



The Role of Machine Learning in Knowledge-Based Response-Adapted Radiotherapy

Huan-Hsin Tseng*, Yi Luo, Randall K. Ten Haken and Issam El Naqa

Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, United States

With the continuous increase in radiotherapy patient-specific data from multimodality imaging and biotechnology molecular sources, knowledge-based response-adapted radiotherapy (KBR-ART) is emerging as a vital area for radiation oncology personalized treatment. In KBR-ART, planned dose distributions can be modified based on observed cues in patients' clinical, geometric, and physiological parameters. In this paper, we present current developments in the field of adaptive radiotherapy (ART), the progression toward KBR-ART, and examine several applications of static and dynamic machine learning approaches for realizing the KBR-ART framework potentials in maximizing tumor control and minimizing side effects with respect to individual radiotherapy patients. Specifically, three questions required for the realization of KBR-ART are addressed: (1) what knowledge is needed; (2) how to estimate RT outcomes accurately; and (3) how to adapt optimally. Different machine learning algorithms for KBR-ART application shall be discussed and contrasted. Representative examples of different KBR-ART stages are also visited.

OPEN ACCESS

Edited by:

John Varlotto,
University of Massachusetts Medical
School, United States

Reviewed by:

Hengyong Yu,
University of Massachusetts Lowell,
United States
Yidong Yang,
University of Miami, United States

*Correspondence:

Huan-Hsin Tseng
thuanhsi@med.umich.edu

Specialty section:

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

Received: 25 February 2018

Accepted: 27 June 2018

Published: 27 July 2018

Citation:

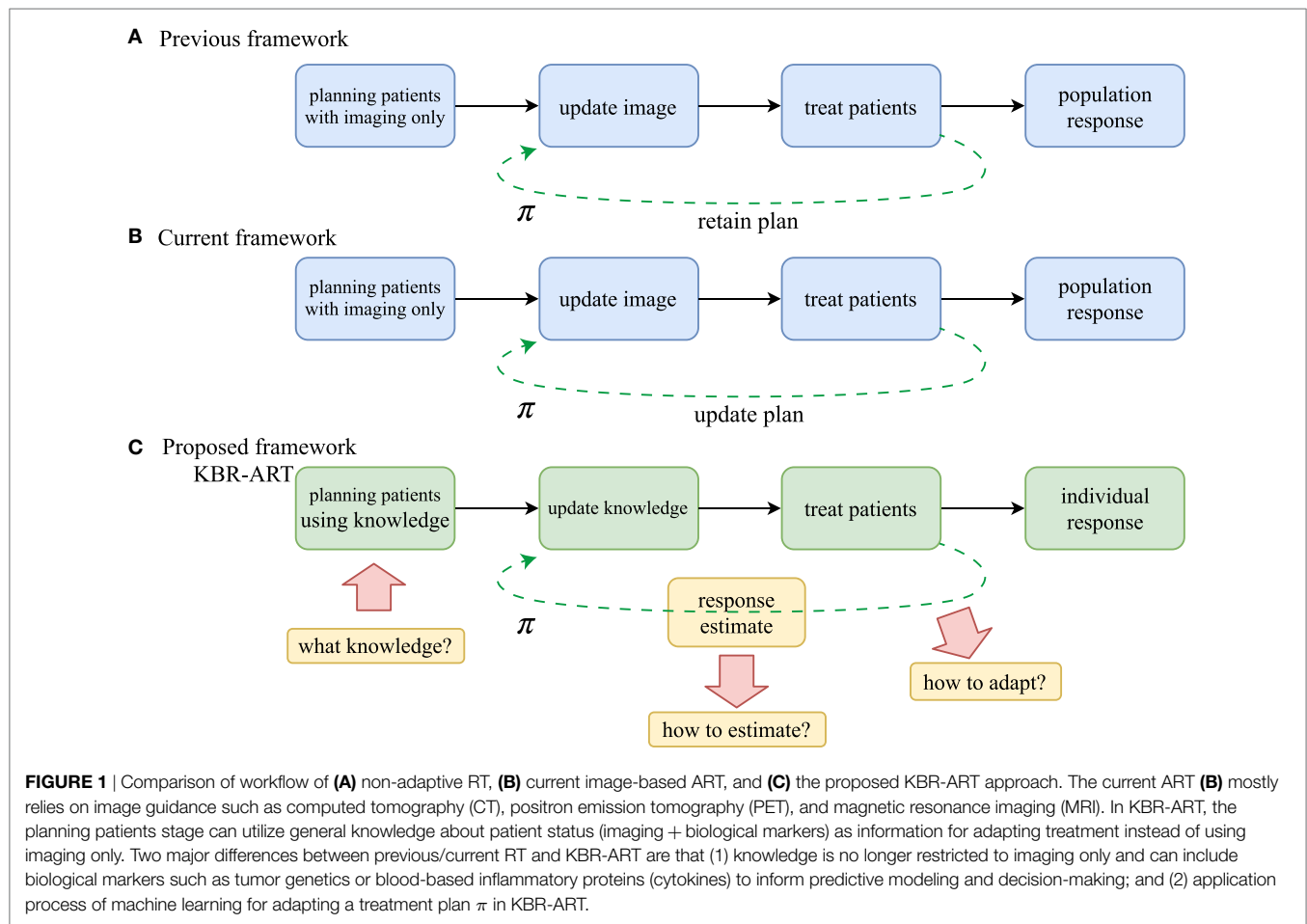
Tseng H-H, Luo Y, Ten Haken RK and
El Naqa I (2018) The Role of Machine
Learning in Knowledge-Based
Response-Adapted Radiotherapy.
Front. Oncol. 8:266.
doi: 10.3389/fonc.2018.00266

Keywords: adaptive radiotherapy, personalized treatment, deep learning, statistical learning, big data

1. INTRODUCTION

Recent advances in cancer multimodality imaging (CT/PET/MRI/US) and biotechnology (genomics, transcriptomics, proteomics, etc.) have resulted in tremendous growth in patient-specific information in radiation oncology, ushering in the new era of Big Data in radiotherapy. With the availability of the individual-specific data, such as clinical, dosimetric, imaging, molecular markers, before and/or during radiotherapy (RT) courses, new opportunities are becoming available for personalized radiotherapy treatment (1, 2).

The synthesis of this information into actionable knowledge to improve patient outcomes is currently a major goal of modern radiotherapy (RT). Subsequently, knowledge-based response-adapted radiotherapy (KBR-ART) has emerged as an important framework that aims to develop personalized treatments by adjusting dose distributions according to clinical, geometrical changes, and physiological parameters observed during a radiotherapy treatment course. The notion of KBR-ART extends the traditional concept of adapted RT (ART) (3, 4), primarily based on imaging information for guidance, into a more general ART framework that can receive and process all relevant patient-specific signals that can be useful for adaptive decision-making. Our goal is to explore in more details the processes involved in the KBR-ART framework that would allow aggregating and analyzing relevant patient information in a systematic manner to achieve more accurate decision making and optimize long-term outcomes.



The proposed KBR-ART framework can be thought of as being comprised of four stages, as depicted in **Figure 1**. These stages include: (1) *planning patients using available knowledge*, or pre-treatment modeling, (2) *updating the prediction models with evolving knowledge through the course of therapy*, or during-treatment modeling, (3) personalizing initial patient's treatments, and (4) *adapting the initial treatment to individual's responses*, where the two middle steps can be repeated at each radiation dose fraction (or few fractions) so that optimal treatment objectives are met and potentially long-term goals are optimized, i.e., long-term tumor control with limited side effects to surrounding normal tissues.

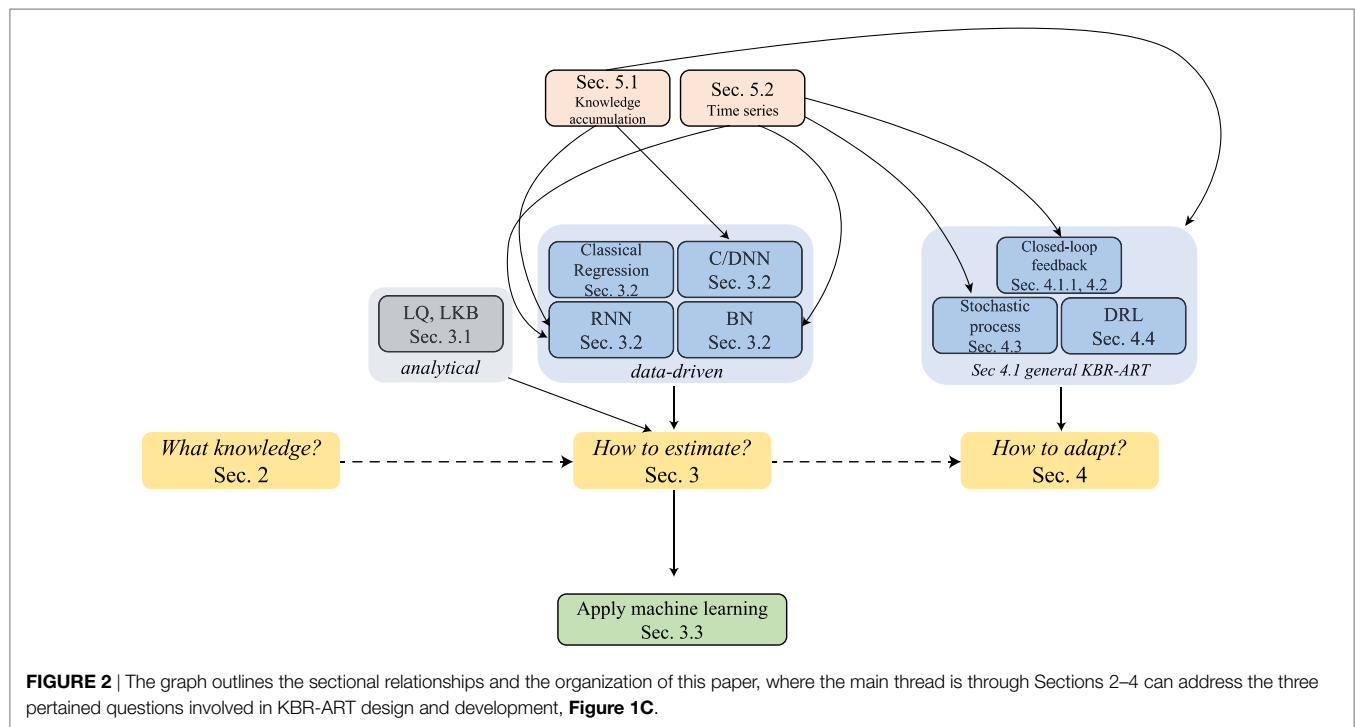
The first step in the implementation of a KBR-ART framework starts at the planning stage of patients by extending the current "image-only patients" into a more general preparation stage that can incorporate all relevant informatics signals for evaluating available treatment options, c.f. **Figures 1A,B**. Thus, the "K" in our KBR-ART refers to any useful knowledge (e.g., imaging (CT/PET/MRI) and biological markers (genomics, transcriptomics, proteomics, etc.) that can potentially aid the process of personalizing treatment to an individual patient's molecular characteristics and is not limited to imaging only as currently is the case. In Section 2, we shall introduce four major categories of data that are relevant to improved knowledge synthesis in RT. As the era of Big data (BD) is upon us, many useful tools applied for BD analytics are being actively developed in the context of modern

machine learning algorithms, where KBR-ART is expected to be a prime beneficiary of this progress toward the development of dynamically personalized radiotherapy treatment leading to better outcomes and improved patients' quality of life. However, there are three essential questions pertaining to the successful development of a KBR-ART framework in radiotherapy that need to be addressed:

- Q1: What knowledge should be synthesized for radiotherapy planning?
 Q2: How can we develop powerful predictive outcome modeling techniques based on such knowledge?
 Q3: How can we use these models in a strategically optimal manner to adapt a patient's treatment plan?

The answers to these three questions are at the core of successful development of the proposed KBR-ART framework and we shall attempt to address them in more depth in Sections 2–4 of this paper. During the process of exploring the answers to these questions, we shed more light on the pivotal role that machine learning algorithms play in the design and development of a modern KBR-ART system in subsequent sections as outlined in **Figure 2**.

A major inherent merit of the KBR-ART framework is that the treatment planning would be designed to dynamically adapt to



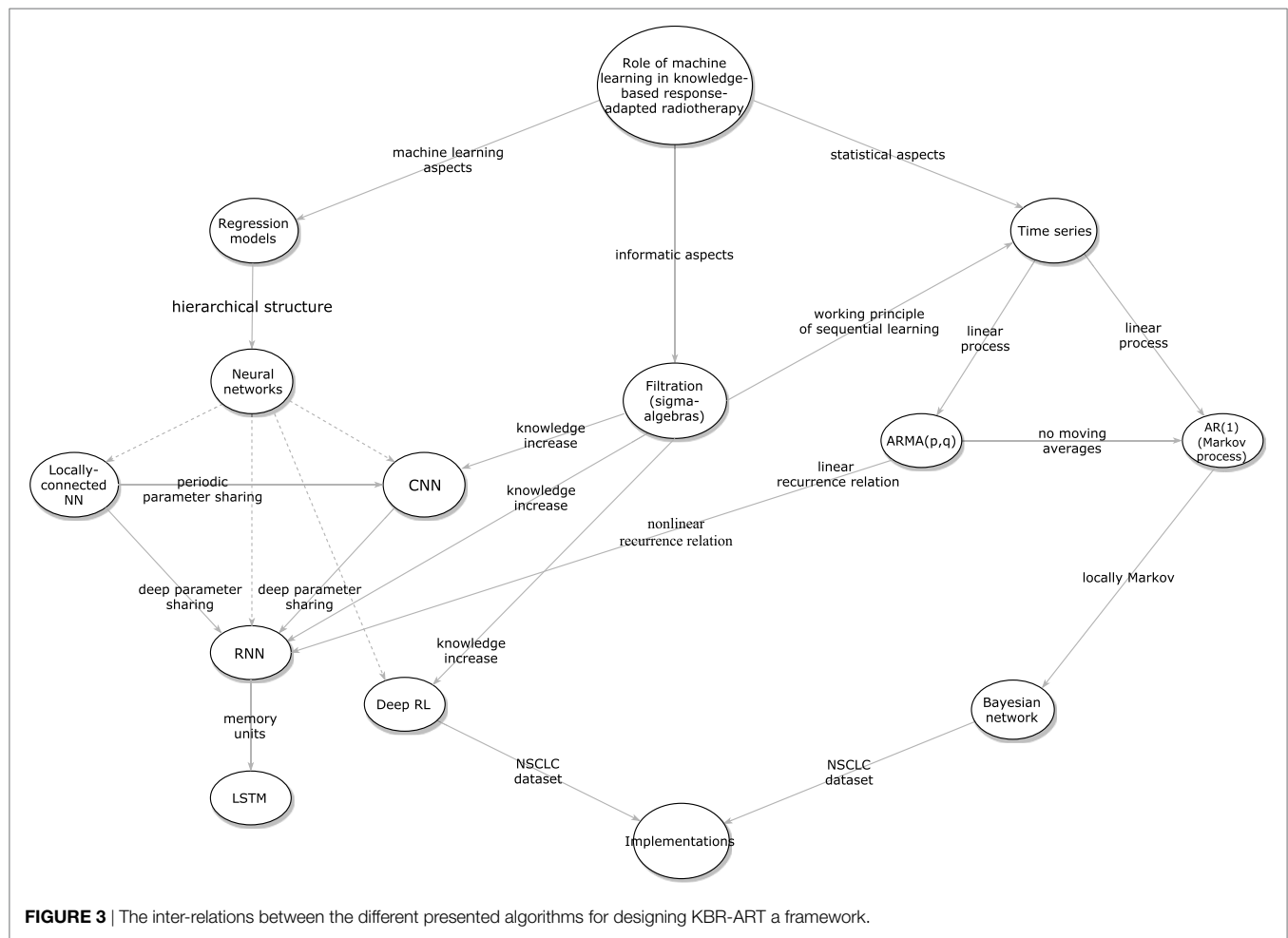
ongoing changes during the course of therapy to optimize radiotherapy goals of eradicating the tumor while minimizing harm to uninvolved normal tissue based on the individual patient's characteristics. As shown in **Figure 1**, adaptation of a treatment plan can be more formally accomplished in accordance to a decision making function π . This is represented in **Figure 1A** for the previous/current framework, where π is a non-varying function but in the case of KBR-ART, **Figure 1B**, π is a time-varying function that depends on the information (knowledge updates) available during the course of therapy. The following scenario may be used as an example on how KBR-ART can be implemented in practice: a given planned radiation course was considered optimal according to an initial population-based model such as traditional dose-based tumor control probability (TCP) and normal tissue complication probability (NTCP) and the goal is to optimize the uncomplicated tumor control [$p^+ = \text{TCP} \cdot (1 - \text{NTCP})$], for instance. Then, through the course of fractionated radiotherapy treatment, the patient did not achieve the predicted TCP value as expected, or worse suffered from unexpected toxicities due to treatment, i.e., NTCP exceeded the designed risk limit. This is where KBR-ART comes into action; to learn from current observations with its previous decisions taking into account available information during therapy and to adjust the course of action [e.g., increase dose to improve TCP or decrease it to specific organ-at-risk (OAR) to limit its NTCP] and develop a better personalized treatment plan based on the updated knowledge (from imaging and biomarkers) of the specific patient under treatment as shown in **Figure 1B**.

