Machine Learning for Radiation Outcome Modeling and Prediction

Yi Luo*[a], Shifeng Chen[b], Gilmer Valdes[c]

[a] *Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan, 48103 United States*
E-mail: yiyiLuo@med.umich.edu

[b] *Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, Maryland, 21201 United States*

[c] *Department of Radiation Oncology, University of California, San Francisco, California, 94158 United States*

*Corresponding Author. Mailing address: 519 W William Street, Argus Building 1, Ann Arbor MI 48103-4943; Email address: yiyiLuo@med.umich.edu.

Abstract

This review paper intends to summarize the application of machine learning to radiotherapy outcome modeling based on structured and un-structured radiation oncology datasets. The most appropriate machine learning approaches for structured datasets in terms of accuracy and interpretability are identified. For un-structured datasets, deep learning algorithms are explored and a critical view of the use of these approaches in radiation oncology is also provided. Moreover, we discuss the challenges in radiotherapy outcome prediction, and suggest to improve radiation outcome modeling by developing appropriate machine learning approaches where both accuracy and interpretability are taken into account.

Key text: Radiation outcome modeling, machine learning, structured and un-structured datasets, accuracy, interpretability.

1. Introduction

Machine learning, which evolved from the study of pattern recognition and computational learning theory in artificial intelligence, intends to explore the study and construction of algorithms that can learn from data [1]. Recently, there is a tremendous increase of the use of machine learning in different areas of

radiation oncology, such as treatment planning optimization [2, 3], segmentation [4], radiation physics quality assurance [5-7], contouring or image-guided radiotherapy [8, 9]. In this paper, we focus on machine learning for radiation outcomes modeling [10], [11], [12].

Radiation outcomes modeling includes survival analysis, local tumor control probability (TCP) and normal tissue complication probability (NTCP), (e.g. radiation pneumonitis, cardiac toxicity, and esophagitis). In the past, these approaches relied on analytical models that assumed a defined relationship between the dose and the specific outcome (e.g. linear quadratic model for tumor control probability [13] or the Lyman model for normal tissue complication probability [14]). Also, they were limited to only considering dosimetric factors. For analytical models, if the underlying assumptions and simplifications are not correct as it is usually the case, then prediction accuracy suffers. On the other hand, with the use of Machine Learning (ML) algorithms, different relationships between dose and outcome can be automatically investigated. Additionally, non dosimetric factors such as comorbidities, age, performance status, imaging and genetic information, etc. can be easily incorporated [10-12], and they have different representations, where age and imaging are described by continuous variables, DVH bins, performance status and single nucleotide polymorphisms (SNPs) are denoted by ordinal variables, and gender, chemotherapy are indicated by binary variables. The radiation outcomes depend on all these characteristics included in them may result in more accurate models but this impose a challenge in terms of data size. Essentially, each patient is being characterized by a high dimensional feature vector but the number of samples available to learn a valid relationship is comparatively smaller in radiation oncology compared to other fields. In order to enhance generalization or reduce overfitting, either redundant or irrelevant features may need to be removed without incurring much loss of the information, which is called “feature selection” in ML. Basically, wrappers, filters and embedded methods are three main categories of feature selection algorithms [15]. While wrapper approaches use a predictive model to score feature subsets and filter approaches use a proxy measure instead of the error rate to score a feature subset, embedded approaches are a catch-all group of techniques to perform feature selection as part of the model construction process. When filters and wrappers are used, a common mistake is to reuse the same observations used to select the variables to build and report the accuracy of the final model. Therefore, recently embedded methods like least absolute shrinkage and selection operator (LASSO) seem to be preferred because they perform feature selection internally and there is smaller chance of information leakage from the training to the testing datasets [16].

For radiation outcome modeling, supervised machine learning is the main approach. In supervised learning, the Machine Learning (ML) algorithm is provided with input features (normally in the form of a matrix with each row corresponding to a patient and each column corresponding to a feature) and

outcome (in the form of a vector) as shown in Table 1. With this data, a specific loss function is written and by minimizing it, the optimal mapping between the input features and the outcome is found. In this paper, we will review the challenges, pros and cons of using these algorithms to model outcome data in radiation oncology.

