

# Radiomic and radiogenomic modeling for radiotherapy: strategies, pitfalls, and challenges

James T. T. Coates<sup>1a,\*</sup>, Giacomo Pirovano<sup>1b</sup>, and Issam El Naqa<sup>1c</sup>

<sup>a</sup>Massachusetts General Hospital & Harvard Medical School, Center for Cancer Research, Boston, Massachusetts, United States

<sup>b</sup>Memorial Sloan Kettering Cancer Center, Department of Radiology, New York, New York, United States

<sup>c</sup>Moffitt Cancer Center and Research Institute, Department of Machine Learning, Tampa, Florida, United States

**Abstract.** The power of predictive modeling for radiotherapy outcomes has historically been limited by an inability to adequately capture patient-specific variabilities; however, next-generation platforms together with imaging technologies and powerful bioinformatic tools have facilitated strategies and provided optimism. Integrating clinical, biological, imaging, and treatment-specific data for more accurate prediction of tumor control probabilities or risk of radiation-induced side effects are high-dimensional problems whose solutions could have widespread benefits to a diverse patient population—we discuss technical approaches toward this objective. Increasing interest in the above is specifically reflected by the emergence of two nascent fields, which are distinct but complementary: radiogenomics, which broadly seeks to integrate biological risk factors together with treatment and diagnostic information to generate individualized patient risk profiles, and radiomics, which further leverages large-scale imaging correlates and extracted features for the same purpose. We review classical analytical and data-driven approaches for outcomes prediction that serve as antecedents to both radiomic and radiogenomic strategies. Discussion then focuses on uses of conventional and deep machine learning in radiomics. We further consider promising strategies for the harmonization of high-dimensional, heterogeneous multiomics datasets (panomics) and techniques for nonparametric validation of best-fit models. Strategies to overcome common pitfalls that are unique to data-intensive radiomics are also discussed. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.8.3.031902](https://doi.org/10.1117/1.JMI.8.3.031902)]

**Keywords:** radiogenomics; radiomics; radiotherapy; predictive modeling; outcomes.

Paper 20229VSSR received Sep. 1, 2020; accepted for publication Jan. 12, 2021; published online Mar. 23, 2021.

## 1 Introduction

Radiotherapy is received by approximately half of all cancer patients.<sup>1</sup> The development of predictive models for determining which patients are likely to benefit from radiotherapy and which are at risk of incurring aberrant toxicities could, therefore, provide benefit to a large population. Early efforts to capture radiation-induced side effects or assign a probability for achieving local control consisted principally of dose–volume effect correlations<sup>2,3</sup>—these approaches yielded some success with classical treatment modalities, but are now recognized to be intrinsically limited due to patient heterogeneity at the biological level.<sup>4–6</sup> Two distinct but complementary strategies have emerged in recent years to overcome this limitation: (i) the integration of patient-specific biological risk factors into dose–volume-based outcome models, which we refer to herein as radiogenomics,<sup>7</sup> and (ii) the integration of imaging correlates together with treatment-related and biological data for outcomes prediction, radiomics.<sup>8</sup> We note that an alternative definition of radiogenomics exists in literature entailing the exclusive use of imaging correlates and genomic data;<sup>9</sup> however, given our focus on radiotherapy outcome modeling, we forgo this

\*Address all correspondence to James T. T. Coates, [jtcoates@mgh.harvard.edu](mailto:jtcoates@mgh.harvard.edu)

definition in preference of the above. Taken together, there is renewed interest in the development of personalized radiotherapy treatment plans and, eventually, their integration into automated treatment planning systems.<sup>10,11</sup>

We review radiogenomic and radiomic modeling strategies herein first by discussing the types of data and derivative metrics (correlates) commonly used, including next-generation datasets. Classes of outcomes associated with either tumor control or normal tissue toxicity are then reviewed. We consider conventional dose–volume-based approaches for outcomes modeling, which sets the stage for discussion of radiogenomic and radiomic modeling approaches and furthermore highlight techniques for the augmentation of classical models allowing them to incorporate biological and/or clinical risk factors. Both conventional and deep learning strategies are reviewed in the context of radiomics. Given the emergence of next-generation technologies and large datasets obtained from unique sources, we discuss potential outcome modeling strategies using the integration of heterogeneous and high-dimensional multiomics datasets (panomics).<sup>12</sup> As they are critical for any modeling approach, we conclude by discussing common pitfalls for data-intensive radiomics or panomics and validation methods that can be used to maximize reproducibility and robustness of best-fit models.

## 2 Definition of Risk

Predicting outcomes for radiotherapy necessitates a definitive understanding of risk (that is to be predicted). In the case of radiotherapy, there are two main categories:<sup>6</sup> (i) tumor control probability (TCP) and (ii) normal tissue complication probability (NTCP). TCP refers to the probability of success for the treatment procedure defined over a time period. Ideally, TCP models would also take into account the possibility of recurrence at longer times after treatment, but data for this can be sparse and furthermore depend on factors that significantly increase model complexity.<sup>5</sup> NTCP by contrast defines the risk of specific aberrant toxicities induced as a consequence of ionizing radiation and are generally attributed to damage of healthy tissues.<sup>6</sup> It has become recognized that better performing NTCP strategies necessitate the inclusion of data beyond dose–volume metrics, such as the functional statuses of genes that significantly modulate normal repair processes, or the use of other biologically relevant data, to better capture inter-patient variabilities.<sup>13,14</sup>

## 3 Radiotherapy Outcomes (Endpoints)

Ionizing radiation induces effects that span nanoseconds (free radical production) to years after delivery (late side effects). Upon delivery, free radicals diffuse and induce a complex cascade of molecular and cellular processes that may only become manifest and clinically detectable days, weeks, months, or years later.<sup>15,16</sup> Every effort is made to spatially target radiation to cancer cells; however, normal cells and critical structures adjacent to target volumes unavoidably receive a portion of the therapeutic dose.<sup>17</sup> Herein, we refer to toxicity from radiotherapy as dose-limiting and this may include acute effects or late adverse events as detailed below.

### 3.1 Early Side Effects

Acute side effects from radiation therapy are often transient, self-limiting events that can typically resolve within a few weeks post-treatment and do not induce severe or long-term morbidity, although this might vary patient-by-patient. An example is the acute inflammation or ulceration of mucosal membranes inducing mucositis in head and neck cancer treatment.<sup>18</sup> Notably, a difficulty in assessing early side effects is the interpretation of data from independent sources that may use different grading schemas or endpoints.<sup>19</sup>

### 3.2 Late Toxicity Events

Late radiation-induced damage to normal tissue occurs empirically >90 days after completion of radiotherapy. Late toxicities can often be difficult to assess, as no quantitative physiological

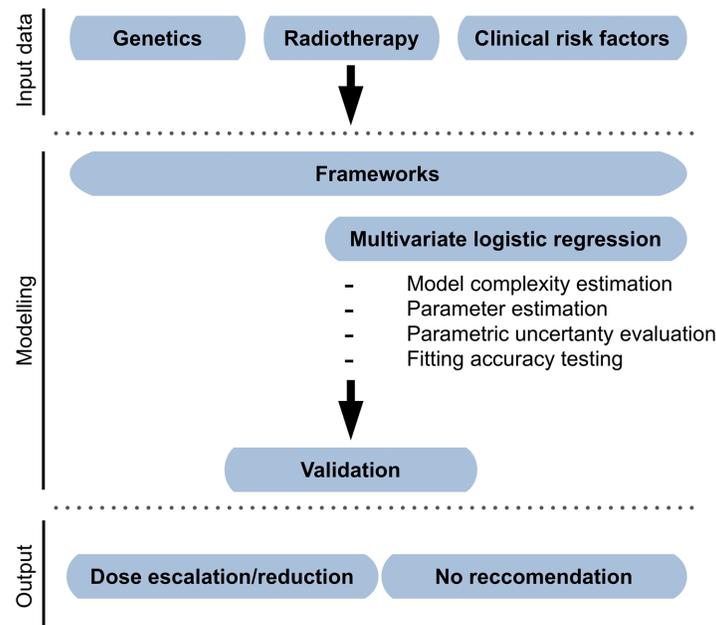
evidence exists or can readily be obtained, and include mild, moderate, severe, and life-threatening morbidities, sometimes requiring additional intervention to be mitigated. Grading schemes have been developed to classify such morbidity by physicians assigning integer values to the induced side effects. These could be self-scoring questionnaires or grades assigned by an attending physician. The correlation between different scoring schemes using a single set of data can be explored utilizing various machine learning (ML) techniques that rely on the accuracy of outcome measures.<sup>20–22</sup>

### 3.3 Local Control Endpoints

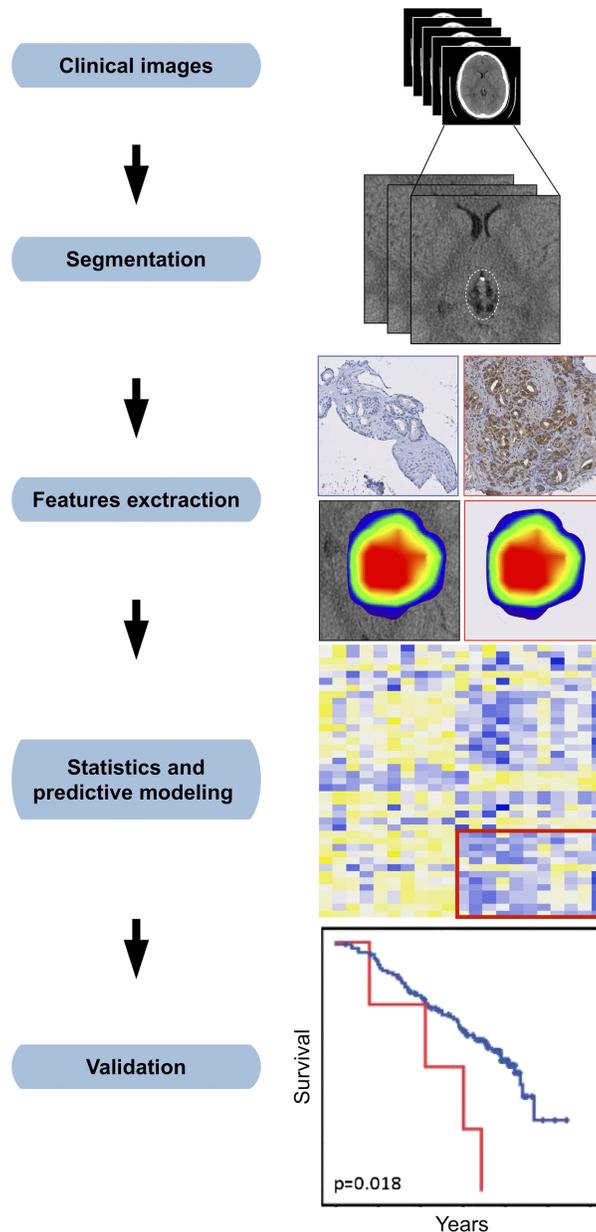
Reporting of local control is usually specific to a given site. For instance, in prostate cancer, clinical studies typically report local control according to the ASTRO-RTOG Phoenix definition of biochemical failure.<sup>23</sup> Nondichotomized metrics such as prostate-specific antigen (PSA) doubling time or PSA scores over time could also be used as endpoints in modeling frameworks if continuous outcomes measures are required but may be less reliable.<sup>24</sup>

## 4 Modeling Workflow

The overarching objective of both radiogenomics and radiomics is to prospectively assess the suitability of a patient for a particular treatment regimen; however, the two approaches have unique requirements. Radiogenomics leverages genetic and/or other biological information for improving the prediction of dose–volume-based models, such as in Fig. 1. Radiomics, however, further requires the processing of imaging records to derive metrics that can be used as input data for modeling (Fig. 2). Herein, we consider radiomics to include all aspects of radiogenomics, but strive to discriminate between the two strategies when required as the introduction of imaging features restricts suitable modeling approaches.