Much effort of this study will be devoted to tackling questions (ii and iii), which requires consideration of some advanced data-driven models that can also incorporate temporal information (i.e., knowledge updates). The steps involved in the development

of a knowledge-adapted plan using the KBR-ART framework will be the main subject of this paper. For this purpose, we will first review pertained modern machine learning algorithms that feature modeling of sequential data. These include efficient *deep-learning* approaches such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and the more recently developed deep reinforcement learning (DRL). The subject of sequential data modeling have been applied in many diverse fields, such as handwriting recognition (5), speech recognition (6), bioinformatics (7), medical care (8, 9), and also high energy physics (10).

The introduced algorithms based on deep learning would require some basic background of neural networks (NNs) which are briefly reviewed in Section 3.2.2. Most of the notations in this paper are self-contained and self-consistent. In addition to the presented advanced data-driven models, we also provide probabilistic and statistical perspectives as a theoretical foundation for sequential machine learning models. In particular, via "*filtration*" we are to describe notions related to "knowledge accumulation" or "growing of knowledge" in more concrete manner. A main part of KBR-ART development relies on constructing a new RT plan prescription based on historical information; thus we would like to address issues related to representing knowledge accumulation in sequential learning models.

Moreover, we recognize that KBR-ART has a close analogy to stock pricing or autonomous car driving, in that it shares the same goal of analyzing acquired information a long a period of time to maximize final rewards (e.g., better radiotherapy treatment outcome in our case). Therefore, techniques derived from time series analysis will be helpful to analyze such sequential data from an analytical perspective, such as the trends and the stationarity of such stochastic (random) processes. In particular, it suffices for



our purpose to revisit main *linear processes*, such as the autoregressive moving average (ARMA) model and its natural descendant the autoregressive (AR) models, which can be linked to Bayesian networks (BNs), another useful approach for dynamical learning as summarized in **Figure 3**. Together, our goal is to provide a comprehensive overview and a frontier survey that covers the major facets for the application of KBR-ART and layout the foundation for this emerging field.

It worth noticing that we organized the sections of this paper so that it follows the necessary building steps for the development of a successful KBR-ART framework as pertained to addressing the three aforementioned questions involved in KBR-ART implementation and review the related literature accordingly. Two implementations using non-small lung cancer (NSCLC) datasets will be presented for illustration.

2. Q1: WHAT KNOWLEDGE TO BE USED FOR KBR-ART PLANNING?

There are four major types of RT data that are potentially useful as part of the knowledge synthesis for KBR-ART: *clinical, dosimetric, imaging radiomics, and biological data*. To understand why and how they can be informative for assessing treatment outcomes, we provide a brief description about these four categories of data.

2.1. Clinical Data

Clinical data refers to cancer diagnostic characteristics (e.g., grade, stage, histology, site, etc.), physiological metrics (e.g., blood cell counts, heart/pulse rates, pulmonary measurements, etc.), and patient-related information (e.g., comorbidities, gender, age, etc.). Due to their nature, clinical data can usually be found in unstructured format such that can be challenging for extracting information directly. Therefore, machine learning techniques for natural language processing could be useful for transforming such data into structured format (e.g., tabulated) before further processing (11).

2.2. Dosimetric Data

Dosimetric data are informative to the treatment planning process in RT, which includes simulated calculation of radiation dose using computed tomography (CT) imaging. In particular, dose-volume metrics obtained out of dose-volume histograms (DVHs) are extensively investigated for outcome modeling (12–16). Useful metrics are typically the volume receiving greater than or equal to a certain dose (V_x), the minimum dose to the hottest $x\%$ of the volume (D_x), mean, maximum, minimum dose, etc. (17). Notably, a dedicated software based on MATLAB™ called “DREES” can derive these metrics automatically and apply them in outcome prediction models of RT response (18).

2.3. Radiomics Data

Radiomics is a field of medical imaging study that aims to extract meaningful quantitative features from medical images and relate this information to clinical and biological endpoints. The most common imaging modality is CT, which has been considered the standard for treatment planning in RT. Other imaging modalities used for improving treatment monitoring and prognosis in various cancer types are also used, such as positron emission tomography (PET), and magnetic imaging resonance (MRI). These modalities can be used individually or combined (19, 20).

2.4. Biological Data

According to (21) a biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention.” Measurements of biomarkers are typically based on tissue or fluid specimens, which are analyzed using molecular biology laboratory techniques (22) and have the following two categories according to their biochemical sources:

- (a) *Exogenous biomarkers*: by injecting foreign substance into patients such as that used in molecular imaging and are used in radiomics applications.
- (b) *Endogenous biomarkers*: there exists two subclasses within this category:
 - (i) *Expression biomarkers*: changes measured in protein levels or gene expression.
 - (ii) *Genetic biomarkers*: measuring variations between the underlying DNA genetic code and tumors or normal tissues.

2.5. Example: Aggregating Relevant Knowledge From a Lung Cancer Dataset

In this paper, we shall apply an institutional non-small cell lung cancer (NSCLC) dataset (23) as an example for implementation of KBR-ART. The first step is to collect relevant knowledge from such dataset that is suitable for the purposes of adapting radiotherapy treatment planning during a fractionated course. These data will be used subsequently for outcome modeling (TCP/NTCP) and plan adaptation as discussed later.

2.5.1. Data Description

The NSCLC dataset was recorded from NSCLC patients, where they have been treated on prospective protocols with standard and dose escalated fractionation under IRB approval (24). Collectively, 125 patients with relatively complete characteristics were selected for predicting TCP (local control) and NTCP (radiation pneumonitis of grade 2 or above (RP2)).

The dataset had over 250 features containing positron emission tomography (PET) imaging radiomics features, circulating inflammatory cytokines, single-nucleotide polymorphisms (SNPs), circulating microRNAs, clinical factors, and dosimetric variables before and during radiotherapy. All features were recorded at three time periods (at baseline, at 2 weeks of treatment, and at 4 weeks). However, certain features were collected only at

baseline such as microRNAs and SNPs. Thus, the data for the purpose of KBR-ART can be represented as forming 3 time blocks:

$$N \text{ samples} \left[\begin{array}{ccc|ccc|ccc} x_{11}^{(0)} & x_{12}^{(0)} & \dots & x_{1n}^{(0)} & x_{11}^{(1)} & x_{12}^{(1)} & \dots & x_{1n}^{(1)} & x_{11}^{(2)} & x_{12}^{(2)} & \dots & x_{1n}^{(2)} \\ x_{21}^{(0)} & x_{22}^{(0)} & \dots & x_{2n}^{(0)} & x_{21}^{(1)} & x_{22}^{(1)} & \dots & x_{2n}^{(1)} & x_{21}^{(2)} & x_{22}^{(2)} & \dots & x_{2n}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N1}^{(0)} & x_{N2}^{(0)} & \dots & x_{Nn}^{(0)} & x_{N1}^{(1)} & x_{N2}^{(1)} & \dots & x_{Nn}^{(1)} & x_{N1}^{(2)} & x_{N2}^{(2)} & \dots & x_{Nn}^{(2)} \end{array} \right], \quad (1)$$

where $x_{ij}^{(k)}$ denotes the value of the j th feature of patient i at time period k .

Values of mean tumor and lung doses were computed in their 2 Gy equivalents (EQD2) by using the linear-quadratic (LQ) model (Section 3.1.1) with $\alpha/\beta = 10$ Gy, 4 Gy for the tumor and the lung, respectively. Generalized equivalent uniform doses (gEUDs) with various a parameters were also calculated for gross tumor volumes (GTVs) and uninvolved lungs (lung volumes exclusive of GTVs).

3. Q2: HOW TO ESTIMATE RADIOTHERAPY OUTCOME MODELS FROM AGGREGATED KNOWLEDGE?

Radiotherapy outcome models are typically expressed in terms of tumor control probability (TCP) and normal tissue complication probability (NTCP) (25, 26). In principle, both TCP and NTCP may be evaluated using analytical and/or data-driven models. Though the former provides structural formulation, it can be incomplete and less accurate due to the complexity of radiobiological processes. On the other hand, data-driven models tend to learn empirically from the data observed, and thus they are capable of considering higher complexities and interactions of irradiation with the biological system. The trade-offs between analytical models and data-driven models can vary in terms of radiobiological understanding and prediction accuracy. In the following, we list examples, more detailed description on treatment outcome models can be found in (27).

3.1. Analytical Models

These models are generally based on simplified understanding of radiobiological processes and can provide a mechanistic formalism of radiation interactions with live tissue.

3.1.1. TCP

The most prevalent TCP models are based on the linear quadratic (LQ) model (28) parametrized by the radiosensitivity ratio α/β derived from clonogenic cell survival curves. The LQ model expresses the survival fraction (SF) after irradiation as follows:

$$SF = e^{-\alpha D - \beta D^2}, \quad (2)$$

where $D \geq 0$ is the total delivered dose. For n fractions of dose d in uniformly delivered fractions is represented by:

$$SF = e^{-n(\alpha d + \beta d^2)}. \quad (3)$$

Many types of TCP models were proposed (28) in the literature such as the birth-death (29) and the Poisson-based (30) models, which are expressed as:

$$TCP = e^{-N \cdot e^{-n(\alpha d + \beta d^2) - t \ln 2 / T_{pot}}}, \quad (4)$$

where N is the initial number of clonogenic cells, and T_{pot} denotes the potential cell doubling time, with t as the time difference within the total treatment elapse T , the lag period before accelerated clonogenic repopulation begins.

3.1.2. NTCP

The most frequently used analytical model is the Lyman–Kutcher–Burman (LKB) model, which is a phenomenological approach (31). In the uniform dose case, NTCP is expressed by a gaussian integral (probit function):

$$NTCP_{m,D_{50}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad \left(x = \frac{D - D_{50}}{m D_{50}}\right), \quad (5)$$

where D_{50} is defined as the dose that corresponds to NTCP probability (curve in Figure 4) of 50% and m is a parameter tuning the shape of the NTCP curve. Typical trade-off between TCP and NTCP to achieve a therapeutic ratio is shown in Figure 4.

To account for dose inhomogeneities in developing TCP/NTCP models, the Equivalent Uniform Dose (EUD) (32) or Generalized EUD (gEUD) (33) are used. Mimicking a weighted sum of doses, gEUD is given by:

$$gEUD = \sqrt[a]{\sum_{i=1}^n v_i D_i^a}, \quad (6)$$

where v_i is the fractional organ volume receiving dose D_i and a is a volume parameter that depends on the tissue type. An $a < 0$ value will correspond to minimum dose effect, which is typically associated with tumor response. An $a > 0$ value will

correspond to maximum dose effect, which is typically associated with serial normal tissue architecture response, while an $a = 1$ will correspond to mean dose effect, which is associated with parallel normal tissue architecture response.

More complex analytical models for toxicity can be developed by incorporating variables other than dose in the LKB model, for instance (34, 35):

$$NTCP_{m,D_{50},DMFs}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \quad (7)$$

with

$$x = \frac{D_{eff} \cdot DMF_1 \cdot DMF_2 \cdots DMF_k - D_{50}}{m D_{50}},$$

where the DMFs are dose modifying factors and represent the impact of covariates other than dose (e.g., single-nucleotide polymorphism (SNPs) genotype, copy number variations (CNVs), smoking status, etc.). Although analytical models are useful, in many circumstances, they are simply approximations of the complex physical and biological processes that are currently beyond such simple formalisms. Therefore, more data-driven approaches are being sought to achieve more accurate predictions of TCP/NTCP.

3.2. Data-Driven Models

By definition, data-driven models are approximations built based on observation of data. However, one drawback is that such modeling is likely not unique even from the same dataset and, therefore, one needs to choose a suitable technique that fits one’s dataset best, which is an open question in the data science world. The purpose main of this section is to present several advanced data-driven techniques that can suite the implementation of predictive outcome modeling component of the KBR-ART framework. Below, we summarize some frequently used data-driven techniques for outcome modeling ranging from classical regression models to more advanced machine learning techniques.

3.2.1. Classical Models

Regression models such as Ridge, LASSO, and Logistic are commonly used for building outcome models and follow conventional statistical approaches (36). They are essentially constructed by minimizing the following objective:

$$\mathcal{L}(\mathbf{w}) = \sum_{i=1}^N [y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b)]^2 + \lambda \cdot h(\mathbf{w}), \quad (8)$$

where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$, $i = 1, \dots, N$, are the data input and outputs, respectively. Here, the weights $\mathbf{w} \in \mathbb{R}^n$ and bias $b \in \mathbb{R}$ are unknown parameters to be fitted by minimizing **regression error**, Equation (8). The second term in Equation (8) represents **penalty**, usually used to suppress possible model’s overfitting. There are several types of penalty corresponding to different model characteristics, such as $h(\mathbf{w}) = \|\mathbf{w}\|$ is called the **LASSO** by Tibshirani (37), $h(\mathbf{w}) = \|\mathbf{w}\|^2$ is called the **Rigid (Tikhonov) regularization** (37), and $h(\mathbf{w}) = \lambda_1 \|\mathbf{w}\| + \lambda_2 \|\mathbf{w}\|^2$ is called the

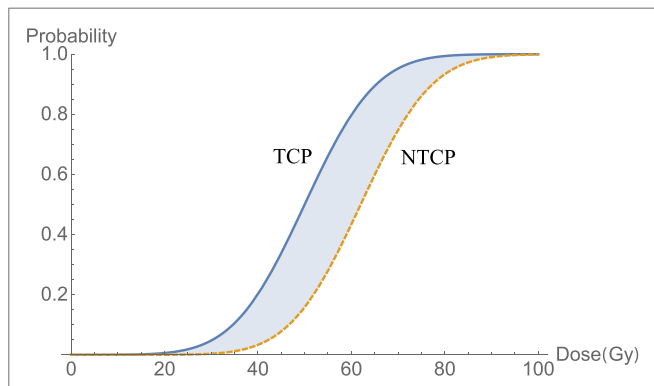


FIGURE 4 | An illustration of a therapeutic ratio showing that the trade-off between TCP and NTCP as delivered dose increases. The blue-shaded area between two curves TCP (blue) and NTCP (orange-dashed) is a best window for dose delivery.

Elastic Net regularization (38). The regularization parameter λ controls the magnitude of the penalty.

Due to the characteristic of L_1 -norm, $\|\cdot\|_1$, the LASSO regularization tends to suppress many parameters to equal zero, so that the parameter vector is sparse, which makes it a natural candidate for relevant *feature selection* (39).

Another benefit of regression models other than their simplicity is the convex optimization property of their loss function, which guarantees optimal fitting parameters $\mathbf{w} = \mathbf{w}_*$. In fact, it can be explicitly solved using simple matrix inversion $\mathbf{w}_* = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \cdot \mathbf{X}^T \mathbf{y}$, for Ridge regression, for instance, where \mathbf{X} is known from the given data:

$$\mathbf{X} \stackrel{\text{def}}{=} \begin{pmatrix} - & \mathbf{x}_1 & - \\ - & \mathbf{x}_2 & - \\ & \vdots & \\ - & \mathbf{x}_N & - \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nn} \end{pmatrix}. \quad (9)$$

3.2.2. Neural Networks

One notable model in machine learning is called **Neural Networks** (NN), which are inspired by the neurobiology of the brain, and hence the name. Mathematically, NNs utilize (repeated) composition of nonlinear transformations in developing their architecture. The definition is fairly simple (40); given a set of data inputs $\mathbf{x}_i \in \mathbb{R}^n$ and labels $\mathbf{y}_i \in \mathbb{R}$, $i = 1, \dots, N$ as defined above, a NN is aimed to approximate a function of the form:

$$f_{\mathbf{w}, \mathbf{b}}(\mathbf{x}) = \sigma_L(\mathbf{w}^{(L)} \cdot \sigma_{(L-1)}(\mathbf{w}^{(L-1)} \cdots \sigma_1(\mathbf{w}^{(0)} \cdot \mathbf{x} + \mathbf{b}^{(0)}) + \mathbf{b}^{(L-2)}) + \mathbf{b}^{(L)}), \quad (10)$$

via adjusting unknown coefficients $\{\mathbf{w}^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell+1}}\}_{\ell=0}^L$ and $\{\mathbf{b}^{(\ell)} \in \mathbb{R}^{n_\ell}\}_{\ell=0}^L$ such that the loss function is minimal between the data and the NN model:

$$\mathcal{L}(\{\mathbf{w}^{(\ell)}\}, \{\mathbf{b}^{(\ell)}\}) = \sum_{i=1}^N g(\mathbf{y}_i, f_{\mathbf{w}, \mathbf{b}}(\mathbf{x}_i)), \quad (11)$$

where in Equation (10), the given functions $\sigma_\ell: \mathbb{R}^{n_\ell} \rightarrow \mathbb{R}^{n_\ell}$ are called *activation functions*, which are fixed for a particular architecture. The integer L of max composition is interpreted as layers with index $\ell = 0, \dots, L$ denoting the layer number as shown in **Figure 5A** and n_ℓ is an integer denotes the number of **nodes (neurons)** in layer ℓ . The function g in Equation (11) should also be fixed depending on data query type. For continuous labels \mathbf{y}_i , such NN is called a regression prediction function with $g(\mathbf{y}, \mathbf{h}(\mathbf{x})) = \|\mathbf{y} - \mathbf{h}(\mathbf{x})\|^2$ typically adopted for an arbitrary loss function $\mathbf{h}: \mathbb{R}^n \rightarrow \mathbb{R}^m$. For discretized labels of multidimensions $\mathbf{y} = (y_1, \dots, y_m)$, such NN is called a classification prediction function with cross entropy loss function $g(\mathbf{y}, \mathbf{h}(\mathbf{x})) = \sum_{k=1}^m [y_k \log(h_k(\mathbf{x})) + (1 - y_k) \log(1 - h_k(\mathbf{x}))]$ typically chosen with $\mathbf{h} = (h_1, \dots, h_m)$.

