabolic liver function after stereotactic body radiation therapy for hepatocellular carcinoma. *Acta Oncol* (2016) 55(7):886–91. doi: 10.3109/0284186X.2015.1137352
47. Tang JY, Ohri N, Kabarriti R, Aparo S, Chuy J, Goel S, et al. Model for end-stage liver disease and sodium velocity predicts overall survival in nonmetastatic hepatocellular carcinoma patients. *Can J Gastroenterol Hepatol* (2018) 2018:5681979. doi: 10.1155/2018/5681979
48. Botta F, Giannini E, Romagnoli P, Fasoli A, Malfatti F, Testa E, et al. MELD scoring system is useful for predicting prognosis in patients with liver cirrhosis and is correlated with residual liver function: a European study. *Gut* (2003) 52(1):134–9. doi: 10.1136/gut.52.1.134
49. Bibault JE, Dewas S, Vautravers-Dewas C, Hollebecque A, Jarraya H, Lacornerie T, et al. Stereotactic body radiation therapy for hepatocellular carcinoma: Prognostic factors of local control, overall survival, and toxicity. *PLoS One* (2013) 8(10):e77472. doi: 10.1371/journal.pone.0077472

50. Wu CT, Hsieh CC, Yen TC, Chen WC, Chen MF. TGF-beta1 mediates the radiation response of prostate cancer. *J Mol Med (Berl)* (2015) 93(1):73–82. doi: 10.1007/s00109-014-1206-6
51. Calon A, Lonardo E, Berenguer-Llargo A, Espinet E, Hernando-Momblona X, Iglesias M, et al. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* (2015) 47(4):320–U62. doi: 10.1038/ng.3225
52. Onderdonk BE, Chmura SJ. The yin and yang of cytoreductive SBRT in oligometastases and beyond. *Front Oncol* (2019) 9. doi: 10.3389/fonc.2019.00706
53. Wang PM, Chung NN, Hsu WC, Chang FL, Jang CJ, Scorsetti M. Stereotactic body radiation therapy in hepatocellular carcinoma: Optimal treatment strategies based on liver segmentation and functional hepatic reserve. *Rep Pract Oncol Radiother* (2015) 20(6):417–24. doi: 10.1016/j.rpor.2015.03.005
54. Tripodi A, Chantarangkul V, Primignani M, Fabris F, Dell'Era A, Sei C, et al. The international normalized ratio calibrated for cirrhosis (INR(liver)) normalizes prothrombin time results for model for end-stage liver disease calculation. *Hepatology* (2007) 46(2):520–7. doi: 10.1002/hep.21732
55. Chan J, Behr S, Pai J, Feng MUS, Chang A, Haas-Kogan DA, et al. Stereotactic body radiotherapy for hepatocellular carcinoma in patients with poor liver function. *J Clin Oncol* (2018) 36(4):397–397. doi: 10.1200/JCO.2018.36.4_suppl.397
56. Huo T, Lin HC, Hsia CY, Huang YH, Wu JC, Chiang JH, et al. The MELD-Na is an independent short- and long-term prognostic predictor for hepatocellular carcinoma: A prospective survey. *Digestive Liver Dis* (2008) 40(11):882–9. doi: 10.1016/j.dld.2008.01.015