**Fig. 1** Example of radiogenomic modeling framework. Clinical, dosimetric, and biological data are regressed using a logistic transformation. Internal resampling is used as internal validation. A recommendation of dose escalation, de-escalation, or no recommendation can be generated according to desired thresholds for risk.



**Fig. 2** Image processing steps for radiomics modeling. Extracted features from a variety of clinical imaging modalities can be used for radiomics, but typically include one or more of MRI, CT, and positron emission tomography (PET). Segmentation entails the delineation of ROI, which could make use of multimodality acquisition. Features are extracted and used as input data for modeling. Predictive modeling is then performed using a selected framework or strategy before validation and assessment of model generalizability.

#### 4.1 Retrieval of Nonimaging Input Data

Dose and volume metrics can be quickly extracted from dose–volume histograms at the treatment planning stage. Clinical data are important in outcomes modeling and can be parameterized as a covariate of interest.<sup>13</sup> An example where clinical information may be key is the use of anticoagulants and the reporting of aberrant bleeding, for instance. The spatial dose–volume histogram (z-DVH) can also be used to compare spatial treatment planning information with outcomes and many resources exist for their derivation.<sup>25,26</sup> Biological variables generally refer to single-nucleotide polymorphisms (SNPs) but can also include copy number variations (CNVs)—these mutations can be quantified by genotyping/sequencing as per our previous work and input into frameworks as dichotomous factors.<sup>27</sup> We discuss each of these in more detail in

the next major section with the exception of datasets that consist of next-generation sequencing (NGS), or similar high-throughput approaches (high-dimensional datasets), which are discussed as panomics strategies later.

## 4.2 Modeling Framework

NTCP or TCP modeling can be divided into two overarching methods: analytical or data-driven. Analytical approaches are based on simplified versions of the underpinning biochemical effects that treatment has on different tissues—analytical models are, therefore, also referred to as quasianalytical due to their recognized limitations discussed in detail elsewhere.<sup>6</sup> In contrast, data-driven models, or multimetric, are phenomenological in nature such that modeling frameworks based on them will seek to find the best performing model to capture the underlying effects regardless of the constituents of the model itself (unbiased).<sup>28</sup> Each of these methods is described in detail in Sec. 6.

## 4.3 Model Validation

The importance of crafting a modeling framework itself is critical, but internal performance evaluation prior to testing on unseen data has increasingly become recognized as an important step in predictive modeling—this is especially the case for more complex models whereby the opportunities for overfitting a particular dataset is significant.

There are several strategies that can be applied to quickly ascertain whether the results of a modeling exercise are suitable to proceed to the next steps or whether the framework requires fine-tuning. Most commonly used are nonparametric statistical measures, such as the Spearman's rank correlation ( $r_s$ , or  $\rho$ ).<sup>28</sup> Used in combination with Chi-squared ( $\chi^2$ ) statistical analyses, the degree of fit between each model parameter and the observed outcomes can be assessed without a need for increased computational power. As we discuss, information theory approaches and, more commonly, resampling techniques (e.g., cross-validation through jack-knifing or bootstrapping) can be used to evaluate a model's complexity relative to its fit to assess the performance of the model on unseen data (generalizability).<sup>29</sup>

## 4.4 Image Processing for Radiomics

Radiomics approaches typically include steps similar to radiogenomic modeling (data retrieval, modeling, validation), but with added caveats relating to image acquisition, tumor segmentation, and feature extraction before predictive modeling and model validation (Fig. 2).<sup>5,9</sup> Because of the amount of data generated in radiomics, the most common approach for modeling is ML. Moreover, in contrast to dose–volume metric modeling, no analytical associations of imaging correlates or derivatives and clinical endpoints currently exist, thus restricting radiomics to data-driven techniques.

### 4.4.1 Image acquisition

Various imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and 18F-fluorodeoxyglucose-positron emission tomography (FDG-PET) are used in the clinics to provide visualize and evaluate the underlying anatomical or physiological properties, which vary inpatient and interpatient. Unfortunately, there is often a lack of standard methods for image investigation and acquisition, which can represent a barrier to the pan-applied radiomics approach.<sup>30</sup> From the radiomics point of view, it is crucial to have homogeneous data (albeit through preprocessing or standardization during acquisition) when dealing with multicenter sources, as it is a necessity for validation. Recent developments in high-performing computing can help by mitigating variability and with harmonizing data.<sup>31,32</sup>

### 4.4.2 Tumor segmentation

Segmentation is the step of distinguishing the tumor area from the acquired images. Treatment plans in radiotherapy patients already provide a tumor delineation that can be used

retrospectively in the radiomic analysis. However, this human-based approach can also present variations among different oncologists or institutes. Radiomic modeling allows testing models for such variations by introducing tumor region perturbations.<sup>9</sup> An alternative to such an approach would be the use of segmentation algorithms, e.g., deep-learning techniques, which can learn the features along with the classification task.<sup>33–36</sup>

#### 4.4.3 Feature extraction

To fully characterize the tumor area, both semantic and agnostic radiomic features can be used. The first is existing radiology tumor descriptors. The latter is purely computational features, which can describe shape, density, and texture to define intertumoral.

## 5 Input Data for Modeling Frameworks

A plethora of metrics exist for use as inputs to modeling frameworks. These can be of different natures and can add complexity to the model to virtually an endless number of degrees. However, larger datasets can introduce high amounts of complexity, which can make it challenging to produce an accurate model. The main data types are described below.

### 5.1 Physical

Outcome modeling starts from dose–volume metrics, i.e., dose to a given volume or volume of tissue receiving a particular dose. Already at the stage of treatment planning, it is possible to derive such parameters. Other variations, such as physiological changes and changes in tumor composition or anatomy, may take place during treatment. For this reason, often, the delivered dose may not necessarily reflect the actual biologically absorbed dose. Intrafractional computed tomography (CT) scan changes could be included to improve predictive accuracy.<sup>37</sup>

The equivalent uniform dose (EUD) can be used to describe inhomogeneous dose distributions [Eq. (1)]<sup>38</sup> and an extension of the EUD for include normal tissue dose is the generalized EUD (gEUD) [Eq. (2)]

$$\text{EUD} = \left( \sum_i v_i D_i^{1/a} \right)^a, \quad (1)$$

$$\text{gEUD} = \left( \sum_i v_i D_i^a \right)^{1/a}, \quad (2)$$

where  $v_i$  represents fractional volume for the tissue exposed to a dose  $D_i$ , and the parameter  $a$  refers to the volume effect of a chosen tissue type. These two metrics serve as excellent tools to summarize dose–volume distributions.

### 5.2 Clinical

Other than demographic data, clinical data can impact significantly the outcome modeling and should be parameterized and used to define the covariates. Anything that might have an impact on patients' health can be incorporated and can potentially extend the model to any degree of complexity.

Covariates of interest can be extracted from clinical data. For example, follow-up therapies such as anticoagulants or androgen deprivation therapy can induce effects that are difficult to distinguish from late toxicities.

### 5.3 Spatial

When comparing spatial treatment planning information with clinical outcomes, z-DVHs can be used.<sup>25,26,39,40</sup> This approach carries the advantage of incorporating spatial information about

the physical location of dose extremes and does not rely solely on volume-averages (or gEUD) approaches. This reduces the risks of undervaluing the contribution of hot or cold spots.

## 5.4 *Biological*

The most popular class of biological variable found in literature today is related to genetic mutations, but frameworks can be augmented to include epigenetic and transcriptomic data—especially pertinent to long-term radiotherapy outcomes.<sup>41,42</sup>

### 5.4.1 *Genetic variables*

Radiotherapy efficacy and toxicity is a fine balance regulated by a number of biological variables and mechanisms. With the aim of maximizing prospective classification performance, modeling frameworks will need to incorporate several genetic variables. Known variables of biological data are related to genetic mutations. The role of epigenetics and transcript expression levels in long-term radiotherapy outcomes should also be taken into account.<sup>43–46</sup> Consider even that recent hypotheses have been raised about the role of the microbiota as a key player in cancer therapy response.<sup>47</sup>

For data-driven modeling, genetic parameters can be included as independent variables and regressed alongside other factors such as clinical risk and dose–volume metrics. For analytical models, genetic parameters are integrated in different ways; dose-modifying factors (DMFs) obtained from clinical risk factors can be used to stratify standard analytical models and generate “mixed” data-type model.<sup>48</sup> To maximize prospective classification performance, this approach can be expanded to include SNPs resulting in significantly improved classification performance.<sup>49</sup> This approach can be further extended using clinical risk factors for logit and EUD models.<sup>50</sup> Further to SNPs, CNVs can represent an effector of radiotherapy outcomes, and therefore, be included.<sup>27,51</sup>

***Augmenting Dose–Volume Models with Genetic Variables.*** Genetic features represent independent variables in data-driven modeling together with dose–volume metrics and clinical risk factors. DMFs obtained from risk factors can be used for standard analytical model stratification to generate mixed data-type models.<sup>52</sup> This approach can be further developed to include clinical risk factors for logit and EUD modeling.<sup>50</sup>

### 5.4.2 *Epigenetics and transcriptomics*

Additional biological variables include epigenetic alterations, with the extra complications of these including therapy intervention as well as an effector of the epigenetic landscape.<sup>53</sup> An epigenetic code is under development<sup>54</sup> and may facilitate the ability to perform and interpret the epigenome-wide association studies results and provide a new class of input data for outcome models.<sup>55</sup>

Even with supervised learning algorithms to preprocess the data, the number of messenger ribonucleic acid (mRNA) transcripts from a single microarray experiment can be challenging and often requires large-scale validations. In such a big-data environment, machine-learning techniques in artificial intelligence (AI) have the ability to process highly structured, high-dimensional data, and control for over- and underfitting at the same time.

### 5.4.3 *Next-generation data*

In the big-data era, numerous methods used to quantify large numbers of biological factors have been pioneered and introduced into mainstream biology research within the last decade. These technologies include well-characterized microarrays and proteomic analysis technologies that can quantify the levels of expression of up to tens of thousands of mRNA transcripts or proteins in a single sample.

After generating large quantities of data, high-throughput modeling frameworks can be used that are able to deal with large numbers of variables.<sup>56</sup> This approach has been used successfully

in clinical oncology to stratify tumor phenotypes and estimate prognoses to help guide optimal therapeutic regimens.<sup>57</sup>

## 5.5 Imaging

There are a variety of textures that can be extracted from medical images and used as correlates.<sup>58–61</sup> Perhaps the most common are qualitative features (semantic) that otherwise can be used to identify potential lesions. In contrast, quantitative features are algorithmically derived from images. These two classes of features can vary widely in complexity; however, both produce metrics that describe the shape and intensity of the voxel histogram and the spatial arrangement of the voxels themselves, i.e., textures.<sup>5</sup> Notably, features can be directly extracted from the images or extracted after transforming the raw imaging data.

There are several types of quantitative features that can be derived: shape features and statistics that are first order, second order, or higher order.

### 5.5.1 Shape features

These delineate the geometry of the regions of interest (ROI) in addition to other related properties such as volume, diameter, surface area, and sphericity. First-order statistics are features agnostic to spatial information and describe individual voxel values. Considering a summation of individual voxels represented as a histogram these might include, for instance, mean, min, max, median, skew, and kurtosis.

### 5.5.2 First-, second-, and higher-order statistics

These can also be referred to as textural features since they pertain distinctly to the relationship of voxels to one another rather than the individual voxels themselves.<sup>62</sup> They, therefore, provide information on the geometry or arrangement of voxels for use in describing intratumoral heterogeneity. Deriving second-order statistics can be done from the gray-level co-occurrence matrix<sup>63</sup> or the gray-level run-length matrix,<sup>64</sup> which both describe voxel intensity distributions across fixed directions. Beyond first- or second-order statistics, there are higher-order statistics, which overarchingly group features that are extracted after filtering or transforming images. Preprocessing of images may involve simple denoising or more complex transformations such as wavelet<sup>65</sup> or Laplacian.<sup>66</sup>

Given the above, it, therefore, becomes apparent that there exist many correlates that can be derived from any one single image. Importantly, however, a majority of extracted features will be redundant or interdependent.<sup>67</sup> As such, it is important to first identify endpoints that are clinically relevant to inform what correlates may be most relevant.