In practice, there are several choices for activation functions σ_i , such as sigmoid, ReLU, eLu, Leaky ReLU function, etc., whose effectiveness usually depends on the nature of the dataset and the problem in question. The terms relating *forward dynamics*, *error backward propagation*, and *weights gradient descent* are

technical procedures for estimating the unknown coefficients $\{\mathbf{w}^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell+1}}\}_{\ell=0}^L$ and $\{\mathbf{b}^{(\ell)} \in \mathbb{R}^{n_\ell}\}_{\ell=0}^L$ from Equation (11). Although the design construction of an NN is relatively simple, the proper optimization of its parameters could be tedious numerically (40, 41).

In general, it is conventionally dubbed a *deep neural network* (DNN) when the number of hidden layers exceed 2, or $L \geq 3$. These neural networks are widely applied and are the foundations for the emerging field of *deep learning*, which is currently overperforming many of the classical machine learning techniques.

3.2.3. Deep learning Models

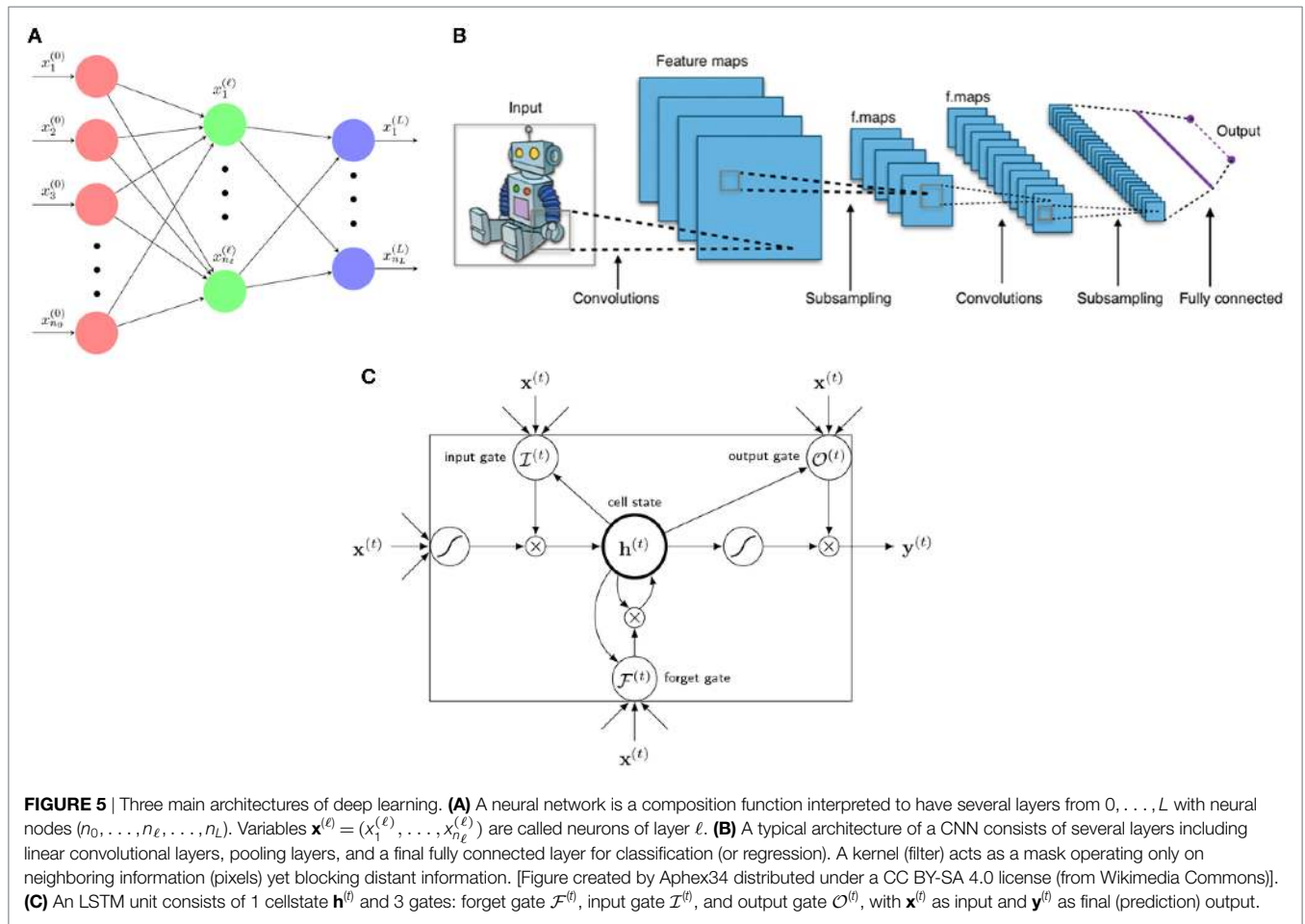
In KBR-ART, one expects that the processes involved in outcome modeling and adaptation procedures can be quite complex in nature for individualizing patient's treatment according to her/his predicted response over the course of fractionated therapy. There are few advanced data-driven models, mostly deep learning based, which can effectively into consideration such temporal information for updating knowledge and interactions between physical and biological variables for adapting therapy. In the following, we will briefly describe some of the main deep learning technologies in the literature.

3.2.3.1. Convolutional Neural Networks (CNNs)

CNNs are best known for image recognition and image-related prediction. The idea of CNN stemmed from the successful application of the signal processing operation of *convolution* in imaging processing, which was then been applied into neural networks for handling image related tasks. A CNN typically consists of several *convolutional layers*, *pooling layers*, *with activation functions* (42), where the convolution layer is the core component that applies an efficient convolutional filter (kernel) to the data in contrast to the tedious matrix operations described earlier with standard NN. In the case of a 2D image of size $L_1 \times L_2$ with multi color channels (C_1), the data are represented by a 3d-tensor $\mathcal{I} = \{\mathcal{I}_{i,j,\alpha}\}_{i=1, j=1, \alpha=1}^{L_1, L_2, C_1} \in \mathbb{R}^{2+1}$, a convolutional layer with stride s renders an output image (also called **feature maps**) $\tilde{\mathcal{I}}$ (of size $\tilde{L}_1 \times \tilde{L}_2$ with C_2 channels) by applying the following convolution process (42).

$$\tilde{\mathcal{I}}_{k,\ell,\beta} = \sum_{m,n,\alpha}^{L_1, L_2, C_1} w_{m,n,\alpha,\beta} \cdot \mathcal{I}_{s(k-1)+m, s(\ell-1)+n, \alpha} \quad (\text{image convolution. } k = 1, \dots, \tilde{L}_1, \ell = 1, \dots, \tilde{L}_2, \beta = 1, \dots, C_2) \quad (12)$$

here, $\mathbf{w} = \{w_{m,n,\alpha,\beta}\}_{m=1, n=1, \alpha=1, \beta=1}^{L_1, L_2, C_1, C_2} \in \mathbb{R}^{2+1+1}$ is a 4-tensor convolutional kernel. Such convolution process with stride is then equivalent to a regular convolution with image downsampling procedure. In fact, one can recognize that CNNs use these kernels in a neural network to "capture" local information within a neighborhood while "blocking" distant information or less related ones, as depicted in **Figure 5B**. Activation functions in CNNs have similar choices as a standard NN, Equation (10), mentioned above. CNNs has been successfully applied for image segmentation (43–46) in radiotherapy and for modeling of rectal toxicity in cervical cancer using transfer learning (47, 48). This will be further discussed in Section 3.3.1.



3.2.3.2. Recurrent Neural Networks (RNNs)

RNNs are another variant of neural networks especially useful for learning sequential data, such as voice, text data, and handwriting. Therefore, it is also considered ideal for sequential adaptive radiotherapy with changing dose fractionations. In this case, suppose that we have sequential data $\{\mathbf{x}^{(t)} \in \mathbb{R}^n | t \in T\}$ as an input and $\{\tilde{\mathbf{y}}^{(t)} \in \mathbb{R}^m | t \in T\}$ as the corresponding labels where T denotes an index set (continuous or discrete) labeling separation across time steps. An important property of a RNN is that it introduces hidden units $\{\mathbf{h}^{(t)} \in \mathbb{R}^k | t \in T\}$ for making neural network deeper in increasing sequential prediction. A RNN is then aimed to learn the relationships between data $\{\mathbf{x}^{(t)} \in \mathbb{R}^n\}$ and labels $\{\tilde{\mathbf{y}}^{(t)}\}$ via hidden units $\{\mathbf{h}^{(t)} \in \mathbb{R}^k\}$ dynamically.

An RNN is designed to model the hidden variables via the recursive function $f_\theta : \mathbb{R}^k \times \mathbb{R}^n \rightarrow \mathbb{R}^k$.

$$\mathbf{h}^{(t)} = f_\theta (\mathbf{h}^{(t-1)}, \mathbf{x}^{(t)}) \in \mathbb{R}^k, \quad (13)$$

where θ usually serves as unknown neural weights to be solved, as $\{\mathbf{w}^{(\ell)}, \mathbf{b}^{(\ell)}\}_{\ell=0}^L$ in Equation (10).

One of the most successful RNN is the Long Short-Term Memory (LSTM). An LSTM is a state-of-the-art RNN model effective in sequential learning utilizing the so-called *gated units*, who learns by itself to store and forget internal memories when needed such that it is capable of creating long-term dependencies and paths

through time, **Figure 5C**. A LSTM is constructed by 3 gates and 1 cell (hidden) state built up by the following equations.

$$\begin{aligned} \mathcal{F}^{(t)} &= \sigma_g + (W_{\mathcal{F}} \cdot \mathbf{x}^{(t)} + U_{\mathcal{F}} \cdot \mathbf{h}^{(t-1)} + \mathbf{b}_{\mathcal{F}}) \in [0, 1] \\ \mathcal{I}^{(t)} &= \sigma_g + (W_{\mathcal{I}} \cdot \mathbf{x}^{(t)} + U_{\mathcal{I}} \cdot \mathbf{h}^{(t-1)} + \mathbf{b}_{\mathcal{I}}) \in [0, 1] \\ \mathcal{O}^{(t)} &= \sigma_g + (W_{\mathcal{O}} \cdot \mathbf{x}^{(t)} + U_{\mathcal{O}} \cdot \mathbf{h}^{(t-1)} + \mathbf{b}_{\mathcal{O}}) \in [0, 1] \\ \mathbf{h}^{(t)} &= \mathcal{F}^{(t)} \circ \mathbf{h}^{(t-1)} + \mathcal{I}^{(t)} \circ \sigma_h (W_h \cdot \mathbf{x}^{(t)} + U_h \cdot \mathbf{h}^{(t-1)} + \mathbf{b}_h) \\ \mathbf{y}^{(t)} &= \mathcal{O}^{(t)} \circ \sigma_y (\mathbf{h}^{(t)}), \end{aligned} \quad (14)$$

where $\sigma_g, \sigma_h, \sigma_y$ are 3 non-linear activation functions depending on one's choice, $\{\mathcal{F}^{(t)}, \mathcal{I}^{(t)}, \mathcal{O}^{(t)}\}$ are called the **forget gate, input gate, and output gate** at time t , respectively.

The 3 gates, with all their numerical values in $[0,1]$, are used to control and determine when and how much should the previous information be kept or forgotten. The unknown parameters of an LSTM are (W_h, U_h, \mathbf{b}_h) and $\{(W_\alpha, U_\alpha, \mathbf{b}_\alpha) | \alpha = \mathcal{F}, \mathcal{I}, \mathcal{O}\}$ and, therefore, an LSTM unit generally possesses four times parameters than a plain neural net in Equation (10) requiring a large amount of data for training. RNNs have been evaluated in radiotherapy for respiratory motion management (49). An interesting approach combining RNN with CNN was used for pancreas segmentation

on both CT and MRI datasets, which mitigated the problem of using spatial smoothness consistency constraints (50, 51).

The previously presented machine learning methods do not allow visualization of the system dynamics and act primarily as a black box mapping from the input to the output data and are referred to as *discriminant* models. Alternatively, system dynamics of mapping input to output data can be revealed using so-called *generative* models. A common example of such models is Bayesian networks, which will be discussed next.

3.2.4. Bayesian Networks

Bayesian networks (BNs) are a class of probabilistic graphical models (GM) corresponding to directed acyclic graphs (DAGs), which are also named as belief networks. BNs combine graph theory, probability theory, computer science, and statistics to represent knowledge in an uncertain domain. They are popular in the societies of statistics, machine learning, and artificial intelligence. Especially, BNs are mathematically rigorous and intuitively understandable, which enable an effective way to represent and compute the joint probability distribution (JPD) over a set of random variables (52).

Each BN includes the sets of nodes and directed edges. While the former indicate random variables represented by circles, the latter display direct dependencies among these variables illustrated by arrows between nodes. In a BN, an arrow from node X_i to node X_j shows a statistical dependence between them, which indicates that a value of variable X_j depends on that of variable X_i , or variable X_i “affects” X_j . Also, their relationship can be described as follows: Node X_i is a parent of X_j and node X_j is the child of X_i . In general, the set of nodes that can be reached on a direct path from the node is named as the set of its descendants, and the set of nodes from which the node can be reached on a direct path is called as the set of its ancestor nodes (53).

The DAG structure guarantees that no node can be its own ancestor or its own descendant, which is of vital importance to the factorization of the JPD of a collection of nodes. A BN is designed to reflect a conditional independence statement, where each variable is independent of its nondescendants in the BN given its parents. This property is used to significantly reduce the number of parameters required to characterize the JPD of the variables. Especially, this reduction leads to an efficient way in computing the posterior probabilities given the evidence (52, 54, 55).

Moreover, the parameters of the BN are described in a manner following a Markovian property, where the conditional probability distribution (CPD) of each node only depends on its parents. These conditional probabilities are often represented by a table for discrete random variables to list the conditional probability that a child node takes on each of the feasible values from each combination of values of its parents. The joint distribution of a collection of variables can be obtained uniquely by these conditional probability tables (CPTs).

Generally, a BN B can be considered as a DAG that represents a joint probability density function over a set of random variables V . The BN is defined by a pair $B = \langle G, \phi \rangle$, where G is the DAG whose nodes X_1, X_2, \dots, X_n denotes random variables, and whose edges indicate the direct dependencies between them. The graph G includes independence assumptions, where each variable X_i is

independent of its nondescendants given its parents in G . The second component ϕ represents the set of parameters of the BN. This set contains the parameter $\theta(x_i|\pi_i) = P_B(x_i|\pi_i)$ for each realization x_i of X_i conditioned on π_i , which is the set of parents of X_i in G . Then, B describes a unique JPD over V :

$$P_B(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P_B(X_i|\pi_i) = \prod_{i=1}^n \theta_{X_i|\pi_i}, \quad (15)$$

where if X_i does not have parents, its probability distribution is considered to be unconditional; otherwise it is conditional. Once the variable indicated by a node is observed, the node is considered as an evidence node; otherwise the node is treated as a hidden or latent node. Because of their generative nature, BNs have been widely applied for modeling radiotherapy errors (56, 57) and outcomes (58–62). This will be further discussed in Section 3.3.2.

3.3. Example Application of Machine Learning to Outcome Modeling

As examples of application of modern machine learning to outcome modeling, in the following, we discuss application of a discriminant modeling approach by CNN of rectal toxicity and a generative modeling approach by BN for lung toxicity.

3.3.1. NTCP Modeling of Rectal Toxicity Using CNN

Zhen et al. (47) studied the possibility of modeling rectal toxicity in cervical cancer using CNNs from unfolded rectum surface dose maps (RSDMs) (63) with the help of transfer learning, as depicted in **Figure 6**. A retrospective data of 42 cervical cancer patients were studied. These patients were treated with external beam radiotherapy (EBRT) and/or brachytherapy (BT). The EBRT was delivered in 25 fractions (2 Gy/fraction) and BT was delivered in 4–6 fractions (6–7 Gy/frac).

For transfer learning, CNN of VGG-16 (64) was chosen as optimal architecture, which consists of 16 convolutional layers of suitable sizes including up to 138 million parameters. The VGG-16 is pretrained using a publicly annotated natural images database (ImageNet). The finetuned VGG-16 on the cervix cancer dataset with ADASYN method for imbalance correction, achieved an AUC of 0.89 on leave-one-out cross validation for rectal toxicity prediction. In addition to a successful model building of relating RSDMs to toxicity, Zhen et al. also attempted to interpret what and how CNNs “view” an RSDM, where the method of Grad-CAM map (65) was utilized to unveil the nature of the CNN learnt features (**Figure 7**). From **Figure 8**, one finds that the Grad-CAM interpreted maps (d, e) (from mapping CNN weights) have high consistency of distinct image patterns with toxicity (b) and non-toxicity (c) that were recognizable by human eyes. Therefore, by visualizing the CNN model, one can have better understanding of the features learned by the machine learning algorithm.