2. Machine learning approaches for radiation outcome modeling

Several ML algorithms have been investigated for radiation therapy outcome modelling. If the input features have meaning by themselves (called in the ML literature structured or tabular data) then machine learning algorithms like logistic regression with penalization, decision trees, Bayesian networks, random forests and gradient boosting are preferred, although algorithms that do representational learning (design features or concepts) like neural networks can still be used in our field and outside for the analysis of the structured data. If the input variables are unstructured (images, text), then neural network based deep learning algorithms are preferred.

2.1. Machine learning approaches for structured data

The observed radiation treatment outcome can be considered the result of the interaction of several dosimetric, clinical, or biological variables [17]. Machine leaning algorithms intend to develop data-driven models by fitting parameters using the collected clinical and dosimetric data [18]. For structured dataset, ML approaches have already been employed for several treatment sites such as lung [19], prostate [20], head & neck cancer [21] or meningioma [11]. In all cases many different ML algorithms, such as linear regression, artificial neural network, support vector machine, Bayesian networks, decision tree, random forests or gradient boosting machine, have been explored. Below we will attempt to cover previous publications using these algorithms (non-exhaustive).

Linear regression models the relationship between the response and one or more independent variables as a linear equation. Systems biology is the computational and mathematical modeling of complex biological systems. In order to develop a systems biology understanding of radio-sensitivity to enhance our ability to identify radiation-specific biomarkers for lung, colon, and breast cancers [22], linear regression was employed to correlate gene expression with survival. In another study, the impact of overall treatment time on patient outcome in non-small cell lung cancer was studied using linear-regression and Cox proportional hazard models [23] .

Artificial neural networks (ANNs) are multi-layer non-linear models that use gradient descent and back propagation to find the optimal coefficients. An ANN is based on a collection of connected nodes called artificial neurons. Each connection between nodes can transmit a signal from one neuron to another.

There is a weight associated with each connection, and the weight increases or decreases the strength of the signal at a connection which is adjusted as learning proceeds. In a study to predict symptomatic lung injury based on pre-radiotherapy, biological and physical data [24], a nonlinear, feedforward ANNs was trained to predict the radiation outcome. The ANN approach was also used to predict radiation pneumonitis following radiotherapy [25]. Three ANNs, each with three layers, were developed for this classification task. A growing/pruning based ANN approach was developed to predict Grade 2+ radiation pneumonitis (RP2), starting from the smallest possible network until a satisfactory solution was found [26]. While the input of the first ANN was limited to the patient lung dose-volume data only, the second study showed that the addition of non-dosimetric features can significantly improve the generalization capability of the network.

Support vector machine (SVM) is a class of machine learning algorithm that attempts to find a feature space where the data can be linearly separated using a kernel function to do the transformation. In a study to predict patient's complication risk for personalized treatment planning [27], a nonlinear kernel based SVM method was used to learn complex interactions between the observed toxicities and treatment, anatomical, and patient-related variables. In the same study, the value of feature transformation using principle component analysis (PCA) to visualize high dimensional data and determine proper kernel type was also demonstrated. It was demonstrated that linear models probably work well and non-linear methods are unnecessary when responses may be separated along a linear ridge in a PCA plot, and nonlinear features generated via kernel/SVM methods may improve the model prediction when there is no such linear ridge under PCA analysis.

Bayesian networks (BNs) are a type of probabilistic graphical model to represent a set of variables and their conditional dependencies via a directed acyclic graph. In a new treatment planning model for more outcome-focused decision making, a BN was employed to model the radiation therapy process of prostate cancer and treatment outcomes such as distant metastasis, rectal and bladder complications using expert opinion, results of clinical trials, and published research. Then, the quality-adjusted life expectancy of a patient can be evaluated based on the BN [28]. In order to explore the biophysical relationship among radiation treatment, patients' characteristics (age, comorbidities, etc), and radiation outcomes, BN were employed to predict radiation pneumonitis [29, 30], local control [31], or both [32]. Figure 1 shows an example of a pre-treatment joint BN to predict lung cancer tumor local control (LC) and radiation pneumonitis toxicity grade two or above (RP2) simultaneously [32]. The most important features for radiation outcome prediction including tumor and lung gEUDs, one SNP (cxcr1-Rs2234671), two miRNAs (miR-20a-5p and miR-191-5p), one pre-treatment PET radiomics feature (MTV) and three pre-treatment cytokines (IL-15, IL-4 and IL-10) were selected from a high-dimensional retrospective dataset

as indicated by the nodes in the BN. The edges of the BN, identified with different colors, represent the biophysical relationships between the variables analyzed. BNs can be used to explore biophysical cause-effect relationship from retrospective dataset, but it is limited by inference complexity, and its computational cost grows exponentially as the number of nodes increases.