Radiomics does not stop here, however, as it may furthermore encompass the use of -omics data to augment the predictive value of such imaging correlates. Together, the integration of NGS data together with imaging correlates should provide a better opportunity to capture underlying biophysical effects of radiation therapy in an unbiased way. Techniques for the integration of -omics data with imaging correlates typically entail the use of MI approaches (see Sec. 6.4). Indeed, applications of panomics in the clinic are expected to be transformative.<sup>68,69</sup>

## 6 Modeling Strategies

### 6.1 Dose–Volume Approaches

Outcome modeling typically consists of the evaluation of dose to a given volume of tissue (dose–volume metrics). These parameters already can be used at the treatment planning stage and extracted from dose–volume histograms. However, because physiological changes are expected to occur during treatment as therapeutic effect and as morbidity, the delivered dose does not necessarily reflect biologically absorbed dose. To induce an improved predictive accuracy, intra-fractional CT scan changes can be incorporated.<sup>37</sup> Typically, to summarize dose distributions

across volumes, inhomogeneous dose distributions are modeled as EUD or gEUD if considering the normal tissue of interest.<sup>38,70</sup>

### 6.1.1 Analytical

Models of the analytical class, also known as mechanistic models, are based on theoretical mechanisms of action of radiobiological intervention. They include growing-in-complexity levels of mechanistic insight into a specific mechanism by which radiotherapy outcomes become evident.

**Linear-quadratic formalisms.** Nearly all analytical dose–volume TCP models make use of the well-known linear-quadratic (LQ) formulation for predicting cell-kill:

$$\text{SF} = e^{-\alpha D - \beta D^2}, \quad (3)$$

where SF is the surviving fraction for a given radiation fraction of size  $D$ . The coefficients  $\alpha$  and  $\beta$  relate to tissue-specificity of single- and dual-track cellular deactivation. More commonly, these coefficients are reported as a ratio ( $\alpha/\beta$ ). When formulated as per Eq. (3), the LQ model has been shown to prove valuable for predicting cellular responses *in vitro* and with particular qualities of radiation only. Consequently, reformulations to the canonical LQ model have been done extensively to take into account multifraction regimens, repopulation kinetics, high linear energy transfer radiation, increased dose rates, and stereotactic-body radiotherapy delivery, just to name a few.<sup>6</sup> Previous works can readily be consulted for further details of LQ formulations<sup>71</sup> but have, to a degree, fallen into disuse on account of data-driven approaches but also because normal tissue toxicities are dose-limiting in modern treatment regimens.

**Lyman–Kutcher–Burman.** The Lyman–Kutcher–Burman (LKB) model is the most used analytical method for predicting NTCP [Eqs. (4) and (5)]

$$\text{NTCP}(D, D_{50}, m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{\left(\frac{-u^2}{2}\right)} du, \quad (4)$$

$$t = \frac{D - TD_{50}(v)}{m \cdot TD_{50}(v)} \quad \text{and} \quad TD_{50}(v) = TD_{50}(1) \times v^{-n}, \quad (5)$$

where  $m$  is the slope of the best-fit NTCP sigmoid,  $TD_{50}(1)$  is the dose at which NTCP = 50% for a specific endpoint, and  $TD_{50}(v)$  is the tolerance dose for a given partial volume with tissue-specific volume exponent  $n$ . The canonical LKB model allows to stratify patient risk based in their EUD relative to the dose at which NTCP is 50% for a specific endpoint.

The canonical LKB model has been shown to be fairly consistent toward different data types and to be able to provide reliable first-order calculations. These calculations are straightforward and the model is able to obey the boundary conditions without additional constraints being needed.<sup>72</sup>

Despite its success in modeling some of the risk factors associated with radiotherapy, the LKB model still has room for improvement, for example, in the case of radiation pneumonitis (RP). Clinical evidence can show that almost all patients undergoing radiotherapy have an RP risk of <50%. However, it is likely that this is a strong underestimation of the actual risk of RP.<sup>49</sup>

**Binomial Models.** This class of analytical NTCP model arises by modeling tissues as composed of function subunits (FSUs). The arrangement of the FSUs in a given tissue is based partially on physiological modeling and partly on empirical evidence. The classical example is that of the gastrointestinal tract, wherein the tissue is considered to have a more parallel architecture such that irradiation of one subsection of the tract does not directly compromise the integrity of the crypt cells that received a subtherapeutic dose.<sup>73</sup> Conversely, the spinal cord is more serial in nature whereby irradiation of one section may compromise inferior compartments.<sup>4</sup> In this way, differences in response according to architecture can better be captured. The most often used formalization is the critical volume (CV) model

$$P_i = \binom{N}{t} P_{\text{FSU}}^t (1 - P_{\text{FSU}})^{N-t}, \quad (6)$$

where the first term of the right-hand side of the equation is the binomial coefficient for  $N$  and  $t$ ,  $P_{\text{FSU}}$  is the probability that  $t$  of  $N$  subunits will be deactivated by ionizing radiation,  $N$  is the total number of subunits, and  $t$  is the number deactivated. The use of the term CV arises since some tissues can be deactivated by a volumetric effect (as discussed). Using binomial statistics, which arise naturally from the consideration of radiation deactivation and the underlying tissue as more parallel or more serial, then the probability that  $M$  subunits will be deactivated can be calculated

$$P = \sum_{t=M+1}^N P_t = \sum_{t=M+1}^N \binom{N}{t} P_{\text{FSU}}^t (1 - P_{\text{FSU}})^{N-t}. \quad (7)$$

Importantly, the CV model and the FSU concept do not preclude tissues that may exhibit both parallel and serial effects to one degree or another. These arrangements of FSUs are termed complex, but may similarly be modeled using Eqs. (6) and (7).

Notably, the CV model has been shown to yield the LKB model with specific factors set as constant, suggesting the latter may be a specific case of the former.<sup>73</sup>

Before extensive use of data-driven modeling, CV models and the FSU concept were widely applied. For instance, applications in prostate cancer yielded some success to predict late toxicities, but were never widely clinically adopted. Indeed, while better incorporating tissue architecture, the CV model does not directly consider biological mutations that predispose individual patients to aberrant toxicities. Nor can CV models readily take advantage of NGS data.

### 6.1.2 Data-driven

Data-driven approaches, also known as phenomenological or statistical techniques, are based on empirical observations and are typically more robust than the analytical approach counterparts. It is not uncommon for this type of modeling to require preprocessing of data as the number of variables to take into account can be large.

**Regression-based approaches.** To fit TCP or NTCP, it is possible to use functions such as logit

$$\pi(X_i) = \Phi[g(X_i)], \quad (8)$$

and probit

$$\pi(X_i) = \frac{e^{g(X_i)}}{1 + e^{g(X_i)}} = \frac{1}{1 + e^{-g(X_i)}}, \quad (9)$$

often used in sequence. The mathematical simplicity of the logit function makes it the most used one.

Regression-based techniques represent to date the most frequently used approaches to data-driven modeling in radiotherapy. Regression link functions are typically sigmoidal to achieve the nonlinear dose-responses seen experimentally and advanced methods in AI that are able to handle nonlinear data complexity more readily are becoming increasingly popular due to superior prospective classification performance in many areas of medicine.<sup>8,74</sup>

The main problem that could arise with phenomenological multivariable models is that many different conceivable models can be developed to be consistent with the used data set. These can describe the present data accurately, but may later turn out to be inconsistent with new data sets.<sup>75</sup>

Overfitting can occur when a model is fitted to a dataset in such specific detail that the result loses its general validity for different datasets.<sup>76</sup> Furthermore, multiple alternative phenomenological models exist and choosing which one and why is more appropriate can be challenging and not theoretically solid. This induces a model selection uncertainty or instability.<sup>28</sup> More generally, it is challenging to guarantee how good the model predictions will be without making

strong assumptions; models that rely on variations and relations of the training data cannot guarantee accuracy in subsequent datasets with a different statistical structure. Statistically, a correlation does not always imply causality. As such, it is important to note that phenomenological models are, therefore, not strictly guaranteed to describe causal relationships.

**Kernel-based methods.** As an alternative to regression-based methods, kernels can be applied to classification problems that depend on the deconvolution of nonlinear datasets. Kernels operations can be applied to generate classifiers such as hyperplanes that differentiate (classify) higher-dimensional datasets accordingly. In radiology, kernels are particularly useful given the nonlinear interaction of input data. Mechanistically, they attempt to maximize the distances between groups (clusters) and so kernels can be considered as extensions to Fischer linear discriminants or principal component analysis (PCA).<sup>77,78</sup>

PCA is a well-known approach used widely both within and outside of radiology and so further details can be found in previous works.<sup>62,79</sup>

Support-vector machines (SVMs) are the most often encountered kernel variant.<sup>80</sup> An SVM seeks to redefine classification steps into maximization problems through the use of quadratic programming. Using such an approach, a computationally inexpensive SVM can be formulated as follows:

$$f(X) = \sum_{i=1}^{n_i} \alpha_i y_i K(S_i, X) + a_o, \quad (10)$$

wherein  $n$  is the number of support vectors,  $K$  is the kernel transformation, and  $\alpha_i$  is the coefficient that is to be optimized by quadratic programming. SVMs are, therefore, nonparametric and are robust tools for higher-dimensional data classification.

## 6.2 Radiogenomics

In data-driven modeling, genetic parameters can be considered as independent variables and, for example, regressed alongside clinical risk factors and dose–volume metrics. For analytical models, genetic parameters can be integrated using different model-dependent methods.

### 6.2.1 Augmented analytical models

Dose and volume-based TCP or NTCP models can be augmented by the integration of biological variables to enhance their prediction performance. The most frequently used example for this is the LKB model, which can be modified to include dichotomous clinical or genetic risk factors using  $\delta_k$  as DMF

$$\text{NTCP}(D, D_{50}, m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{\left(\frac{-u^2}{2}\right)} du, \quad (11)$$

$$t = \frac{D - TD_{50}(v) \cdot e^{\delta_1 R_1} \cdot e^{\delta_2 R_2} \cdot \dots \cdot e^{\delta_k R_k}}{m \cdot TD_{50}(v) \cdot e^{\delta_1 R_1} \cdot e^{\delta_2 R_2} \cdot \dots \cdot e^{\delta_k R_k}}, \quad (12)$$

where  $\delta_k$  is the weighted risk-factor coefficient for risk-factor  $k$ . The  $R_k$  parameter is binary and indicates the presence ( $R_k = 1$ ) or absence ( $R_k = 0$ ) of the risk-factor for a given patient.

In our previous work, we used a modified LKB model to integrate clinical risk factors together with SNPs and CNVs for predicting late effects of hypofractionated prostate radiotherapy.<sup>27</sup> We found there to be benefit in the inclusion of select deoxyribose nucleic acid (DNA)-repair associated genetic mutations, but no clinical parameters improve the fit of either severe late rectal bleeding (RB) or late erectile dysfunction models. Similarly, a separate group demonstrated the value of including a clinical risk factor related to abominable surgery (yes/no) for the prediction of late RB toxicity in prostate cancer patients that underwent surgery.<sup>81</sup>

Imaging information can similarly be used as per the above examples. For example, in the context of patients with hepatocellular carcinoma having undergone liver irradiation, the

addition of imaging information related to perfusion of their portal veins was found to improve the classification performance of the dosimetric LKB model for predicting severe (grade  $\geq 3$ ) enzymatic changes.<sup>82</sup> Such parallel approaches are often overlooked in favor of more complex modeling strategies, but are recommended in the first instance since they may be powerful yet robust. Indeed, understanding of the physiology at-hand is crucial for their coherence and it is possible that data-driven approaches may eventually lead to the identification of similar models but at the cost of time and effort.