3.3.2. NTCP Modeling of Lung Toxicity Using Bayesian Networks

Radiation pneumonitis of grade 2 or above (RP2) is a major radiation-induced toxicity in NSCLC radiotherapy, and it may depend on radiation dose, the patients’ clinical, biological, and genomic characteristics. In order to find appropriate treatment

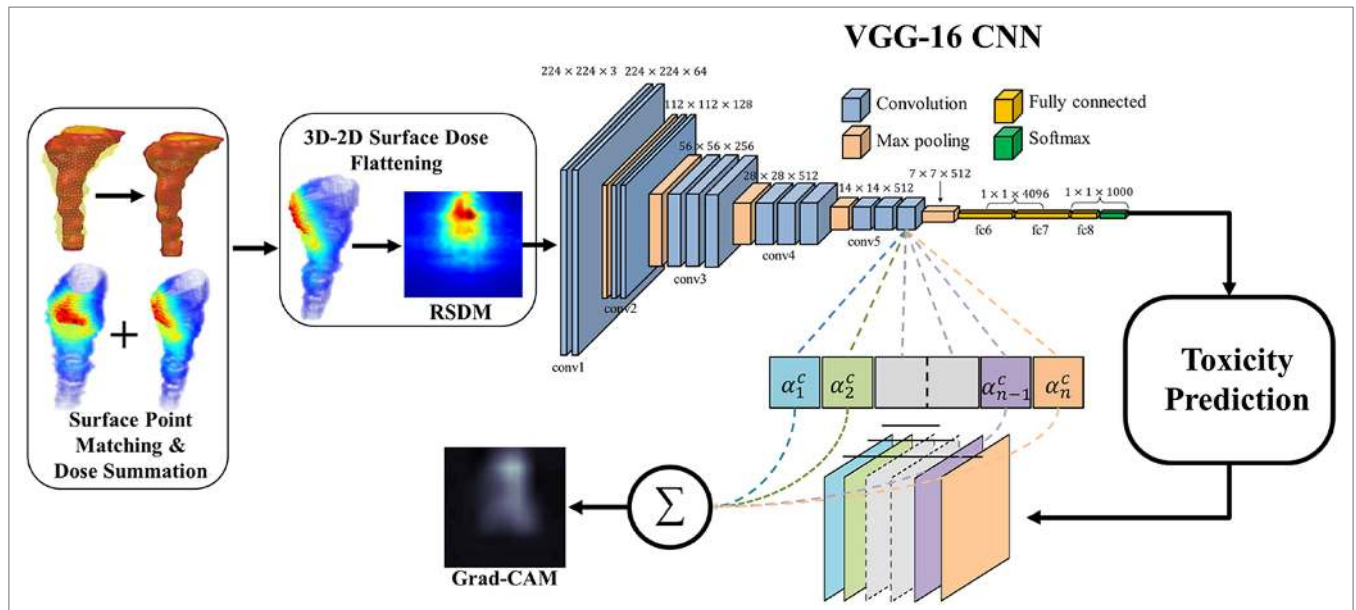


FIGURE 6 | The workflow of the rectum toxicity study in (47) using VGG-16 receiving 2D RSDM image input with Grad-CAM map as interpretation of CNN weights. [© Institute of Physics and Engineering in Medicine. Reproduced by permission of IOP Publishing. All rights reserved.]

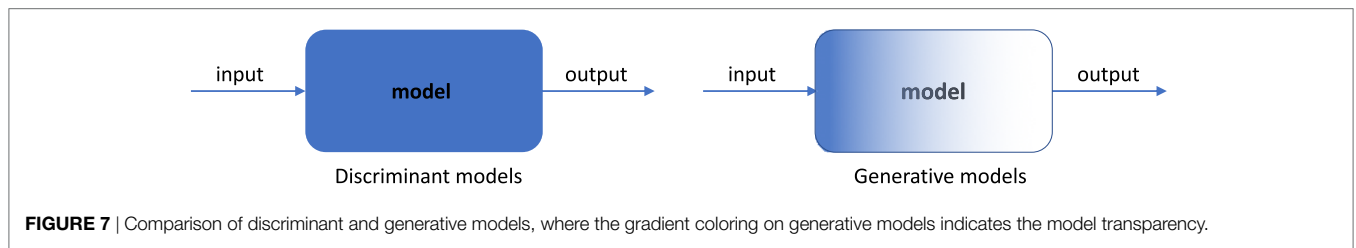


FIGURE 7 | Comparison of discriminant and generative models, where the gradient coloring on generative models indicates the model transparency.

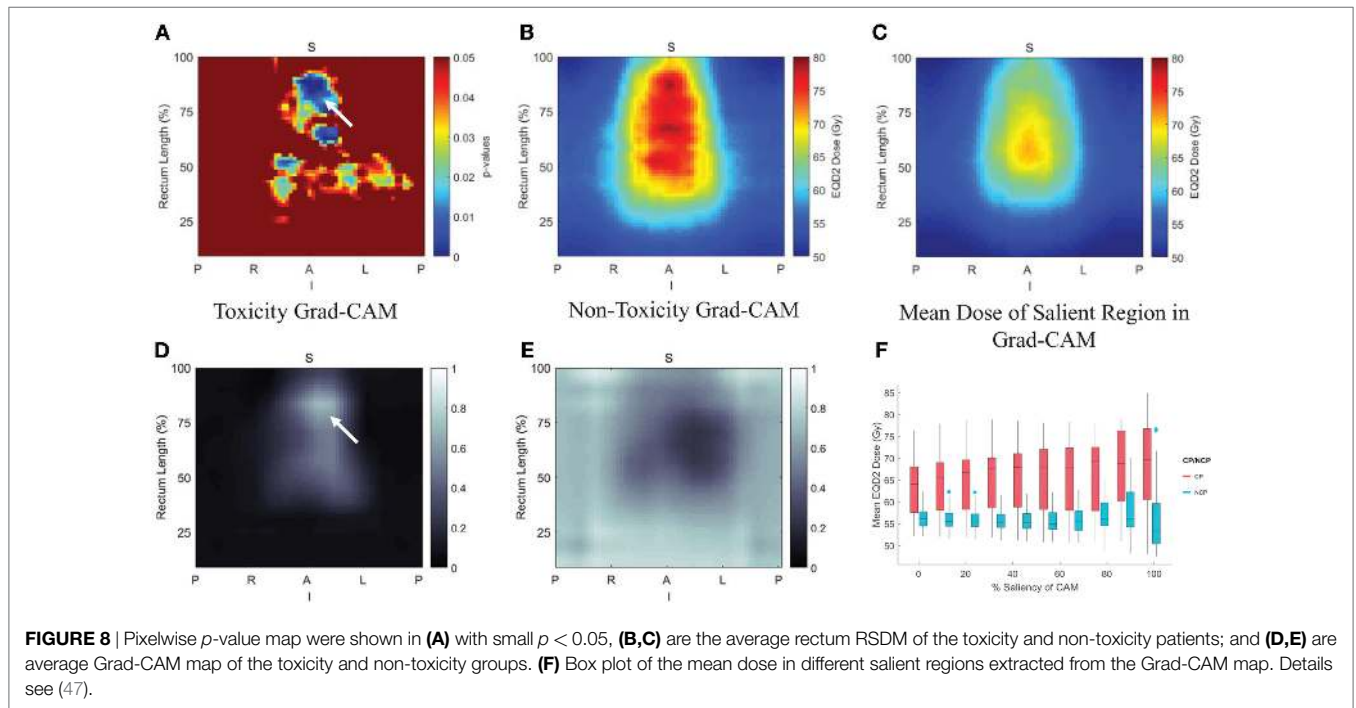


FIGURE 8 | Pixelwise p -value map were shown in (A) with small $p < 0.05$, (B,C) are the average rectum RSDM of the toxicity and non-toxicity patients; and (D,E) are average Grad-CAM map of the toxicity and non-toxicity groups. (F) Box plot of the mean dose in different salient regions extracted from the Grad-CAM map. Details see (47).

plans and improve patients' therapeutic satisfaction, a systematic machine learning approach needs to be developed to find the most important features from the high dimensional dataset and to discover the relationships between them and RP2 for clinical decision-making. Thus, a BN approach was developed to explore interpretable biophysical signaling pathways influencing RP2 from a heterogeneous dataset including single nucleotide polymorphisms (SNPs), micro RNAs (miRNAs), cytokines, clinical data, and radiation treatment plans before and during the course of radiotherapy of NSCLC patients.

In this BN implementation, the dataset described in Section 2.5.1 with 79 patients (21 cases of RP2) was used for model building and 46 additional patients were reserved for independent model testing. The BN approach mainly included a large-scale Markov blanket (MB) method to select relevant predictors, and a structure learning algorithm to find the optimal BN structure based on Tabu search and the performance evaluation of outcome prediction (24). *K*-fold cross-validation was used to guard against over-fitting, and the area under the receiver-operating characteristics (AUC) curve was utilized as a prediction metric.

The large-scale MB method intends to identify the most relevant variables of RP2 before or during the course of radiotherapy. **Figure 9A** shows the extended MB neighborhoods of RP2 before

radiation treatment, where the MB of RP2 based on pretreatment training data is formed from "Mean_Lung_Dose," "pre_MCP_1," "pre_TGF_alpha," and "pre_eotaxin." In the meantime, each of these variables has its own MB neighborhood as shown in **Figure 9A**. For example, "V20," "nos3_Rs1799983," "stage," and "RP2" form the MB of "Mean_Lung_Dose." In this study, potential variables of the BN were identified from the extended MB neighborhoods within two layers of RP2. **Figure 9B** indicates the updated extended MB neighborhoods in an extended model after incorporating the slopes of cytokine levels before and during-treatment (SLP) as the patients' responses during the radiation treatment. Although the MB of RP2 during the radiation treatment based on the whole training dataset keeps the same as that in **Figure 9A**, the MB of "Mean_Lung_Dose" has been updated, and it includes patients' cytokine responses such as "SLP_IL_17," "SLP_GM-CSF." **Figures 9C,D** illustrate biophysical signaling pathways from the patients' relevant variables to RP2 risk based on pretreatment and during BN model building, respectively. The results of internal cross-validation show that the performance of the BN yielded an AUC = 0.82, and it was improved by incorporating during treatment cytokine changes to AUC = 0.87. In the testing dataset, the pre- and during AUCs were 0.78 and 0.82, respectively. It turns out that the BN approach allows for

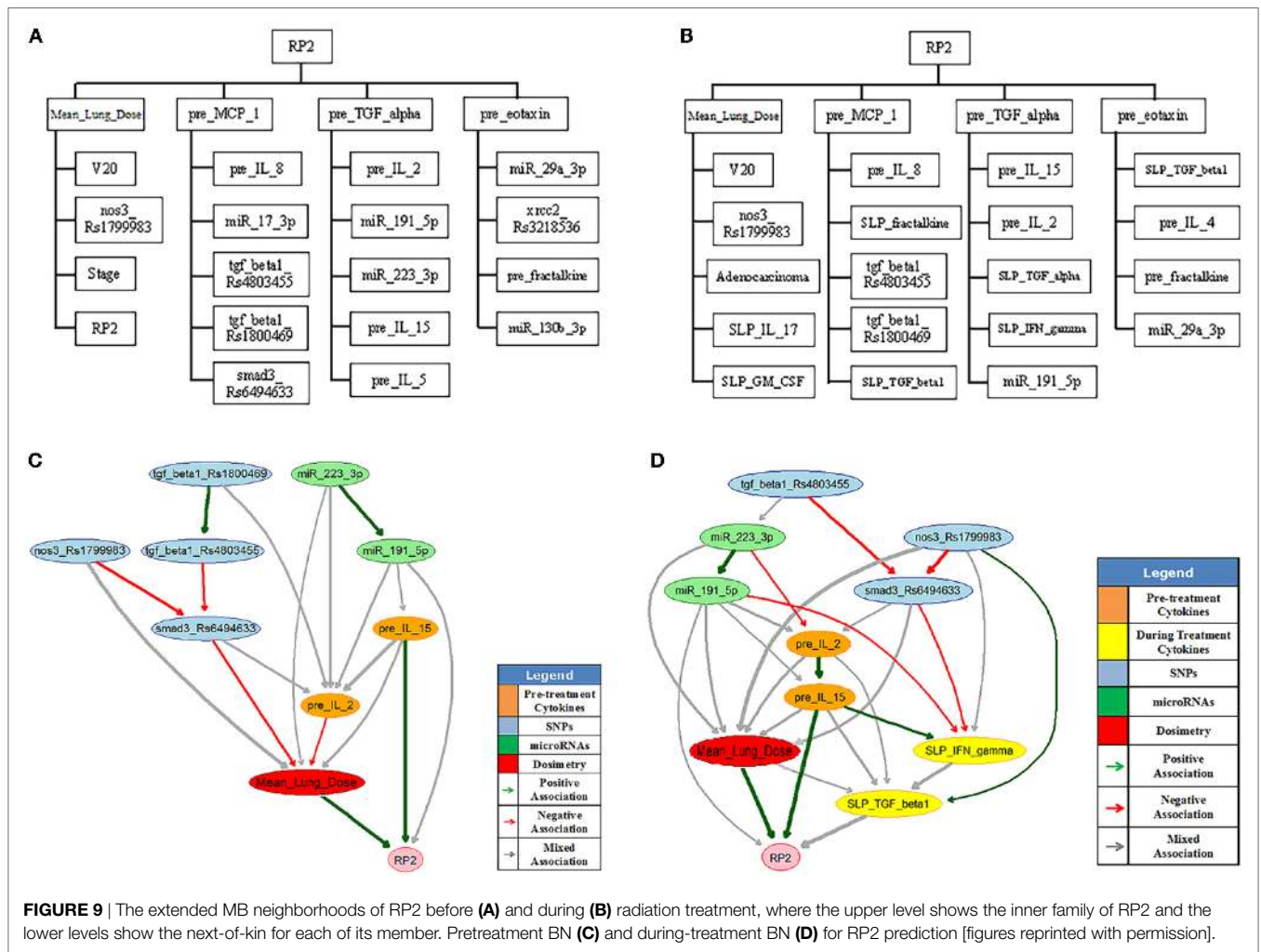


FIGURE 9 | The extended MB neighborhoods of RP2 before (A) and during (B) radiation treatment, where the upper level shows the inner family of RP2 and the lower levels show the next-of-kin for each of its member. Pretreatment BN (C) and during-treatment BN (D) for RP2 prediction [figures reprinted with permission].

unraveling of relevant biophysical features that lead to RP2 risk and prediction of RP2, and this prediction improved by incorporating during treatment information (24).

4. Q3: HOW TO ADAPT PLANS IN KBR-ART?

The precise estimation of treatment outcome is necessary step before deciding on the right course of action, since we desire to evaluate potential outcomes effects beforehand as we weigh the different alternatives for the best possible strategy (i.e., set of actions) to optimize the individual's treatment response. This is in a simplistic sense no different than playing board games or chess when a player may evaluate a dozen of options before carrying out a move. Therefore, by assuming one can attain accurate prediction estimates of TCP and NTCP, as discussed in the previous section, then, the final question to address in the context of KBR-ART is how to optimally adapt the plan (e.g., increase the tumor fraction dose) to achieve improved outcomes.

A utility function is usually required to estimate the total effect of a treatment plan weighting on both positive outcomes and the possible side effects caused. In RT, an example utility function called *complication-free tumor control* (P^+) can be used. The P^+ measures the performance of a treatment at each stage based on combined TCP and NTCP under the form $P^+ = U(\text{TCP}, \text{NTCP}; \theta)$ where P^+ indicates *probability of a positive treatment outcome*. One linear form is particularly simple and effective (66) where:

$$P^+ = \text{TCP} \times (1 - \text{NTCP}) \quad (16)$$

Notably, some other functional forms may be used as well, such as Equation (42).

In the practice of KBR-ART, if one has already synthesized relevant knowledge (clinical, dosimetric data, . . . etc.) from Section 2 with variables x_1, \dots, x_n as predictors and applied analytical/data-driven models in Section 3, then we can derive models of TCP and NTCP in the form $\text{TCP} = f_{\text{TCP}}(x_1, \dots, x_n)$ and $\text{NTCP} = f_{\text{NTCP}}(x_1, \dots, x_n)$ based on retrospective data such that the P^+ response estimation function reads:

$$P^+ = U(f_{\text{TCP}}(x_1, \dots, x_n), f_{\text{NTCP}}(x_1, \dots, x_n); \theta) \quad (17)$$

With the response estimation defined by the P^+ utility functions, next, we design a scheme for treatment adaptation. Machine learning based on reinforcement learning (RL) is a suitable approach for realizing plan adaptation as it can search over all possible decisions to maximize the P^+ function as *rewards* and identify the best *policy* (e.g., dose per fraction) for the treatment planning.

4.1. Generalized KBR-ART Framework

The KBR-ART can be described by the following general formulation:

$$\begin{aligned} \{\mathbf{x}^{(t)} \in \mathbb{R}^n | t \in T\}, \quad \{\mathbf{y}^{(t)} \in \mathbb{R}^m | t \in T\}, \quad \{\mathbf{u}^{(t)} \in \mathbb{R}^p | t \in T\} \\ \mathcal{L}(\{\mathbf{x}^{(t)}\}, \{\mathbf{y}^{(t)}\}, \{\mathbf{u}^{(t)}\}; \theta), \quad \mathcal{C}(\{\mathbf{x}^{(t)}\}, \{\mathbf{y}^{(t)}\}, \{\mathbf{u}^{(t)}\}; \phi), \end{aligned} \quad (18)$$

where $\mathbf{x}^{(t)}$ is the state of a system at time $t \in T$, $\mathbf{y}^{(t)}$ is the observation of state $\mathbf{x}^{(t)}$, $\mathbf{u}^{(t)}$ is the controls for the system to influence next states $\mathbf{x}^{(t+1)}$, and $\mathcal{L}(\{\mathbf{x}^{(t)}\}, \{\mathbf{y}^{(t)}\}, \{\mathbf{u}^{(t)}\}; \theta)$ is a loss function serving a specific purpose for the system to be minimized over temporal information $\mathbf{x}^{(t)}$, $\mathbf{y}^{(t)}$, and $\mathbf{u}^{(t)}$ along with some constraints $\mathcal{C}(\{\mathbf{x}^{(t)}\}, \{\mathbf{y}^{(t)}\}, \{\mathbf{u}^{(t)}\}; \phi)$. Any of the vectors $\mathbf{x}^{(t)}$, $\mathbf{y}^{(t)}$, and $\mathbf{u}^{(t)}$ can be real-valued vectors or vectors of random variables such that the temporal sequences can be deterministic or a random processes adaptation. Although dimensions of n, m, p may be infinite in Equation (18), almost all real-life implementations are finite dimensions. Equation (18) may apply to many legacy ART approaches in different manners. In the following, we provide a brief overview for alternative ART approaches.

4.1.1. Linear Feedback ARTs

Traditionally, linear feedback (loop) control systems are considered as viable implementations of ART, where most of the adaptive feedback is based on imaging information such as CT and/or MRI. Generally, there are two types of control systems: *open-loop* and *closed-loop*.

With notations in Equation (18), a *linear loop control* is generally described by two sets of linear equations:

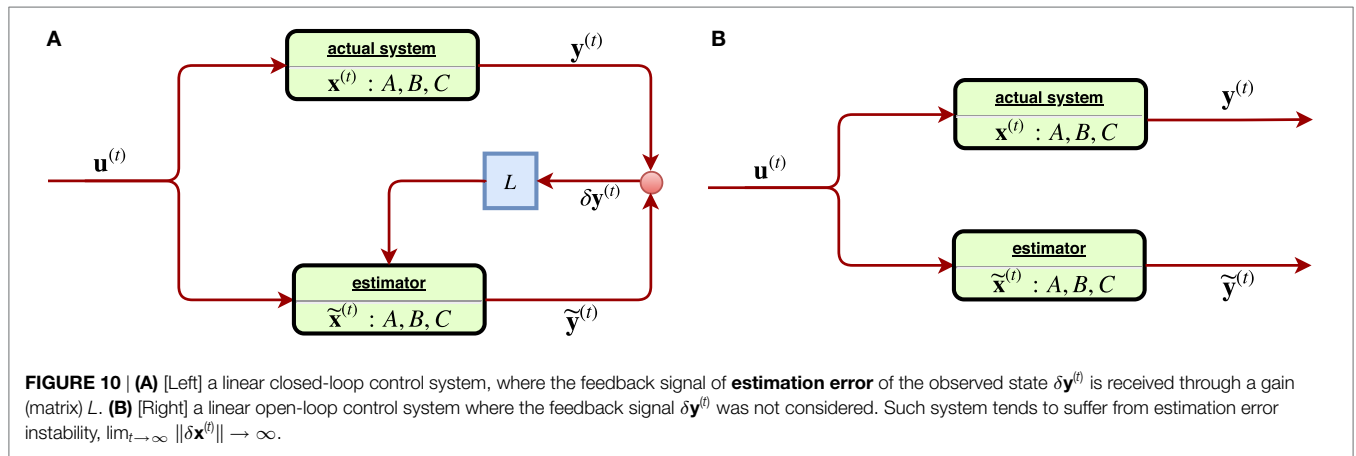
$$\dot{\mathbf{x}}^{(t)} = A\mathbf{x}^{(t)} + B\mathbf{u}^{(t)}, \quad \mathbf{y}^{(t)} = C\mathbf{x}^{(t)} \quad (19)$$

$$\dot{\tilde{\mathbf{x}}}^{(t)} = A\tilde{\mathbf{x}}^{(t)} + B\mathbf{u}^{(t)} + L\delta\mathbf{y}^{(t)}, \quad \tilde{\mathbf{y}}^{(t)} = C\tilde{\mathbf{x}}^{(t)}, \quad (20)$$

where in Equation (19), A, B, L, C are linear operators, $\mathbf{y}^{(t)}$ is the *observation* of the actual state $\mathbf{x}^{(t)}$, and $\{\mathbf{u}^{(t)} | t \in T\}$ represents controls of the system as adaptations for the treatment of a radiotherapy. Equation (20) as a similar copy of Equation (19) describes the **estimation** $\tilde{\mathbf{x}}^{(t)}, \tilde{\mathbf{y}}^{(t)}$ of the corresponding variables $\mathbf{x}^{(t)}, \mathbf{y}^{(t)}$ of the system, with $\delta\mathbf{y}^{(t)} \stackrel{\text{def}}{=} \mathbf{y}^{(t)} - \tilde{\mathbf{y}}^{(t)}$ as the **estimation error** of the observed state and in turn shall be used as the **feedback** in the subsequent iterations. With $L \neq 0$, the system constantly receiving the estimation error shall adjust itself accordingly, and thus such is called a **closed-loop** control system, **Figure 10A**.

Incidentally, in the perfect case, the three characters $\mathbf{x}^{(t)}, \tilde{\mathbf{x}}^{(t)}, \tilde{\mathbf{y}}^{(t)}$ shall coincide into one with $C = I, \delta\mathbf{y}^{(t)} = 0$ and thus the Equations (19) and (20) reduce to one. However, in most of cases, they tend to split. In a system, where the matrix L vanishes, it becomes an **open-loop** control system since any feedback signal $\delta\mathbf{y}^{(t)}$ from the system is not considered, **Figure 10B**. An obvious drawback of the open-loop system is the estimation instability, which can be easily seen from Equations (19) and (20) as the quantity $\delta\mathbf{x}^{(t)} \stackrel{\text{def}}{=} \mathbf{x}^{(t)} - \tilde{\mathbf{x}}^{(t)}$ describing the **estimation error** is subject to the state equation $d/dt(\delta\mathbf{x}^{(t)}) = A \cdot \delta\mathbf{x}^{(t)}$ with $L \equiv 0$. The solution $\delta\mathbf{x}^{(t)} = e^{AT} \cdot \delta\mathbf{x}^{(0)}$ indicates that the error has exponential growth as time elapses such that soon an open-loop system easily becomes unreliable. On the other hand, by receiving a feedback signal due to a close-loop system ($L \neq 0$) can improve reliability, as the evolution $\delta\mathbf{x}^{(t)} = e^{(A-LC)t} \cdot \delta\mathbf{x}^{(0)}$ will converge by suitable choice of a **gain** L such that the eigenvalues $|\lambda_i(A-LC)| < 1$. In a linear control problem, the control is modeled by $\mathbf{u}^{(t)} = -K\mathbf{x}^{(t)}$ with a constant matrix K such that Equation (20) reads:

$$\dot{\tilde{\mathbf{x}}}^{(t)} = (A-LC)\tilde{\mathbf{x}}^{(t)} - BK\mathbf{u}^{(t)} \quad (21)$$



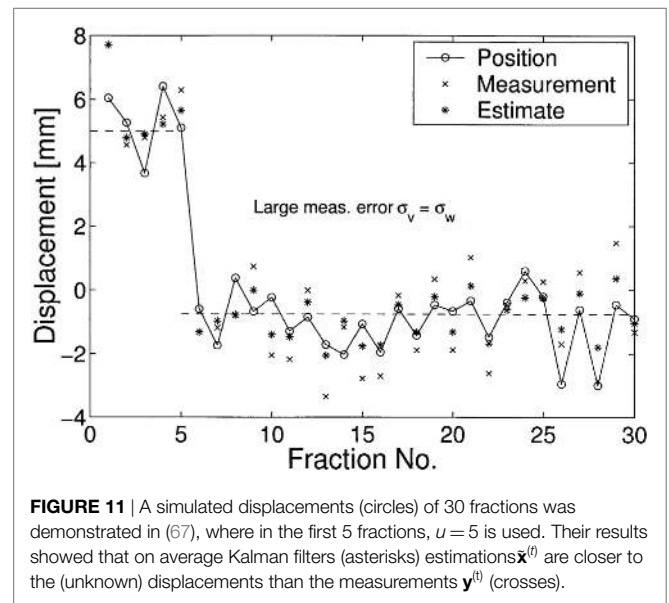
In control theory, one may also consider a loop-control system with small uncertainty. Typically, by considering stochasticity, the system can become more stable and robust. The (time) discretized linear control with a random process starts with extension of Equation (19) as:

$$\begin{aligned} \mathbf{x}^{(t)} &= A \cdot \mathbf{x}^{(t-1)} + B \cdot \mathbf{u}^{(t)} + \mathbf{w}^{(t)} \\ \mathbf{y}^{(t)} &= C \cdot \mathbf{x}^{(t-1)} + \mathbf{v}^{(t)}, \end{aligned} \tag{22}$$

where the two random processes $\{\mathbf{w}^{(t)}\}$ and $\{\mathbf{v}^{(t)}\}$ denote the noise of state $\mathbf{x}^{(t)}$ and observation $\mathbf{y}^{(t)}$ assumed multivariate Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{Q}^{(t)})$ and $\mathcal{N}(\mathbf{0}, \mathbf{R}^{(t)})$, respectively. Kalman filters are then a common analysis for deriving optimal estimation of $\delta \mathbf{x}^{(t)}$. In (67), Keller et al. established a linear stochastic closed-loop system that utilized Kalman filters (68) to derive optimal control law. They assumed an image-guided radiotherapy, which attempts to provide optimal correction strategies for setup errors, which can also take the measurement uncertainties into account. Let $\mathbf{x}^{(t)} = \mathbf{x}_1^{(t)} + \mathbf{x}_2^{(t)} \in \mathbb{R}^3$ denote the difference between the actual and planned positions of the center-of-mass of the clinical tumor volume (CTV), i.e., the daily displacement $\mathbf{x}^{(t)}$ containing (1) the setup error $\mathbf{x}_1^{(t)}$ (displacement of bony structures) and (2) the organ motion (displacement $\mathbf{x}_2^{(t)}$ with respect to the bony structures). Decompose $\mathbf{x}^{(t)}$ into two parts $\mathbf{x}^{(t+1)} = \mathbf{u}^{(t)} + \mathbf{w}^{(t)}$ with $\mathbf{u}^{(t)} = \mathbf{u}_1^{(t)} + \mathbf{u}_2^{(t)}$ called the **systematic component** and $\mathbf{w}^{(t)} = \mathbf{w}_1^{(t)} + \mathbf{w}_2^{(t)}$ called the **random component**, where the subindex “1” and “2” refer to setup errors and organ motion, respectively. Together, they modeled the ART displacement with a stochastic linear system:

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \mathbf{u}^{(t)} + \mathbf{w}^{(t)} \\ \mathbf{y}^{(t)} &= \mathbf{x}^{(t)} + \mathbf{v}^{(t)} \end{aligned} \tag{23}$$

where $\mathbf{y}^{(t)}$ is the observation of $\mathbf{x}^{(t)}$. By defining the estimation of state $\mathbf{x}^{(t)}$ as $\tilde{\mathbf{x}}^{(t)} \stackrel{\text{def}}{=} P(\mathbf{x} | \mathbf{y}^0, \dots, \mathbf{y}^{(t-1)})$ based on previous observations $\mathbf{y}^0, \dots, \mathbf{y}^{(t-1)}$ as in Equation (20), Kalman filters are able to provide an optimal estimation of $\tilde{\mathbf{x}}^{(t)}$ such that the estimation error $\tilde{\mathbf{x}}^{(t+1)} \stackrel{\text{def}}{=} \mathbf{x}^{(t+1)} - \tilde{\mathbf{x}}^{(t)}$ is minimal. Immediately, they derived



the *optimal control law* $\mathbf{u}_{c*}^{(t)} = -\tilde{\mathbf{x}}^{(t)}$, which seems to be an intuitive result. A comparison was made with respect to an obvious control law that is “suboptimal” $\mathbf{u}_c^{(t)} = -\tilde{\mathbf{y}}^{(t)}$, which is merely the correction of observation itself. Subsequently, they attempted to measure the effectiveness of decisions given by Kalman filters \mathbf{u}_{c*} and the observation \mathbf{u}_c by computing

$$e \stackrel{\text{def}}{=} \frac{\sigma_{\mathbf{x}-\tilde{\mathbf{x}}}^2}{\sigma_{\mathbf{x}-\mathbf{y}}^2} \tag{24}$$

where $\sigma_{\mathbf{x}-\tilde{\mathbf{x}}}^2$ and $\sigma_{\mathbf{x}-\mathbf{y}}^2$ are two residue variances of different estimation toward the state $\mathbf{x}^{(t)}$. One simulated result was made to demonstrate the performance of Kalman filters in predictions of stochastic linear control system, as shown in **Figure 11** where a treatment of 30 fractions were simulated with the first 5 fractions, a random systematic error $u = +5$ mm and measurement noise $\sigma_v = \sigma_w = 1$ mm were imposed, which means the correction started only at the sixth fraction. Their results showed that on average Kalman filter estimations $\tilde{\mathbf{x}}^{(t)}$ are closer to the (unknown) displacements than the measurements $\mathbf{y}^{(t)}$, where in the first fraction the estimate equals the value of the measurement.

4.2. Nonlinear Feedback ARTs

It is natural to consider nonlinear feedback control for ARTs due to inherent complexity. In (69), Zerda et al. developed a nonlinear closed-loop ART for treatment planning. In particular, they proposed two algorithms: *Immediately Correcting Algorithm* and *Prudent Correcting Algorithm*. With the following notation corresponding to Equation (18),

$$\begin{aligned} \mathbf{x}^{(t)} = \mathbf{y}^{(t)} &\rightarrow \psi^{(t)} = (\psi_{\text{geometry}}^{(t)}, \psi_{\text{cumdose}}^{(t)}), \\ u^{(t)} = \xi^{(t)}(\{\mathbf{x}^{(t)}\}) &\rightarrow \beta^{(t)} = \xi^{(t)}(\psi^{(t)}), \\ \mathcal{L}(\{\mathbf{x}^{(t)}\}, \{\mathbf{y}^{(t)}\}, \{\mathbf{u}^{(t)}\}; \theta) &\rightarrow \sum_{v \in V} \alpha(v) (D_{\text{prescribed}}(v) - \psi_{\text{cumdose}}^{(T)}(v))^2, \end{aligned} \tag{25}$$

where $v \in V$ is a voxel under consideration, $v \mapsto \alpha(v)$ is the importance factor, and the control is promoted as a nonlinear function of states, $\mathbf{u}^{(t)} = \xi^{(t)}(\{\mathbf{x}^{(t)}\})$ rather than the linear form $\mathbf{u}^{(t)} = -K \cdot \mathbf{x}^{(t-1)}$. With $\psi^{(t)}$ denoting the state of the ART system, it was assumed to consists of two parts: (1) **cumulative dose** $\psi_{\text{cumdose}}^{(t)}$ after $t \in T$ and (2) **patient's geometric model** obtained from conebeam CT (CBCT images) $\{\psi_{\text{geometry}}^{(t)} \mid t \in T\}$, where it was further assumed the geometry information interacts with the cumulative dose by the relation

$$\begin{aligned} \psi_{\text{cumdose}}^{(t)} &= \psi_{\text{cumdose}}^{(t-1)} \\ &+ D(v; \{\beta^{(t)} \mid t \in T\}, \{\epsilon^{(t)} \mid t \in T\}, \{\psi_{\text{geometry}}^{(t)} \mid t \in T\}) \end{aligned} \tag{26}$$

with the *dose delivery* function $D(v; \{\beta^{(t)}\}, \{\epsilon^{(t)}\}, \{\psi_{\text{geometry}}^{(t)}\})$ related to delivery errors $\{\epsilon^{(t)}\}$, where it is always assumed vanishing throughout the paper (69). In other words, from Equations (25) and (26), the objective of the Immediately Correcting

Algorithm is to minimize the following loss:

$$\begin{aligned} \mathcal{L}(\beta^1, \dots, \beta^{|T|}) &= \sum_{v \in V} \alpha(v) \left(D_{\text{prescribed}}(v) \right. \\ &\left. - \sum_{t \in T} D(v; \{\beta^{(t)}\}, \{\epsilon^{(t)}\}, \{\psi_{\text{geometry}}^{(t)}\}) \right)^2 \end{aligned} \tag{27}$$

via an optimal sequence of dose fractionation (controls) $(\beta_1, \dots, \beta_{|T|})$ to be found, and thus it is regarded as a special realization of the general scheme **Figures 12A,B**.