While the above approaches have provided benefit in the context of statistical validation, we found the robustness of augmented analytical models to be reduced when compared directly with data-driven models of similar complexity.<sup>27</sup>

### 6.2.2 Data-driven

Data-driven radiogenomic models function similarly to data-driven dose–volume models – see dose-volume-based approaches, data-driven for further details. They are most often based on regression (Fig. 1) so as to be able to handle almost any type of data, but kernel approaches can also be used. Regression frameworks allow stratification of risk factors through the use of dichotomous labels. For instance, the presence or absence of a genetic mutation for each patient can be incorporated into the input data as a 1 or 0, respectively.<sup>83</sup> There is furthermore no limit to the number of such factors that can be included.

### 6.2.3 Application of radiogenomics

Radiogenomic modeling consists of diverse approaches, making radiogenomic specific techniques difficult to define beyond the common integration of dosimetric, clinical, and biological data. We suggest to readers *a priori* work in which we directly compared the improvement in prospective classification achieved for predicting late side effects of prostate radiotherapy using dosimetry versus dosimetry plus common genetic risk factors.<sup>27</sup> We also compared the improvement in regression-based models' classification performance with those in quasianalytical models, such as the LKB, which thus provides an elegant introduction to the core principles of radiogenomics (dosimetric models, radiogenomic models, data-driven techniques, and analytical approaches).

An example of radiogenomics in the clinical setting is one that sought to identify predictive biomarkers of radiosensitivity in breast cancer patients toward classifying patients that are most likely to benefit from radiotherapy.<sup>84</sup> The authors developed a signature for radiotherapy benefit based on two distinct biological processes: antigen processing/presentation (immune component) and intrinsic radiosensitivity. Using a validation cohort of 1439 patients, they found that radiation-sensitive patients who did not receive radiotherapy had a worse prognosis than otherwise. Patients in the immune-effective group that did receive radiotherapy had better disease-specific survival. To extend this study, the authors could have included treatment-related metrics in their analysis *vis-à-vis* radiotherapy treatment plan information; however, it is noted that there are relatively few critical structures adjacent to their target volumes and so may be unlikely to provide addition insight.

## 6.3 Machine Learning for Radiomics

Imaging-based approaches for outcome modeling typically take advantage of AI and ML techniques, which are able to mimic selected human behaviors to a degree, but must first be trained to be able to learn the patterns of interest.<sup>61,85</sup> After training, such frameworks need to be tested on recognizing the pattern in a prospective setting (testing phase). Broadly speaking, there are two categories of frameworks that can be used: supervised (with labeled outcomes) and unsupervised (without labeled outcomes)—typically it is considered that validation is crucial for either approach but more so for the former.<sup>86</sup> The use in oncology of conventional ML techniques, such as artificial neural networks (ANNs), has been successful<sup>87–91</sup> but remains very much in its infancy in light of its potential.

### 6.3.1 Conventional

Classical MI techniques in radiomics typically follow a top-down approach,<sup>92</sup> where knowledge of interactions of radiation with tissue and biological systems is often ignored and radiotherapy outcomes are modeled through complex statistical analysis. For this reason, supervised learning would be preferred, i.e., models aimed at constructing a genotype–phenotype relationship by learning such genetic patterns from a labeled set of training examples. Nevertheless, an unsupervised approach (e.g., PCA or clustering) is sometimes used to reduce the dimensionality of datasets, extract a subset of relevant features, or construct features to be later included in the chosen learning method.<sup>93</sup> The best way to model large datasets such as those containing genomic information is through support vector machines (SVMs).<sup>80</sup> SVMs are tailored to identify a potentially complex unknown correlation structure between the phenotyping outcome and high-dimensional genotyping data.

Feedforward neural networks (FFNNs) are used frequently to train AI to basic pattern recognition problems. FFNNs do not include any recurring nodal inputs (memories) and are fully defined by their architecture such that arrangements of nodes into different topologies can induce different system responses. FFNNs have been used in radiation oncology due to their robustness.<sup>94,95</sup> Typical activation functions are logit or probit functions, but also the radial basis function. Outputs from nodes are stratified by weights. Weights are the elements of the FFNN that are trained when building an FFNN. To adjust node input and output weights, the *delta rule* can be used via backpropagation until classification performance is optimized.<sup>96</sup> Datasets for training can be used all at once (batch) or can be segregated into pattern-based subgroups (sequential).

### 6.3.2 Deep learning

Deep learning can be contrasted with conventional learning by the fact that it does not require engineered features to be used as input to the framework.<sup>97</sup> In other words, if found to be valuable for optimizing the objective, such inferences or associations can be synthetically derived by the framework itself and used to improve classification performance without declaring them *a priori*. In terms of probabilities, deep learning strategies can be categorized as either generative, whereby the joint probability  $(x, y)$  is used to estimate the posterior probability of  $(x, y)$ , or as discriminative, wherein estimates of the posterior probability are made without calculating the joint distribution (direct mapping).

The decision to use one or the other is contingent upon the nature of the question (objective) of the framework itself. If the classification of data according to specific labels is all that is required, then discriminative approaches may be most relevant. However, if the dataset requires manipulation (e.g., is incomplete), then a generative approach would be indicated to manipulate the data into an acceptable form. Moreover, generative models, as the name suggests, allow higher-order parameters to be synthesized and used as input to the framework, or used to capture internal relationships of variables within the dataset. Extending the above to radiomics specifically, an example of discriminative learning may be the categorization of a volume as benign or malignant while the evaluation of intratumoural heterogeneity may warrant the use of generative strategies.

**Discriminative deep learning.** Convolutional neural networks (CNNs) are the most common type of network architecture in use today for image-intense processing,<sup>98</sup> which is likely a consequence of increased accessibility to platforms that can generate such models (e.g., MATLAB®). They are used preferentially for the processing of especially large images.<sup>99</sup> CNNs use local weights to control adjacent connections and then pooling to generate features that are spatially invariant. An important characteristic of CNNs is a fewer number of parameters that are trainable compared with an equal-sized ANN. Local filters can be used to further reduce the number of independent weights that are required. Transfer learning can be used to train a many-layered CNN to a rough approximation either to extract features or as a basis for an entire modeling process.<sup>100</sup> As is the case with conventional learning, CNNs can make use of supervised or unsupervised methods (or a mix).

Another class of discriminative deep learning model is the autoencoder. Autoencoders are unsupervised neural networks that reconstruct underlying data through identification of intrinsic relationships between the input variables.<sup>101</sup> Consequently, autoencoders reduce the dimensionality of input to a lower dimensional representation than is then input. When used alongside supervised frameworks, autoencoders offer a robust method to parse incomplete datasets. Similarly, encoders offer a way to resolve issues pertaining to sparseness when variance within training data is significantly increased compared with testing. Autoencoders can also be used in various architectures for classification or denoising of image data. We discuss in the next section a method for the use of autoencoders as generative models.

**Generative deep learning.** Variants of autoencoders can be used in the context of generative models (deep encoders) and have only recently found applications in feature extraction-based frameworks as well as for segmentation. The variational autoencoder consists of an autoencoder that produces latent vectors in a Gaussian distribution. The loss function includes both the Kullback–Leibler (KL) divergence between the latent vector and the Gaussian distribution as well as the squared error (mean) between the input and output. Other variants of deep autoencoders include those that make use of Bayesian methods for probability distribution that represents the data in question or convolutional autoencoders for the preservation of spatial locality, both of which are discussed in further detail in previous work.<sup>102</sup>

Another type of generative neural network is the deep belief network (DBN), which is several layers deep and consists of stochastic and latent variables.<sup>103</sup> Every internal layer of a DBN serves as a hidden layer for the proceeding layer and as input to the successive layer. Consequently, a DBN can also be defined as an unsupervised rendition of a restricted Boltzman machine (RBM) or autoencoder. Importantly, DBNs are trained in a greedy manner using input from each previous layer. Relatedly, RBMs use both hidden and visible layers with forward passes learning the activation probability and backward passes the probabilities of inputs according to activations. In contrast to autoencoders, RBMs use stochastic units according to specific distributions rather than deterministic units. In other words, RBMs estimate joint probability distributions for inputs and activations and DBNs can, therefore, be defined as a stack of RBMs. In radiomics, labeled datapoints are often unavailable, and therefore, a combination of supervised and unsupervised techniques is often selected. For instance, a supervised network can be trained with the loss component estimated from unsupervised models such as RBMs.

Reinforcement learning (RL) is defined by an algorithm that seeks to maximize defined criteria leading to a reward given specific tradeoffs.<sup>102,104</sup> The requirements for the use of RL in the context of deep learning include adequate knowledge of the environment, a defined reward, a defined value function, and governing policies (rules). In other words, RL approaches are goal oriented. For instance, RL approaches could be used to maximize the local control probability for a patient weighing the relevant risks of aberrant normal tissue damage. In this sense, RL learning provides an avenue for the integration of TCP and NTCP models. The use of RL in the context of deep learning has been used successfully to a degree for imaging-based landmark detection and, to a more limited degree, for radiotherapy treatment response predictions.<sup>105</sup> While the case for the use of RL in the context of landmark detection in imaging is relatively straightforward provided one has an adequate training set, radiotherapy responses are an intrinsically more complex, multiscale phenomenon subject to inter- and inpatient heterogeneity. Attempts to improve response predictions in radiomics, therefore, may necessitate the use of multiomics strategies, for which deep RL may provide benefit.

### 6.3.3 Applications of radiomics

We briefly consider here three examples of radiomics applied clinically. The first example by Wu et al. sought to identify breast cancer subtypes.<sup>106</sup> The authors retrospectively analyzed dynamic contrast-enhanced MRI data of breast cancer patients at their institution and extracted features from their dataset. Using an outside, multiinstitutional validation cohort, a gene expression-based classifier of imaging subtypes was developed and further tested against publicly available datasets. The result was a three-way classifier for stratifying recurrence-free survival of breast

cancer patients. Indeed, in such a case, radiomic analysis would likely prove complementary to classical clinical histologic/molecular subtype categorization.

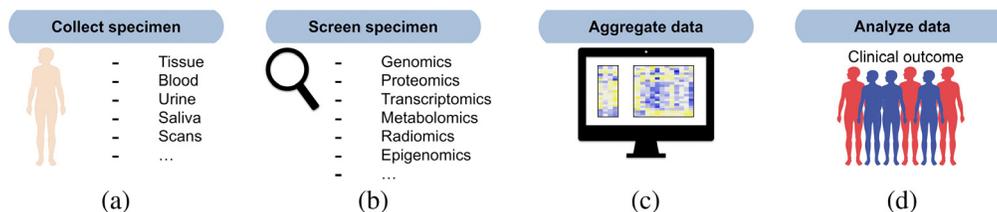
The second, more recent example by Jiang et al. developed a noninvasive radiomic signature predictive of gastric cancer ImmunoScore, which is a classifier of 27 immune cell features shown previously to effectively predict reoccurrence.<sup>107,108</sup> Using four independent cohorts consisting of  $N = 1778$  patients, the authors extracted  $>500$  quantitative features from contrast enhanced CT images. To correlate imaging features/metrics with the ImmunoScore results, a logistic regression model was then used. The result was a relatively robust signature of 13 imaging features that was found to strongly correlate with ImmunoScore, thereby providing a noninvasive surrogate marker.