4.3. Stochastic ARTs

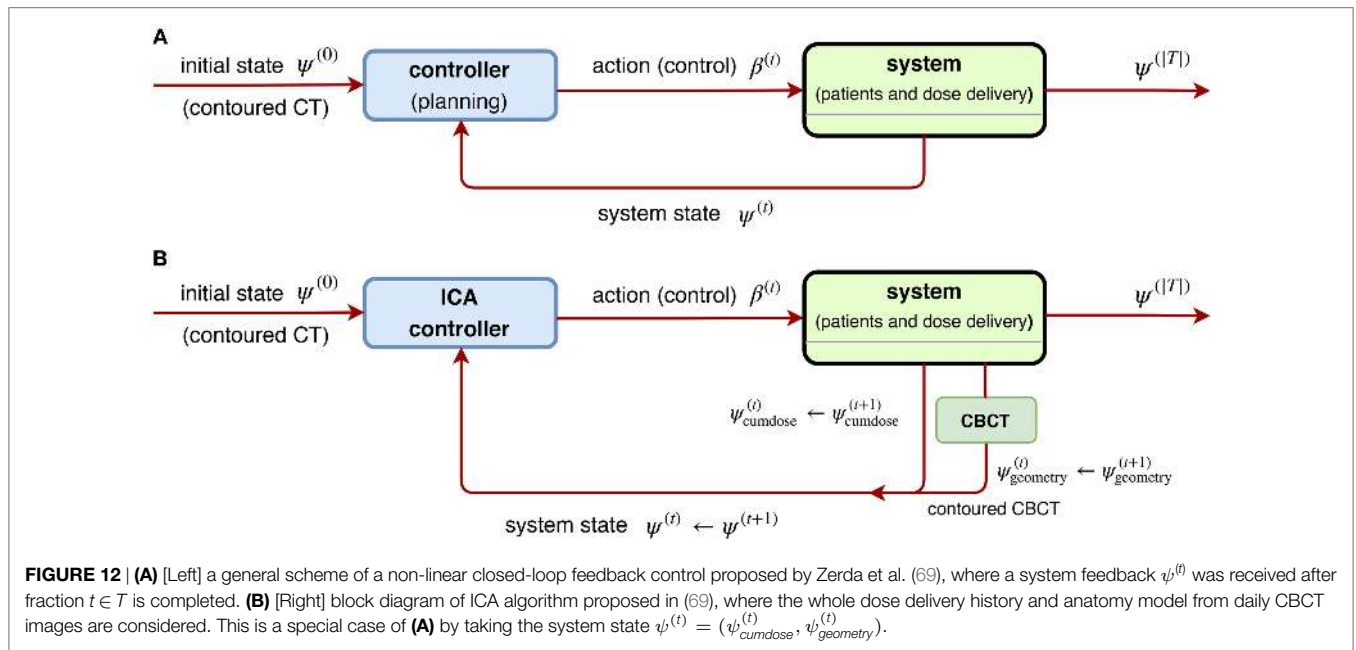
In (70), Bortfeld et al. developed a static *robust* optimization by treating the dose delivery problem of intensity modulated RT (IMRT) as a probabilistic problem with uncertainties. Using the notations in Equation (18) and letting $\mathbf{x}^{(t)}$ as a **breathing phase** (state) at time t , $\mathbf{u}^{(t)}$ as a **control probability function** over all breathing states, the **observed state** $\mathbf{y}^{(t)} = \mathbf{x}^{(t+1)}$:

$$\mathbf{x}^{(t)} \rightarrow x, \quad \mathbf{u}^{(t)} \rightarrow p(x), \quad \theta \rightarrow \{\Delta_{v,b,x}, w_b, \gamma, \theta_v\}, \tag{28}$$

we arrive at the loss function and constraints proposed by Bortfeld et al.

$$\begin{aligned} \text{minimize } \mathcal{L} &= \sum_{v \in V} \sum_{x \in X} \sum_{b \in B} \Delta_{v,b,x} p(x) w_b \\ \text{subject to } C_1 &= \sum_{v \in V} \sum_{x \in X} \sum_{b \in B} \Delta_{v,b,x} \tilde{p}(x) w_b \geq \theta_v, \quad \forall v \in T, \tilde{p} \in P_U \\ C_2 &= \sum_{v \in V} \sum_{x \in X} \sum_{b \in B} \Delta_{v,b,x} \tilde{p}(x) w_b \leq \gamma \theta_v, \quad \forall v \in T, \tilde{p} \in P_U. \end{aligned} \tag{29}$$

Essentially, they considered the dose (to be delivered) as an expectation value following a predefined probability distribution (PDF) over all breathing phases, $D_{v,b} = \mathbb{E}_x[\Delta_{v,b,x}] = \sum_{x \in X} \Delta_{v,b,x} p(x)$, where $v \in V$ denotes a voxel, $b \in B$ denotes



a beamlet, $\Delta_{v,b,x}$ is a matrix computed for the snapshots of the anatomy in each phase, and θ_v, γ are some constants specific to the problem in question. The main purpose is to learn an optimal probability $p(x)$ as a stochastic control overall breathing phases $x \in X$ via Equation (29). The motion p.d.f. searched in the infinite-dimensional controls was actually approximated by the discretized set,

$$P = \{p \in \mathcal{F}(X; \mathbb{R}) \cong \mathbb{R}^{|X|} \mid p(x) \geq 0, \sum_{x \in X} p(x) = 1\} \quad (30)$$

such that this problem is tractable. They further required the **realization** \tilde{p} of p in Equation (29) during a treatment to be constrained within certain error bounds ℓ and u :

$$P_U := \{\tilde{p}(x) \in P \mid \underbrace{p(x) - \underline{p}(x)}_{\ell(x)} \leq \tilde{p}(x) \leq \underbrace{p(x) + \tilde{p}(x)}_{u(x)}, \forall x \in U \subseteq X\} \quad (31)$$

As a result, the experiments by Bortfeld et al. showed that even when they allowed an unaccepted underdosage in the tumor anywhere between 6 and 11%, their proposal Equation (29) still offered same level of protection as the margin solution within 1% under dosage on average. Their approach proves that using stochastic controls helps stabilize the system with uncertainty over time. Later in (71), Chan and Mišić further improved the previous adaptive approach by extending the static probability distribution $\{p\}$ into a temporal sequence of PDF ($p^{(1)}, p^{(2)}, \dots, p^{(k)}$) by incorporating uncertainty set updated each time for ART, which corresponds to the sequential control $\{u^{(t)} \mid t \in T\}$ in Equation (18). The proposal in (71) essentially replaces the uncertainty p.d.f. $p \in P_U$ of Equation (29) by $p^{(k)} \in P_U^{(k)}$ iteratively to take care of patient's breathing motions.

$$p^{(k+1)} \leftarrow p^{(k)}, \quad \text{with} \\ p^{(k)} \in P_U^{(k)} := \{\tilde{p}(x) \in P \mid \ell^{(k)}(x) \leq \tilde{p}(x) \leq u^{(k)}(x), \forall x \in U \subseteq X\} \quad (32)$$

Two versions of uncertainty updates are proposed,

$$\ell^{(k+1)} = (1 - \alpha) \ell^{(k)} + \alpha p^{(k)}, \quad u^{(k+1)} = (1 - \alpha) u^{(k)} + \alpha p^{(k)} \quad (33)$$

$$\ell^{(k+1)} = \frac{1}{k+1} \left(\ell^{(k)} + \sum_{i=1}^k p^{(i)} \right), \\ u^{(k+1)} = \frac{1}{k+1} \left(u^{(k)} + \sum_{i=1}^k p^{(i)} \right) \quad (34)$$

where the first version is called the *exponential smoothing* update and the second is called the *running average* update. Together, Equations (29) (32), (33), or (34) constituted their proposal in (71) and suggested that their method does not require accurate information to exist before a treatment commences. Their evaluation further stressed its clinical value as it allows for the tumor dose

to be safely escalated without leading to additional healthy tissue toxicity, which may ultimately improve the rate of patient survival. Subsequently, Mar and Chan (72) further proposed an extension to the adaptive robust ART mentioned above (70, 71) by adding drift component using the Lujan model (73) of patients' breathing patterns.

Another related approach utilizing the formulation Equation (18) is found in (74), where Löf et al. developed statistical models for ART. Their design used *stochastic optimization* to handle two kinds of errors: (1) errors due to internal motion and change of organs (or tissues) and (2) errors due to the uncertainty in the geometrical setup of a patient. They attempted to compensate for the systematic errors by couch corrections and for the random error by modulation of the fluence profiles. This system was further modified by Rehbindler et al. using a linear-quadratic regulator (LQR) (75).

4.4. Reinforcement Learning (RL) for ART

RL is a set of machine learning algorithms that can interact with an "environment" (e.g., radiotherapy). Usually, there is a goal set for the RL, acting as an agent, to reach. Examples could be, winning a chess/board game or driving safely through a trip in an autonomous driving vehicle. Such a procedure is usually done by collecting the so-called *reward* designed by humans. RL serves as an independent machine learning area besides the common supervised or unsupervised learning mentioned earlier. RL is based on the environment defined by a Markov decision process (MDP).

An MPD is a 5-tuple $(S, \mathcal{A}, P, \gamma, R)$, where

- $S = \{(x_1, \dots, x_n) \in \mathbb{R}^n\}$ is the space of all possible states,
- \mathcal{A} is a finite collection of all (discrete) actions,
- $R : \Omega \rightarrow \mathbb{R}$ is the reward function given on the product space $\Omega = S \times \mathcal{A} \times S$,
- $\gamma \in [0, 1]$ is the discount factor, representing the importance (rewards) that propagates from the future back to the present,
- $P : \mathcal{F} \rightarrow [0, 1]$ is a probability measure on Ω with $\mathcal{F} = 2^\Omega$ the power set (σ -algebra) of Ω , whose probability mass function (pmf) $(s, a, t) \mapsto P(s, a, t)$ denotes the transition probability from state $s \in S$ to another $t \in S$ under an action $a \in \mathcal{A}$. Consequently, this induces the condition probability

$$P_{sa}(t) \equiv \text{Prob}(t \mid s, a) \equiv P(s, a, t) / P(s, a), \quad (35)$$

on space of next states t conditioned on previous state s and current action a .

As an example, in chess, each $s_i \in S$ will stand for a configuration of the chess board and action $a_i \in \mathcal{A}$ corresponds to a move given by a player. The purpose of an agent in the RL is to find a sequence of actions $\{a_0, a_1, \dots\}$ (acting on an initial state $s_0 \in S$) such that a path in S collects maximum rewards (and hence winning the goal/game):

$$s_0 \xrightarrow{\pi} s_1 \xrightarrow{\pi} s_2 \xrightarrow{\pi} s_3 \dots \quad (36)$$

An agent is, by itself, a policy function $\pi : S \rightarrow \mathcal{A}$ who determines an action $a = \pi(s)$ under a state s , as described in Equation

(36). There are mainly two ways to construct a policy function by *policy-based* and *value-based* methods in RL: the former parametrizes a policy function directly (76) via $\pi^{(\theta)}$ while the latter builds one implicitly via Q-functions, and hence is also called Q-learning. The policy-based method is usually applied in continuous controls where $\mathcal{A} \cong \mathbb{R}$ or large cardinality $|\mathcal{A}| \rightarrow \infty$. In this study, we shall focus more on the Q-learning and its application in radiotherapy.

An optimal policy $\pi^*: S \rightarrow \mathcal{A}$ is derived from maximizing the Q-function in the Q-learning, such that $Q^{\pi^*} = \max_{\pi} Q^{\pi}$, where the Q-function is defined by evaluating the value at $(s, a) \in S \times \mathcal{A}$ via rewards collected in all possible paths:

$$Q^{\pi}(s, a) = \mathbb{E} \left[\sum_{k=0}^{\infty} \gamma^k R(s_k, \pi(s_k)) \mid \pi, s_0 = s, a_0 = a \right] \quad (37)$$

However, this definition Equation (37) is ideal for comprehension, yet, difficult for actual computation. A practical realization of computing the Q-function is via the following Bellman's iteration, whose optimal value Q^{π^*} is computed by an iterative (functional) sequence $\{\tilde{Q}_i\}_{i=1}^{\infty}$ instead,

$$\tilde{Q}_{i+1}(s, a) = \mathbb{E}_{t \sim P_{sa}} \left[R(s, a) + \gamma \max_{b \in \mathcal{A}} \tilde{Q}_i(t, b) \right]. \quad (38)$$

Such an iteration Equation (38) is guaranteed to converge by the contraction mapping theorem (77) of the uniquely fixed point as $\{\tilde{Q}_i\}_{i=1}^{\infty} \rightarrow Q^{\pi^*}$ if $i \rightarrow \infty$ (78) such that

$$\tilde{Q}^*(s, a) = \mathbb{E}_{t \sim P_{sa}} \left[R(s, a) + \gamma \max_{b \in \mathcal{A}} \tilde{Q}^*(t, b) \right]. \quad (39)$$

The calculation soon becomes intractable when either the cardinality $|S|$ or $|\mathcal{A}|$ is large. A possible solution to this is utilizing deep learning methods for evaluating the Q-function proposed by Google DeepMind (79, 80), hence the name Deep Q-network (DQN). By taking advantages of neural networks, the convergence of the Q-function with Equation (38) becomes more efficient and accurate. DQN proposes $\tilde{Q}_i = Q_{\text{DNN}}^{\Theta_i}$, where Θ_i denotes the parametrization (weights) of the DNN at i th iteration and requires

the following loss function being optimized:

$$\mathcal{L}_i(\Theta_i) = \mathbb{E}_{(s,a) \sim \rho} \times \left[\left(\mathbb{E}_{t \sim P_{sa}} \left[R(t, a) + \gamma \max_{b \in \mathcal{A}} Q_{\text{DNN}}^{\Theta_{i-1}}(t, b) \right] - Q_{\text{DNN}}^{\Theta_i}(s, a) \right)^2 \right]. \quad (40)$$

In short, Equations (38) and (40), and $\tilde{Q}_i = Q_{\text{DNN}}^{\Theta_i}$ together makes the DQN.

4.5. Example: Adapting RT Plans Using Deep Reinforcement Learning

Using the NSCLC dataset from Section 2.5.1, we attempt to apply a DQN to provide automatic dose escalation at the 2/3 period (about 4 weeks) into a treatment as illustrated in **Figure 13**, where the dose escalation is the action to be submitted by the DQN. The main goal of the study is to compare the automatic decision made by the DQN to that established by a clinical protocol (81). This will be described briefly in the following, details can be consulted in (23).

That work explicitly presented a suitable MDP. In particular, a state space chosen to be useful for prediction of local control (LC) and RP2 based on the BN formalism introduced in Section 3.3.2.

By defining the state space as $S = \{(x_1, \dots, x_n) \in \mathbb{R}^n\}$ with $n = 9$ and

$$\begin{aligned} x_1 = \text{IL4} \quad x_2 = \text{IL15}, \quad x_3 = \text{GLSZM.GLN}, \quad x_4 = \text{GLRLM.RLN}, \\ x_5 = \text{MCPI}, \quad x_6 = \text{TGF}\beta 1, \quad x_7 = \text{Lung gEUD}, \quad x_8 = \text{Tumor gEUD}, \quad x_9 = \text{MTV} \end{aligned} \quad (41)$$

where x_1, x_2, x_5, x_6, x_9 are cytokines, x_3, x_4 are of PET radiomics, and x_7, x_8 are doses, and here, the allowed action set will be $\mathcal{A} = \{a_1 = \text{dose/frac}\} \subseteq \mathbb{R}^+$. One notices that such a choice of a MDP for dose automation is not unique; there may exist other environments to attain the same or even better performance (82).

A tricky problem is that the transition probability in Equation (35) is intractable to the real world (radiotherapy environment); therefore, DNNs were utilized to model the radiotherapy environment. Thus, a DNN provided an approximate transition probability $\tilde{P}(s, a; t) := \tilde{P}_{sa}(t) := \widetilde{\text{Prob}}(t \mid s, a)$ modeled from the observed data, where the transition takes place $s \xrightarrow{a} t$ under

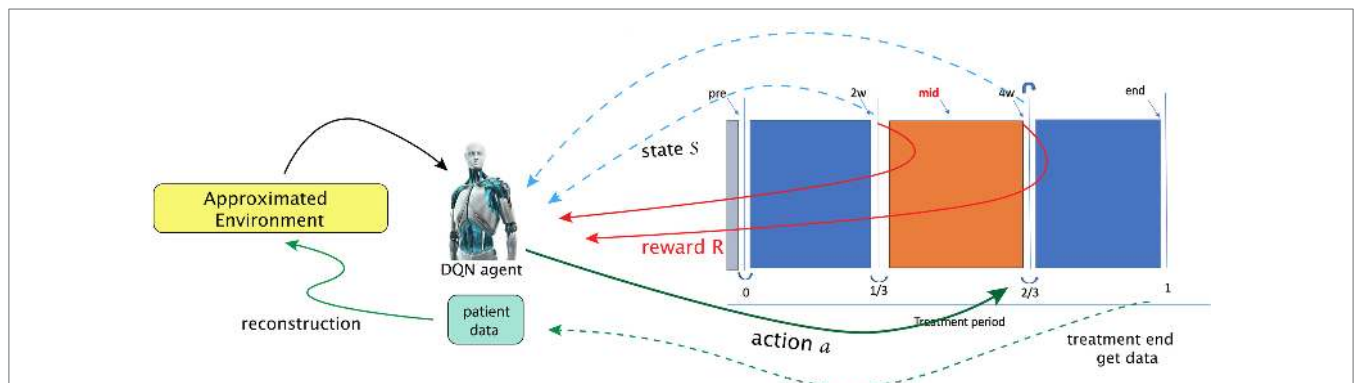


FIGURE 13 | In the paper (23), Tseng et al. proposed to utilize reinforcement learning for making decisions at 2/3 period of a treatment (right solid-green arrow). A first step in their framework is to learn transition functions from the historical data of two transitions recorded (RHS figure) so that the radiotherapy environment can be reconstructed (called *approximated environment*). With the transitions simulated, a DQN agent can then search for optimal dose at each stage [figures reprinted with permission].