The third example we discuss using radiomics incorporates deep learning for the prediction of survival in patients with glioblastoma.<sup>109</sup> In this work, the authors used transfer learning to extract deep features and thereupon build signatures for survival. Using nearly 100,000 deep features alongside a handful of handcrafted features extracted from preoperative multimodality MR images, a six-deep-feature was constructed using least absolute shrinkage and selection operator (LASSO) Cox regression. The proposed signature was found to have significantly better classification performance of patients (prognostically). At the same time, the work demonstrated the feasibility of using transfer learning with deep features, a now frequently used technique for clinical radiomic studies.<sup>110-112</sup>

## 6.4 Panomics

The term panomics encompasses the use and integration of datasets relating to the multiomics technologies, such as genomics and proteomics,<sup>12</sup> but also for our purposes here includes imaging correlates, clinical data, and data produced from other NGS platforms not specifically mentioned (Fig. 3). Given the heterogeneity of such datasets, their potential sparseness, and inconsistencies in their constituent data (missing values), working with them can pose unique challenges but also provide valuable insight. The favored approach for leveraging panomics datasets is, for now, one of systems biology.<sup>113</sup>

A systems biology approach recognizes that radiotherapy occurs at the intersection of physical processes and biological and clinical effects, each of which can take place over very different timescales. For instance, atomic-level effects of free radicals induced by ionizing radiation through to clinical scoring of late radiotherapy responses can span  $10^{-9}$  to  $10^7$  s, for which the entirety of events are unlikely to adequately captured by any one single approach.<sup>114</sup> Radiotherapy is also data intensive in that it encompasses functional imaging data and requires anatomical information, which, when coupled with biological markers from tumor and peripheral bloods, lends itself to the use of informatics approaches that can cope with high-dimensionality and heterogeneity.<sup>115</sup>



**Fig. 3** Workflow for panomics modeling. (a) Significant data can be generated in the clinical setting for patients undergoing radiotherapy through the careful analysis of specimens collected from certain sites and at certain times (pre/posttreatment). (b) Specimens can be used to generate data such as transcriptomics, genomics, proteomics, metabolomics, imaging correlates, and more. These datasets are usually high-dimensional and generated through NGS techniques. (c) Together with dosimetry and other clinical risk factors used in radiotherapy, datasets are preprocessed (normalization, filtering, for instance) and (d) applied to modeling of treatment outcomes (e.g., tumor response, toxicity).

### 6.4.1 Systems biology

The use of systems biology approaches for prediction of radiotherapy outcomes has to date been limited, but as discussed, remains a promising avenue for future works that make use of panomics datasets. Here, we review strategies for the integration of unique datasets but recognize that this list is by no means complete. In truth, the strategies adopted for working with panomics datasets are likely to vary to a degree between applications owing to the specific differences of each dataset at hand; however, several approaches described below can be used in the first instance.

**Graph networks.** In the case of the integration of proteomic and genetic data, graphical networks can be used wherein nodes represent genes/proteins and network edges their interactions. Such approaches make use of Bayesian networks, which our team has previously applied to the prediction of radiation-induced pneumonitis.

**Similarity network fusion.** The similarity network fusion (SNF) approach uses a network fusion method to integrate multiomics datasets (align datasets).<sup>116,117</sup> In the first instance, an individual network is created for each dataset and then a nonlinear fusion network is used to combine them. The resulting network is coined a similarity network and the step by which the networks are combined (fusion step) uses belief propagation to accomplish this.<sup>118,119</sup> The advantage of using an SNF approach is that noise can readily be identified as specific to each dataset and dissolves through the integration of the networks. On the other hand, true signals that are identified are propagated through the similarity network.

**Joint Bayesian factor.** This approach is nonparametric and uses joint Bayesian factors (JBFs) to integrate multiple data sets. Using this approach, the entire feature space is factorized into shared and unique components using a beta-Bernoulli process.<sup>120,121</sup> The product is a joint factor model with two constituents: individual factors unique to each dataset and shared features, both of which can include noise (noise specific to a dataset and noise common to multiple datasets).<sup>122</sup> Relative sparseness of each dataset can be evaluated and included in the analysis using a Student's *t*-test sparseness factor. The advantage of using a JBF approach to integrate datasets is that it allows identification of a subset of factors that maximizes the overall objective function of the framework in an unbiased way.

**Multiple coinertia analysis (MCIA).** MCIA attempts to capture relationships across high-dimensional datasets and has previously been used for gene expression and transcriptomics.<sup>123,124</sup> One advantage of MCIA is that features do not need to necessarily be present for all subjects; however, the number of samples in each dataset must match. Covariance optimization criterion is used to transform the individual datasets and normalize them to the same scale. The datasets are projected onto the same unit space, which can then be searched. Applications of graph theory can facilitate this process and allow the extraction of heterogeneous features that are partial members of multiple datasets.<sup>125</sup>

### 6.4.2 Challenges specific to panomics

The heterogeneity of large panomics datasets, their size, the requirement for intensive computational power, and also the lack of clarity as to the most appropriate methodology for their analysis makes working with such datasets challenging.<sup>12,126</sup> Indeed, efforts have been made to resolve these challenges, but few have yet been widely adopted. For instance, the use of specific dataset formats for analysis is critical and *Feature X Sample* matrices have emerged as candidates for this purpose,<sup>127</sup> but successive steps in processing panomics datasets remain unclear. Indeed, preprocessing and batch effect interrogation is paramount to the reliability of input data, and therefore, to the entire framework. Further processing prior to modeling includes normalization, filtering, and quality assurance, any one of which may perturb or obscure the underlying signal being sought. In a broader sense, panomic analyses must first strive to use the appropriate datasets and formats as they pertain to the question at hand. For example, it is straightforward to normalize datasets that originate from very different sources, but it is more difficult to take into

account unequivocal outcomes or assay readouts.<sup>128</sup> Moreover, interpretation of clinical information in the context of panomic datasets poses a unique challenge such that clinical metadata may be less easily quantified compared with NGS-generated data, both of which may change over time.

## 6.5 Evaluation of Model Performance

There is a diverse list of techniques that can be used to evaluate the performance of a model prior to independent validation on unseen data, which is typically the final step in a modeling framework. The techniques discussed herein are some of the most commonly employed methods but are by no means exhaustive. While they classically are applied to non-MI techniques, they can occasionally be found integrated into MI-based frameworks throughout radiology.<sup>39,129,130</sup>

### 6.5.1 Estimation of error, optimal parameters, and model order

**Statistical resampling.** Resampling can be used to estimate optimal model order, best-fitting parameters, and/or error.<sup>29</sup> Regardless of when it is used, statistical resampling divides a dataset into smaller sets in the first instance. When done with replacement, this can be repeated to produce a much larger pseudoset that can then be used to provide estimations of parameters without having to fully elucidate the true underlying distributions of the individual parameters themselves. The method is, therefore, nonparametric. Certain variations can be extremely quick as well, serving as easy-to-implement internal validations throughout the modeling process.

**Jack-Knifing.** Jack-knifing is a special case of cross-validation, whereby the first datapoint (e.g., patient) is left out and modeling is performed on the remaining cohort ( $N-1$ ). The datapoint is then replaced and the second datapoint is removed and a new model is fit using all but the second datapoint. This is repeated until each successive datapoint is removed, generating a new model at each turn for a total of  $N$ . By examining the resulting models, insight into the robustness of the best-fit model obtained using all of the training data can then be analyzed. Jack-knifing is quick, but provides only an estimate of the more computationally intensive bootstrap technique. A more general version is to use  $k$ -fold cross validation.

**Bootstrapping.** Similar to jack-knifing, bootstrapping is a resampling technique done with replacement that can provide estimation of parameters, model orders, and errors. In contrast to jack-knifing, however, bootstrapping randomly resamples a select number of parameters. These subgroups can then be used to generate estimates of interest—a technique known as bagging. By analyzing their relative variability or averaging them, estimates for best-fit values can be produced. As a resampling technique, bootstrapping is nonparametric and especially finds use when analytical estimation of errors is not feasible. However, it should be noted that bootstrapping assumes independence of each datapoint from one another—this is usually not an issue for outcomes modeling given the independence of each patient from one another.

**Applied information theory.** In contrast to resampling techniques, information theory can be used to provide insight into the balance of model complexity and its fit. These approaches are quick but do not provide insight into the quality of a model nor of its interpretability.<sup>131</sup>

#### *Akaike information criterion (AIC)*

The AIC is a technique based on goodness of fit and penalizes a model for underfitting or over-fitting

$$\text{AIC} = 2k - \ln(L), \quad (13)$$

where the parameter  $k$  is the number of variables in a model and  $L$  is the log-likelihood. By using the KL distance to quantify the relative proximity between the two real and predicted probability distributions, an estimate of the divergence can be made. The optimal AIC value is found by minimizing the log-likelihood term on the right-hand side. AIC rewards models with fewer parameters that can adequately explain the data over those with increased complexity. The

$2k$  is the penalty factor, which increases the AIC and whose contribution to the AIC cannot be overcome with better fit (lower log-likelihood). A drawback of the AIC is that it fails to perform adequately when  $k$  is large and numerous comparisons are being made. For instance, in the context of radiomics models that use NGS data, AIC would likely not be suitable for the evaluation of goodness-of-fit.

*Bayesian information criterion (BIC)*

The BIC is a method similar to that of the AIC but differs in the arrangement of the constituent variables<sup>132</sup> and furthermore includes a parameter  $n$ , which is the number of datapoints contained within the input data (sample size)

$$\text{BIC} = k \cdot \ln(n) - 2 \cdot \ln(L). \quad (14)$$

The penalty term of the BIC is larger than that of the AIC and so penalizes overfitting more.

## 6.5.2 Model performance evaluation

There exist numerous methods in literature to evaluate the ability of a given model to classify data in a prospective sense. Oftentimes, frameworks will employ more than one validation technique to explore the shortcomings of outputted models.

*Validation coefficients and metrics.* Metrics and coefficients are the most readily available tools for calculating the prediction or classification performance of outcome models. Their simplicity is amenable to quick understanding of model behavior and, when several are used together, can yield insightful information. The linear Pearson's correlation is an example of a nonparametric coefficient that is used frequently for estimating the linearity of a relationship between two variables. More often employed in outcome models is the Spearman rank coefficient, which does not assume linearity and instead yields an estimate on the direction of trend between two parameters. Alternatively, receiver-operating characteristic (ROC) values can be summed from ROC plots to readily convey classification performance alongside sensitivity and specificity for the desired classification cut-off value. The selection of such cutpoints can be optimized using Youden indices, for instance.

*Cross-validation by resampling.* Resampling with replacement can be used to quantify the classification performance of models as well as estimate confidence intervals on model performance or provide estimates on the error of classification statistics. In our experience, leave-one-out cross-validation on finalized models serves as an excellent method to quickly estimate how robust a given model is without having to rely on more computationally expensive methods, such as bootstrapping.

## 7 Common Challenges and Pitfalls in Radiogenomics

Radiogenomics modeling requires the accurate processing, analysis, integration, and interpretation of large data sets, often simultaneously. Therefore, there are many challenges that need to be addressed to avoid errors and their propagation. Each set of data can involve a wide range of data types, including nominal, ordinal, and quantitative data. As a result, radiogenomic analysis demands efficient and objective dissection of image features (radiomics) so that relevant features and information can be extracted reliably and measured.<sup>133,134</sup> Current challenges include AI interpretation often compared with the expert opinion and image interpretation.<sup>135</sup> Therefore, the understanding and representation of each input data type is of great importance to be able to perform any accurate analysis.

### 7.1 Curse of Dimensionality

The curse of dimensionality refers to the difficulty associated with analyzing and characterizing large data sets where potential data space increases nonlinearly with the number of dimensions (exponentially increasing sparsity).<sup>136</sup> This is a difficult to avoid cost associated with the ability

of working with large datasets and can lead to the overgeneralization of results from the unjustified application of certain methods. For this reason, appropriate methods must be applied with judicious interpretation of the results. For example, the false discovery rate represents the expected rate of the inappropriately rejected null hypothesis (the expected proportion of false-positive results) and is frequently used in genomic analyses. The most commonly recognized challenge in the analysis of large data sets is the need for multiple hypothesis testing corrections, of which the Hochberg–Bonferroni  $P$  value correction is the simplest and perhaps the most well-known.<sup>137</sup>

The challenges of utilizing a large number of variables in outcome models are well summarized by the multiple testing dilemma: too few samples relative to a large number of variables being tested can lead to spurious correlations. Even after utilizing simple supervised learning algorithms to preprocess the data, the number of mRNA transcripts that a single microarray experiment can yield is often in the thousands.<sup>81</sup> This issue can be mitigated by large-scale validation studies, but these are expensive, time-consuming, and patient accrual can limit achieving the necessary sample size.