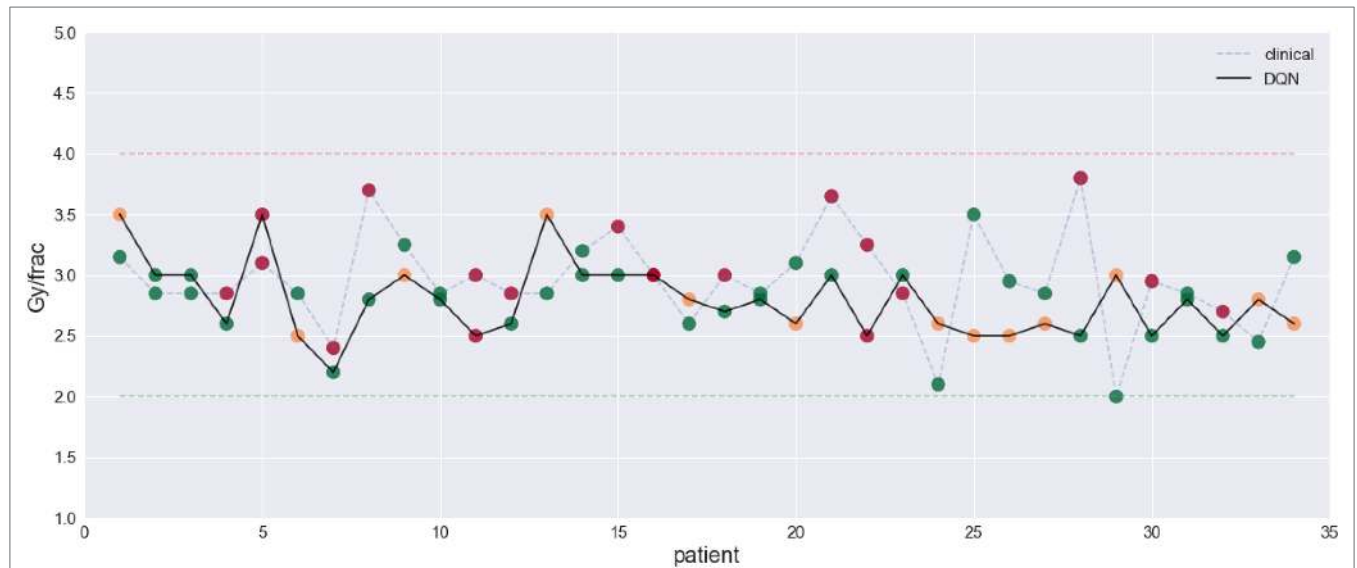


FIGURE 14 | This figure visualizes the dose fraction recommended by the clinicians (blue dashed line) and the autonomous DQN (black solid line). Differences and similarities can thus be compared, with RMSE = 0.5 Gy. An evaluation of good (green dots), bad (red dots), and *potentially good* decisions (orange dots) (23) [figures reprinted with permission].

action a . Another problem to solve in that the sample size was small relative to the DNNs. Hence, a Generated Adversarial Network (GAN) technique was used to alleviate this problem.

After proper choice of actions, $\mathcal{A} = \{1, 1.1, 1.2, \dots, 5\}$ Gy, and a reward function looking upon to higher LC than usual P^+ baseline function:

$$R(s) = \frac{1}{2} \sqrt{\text{Prob}(\text{LC} | s)} \cdot (1 - 0.8 \cdot \text{Prob}(\text{RP2} | s)) \cdot (1 + \text{sgn}(17.2\% - \text{Prob}(\text{RP2} | s))), \quad (42)$$

The results demonstrated the feasibility to derive automated dose levels (black solid line) that are similar to or compatible with the clinical protocol (blue dashed line) as shown in **Figure 14** with the corresponding statistics shown in **Table 1**.

5. DISCUSSION

5.1. Statistical and Probabilistic Aspects

Here, we attempt to provide a fundamental statistical and probabilistic interpretation for sequential machine learning algorithms to help understand their roles in KBR-ART. This will be done with the specific focus on how knowledge can accumulate in such a KBR-ART system when the known information in the system is growing with time. First, we characterize the probability space as: $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{F} is a σ -algebra¹ of a sample space Ω and $P : \mathcal{F} \rightarrow \mathbb{R}^+$ is the probability measure defined on Ω , see (83, 84). In this setting, Ω denotes the set of all possible outcomes and \mathcal{F} as the space of all events. A (multi-dimensional) random variable \mathbf{X} is a then \mathcal{F} -measurable function $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ on a probability space (Ω, \mathcal{F}, P) . Roughly speaking, the σ -algebra corresponds to

¹ \mathcal{F} as a collection of subsets of a set Ω is called a σ -algebra if the following three is satisfied: (1) $\Omega \in \mathcal{F}$, (2) if $A \in \mathcal{U}$ implies $(\Omega \setminus A) \in \mathcal{F}$, and (3) arbitrary union $A = \cup_{k=1}^{\infty} A_k \in \mathcal{F}$ if $A_k \in \mathcal{F}$.

TABLE 1 | Summary for the evaluation on clinicians' and the DQN decisions extracted from (23).

Summary	Good	Bad	Potentially good
Clinicians	19 (55.9%)	15 (44.1%)	0
DQN	17 (50%)	4 (11.8%)	13 (38.2%)

the “information” useful (and related) to the random variable \mathbf{X} . Furthermore, if $\{\mathbf{X}^{(t)} | t \in T\}$ is a sequence of random variables (or a process), a natural σ -algebra induced by the process is defined by:

$$\begin{aligned} \mathcal{U}(t) &:= \mathcal{U}(\mathbf{X}(s) | s \in [0, t]) \\ &:= \left\{ \left(\mathbf{X}^{(s)} \right)^{-1} (B) \subseteq \Omega | \forall \text{ Borel set } B \subseteq \mathbb{R}^n, \forall s \in [0, t] \right\}, \end{aligned} \quad (43)$$

which is interpreted as the *history* of the process up to time t . Therefore, under a process $\{\mathbf{X}^{(t)} | t \in T\}$, one can regard the σ -algebra $\mathcal{U}(t)$ as accumulating information from the observed variable $\mathbf{X}^{(s)}$ along the times $s \in [0, t]$. Thus, a one liner may be best to represent the message we try to deliver:

- a σ -algebra = information;
- a “growing” σ -algebra = more information coming in.

In fact, the idea of considering growing information, such as weather forecasting, stock pricing prediction, or daily CT changes, can be understood by a growing σ -algebra called a *filtration*, **Figure 15**. Such tool for analysis is commonly seen in quantitative finance (85, 86), which we believe it shares the same nature as a treatment in radiotherapy. The following concept describes growing (accumulating) information.

A sequence of σ -algebras $\{\mathcal{F}_t\}_{t \geq 0}$ on a measurable space (Ω, \mathcal{F}) with $\mathcal{F}_t \subseteq \mathcal{F}$ is called a *filtration* if $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$ whenever $t_1 \leq t_2$.

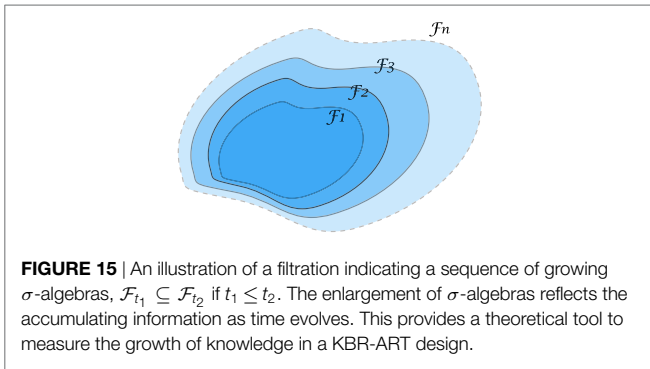


FIGURE 15 | An illustration of a filtration indicating a sequence of growing σ -algebras, $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$ if $t_1 \leq t_2$. The enlargement of σ -algebras reflects the accumulating information as time evolves. This provides a theoretical tool to measure the growth of knowledge in a KBR-ART design.

The labeling index t is usually referred to “time” or a similar concept, where in the radiotherapy case it may be treatment fractions, stages, or phases. If we consider a filtration generated from a stochastic process via $\mathcal{F}_t = \mathcal{U}(t)$, then, intuitively, this filtration is interpreted as containing all history available up to time t , but not future information available about the process. Due to this nature, a process adapted to a filtration \mathcal{F} is also called *non-anticipating*, indicating that one cannot see into the future.

Therefore, a KBR-ART system would rely on machine learning algorithms (such as CNN, RNN, DRL, . . . etc.) to explore non-anticipating filtrations and to learn from accumulating knowledge or information, such as the examples given in Sections 3.3 and 4.

To demonstrate the concept of filtrations more concretely in our setting, the following example is provided. Suppose a sequence of independent random variables $\{X^{(i)}\}_{i=1,2,3,\dots}$ denotes the *growth* in GTV size at stage i with $\mathbb{E}(X^{(i)}) = d_i$ for all i . If we have measured total growth up to stage k , i.e., $S^{(k)} := X^{(1)} + \dots + X^{(k)}$, we like to know what is our best guess for the growth after n more stages $S^{(k+n)}$, given the information of the past $S^{(1)}, \dots, S^{(n)}$?

Some computation reveals that

$$\begin{aligned} &\mathbb{E}(S^{(k+n)} | S^{(1)}, \dots, S^{(k)}) \\ &= \mathbb{E}(X^{(1)} + \dots + X^{(k+n)} | S^{(1)}, \dots, S^{(k)}) = S^{(k)} + \sum_{i=k+1}^n d_i, \end{aligned} \tag{44}$$

which indicates that the best surmise for the future value $S^{(k+n)}$, given the knowledge (history) up to stage k , is $S^{(k)}$ plus empirical understanding (averages), reflecting the information cease to grow after time step k . The computation Equation (44) relies on the following fact:

1. If X is \mathcal{F} -measurable, then $\mathbb{E}(X|\mathcal{F}) = X$ almost surely.
2. If X is independent of \mathcal{F} , then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(x)$ almost surely

After the above discussion of how information can be accumulated using σ -algebras, next, we discuss how to analyze sequential random variables from a more theoretical perspectives using time series.

5.1.1. Time Series

Due to the nature of sequential data, an KBR-ART is naturally related to time series, which are applied comprehensively in forecasting, such as econometrics, quantitative finance, seismology, and signal processing, etc. Quoting from (87):

A time series model for the observed data $\{x^{(t)} | t \in T\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X^{(t)} | t \in T\}$ of which $\{x^{(t)}\}$ is postulated to be a realization.

Incidentally, a time series is a special case of **stochastic processes** $\{X^{(t)} | t \in T\}$, where the time labeling set T can be an infinite set. In a very general case, a process $\{X^{(t)} | t \in \mathbb{Z}\}$ can have *Volterra expansion*

$$\begin{aligned} X^{(t)} = &c + \sum_{j=0}^{\infty} \vartheta_j Z^{(t-j)} + \sum_{j,k}^{\infty} \vartheta_{jk} Z^{(t-j)} Z^{(t-k)} \\ &+ \sum_{j,k,\ell}^{\infty} \vartheta_{jkl} Z^{(t-j)} Z^{(t-k)} Z^{(t-\ell)} + \dots, \end{aligned} \tag{45}$$

where high order terms can be considered. Usually, the modeling of time series is divided by two main categories, linear and non-linear methods.

In particular, there are three classes of linear models that carry practical importance, namely autoregressive models $AR(p)$, the moving average models $MA(q)$, and the integrated (I) models.

(The ARMA (p,q) process with mean μ) The process $\{X^{(t)} | t \in \mathbb{Z}\}$ is called an ARMA (p,q) process if it is stationary and satisfies for all t ,

$$\begin{aligned} &(X^{(t)} - \mu) - \varphi_1 (X^{(t-1)} - \mu) - \dots - \varphi_p (X^{(t-p)} - \mu) \\ &= Z^{(t)} - \vartheta_1 Z^{(t-1)} - \dots - \vartheta_q Z^{(t-q)}, \end{aligned} \tag{46}$$

where $\mu, \varphi_i, \vartheta_i \in \mathbb{R}$ and $\{Z^{(t)}\} \simeq \text{WN}(0, \sigma^2)$ are white noise (error terms).

Here, the ARMA(p, q) process refers to the model with p autoregressive terms and q moving-average terms. Especially, $p = 0$ and $q = 0$ in the ARMA(p, q) process corresponds to two useful linear cases called $AR(p)$ and $MA(q)$ models, respectively. The aim of studying the behavior of a time series $\{X^{(t)}\}$ can be done *via* the analysis of the depending coefficients φ_i, ϑ_i and its autocorrelation function (88), which we will not go through. An interesting fact is that one can study the causality of an ARMA(p, q) process via the following fact:

Let $\{X^{(t)}\}$ be an ARMA(p, q) process with $\varphi(z) := (1 + \varphi_1 z + \dots + \varphi_p z^p)$, $\vartheta(z) := (1 + \vartheta_1 z + \dots + \vartheta_q z^q)$ have no common zeros. Then $\{X^{(t)}\}$ is causal if and only if $\varphi|_{\mathcal{D}} \neq 0$ with $\mathcal{D} = \{z \in \mathbb{C} | \|z\| \leq 1\}$.

Thus $AR(1)$ process with $\mu = 0$ is only a simple case given by $X^{(t)} = Z^{(t)} + X^{(t-1)}$ from Equation (46). Since $\varphi(z) = 1 - \varphi_1 z$, it follows that $\{X^{(t)}\}$ is causal if $|\varphi_1| < 1$ and non-stationary when $|\varphi_1| = 1$. This $AR(1)$ case demonstrates that we may actually learn the behavior of a time series by analyzing the dependent coefficients. In fact, the heuristic $AR(1)$ process is directly related to the Markov process due to a fact (see Proposition 7.6 in (89)). Simply stated, for a process $\{X^{(t)} | t \in \mathbb{Z}\}$ taking values in a Borel space S , Z_1, Z_2, \dots are independent taking values in E and if there exist functions $f_i: S \times E \rightarrow S$, $t \in \mathbb{Z}$, such that $X^{(t)}$ is recursively defined by

$$X^{(t)} = f_t \left(X^{(t-1)}, Z^{(t)} \right), \quad X^{(0)} = x_0 \in S, \quad (47)$$

then the process $\{X^{(t)} | t \in \mathbb{Z}\}$ is Markov. This result Equation (47) then justifies the claim that the AR(1) is a Markov process as the transition functions simply indicate $f_t(X^{(t-1)}, Z^{(t)}) = Z^{(t)} + \varphi_1 X^{(t-1)}$ from Equation (47). Moreover, it is time-homogeneous since $\{Z^{(t)}\}$ are i.i.d. and f_t is fixed across all t . As one recalls that the Markov process is defined under the property

$$\begin{aligned} &P \left(\left(X^{(t)} \right)^{-1} (B) | \mathcal{U}(s) \right) \\ &= P \left(\left(X^{(t)} \right)^{-1} (B) | X^{(s)} \right) \quad (\forall \text{ Borel } B \subseteq \mathbb{R}^n, t \geq s \geq 0) \end{aligned} \quad (48)$$

where $\mathcal{U}(s)$ is as defined in Equation (43). At the prediction level, AR(1) or Markov process then indicates that one can estimate the probabilities of future values $X^{(t)}$ just as well as if one was aware of the entire history of the process $\mathcal{U}(s)$ prior to time s . The Markov property Equation (48) serves as a simplifying assumption to reduce complexities in variables involved. Therefore, it is one of our reasons to introduce the Bayesian Networks modeling based on Markov process in Section 3.2.4.

5.2. Comparison of Varying Data-Driven Models

There are a large number of statistical models in the area of machine learning. They can be basically divided into 3 categories: *supervised*, *unsupervised*, and *reinforcement learning*, where supervised models are mainly used for data *prediction*, unsupervised models are usually used to explore intrinsic data structure such as probability and location distribution, and the reinforcement learning, which we will introduce in Section 4.4 is to learn best controls within certain circumstances. All the methods introduced in Section 3.2 belong to the supervised learning category, which is the cornerstone for KBR-ART system implementation. It is essential for a KBR-ART to have an accurate model for future prediction in patients' status, e.g., organ geometry and shape changing, whether the model is analytical or statistical. Statistical modeling is typically a handy choice over analytical one to overcome the modeling complexity involved in mechanistic realizations of radiotherapy interactions.

Comparison of the merits of several classical methods such as linear regression Section 3.2.1, Bayesian networks Section 3.2.4, decision trees, and SVMs can be found in (90–92). Generally speaking, the pros of classical data-driven models such as linear regression and Bayesian networks is that they are interpretable, numerically stable, computationally efficient, and work even on small sample-sized dataset, but the cons are that they lack versatility in tasking (e.g., no one uses regressions for image segmentation or contouring) and do not possess the ability to handle complex and high variety of data, such as images, video, sequences, languages, and mixture data. For complex data such as the RT data, one can rely on more modern techniques such as deep learning, particularly DNN, CNN, and RNN-based structures. For intensive review regarding deep learning and their merits, one may refer to (42, 93). The trade-off between handling complex data and data interpretability may drive one to choose between classical and

deep machine learning methods. Moreover, deep learning techniques typically require larger amount of observations compared to classical statistical learning techniques. This is a main reason that deep learning is not yet as prominent in medical and biological field compared to its current dominant in computer science and engineering. The bottom line here is that there yet no universal recognition for which classifier can do the best job in biomedicine or oncology. The development of KBR-ART is foreseeable to rely more deep learning approaches for outcome modeling and variety tasks of (image, sequential) data processing and decision-making.

6. CONCLUSION

In this study, we presented a framework for comprehensive KBR-ART design and implementation based on machine learning and explored some of its main characteristics. First, in Section 2, we analyzed the characteristics and types of features in clinical data as effective choice of data for feeding knowledge into KBR-ART. Second, in Section 3, we visited a few promising and powerful techniques of modern machine learning development, such as DNNs, CNNs, RNNs as well as the classical linear regression-type models. The KBR-ART framework we proposed here rely on machine learning techniques, which are capable of accurate prediction and sequential learning, which are the cornerstones for building up a KBR-ART system. There are three pertained questions to the design and realization of KBR-ART, which we addressed in this paper and we presented illustrative examples for each case highlighted by the application RL/BN onto a NSCLC radiotherapy dataset. In Section 4, we provided a unifying formulation in Section 4.1 for designing a KBR-ART system (Equation 18). The purpose was twofold: (1) to clearly understand the essence of previous constructed ARTs of last generation, (2) to provide a guiding principle for designing next generation algorithms.