Alternatively, methods in AI are becoming increasingly popular to explore the complex, hidden relationships between outcomes and biological variables.<sup>82</sup> In contrast to brute-force estimating of correlations, machine-learning techniques have the ability to process highly structured, high-dimensional data while controlling for over- and underfitting by drawing on methods from control, probability, and information theory. Outside the applied field of radiomics, one group has provided a mathematical formulation to guide the detection of dimensionality anomalies as they are applied to deep learning and can be visited for further details.<sup>138</sup>

An important point is that neural networks and especially deep network learning strategies may be impervious to the curse of dimensionality by virtue of their architecture.<sup>139</sup> Instead, they may suffer from challenges relating to dimensionality reduction during the data representation process.

## 7.2 Dimensionality Reduction

While complex methods are often needed by large data sets analysis, there is often a dangerous temptation to overfit the data to inordinately complex or overly parameterized models.<sup>140</sup> The most common statistical tests assume normal distributions for measured variables. Despite this possibly being a fair assumption, nonparametric approaches can be used when the variables are being analyzed as an alternative. These include Bayesian-based approaches, which can prove particularly useful, given that clinical data are often complex and nonlinearly distributed. It is of paramount importance to note that just because a modeling calculation can be performed does not mean that it should be performed, and that underlying assumptions for use of the methods always need to be satisfied by the data to which they are being applied. We again note that validation is always the most important test to avoid under- or overfitting.

## 7.3 Data Preparation

Preprocessing of input data for radiomics frameworks can be critical, especially in the context of MI and panomics techniques.<sup>141</sup> Data preparation can have a significant effect on the resulting classification result and so cannot be taken for granted. During this phase, several steps must be considered, which may include randomization, controls for under/oversampling, discretization of continuous data, scaling of features, and so on. As they pertain to all datasets, we focus our discussion here on randomization, feature scaling, and class balancing (sampling artifacts).

Features that are extracted from images during radiomic modeling can have very different scales, which, if not properly considered, may preclude interrogation of a subset of the feature space in any reasonable amount of time and furthermore risks destabilizing the framework. Specific examples of this include the weights in neural networks, which should be tuned to common scales across all nodes. By using a shared numerical scale, distortion of distributions through the networks can be avoided. It is, however, important to distinguish normalization from standardization. The former specifically refers to a scaling of between 0 and 1 while the latter is typically used to refer to  $Z$ -scoring of a dataset, that is, setting the mean to 0 and the standard

deviation to 1. Ultimately, the choice of one or the other (or both) is highly context dependent and should be considered ahead of time.

Randomization is similarly important for modeling in radiomics since the overall performance of the algorithm may have local optimums that can be mistaken for global extrema. These instances are typically straightforward to overcome using pseudorandom seeding factors or start points for initiation. When coupled with normalization, the scales of parameters within a framework can be contained to within reasonable values. In the event that randomization techniques are not applied at the outset of modeling, certain motifs within the dataset may overwhelmingly influence the resulting model.

Class balancing is critical to ensure that the resulting model does not overtly misrepresent the signal of interest vis-à-vis under- or overfitting. Resampling techniques can resolve this challenge. In the context of radiogenomics, bootstrapping and jack-knifing provide quick and robust methods to achieve this, but for MI strategies, solutions can be computationally expensive and more complex. The preferred technique for class balancing with MI approaches is the synthetic minority oversampling technique (SMOTE). SMOTE generates subsets of underrepresented data with added variances to facilitate interrogation of what effect their increased representation may have on the model in question. A more recent version of SMOTE, coined adaptive synthetic sampling approach for imbalanced learning (ADASYN), also exists specifically for the learning from imbalanced datasets.<sup>142</sup> In addition, generative adversarial network (GAN)-based approaches are becoming more widely used for generating synthetic compensations for such imbalances.

#### 7.4 Roadblocks to Translation and Explainable Artificial Intelligence

A limitation in the ability to broadly apply radiomics/radiogenomics to the wider field of radiology/oncology is translatability. Without a comprehensive understanding of how and why an algorithm performs a classification, hesitation to adopt it as a standard approach is likely to be very real and enduring. Interpretability of deep learning algorithms in particular is often questioned and, indeed, many have shown how such networks can be tricked into misclassifying relatively straightforward cases. For radiomics, such misclassification could eventually be lethal. Oftentimes, deep learning strategies are referred to as black boxes; however, this is likely unhelpful and probably dissuades many seeking to understanding the reasoning behind certain decisions that are made. Some efforts to tackle this problem—that is, to open the black box—have, however, yielded some success in recent years. For instance, through techniques such as deconvolution networks, network inversions, and activation maximization, we increase our understanding of how relationships among variables are managed internally. Nonetheless, translatability of deep learning strategies remains a challenge and an area of active research may be critical for widespread adoption. From a different perspective, ML approaches can broadly be classified into interpretable and noninterpretable.<sup>135</sup> While techniques of the former are conducive to understanding, and therefore, adoption, their performance is often limited, necessitating the use of the more abstruse (noninterpretable) methods. One of the most active cores researching explainable artificial intelligence (XAI) is at the Defense Advanced Research Projects Agency (DARPA).<sup>143–145</sup> In this context, explainable rather than interpretable AI was chosen to highlight their desire to maximize human interactivity of their platforms. Medical XAI (mXAI) is a more recent development with a similar objective of opening the black box.<sup>146,147</sup> In turn, mXAI seeks to provide discrete, high-level accountability, and therefore, transparency of deep learning algorithms in the medical sciences.

#### 7.5 Rare Variants Role

The inclusion of biological risk factors can generate challenges for outcome modeling, regardless of their value for classification performance.<sup>148</sup> In the case of variants that are uncommon across a dataset but valuable for classification purposes, an imbalance can manifest itself during the modeling process if identified as important factors during the modeling process. This may be especially true when dichotomous biological risk factors are included within a framework as they risk overperturbing the real increase in risk for the variant (0% or 100% additional risk). By definition, deleterious genetic mutations are rare variants when compared with the much larger

number of noncoding and innocuous mutations. The role of rare variants is particularly challenging to study because MI approaches are often used to try to identify them, but they would ideally be known *a priori* and compensated for once the relevant pattern has been identified (chicken and egg). Currently, no techniques are widely adopted to take into account challenges associated with rare variants.

### 7.6 Echo Chamber Effect

The echo chamber effect occurs when working with large datasets, such as for radiomics, when an internal relationship in the data is cyclically amplified through the process of data aggregation and is, therefore, considered a variation of selection bias such that the training dataset is not representative of the population.<sup>149,150</sup> An example is a meta-analysis, which, when performed in the context of biomarkers, has an implicitly lower probability of a negative result and thus deviates from real-world conditions.<sup>151</sup>

The opposite of the echo chamber effect is the Yule-Simpson paradox whereby true associations are identified and reported in smaller cohorts, but are lost in larger datasets.<sup>152</sup> Example and issuing controversies from real-life instances of both the above are described in our previous work.<sup>149</sup>

### 7.7 Data Injustices (Bias and Mitigation)

Biased decision making is not a trait inherently unique to MI nor to radiomics, but is none the less a real risk.<sup>153</sup> Dissecting the origin of biases that are observed in testing and validation is likely to point toward unrepresentative biases within the training dataset itself. Unfortunately, as a result, cross-validation techniques and resampling methods discussed earlier cannot overcome these hurdles and independent validation on unseen data is usually required. Within and outside the field of radiomics, these training data biases can occur due to geography of cohort data or any other number of ways, but similarly converge to produce suboptimal models, especially in the context of learning algorithms with memory. Injustices within datasets differ from small sample sizes by passing internal validation and potentially even on unseen data. Rather, they are more cryptic, and therefore, more challenging to resolve. For example, geographical biases may overexpose a training dataset toward specific ethnic or racial backgrounds, and therefore, be unapplicable on a wider population. These challenges can similarly be institutional or social. One method to overcome this challenge is the annotation of the methodology used for the aggregation of datasets in questions. This produces yet another challenge for standardization of, for example, experimental protocols, and may be challenging if identifiable patient metrics are necessitated. An alternative strategy is the use of dedicated MI algorithms to seek out biases unknown to operators. This could either be interpreted or fed directly into the core modeling framework to systemically audit the underlying strategy and maximize its applicability.

## 8 Conclusion and Perspectives

In this work, we discuss two complementary but distinct strategies for predicting radiotherapy outcomes: radiogenomics and radiomics. Both techniques take advantage of biological variables to augment dosimetric risk factors; however, radiogenomics entails the integration of biological, dosimetric, and clinical factors whereas we considered radiomics to further include the use of imaging correlates. As discussed, radiomics quickly necessitates the use of MI strategies for implementation since it is not evident which textures or features may be relevant to a given outcome (nor even in which dimensions they reside). Conversely, radiogenomics consists of a more diverse set of strategies, some more straightforward than others. For this reason, it is often practical to first familiarize oneself with techniques for radiogenomic modeling before generating and integrating imaging correlates for use alongside biological data in predicting radiotherapy outcomes.

Going forward, we feel it is likely that these two techniques (radiomics and radiogenomics) will intersect, but for this to occur will necessitate the use of high-throughput amenable methods—high dimensional datasets from diverse sources can provide unique insight to a broad patient

population but often present a multitude of challenging hurdles at the analysis stage. Thus, we expect that clinically viable predictive frameworks for radiotherapy will trend toward panomics strategies, which are designed to shift through highly heterogeneous/incomplete datasets inclusive of imaging correlates and multiscale biological data. It is our opinion that the noninvasiveness of imaging and the potential capabilities of next-generation platforms together with high spatial resolution will continue to motivate and support the use of radiomics/panomics going forward.

More broadly, we expect that AI and MI techniques will overhaul most aspects of radiation oncology, from delivery, treatment planning, and optimization to outcomes. The integration of such strategies will necessitate further training for physicians and physicists alike as well as development of quality assurance strategies and culpability schemas. Given the complexity of such modeling frameworks, it is possible that poorly trained personal or avoidable misinterpretations may do more harm than good. Thus, continuing to draft and refine policy for potential clinical pipelines using radiomics/panomics is critical. In the interim, we expect imaging biomarkers to continue to be useful for diagnosis and treatment.

## Disclosures

The authors declare they have no conflicts of interest.

## Acknowledgments

IEN acknowledges support from the National Institutes of Health (NIH), Grants No. R37-CA222215, No. R01-CA233487, and No. R41 CA243722, and Contract No. 75N92020D00018.

## References

1. G. Delaney et al., "The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines," *Cancer* **104**(6), 1129–1137 (2005).
2. S. M. Bentzen et al., "Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues," *Int. J. Radiat. Oncol. Biol. Phys.* **76**(3), 3–9 (2010).
3. M. Zaider and L. Hanin, "Tumor control probability in radiation treatment," *Med. Phys.* **38**, 574–583 (2011).
4. L. B. Marks et al., "Use of normal tissue complication probability models in the clinic," *Int. J. Radiat. Oncol. Biol. Phys.* **76**(3), S10–S19 (2010).
5. J. C. Peeken, F. Nüsslin, and S. E. Combs, "'Radio-oncomics': the potential of radiomics in radiation oncology," *Strahlenther. Onkol.* **193**, 767–779 (2017).
6. H. Lin, *A Guide to Outcome Modeling in Radiotherapy and Oncology: Listening to the Data*, I. Naqa, Ed., 1st ed., Series in Medical Physics and Biomedical Engineering, Series Editor: John G. Webster, E. Russell Ritenour, Slavik Tabakov, and Kwan-Hoon. Med Phys., CRC Press, Boca Raton, Florida (2018).
7. B. S. Rosenstein et al., "Radiogenomics: radiobiology enters the era of big data and team science," *Int. J. Radiat. Oncol. Biol. Phys.* **89**(4), 709–713 (2014).
8. P. Lambin et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer* **48**(4), 441–446 (2012).
9. J. Wu et al., "Radiomics and radiogenomics for precision radiotherapy," *J. Radiat. Res.* **59**, i25–i31 (2018).
10. H. J. W. L. Aerts, "The potential of radiomic-based phenotyping in precision medicine a review," *JAMA Oncol.* **2**, 1636–1642 (2016).
11. B. Theek et al., "Automation of data analysis in molecular cancer imaging and its potential impact on future clinical practice," *Methods* (2020).
12. I. El Naqa, "Biomedical informatics and panomics for evidence-based radiation therapy," *Wiley Interdiscip. Rev. Data Mining Knowl. Discov.* **4**(4), 327–340 (2014).

13. J. Coates, "Motivation for the inclusion of genetic risk factors of radiosensitivity alongside dosimetric and clinical parameters in predicting normal tissue effects," *Acta Oncol.* **54**(8), 1230–1231 (2015).
14. K. Unger, "Integrative radiation systems biology," *Radiat. Oncol.* **9**(1), 21 (2014).
15. L. G. M. Pettersson, "Radiation chemistry: radical water," *Nat. Chem.* **5**(7), 553–554 (2013).
16. W. Dörr and J. H. Hendry, "Consequential late effects in normal tissues [Internet]," *Radiother. Oncol.* **61**, 223–231 (2001).
17. H. B. Stone et al., "Effects of radiation on normal tissue: consequences and mechanisms," *Lancet Oncol.* **4**, 529–536 (2003).
18. T. Rancati et al., "Radiation dose-volume effects in the larynx and pharynx," *Int. J. Radiat. Oncol. Biol. Phys.* **76**, S64–S69 (2010).
19. V. Dilalla et al., "Radiotherapy side effects: integrating a survivorship clinical lens to better serve patients," *Curr. Oncol.* **27**(2), 107–112 (2020).
20. U. Hoeller et al., "Increasing the rate of late toxicity by changing the score? A comparison of RTOG/EORTC and LENT/SOMA scores," *Int. J. Radiat. Oncol.* **55**(4), 1013–1018 (2003).
21. S. L. Faria et al., "The challenge of scoring radiation-induced lung toxicity," *Clin. Oncol.* **21**(5), 371–375 (2009).
22. F. Denis et al., "Late toxicity results of the GORTEC 94-01 randomized trial comparing radiotherapy with concomitant radiochemotherapy for advanced-stage oropharynx carcinoma: comparison of LENT/SOMA, RTOG/EORTC, and NCI-CTC scoring systems," *Int. J. Radiat. Oncol.* **55**(1):93–98 (2003).
23. M. Roach et al., "Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: recommendations of the RTOG-ASTRO phoenix consensus conference," *Int. J. Radiat. Oncol. Biol. Phys.* **65**, 965–974 (2006).
24. J. Y. Bruce et al., "Current controversies in the management of biochemical failure in prostate cancer," *Clin. Adv. Hematol. Oncol.* **10**(11), 716–722 (2012).
25. C. W. Cheng and I. J. Das, "Treatment plan evaluation using dose-volume histogram (DVH) and spatial dose-volume histogram (zDVH)," *Int. J. Radiat. Oncol. Biol. Phys.* **43**, 1143–1150 (1999).
26. G. Luxton, P. J. Keall, and C. R. King, "A new formula for normal tissue complication probability (NTCP) as a function of equivalent uniform dose (EUD)," *Phys. Med. Biol.* **53**, 23–36 (2008).
27. J. Coates et al., "Contrasting analytical and data-driven frameworks for radiogenomic modeling of normal tissue toxicities in prostate cancer," *Radiother. Oncol.* **115**(1), 107–113 (2015).
28. I. El Naqa et al., "Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors," *Int. J. Radiat. Oncol.* **64**(4), 1275–1286 (2006).
29. A. M. Molinaro, R. Simon, and R. M. Pfeiffer, "Prediction error estimation: a comparison of resampling methods," *Bioinformatics* **21**, 3301–3307 (2005).
30. E. J. Limkin et al., "Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology," *Ann. Oncol.* **28**(6), 1191–1206 (2017).
31. G. Prax and L. Xing, "GPU computing in medical physics: a review," *Med. Phys.* **38**(5), 2685–2697 (2011).
32. A. Eklund et al., "Medical image processing on the GPU: past, present and future," *Med. Image Anal.* **17**, 1073–1094 (2013).
33. S. Ö. Arik, B. Ibragimov, and L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks," *J. Med. Imaging* **4**(1), 014501 (2017).
34. B. Ibragimov et al., "Segmentation of pathological structures by landmark-assisted deformable models," *IEEE Trans. Med. Imaging* **36**(7), 1457–1469 (2017).
35. B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks," *Med. Phys.* **44**(2), 547–557 (2017).
36. C. Parmar et al., "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLoS One* **9**(7), e102107 (2014).

37. J. L. Barker, Jr., et al., "Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system," *Int. J. Radiat. Oncol.* **59**(4), 960–970 (2004).
38. A. Niemierko, "Reporting and analyzing dose distributions: a concept of equivalent uniform dose," *Med. Phys.* **24**(1), 103–110 (1997).
39. A. J. Hope et al., "Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters," *Int. J. Radiat. Oncol. Biol. Phys.* **65**(1), 112–124 (2006).
40. O. Acosta et al., "Atlas based segmentation and mapping of organs at risk from planning CT for the development of voxel-wise predictive models of toxicity in prostate radiotherapy," *Lect. Notes Comput. Sci.* **6367**, 42–51 (2010).
41. A. A. Hakimi et al., "Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network," *Clin. Cancer Res.* **19**(12), 3259–3267 (2013).
42. J. P. Svensson et al., "Analysis of gene expression using gene sets discriminates cancer patients with and without late radiation toxicity," *PLoS Med.* **3**(10), e422 (2006).
43. C. Weigel et al., "Epigenetics in radiation-induced fibrosis," *Oncogene* **34**, 2145–2155 (2014).
44. M. Merrifield and O. Kovalchuk, "Epigenetics in radiation biology: a new research frontier," *Front. Genet.* **4**, 40 (2013).
45. D. S. Kanakoglou et al., "Effects of high-dose ionizing radiation in human gene expression: a meta-analysis," *Int. J. Mol. Sci.* **21**, 1938 (2020).
46. O. Azimzadeh et al., "Integrative proteomics and targeted transcriptomics analyses in cardiac endothelial cells unravel mechanisms of long-term radiation-induced vascular dysfunction," *J. Proteome. Res.* **14**, 1203–1219 (2015).
47. S. Roy and G. Trinchieri, "Microbiota: a key orchestrator of cancer therapy," *Nat. Rev. Cancer* **17**(5), 271–285 (2017).
48. G. Defraene et al., "The benefits of including clinical factors in rectal normal tissue complication probability modeling after radiotherapy for prostate cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **82**(3), 1233–1242 (2012).
49. S. L. Tucker et al., "Incorporating single-nucleotide polymorphisms into the Lyman model to improve prediction of radiation pneumonitis," *Int. J. Radiat. Oncol. Biol. Phys.* **85**(1), 251–257 (2013).
50. T. Rancati et al., "Inclusion of clinical risk factors into NTCP modelling of late rectal toxicity after high dose radiotherapy for prostate cancer," *Radiother. Oncol.* **100**(1), 124–130 (2011).
51. G. C. Barnett et al., "Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype," *Nat. Rev. Cancer* **9**(2), 134–142 (2009).
52. J. Coates, L. Souhami, and I. El Naqa, "Big data analytics for prostate radiotherapy," *Front. Oncol.* **6**, 149 (2016).
53. K. M. Smits et al., "Epigenetics in radiotherapy: where are we heading?" *Radiother. Oncol.* **111**(2), 168–177 (2014).
54. B. M. Turner, "Defining an epigenetic code," *Nat. Cell Biol.* **9**(1), 2–6 (2007).
55. V. K. Rakyan et al., "Epigenome-wide association studies for common human diseases," *Nat. Rev. Genet.* **12**(8), 529–541 (2011).
56. P. Lambin et al., "Predicting outcomes in radiation oncology: multifactorial decision support systems," *Nat. Rev. Clin. Oncol.* **10**(1), 27–40 (2013).
57. D. M. Abd El-Rehim et al., "High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses," *Int. J. Cancer* **116**(3), 340–350 (2005).
58. R. T. Larue et al., "Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures," *Br. J. Radiol.* **90**(1070), 20160665 (2017).
59. M. Avanzo et al., "Machine and deep learning methods for radiomics," *Med. Phys.* **47**(5), e185–e202 (2020).
60. A. Zwanenburg et al., "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology* **295**, 191145 (2020).

61. L. Wei et al., "Machine learning for radiomics-based multimodality and multiparametric modeling," *Q. J. Nucl. Med. Mol. Imaging* **63**, 323–338 (2019).
62. A.-R. Ali et al., "Applications of intelligent optimization in biology and medicine," *Intell. Syst. Ref. Libr.* (2016).
63. A. Gebejes and E. M. Master, "Texture characterization based on grey-level co-occurrence matrix," in *Conf. Inf. Manage. Sci.* (2013).
64. A. B. Tosun and C. Gunduz-Demir, "Graph run-length matrices for histopathological image segmentation," *IEEE Trans. Med. Imaging* **30**, 721–732 (2011).
65. H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Model. Image Process.* **57**, 235–245 (1995).
66. S. Paris, S. W. Hasinoff, and J. Kautz, "Local Laplacian filters: edge-aware image processing with a Laplacian pyramid," *ACM Trans. Graphics* **58**, 81–91 (2011).
67. A. Traverso et al., "Repeatability and reproducibility of radiomic features: a systematic review," *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 1143–1158 (2018).
68. J. J. Waninger et al., "Integrating radiomics into clinical trial design," *Q. J. Nucl. Med. Mol. Imaging* **63**, 339–346 (2019).
69. K. Nie et al., "NCTN assessment on current applications of radiomics in oncology," *Int. J. Radiat. Oncol. Biol. Phys.* **104**, 302–315 (2019).
70. A. Niemierko, "A generalized concept of equivalent uniform dose (EUD)," *Med. Phys.* **26**(6), 1100 (1999).
71. B. Jones and R. G. Dale, "Mathematical models of tumour and normal tissue response," *Acta Oncol.* **38**(7), 883–893 (1999).
72. G. J. Kutcher and C. Burman, "Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method," *Int. J. Radiat. Oncol. Biol. Phys.* **16**(6), 1623–1630 (1989).
73. A. Niemierko and M. Goitein, "Modeling of normal tissue response to radiation: the critical volume model," *Int. J. Radiat. Oncol. Biol. Phys.* **25**(1), 135–145 (1993).
74. S. Lee et al., "Bayesian network ensemble as a multivariate strategy to predict radiation pneumonitis risk," *Med. Phys.* **42**(5), 2421–2430 (2015).
75. A. van der Schaaf et al., "Embracing phenomenological approaches to normal tissue complication probability modeling: a question of method," *Int. J. Radiat. Oncol. Biol. Phys.* **91**(3), 468–471 (2015).
76. M. A. Pitt and I. J. Myung, "When a good fit can be bad," *Trends Cognit. Sci.* **6**(10), 421–425 (2002).
77. B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," *Lect. Notes Comput. Sci.* **1327**, 583–588 (1997).
78. S. Mika et al., "Kernel PCA and de-noising in feature spaces," *Analysis* **11**(i), 536–542 (1999).
79. J. Coates et al., "Evaluation and visualization of radiogenomic modeling frameworks for the prediction of normal tissue toxicities," in *World Congr. Med. Phys. and Biomed. Eng.*, Toronto, Springer International Publishing, pp. 517–520 (2015).
80. A. Guggenberger, *Another Introduction to Support Vector Machines* (2012).
81. S. T. H. Peeters et al., "Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling," *Int. J. Radiat. Oncol. Biol. Phys.* **66**(1), 11–19 (2006).
82. I. El Naqa et al., "Modeling of normal tissue complications using imaging and biomarkers after radiation therapy for hepatocellular carcinoma," *Int. J. Radiat. Oncol. Biol. Phys.* **100**, 335–343 (2018).
83. A. van der Schaaf et al., "Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size," *Radiother. Oncol.* **105**(1), 115–121 (2012).
84. Y. Cui et al., "Integrating radiosensitivity and immune gene signatures for predicting benefit of radiotherapy in breast cancer," *Clin. Cancer Res.* **24**, 4754–4762 (2018).
85. J. Kang et al., "Genomics models in radiotherapy: From mechanistic to machine learning," *Phys. Med.* **47**(5), e203–e217 (2019).