The application of the presented technologies here provides great promise for the field of KBR-ART, yet there are still numerous challenges ahead. First, there is highly complex nature of radiation interaction with human biology that we are still trying to develop a better understanding. Second, medical datasets typically suffer from small sizes and often incomplete. Several efforts between nations and domestic institutes are being carried out to consolidate larger datasets for oncology studies, for the purpose of statistical model training and validations, but many are still in the infancy. Nevertheless, this paper still serves as a blueprint laying the foundation for the establishment and applicability of KBR-ART using modern machine learning techniques.

AUTHOR CONTRIBUTIONS

H-HT writes up and collects main materials related to the study; YL also collects materials related to the study; RTH plots and organizes the study; IEN directs and organizes the study.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Jolly and Dr. Kong for the help of providing the lung testing datasets for the study. H-HT thanks Sunan Cui for numerous fruitful discussions. This work was supported in part by the National Institutes of Health P01 CA059827.

REFERENCES

- Stanley HB, El Naqa I, Klein EE. Introduction to big data in radiation oncology: exploring opportunities for research, quality assessment, and clinical care. *Int J Radiat Oncol Biol Phys* (2016) 95(3):871–2. doi:10.1016/j.ijrobp.2015.12.358
- El Naqa I. Perspectives on making big data analytics work for oncology. *Methods* (2016) 111:32–44. doi:10.1016/j.jymeth.2016.08.010
- Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *Int J Radiat Oncol Biol Phys* (2017) 99(4):994–1003. doi:10.1016/j.ijrobp.2017.04.023
- Xing L, Siebers J, Keall P. Computational challenges for image-guided radiation therapy: framework and current research. *Semin Radiat Oncol* (2007) 17(4):245–57. doi:10.1016/j.semradonc.2007.07.004
- Graves A, Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in Neural Information Processing Systems*. (2009). p. 545–52.
- Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks. *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*. (2014). p. 1764–72.
- Baldi P, Brunak S, Frasconi P, Soda G, Pollastri G. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* (1999) 15(11):937–46. doi:10.1093/bioinformatics/15.11.937
- Übeyli ED. Recurrent neural networks with composite features for detection of electrocardiographic changes in partial epileptic patients. *Comput Biol Med* (2008) 38(3):401–10. doi:10.1016/j.compbiomed.2008.01.002
- Übeyli ED. Combining recurrent neural networks with eigenvector methods for classification of ecg beats. *Digit Signal Process* (2009) 19(2):320–9. doi:10.1016/j.dsp.2008.09.002
- Shen H, George D, Huerta E, Zhao Z. *Denoising gravitational waves using deep learning with recurrent denoising autoencoders*. (2017). *arXiv preprint arXiv:1711.09919*.
- Wu JT, Derroncourt F, Gehrmann S, Tyler PD, Moseley ET, Carlson ET, et al. Behind the scenes: a medical natural language processing project. *Int J Med Inform* (2018) 112:68–73. doi:10.1016/j.ijmedinf.2017.12.003
- Marks LB. Dosimetric predictors of radiation-induced lung injury. *Int J Radiat Oncol Biol Phys* (2002) 54(2):313–6. doi:10.1016/S0360-3016(02)02928-0
- Levegrün S, Jackson A, Zelefsky MJ, Skwarchuk MW, Venkatraman ES, Schlegel W, et al. Fitting tumor control probability models to biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer: pitfalls in deducing radiobiologic parameters for tumors from clinical data. *Int J Radiat Oncol Biol Phys* (2001) 51(4):1064–80. doi:10.1016/S0360-3016(01)01731-X
- Hope AJ, Lindsay PE, El Naqa I, Alaly JR, Vicic M, Bradley JD, et al. Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters. *Int J Radiat Oncol Biol Phys* (2006) 65(1):112–24. doi:10.1016/j.ijrobp.2005.11.046
- Bradley J, Deasy JO, Bentzen S, El Naqa I. Dosimetric correlates for acute esophagitis in patients treated with radiotherapy for lung carcinoma. *Int J Radiat Oncol Biol Phys* (2004) 58(4):1106–13. doi:10.1016/j.ijrobp.2003.09.080
- Blanco AI, Chao KC, El Naqa I, Franklin GE, Zakarian K, Vicic M, et al. Dose-volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy. *Int J Radiat Oncol Biol Phys* (2005) 62(4):1055–69. doi:10.1016/j.ijrobp.2004.12.076
- Deasy JO, El Naqa I. *Image-Based Modeling of Normal Tissue Complication Probability for Radiation Therapy*. Boston, MA: Springer (2008). p. 211–52.
- El Naqa I, Suneja G, Lindsay P, Hope AJ, Alaly J, Vicic M, et al. Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. *Phys Med Biol* (2006) 51(22):5719. doi:10.1088/0031-9155/51/22/001
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48(4):441–6. doi:10.1016/j.ejca.2011.11.036
- Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med* (2017) 38:122–39. doi:10.1016/j.ejmp.2017.05.071
- Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* (2001) 69(3):89–95. doi:10.1067/mcp.2001.113989
- El Naqa I, Craft J, Oh J, Deasy J. Biomarkers for early radiation response for adaptive radiation therapy. *Adapt Radiat Ther* (2011) 53–68.
- Tseng H-H, Luo Y, Cui S, Chien J-T, Ten Haken RK, El Naqa I. Deep reinforcement learning for automated radiation adaptation in lung cancer. *Med Phys* (2017) 44(12):6690–705. doi:10.1002/mp.12625
- 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
- Webb S. *The Physics of Three Dimensional Radiation Therapy: Conformal Radiotherapy, Radiosurgery and Treatment Planning*. CRC Press (1993). Available from: <https://www.taylorfrancis.com/books/9781420050363>
- Joiner MC, Van der Kogel A. *Basic Clinical Radiobiology*. (Vol. 2). CRC Press (2016). Available from: <https://www.crcpress.com/Basic-Clinical-Radiobiology-Fourth-Edition/Joiner-van-der-Kogel/p/book/9780340929667>
- El Naqa I. *A Guide to Outcome Modeling in Radiotherapy and Oncology: Listening to the Data*. CRC Press (2018).
- Hall E, Giaccia A. *Radiobiology for the Radiologist*. Philadelphia: Lippincott Williams & Wilkins (2006).
- Zaider M, Minerbo G. Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Phys Med Biol* (2000) 45(2):279. doi:10.1088/0031-9155/45/2/303
- Goitein M. Tumor control probability for an inhomogeneously irradiated target volume. *Eval Treat Plan Part Beam Radiother* (1987).
- Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res Suppl* (1985) 8:S13–9. doi:10.2307/3576626
- Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. *Med Phys* (1997) 24(1):103–10. doi:10.1118/1.598063
- Niemierko A. A generalized concept of equivalent uniform dose (eud). *Med Phys* (1999) 26(6):1100.
- Coates J, Jayaseelan AK, Ybarra N, David M, Faria S, Souhami L, et al. Contrasting analytical and data-driven frameworks for radiogenomic modeling of normal tissue toxicities in prostate cancer. *Radiother Oncol* (2015) 115(1):107–13. doi:10.1016/j.radonc.2015.03.005
- Tucker SL, Li M, Xu T, Gomez D, Yuan X, Yu J, et al. Incorporating single-nucleotide polymorphisms into the lyman model to improve prediction of radiation pneumonitis. *Int J Radiat Oncol Biol Phys* (2013) 85(1):251–7. doi:10.1016/j.ijrobp.2012.02.021
- James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: with Applications in R*. (Vol. 112). New York: Springer (2013).
- Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Methodol* (1996) 58(1):267–88.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* (2005) 67(2):301–20. doi:10.1111/j.1467-9868.2005.00527.x
- Ng AY. Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the Twenty-First International Conference on Machine Learning*. New York, NY: ACM (2004). 78 p.
- Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw* (1989) 2(5):359–66. doi:10.1016/0893-6080(89)9020-8
- Bovier A, Picco P. *Mathematical Aspects of Spin Glasses and Neural Networks*. (Vol. 41). Boston: Springer Science & Business Media (2012).
- Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*. (Vol. 1). Cambridge: MIT press (2016).
- Bulat I, Lei X. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys* (2016) 44(2):547–57. doi:10.1002/mp.12045
- Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol* (2017) 7:315. doi:10.3389/fonc.2017.00315
- Qin W, Wu J, Han F, Yuan Y, Zhao W, Ibragimov B, et al. Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation. *Phys Med Biol* (2018) 63(9):095017. doi:10.1088/1361-6560/aabd19
- Wang Y, Zu C, Hu G, Luo Y, Ma Z, He K, et al. Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Process Lett* (2018). doi:10.1007/s11063-017-9759-3

47. Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. *Phys Med Biol* (2017) 62(21):8246. doi:10.1088/1361-6560/aa8d09
48. 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 Transactions on Radiation and Plasma Medical Sciences*. (2018). p. 1–1. Available from: <https://ieeexplore.ieee.org/abstract/document/8353476/>
49. Ogunmolu OP, Gu X, Jiang SB, Gans NR. Nonlinear systems identification using deep dynamic neural networks. *CoRR* (2016) abs/1610.01439.
50. Cai J, Lu L, Xie Y, Xing F, Yang L. Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer International Publishing (2017). p. 674–82.
51. Roth HR, Lu L, Farag A, Sohn A, Summers RM. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing (2016). p. 451–9.
52. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann Publishers Inc (1988).
53. Griffiths T, Yuille A. A primer on probabilistic inference. *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford: Oxford University Press (2008) p. 33–57. doi:10.1093/acprof:oso/9780199216093.001.0001
54. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search*. New York: Springer (2012).
55. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn* (1997) 29(2–3):131–63. doi:10.1023/A:1007465528199
56. Kalet AM, Gennari JH, Ford EC, Phillips MH. Bayesian network models for error detection in radiotherapy plans. *Phys Med Biol* (2015) 60(7):2735. doi:10.1088/0031-9155/60/7/2735
57. Gomes E, Duarte J, Frutuoso e Melo PF. Human reliability modeling of radiotherapy procedures by bayesian networks and expert opinion elicitation. *Nucl Technol* (2016) 194(1):73–96. doi:10.13182/NT15-29
58. Jayasurya K, Fung G, Yu S, Dehing-Oberije C, De Ruyscher D, Hope A, et al. Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Med Phys* (2010) 37(4):1401–7. doi:10.1118/1.3352709
59. Oh JH, Craft J, Lozi RA, Vaidya M, Meng Y, Deasy JO, et al. A bayesian network approach for modeling local failure in lung cancer. *Phys Med Biol* (2011) 56(6):1635. doi:10.1088/0031-9155/56/6/008
60. Lee S, Ybarra N, Jeyaseelan K, Faria S, Kopeck N, Brisebois P, et al. Bayesian network ensemble as a multivariate strategy to predict radiation pneumonitis risk. *Med Phys* (2015) 42(5):2421–30. doi:10.1118/1.4915284
61. Luo Y, El Naqa I, McShan D, Matuszak M, Jolly S, Haken RKT. Simultaneous prediction of specific radiotherapy outcomes using a multi-objective bayesian network (moBN) approach. *Int J Radiat Oncol Biol Phys* (2017) 99(2, Suppl):S35. doi:10.1016/j.ijrobp.2017.06.094
62. Jochems A, Deist TM, El Naqa I, Kessler M, Mayo C, Reeves J, et al. Developing and validating a survival prediction model for nscl patients through distributed learning across 3 countries. *Int J Radiat Oncol Biol Phys* (2017) 99(2):344–52. doi:10.1016/j.ijrobp.2017.04.021
63. Tucker SL, Zhang M, Dong L, Mohan R, Kuban D, Thames HD. Cluster model analysis of late rectal bleeding after imrt of prostate cancer: a case-control study. *Int J Radiat Oncol Biol Phys* (2006) 64(4):1255–64. doi:10.1016/j.ijrobp.2005.10.029
64. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *CoRR* (2014).
65. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. *IEEE International Conference on Computer Vision (ICCV)*. (2017). p. 618–26. Available from: <https://ieeexplore.ieee.org/document/8237336/>
66. Ågren A, Brahme A, Turesson I. Optimization of uncomplicated control for head and neck tumors. *Int J Radiat Oncol Biol Phys* (1990) 19(4):1077–85. doi:10.1016/0360-3016(90)90037-K
67. Keller H, Ritter MA, Mackie T. Optimal stochastic correction strategies for rigid-body target motion/optimal stochastic correction strategies for rigid-body target motion. *Int J Radiat Oncol Biol Phys* (2003) 55(1):261–70. doi:10.1016/S0360-3016(02)03867-1
68. Humpherys J, Redd P, West J. A fresh look at the kalman filter. *SIAM Rev* (2012) 54(4):801–23. doi:10.1137/100799666
69. de la Zerda A, Armbruster B, Xing L. Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework. *Phys Med Biol* (2007) 52(14):4137. doi:10.1088/0031-9155/52/14/008
70. Bortfeld T, Chan TCY, Trofimov A, Tsitsiklis JN. Robust management of motion uncertainty in intensity-modulated radiation therapy. *Oper Res* (2008) 56(6):1461–73. doi:10.1287/opre.1070.0484
71. Chan TC, Mišić VV. Adaptive and robust radiation therapy optimization for lung cancer. *Eur J Oper Res* (2013) 231(3):745–56. doi:10.1016/j.ejor.2013.06.003
72. Mar PA, Chan TCY. Adaptive and robust radiation therapy in the presence of drift. *Phys Med Biol* (2015) 60(9):3599. doi:10.1088/0031-9155/60/9/3599
73. Lujan AE, Larsen EW, Balter JM, Ten Haken RK. A method for incorporating organ motion due to breathing into 3d dose calculations. *Med Phys* (1999) 26(5):715–20. doi:10.1118/1.598577
74. Löf J, Lind BK, Brahme A. An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion. *Phys Med Biol* (1998) 43(6):1605. doi:10.1088/0031-9155/43/6/018
75. Rehbinder H, Forsgren C, Löf J. Adaptive radiation therapy for compensation of errors in patient setup and treatment delivery. *Med Phys* (2004) 31(12):3363–71. doi:10.1118/1.1809768
76. Sutton RS, McAllester DA, Singh SP, Mansour Y. Policy gradient methods for reinforcement learning with function approximation. In: Solla SA, Leen TK, Müller K, editors. *Advances in Neural Information Processing Systems*. MIT Press (2000). p. 1057–63. Available from: <http://papers.nips.cc/paper/1713-policy-gradient-methods-for-reinforcement-learning-with-function-approximation.pdf>
77. Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning*. MIT Press (2012).
78. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. (Vol. 1). Cambridge: MIT press (1998).
79. Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. *Playing atari with deep reinforcement learning*. (2013). *arXiv preprint arXiv:1312.5602*.
80. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature* (2015) 518(7540):529–33. doi:10.1038/nature14236
81. Kong FM, Ten Haken RK, Schipper M, Frey KA, Hayman J, Gross M, et al. Effect of midtreatment PET/CT-adapted radiation therapy with concurrent chemotherapy in patients with locally advanced non-small-cell lung cancer: a phase 2 clinical trial. *JAMA Oncol* (2017) 3(10):1358–65. doi:10.1001/jamaoncol.2017.0982
82. Schuck NW, Wilson RC, Niv Y. *A State Representation for Reinforcement Learning and Decision-Making in the Orbitofrontal Cortex*. (2017). *bioRxiv*.
83. Chung KL, editor. *Probability and mathematical statistics. A Course in Probability Theory (Second Edition)*. San Diego: Academic Press (1974). Available from: <http://www.sciencedirect.com/science/article/pii/B978008057040250001X>
84. Durrett R. *Probability: Theory and Examples*. Cambridge University Press (2010). Available from: <http://www.cambridge.org/us/academic/subjects/statistics-probability/probability-theory-and-stochastic-processes/probability-theory-and-examples-4th-edition?format=HB&isbn=9780521765398>
85. Davis M. *Mathematics of financial markets*. In: Engquist B, Schmid W, editors. *Mathematics Unlimited—2001 and Beyond*. Berlin, Heidelberg: Springer (2001). p. 361–80.
86. Privault N. *Stochastic Analysis in Discrete and Continuous Settings: With Normal Martingales*. Berlin, Heidelberg: Springer (2009).
87. Brockwell PJ, Davis RA. *Introduction to Time Series and Forecasting*. Switzerland: Springer International Publishing (2016).
88. Brockwell PJ, Davis RA, Fienberg SE, Berger JO, Gani J, Krickeberg K, et al. *Time Series: Theory and Methods*. New York: Springer (1991).
89. Kallenberg O. *Foundations of Modern Probability*. New York: Springer (2006).
90. Kotsiantis SB. Supervised machine learning: a review of classification techniques. *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications*

- in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Amsterdam, Netherlands: IOS Press (2007). p. 3–24.
91. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Knowl Inform Syst* (2008) 14(1):1–37. doi:10.1007/s10115-007-0114-2
 92. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*. New York, NY: ACM (2006). p. 161–8.
 93. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* (2015) 521(7553):436. doi:10.1038/nature14539

Conflict of Interest Statement: 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.

Copyright © 2018 Tseng, Luo, Ten Haken and El Naqa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner 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.