86. M. Liang et al., "Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach," *IEEE/ACM Trans. Comput. Biol. Bioinf. IEEE/ACM* **12**, 928–937 (2015).
87. T. Anagnostou et al., "Artificial neural networks for decision-making in urologic oncology," *Eur. Urol.* **43**(6), 596–603 (2003).
88. J. M. Jerez-Aragonés et al., "A combined neural network and decision trees model for prognosis of breast cancer relapse," *Artif. Intell. Med.* **27**(1):45–63 (2003).
89. B. Djavan et al., "Novel artificial neural network for early detection of prostate cancer," *J. Clin. Oncol.* **20**(4), 921–929 (2002).
90. G. Campanella et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nat. Med.* **25**(8), 1301–1309 (2019).
91. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature* **542**(7639), 115–118 (2017).
92. M. Vallières et al., "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer," *Sci. Rep.* **7**, 10117 (2017).
93. J. Kang et al., "Machine learning and radiogenomics: lessons learned and future directions," *Front. Oncol.* **8**, 228 (2018).
94. S. Chen et al., "A neural network model to predict lung radiation-induced pneumonitis," *Med. Phys.* **34**(9), 3420–3427 (2007).
95. M. Su et al., "An artificial neural network for predicting the incidence of radiation pneumonitis," *Med. Phys.* **32**(2), 318–325 (2005).
96. K. Gurney, "The delta rule," in *An Introduction to Neural Networks*, UCL Press (2010).
97. F. Jiang et al., "Artificial intelligence in healthcare: past, present and future," *Stroke Vasc. Neurol.* **2**, 230–243 (2017).
98. R. Yamashita et al., "Convolutional neural networks: an overview and application in radiology," *Insights Imaging* **9**, 611–629 (2018).
99. C. Dong et al., "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2016).
100. S. Ruder et al., "Transfer learning – machine learning's next frontier," *PPT* (2017).
101. A. Hosny et al., "Artificial intelligence in radiology," *Nat. Rev. Cancer* **18**, 500–510 (2018).
102. S. Cui et al., "Introduction to machine and deep learning for medical physicists," *Med. Phys.* **47**, e127–e147 (2020).
103. D. P. Kingma et al., "Semi-supervised learning with deep generative models," in *Adv. Neural Inf. Process. Syst.* (2014).
104. J. Sühmidhuber, "A local learning algorithm for dynamic feedforward and recurrent networks," *Connect. Sci.* **1**, 403–412 (1989).
105. H. H. Tseng et al., "Deep reinforcement learning for automated radiation adaptation in lung cancer," *Med. Phys.* **44**, 6690–6705 (2017).
106. J. Wu et al., "Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways," *Clin. Cancer Res.* **23**, 3334–3342 (2017).
107. Y. Jiang et al., "Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer," *Ann. Oncol.* **31**, 760–768 (2020).
108. Y. Jiang et al., "ImmunoScore signature: a prognostic and predictive tool in gastric cancer," *Ann. Surg.* **267**, 504–513 (2018).
109. J. Lao et al., "A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme," *Sci. Rep.* **7**, 10353 (2017).
110. J. Yu et al., "Lymph node metastasis prediction of papillary thyroid carcinoma based on transfer learning radiomics," *Nat. Commun.* **11**, 4807 (2020).
111. W. Han et al., "Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas," *Am. J. Neuroradiol.* **41**, 40–48 (2020).
112. L. Y. Xue et al., "Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis," *Eur. Radiol.* **30**, 2973–2983 (2020).
113. L. Feinendegen et al., "Systems biology and its potential role in radiobiology," *Radiat. Environ. Biophys.* **47**(1), 5–23 (2008).

114. O. G. Garzon and M. C. Plazas, "Evolution of physico-mathematical models in radiobiology and their application in ionizing radiation therapies," *TECCIENCIA* **9**, 17–26 (2014).
115. F. Cheng et al., "Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era," *Brief Bioinform.* **18**, 682–697 (2017).
116. B. Wang et al., "Similarity network fusion for aggregating data types on a genomic scale," *Nat. Methods* **11**, 333–337 (2014).
117. V. Gligorijević, N. Malod-Dognin, and N. Prulj, "Fuse: multiple network alignment via data fusion," *Bioinformatics* **32**, 1195–1203 (2016).
118. X. Cheng, L. Zhang, and Y. Zheng, "Deep similarity learning for multimodal medical images," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **6**, 248–252 (2018).
119. S. Pai and G. D. Bader, "Patient similarity networks for precision medicine," *J. Mol. Biol.* **430**, 2924–2938 (2018).
120. P. Ray et al., "Bayesian joint analysis of heterogeneous genomics data," *Bioinformatics* **30**, 1370–1376 (2014).
121. Y. Liu et al., "Multilevel Omic data integration in cancer cell lines: advanced annotation and emergent properties," *BMC Syst. Biol.* **7**, 14 (2013).
122. A. Bhattacharya and D. B. Dunson, "Sparse Bayesian infinite factor models," *Biometrika* **98**, 291–306 (2011).
123. C. Meng et al., "A multivariate approach to the integration of multi-omics datasets," *BMC Bioinf.* **15**, 162 (2014).
124. G. Tini et al., "Multi-omics integration: a comparison of unsupervised clustering methodologies," *Brief Bioinform.* **20**, 1269–1279 (2018).
125. Z. Zhou et al., "Novel radiomic features based on graph theory for PET image analysis," in *Proc.- Int. Symp. Biomed. Imaging* (2019).
126. M. Kohl et al., "A practical data processing workflow for multi-OMICS projects," *Biochim. Biophys. Acta* **1844**, 52–62 (2014).
127. I. Subramanian et al., "Multi-omics data integration, interpretation, and its application," *Bioinf. Biol. Insights* **14**, 117793221989905 (2020).
128. C. Sandhu, A. Qureshi, and A. Emili, "Panomics for precision medicine," *Trends Mol. Med.* **24**, 85–101 (2018).
129. J. O. Deasy, A. I. Blanco, and V. H. Clark, "CERR: a computational environment for radiotherapy research," *Med. Phys.* **30**(5), 979–985 (2003).
130. I. El Naqa et al., "Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships," *Phys. Med. Biol.* **51**(22), 5719–5735 (2006).
131. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Int. Symp. Inf. Theory*, pp. 267–281 (1973).
132. D. J. Spiegelhalter et al., "Bayesian measures of model complexity and fit," *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).
133. E. Segal et al., "Decoding global gene expression programs in liver cancer by noninvasive imaging," *Nat. Biotechnol.* **25**(6), 675–680 (2007).
134. O. Gevaert et al., "Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data: methods and preliminary results," *Radiology* **264**(2), 387–396 (2012).
135. Y. Luo et al., "Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling," *BJRiOpen* **1**, 20190021 (2019).
136. R. E. Bellman, *Adaptive Control Processes: A Guided Tour*, Vol. **2045**, Princeton University Press (2015).
137. M. D. Kuo and N. Jamshidi, "Behind the numbers: decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations," *Radiology* **270**(2), 320–325 (2014).
138. A. Zimek, E. Schubert, and H. P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Mining* **5**, 363–387 (2012).
139. B. Bauer and M. Kohler, "On deep learning as a remedy for the curse of dimensionality in nonparametric regression," *Ann. Stat.* **47**, 2261–2285 (2019).

140. S. E. Viswanath et al., "Dimensionality reduction-based fusion approaches for imaging and non-imaging biomedical data: concepts, workflow, and use-cases," *BMC Med. Imaging* **17**, 2 (2017).
141. B. Koçak et al., "Radiomics with artificial intelligence: a practical guide for beginners," *Diagn. Interv. Radiol.* **25**, 485–495 (2019).
142. H. He et al., "ADASYN: adaptive synthetic sampling approach for imbalanced learning," in *Proc. Int. Joint Conf. Neural Networks* (2008).
143. J. You, "Darpa sets out to automate research," *Science* **347**(6221), 465 (2015).
144. D. Gunning, "Explainable artificial intelligence program update," in *DARPA/I2O* (2017).
145. D. Gunning and D. W. Aha, "DARPA's explainable artificial intelligence program deep learning and security," in *AI Mag. Deep Learn. Secur.* (2019).
146. E. Tjoa and C. G. Fellow, "A survey on explainable artificial intelligence (XAI): towards medical XAI," *IEEE Trans. Neural Networks and Learn. Syst.* (2019).
147. A. Holzinger et al., "What do we need to build explainable AI systems for the medical domain?" arXiv:1712.09923 (2017).
148. M. Schubach et al., "Imbalance-aware machine learning for predicting rare and common disease-associated non-coding variants," *Sci. Rep.* **7**, 2959 (2017).
149. I. El Naqa, "Perspectives on making big data analytics work for oncology," *Methods* **111**, 32–44 (2016).
150. P. Lake and R. Drake, "Information systems management in the big data era," in *Advanced Information and Knowledge Processing*, L. C. Jain and X. Wu, Eds., Springer (2014).
151. F. Andre et al., "Biomarker studies: a call for a comprehensive biomarker study registry," *Nat. Rev. Clin. Oncol.* **8**, 171–176 (2011).
152. E. H. Simpson, "The interpretation of interaction in contingency tables," *J. R. Stat. Soc. Ser. B* **13**, 238–241 (1951).
153. J. Zou and L. Schiebinger, "AI can be sexist and racist: it's time to make it fair," *Nature* **559**, 324–326 (2018).

**James T. T. Coates** is a research fellow at Harvard Medical School and Massachusetts General Hospital Cancer Center in Boston. He focuses on developing techniques for the integration of high-throughput data together with patient-specific treatment metrics on toward improving clinical outcomes. Recently, his interests include the development of deep learning strategies that integrate highly heterogeneous datasets together for real-time optimization of patient-specific treatment regimens.

**Giacomo Pirovano** is a research fellow at Memorial Sloan Kettering Cancer Center in New York. His main research interests include the development of innovative approaches to improve radiotherapy with a particular focus on the development of novel methods for targeted radiotherapy, and radiotheranostics. His research combines different disciplines to allow more personalized cancer treatment plans.

**Issam El Naqa's** research interests are in the areas of oncology bioinformatics, multimodality image analysis, and treatment outcome modeling. His lab operates at the interface of physics, biology, and engineering with the primary motivation to design and develop approaches to unravel cancer patients' responses to chemoradiotherapy treatment by integrating physical, biological, and imaging information into advanced mathematical models and by applying techniques of machine learning and complex systems analysis to first principles and evaluating their performance in clinical and preclinical data. He is currently the Chair of Machine Learning at Moffitt and is an AAPM and IEEE fellow.