As mentioned above, one problem that limits the exploration of biological connections with BN is dataset size. Since most datasets in radiation oncology tend to be small, sharing data among different institutions might be necessary. Distributed learning is a technique that can allow the learning of predictive models on data originating from multiple hospitals without the data leaving the hospital. A BN model was adapted for distributed learning to predict dyspnea by combining data originating from multiple hospitals. It was shown that it is possible to use the distributed learning approach to train the BN model without the specific datasets leaving the individual hospital [33]. This results open a venue for data sharing among multiple hospitals without violating privacy laws [33].

Decision trees (DTs) are constructed using recursive partitioning analysis to optimize successive dichotomisation of input variables. The resulted tree-like structure had been used to augment prediction of the classic Lyman NTCP [14], pneumonitis [10], [34], chest wall pain [12], salvage high-dose-rate brachytherapy (sHDRB) [35] and meningioma [11]. An advantage of DTs is that they are highly interpretable. Their biggest disadvantage is that they are not the most accurate algorithm on expectation. MediBoost is a single tree technique based on boosting that attempts to improve accuracy while still building interpretable trees. Since it uses weighted versions of all cases to derive splits at each point in the tree, it can be advantageous for relatively small datasets. MediBoost has been used to predict both pneumonitis and biochemical failure after sHDRB surpassing random forest performance in some cases but not all [11, 35].

Random forest (RF) is considered one of the best out of the shelves ML algorithms. It improves trees' accuracy by building a large number of deep trees (ensemble) and averaging them which reduces variance and improves accuracy. In addition to predicting the above mentioned outcomes [10], [12], [11, 35], the RF approach was also employed to predict rectal toxicity following prostate cancer radiotherapy [36], and xerostomia after IMRT treatments [37]. Random survival forest (RSF) is a random forests method for the analysis of right-censored survival data. Not only it incorporates all univariate and multivariate effects automatically, but also it can find influential covariates in highly correlated subsets, which is particularly useful in high-dimensional covariate selection problems [38, 39]. The RSF had been employed to explore single and multivariable models of overall survival (OS) and progression-free survival (PFS) [40].

Gradient boosting machines (GBM) is another popular ensemble machine-learning technique usually use with decision trees. It uses gradient boosting, a way to improve any machine-learning model by iteratively training new models that specialize on the weak points of the previous ones. GBMs have been used in radiotherapy to predict radiation induced pneumonitis [10], and meningioma local failure [11]. A version of GBM, XGBoost, consistently wins most Kaggle competitions that involve structured data today [10, 41]. Kaggle is an online platform where private and public entities open data science projects for third parties to compete.

With all these algorithms previously used, an important question is which algorithm or algorithms should be preferred? It is important to note that it has been proven mathematically that no algorithm is guaranteed to be best in all problems [42]. However, general recommendations can be made to point out which algorithms are more likely to work better on average. If accuracy is the goal, and provided that there is enough data, RF and GBM are on average the best algorithms for the analysis of structured data [43, 44]. On the contrary, if interpretation and understanding are the goals, linear models, DTs and BNs are inherently visual, and relationship between variables can be possibly interpreted [44, 45]. In the case of radiation oncology datasets, which tend to be smaller than those analyzed in other fields, it was shown than on average linear models are competitive to RF and GBM [46]. Therefore, it is our advised that at least these three algorithms should be explored.

2.2. Machine learning approaches for unstructured data (images and text)

Currently, many parts of the medical datasets collected in radiation oncology have unstructured format (e.g. medical notes, images). Understanding how much knowledge can be extracted from them is a challenging and exciting task [47-51]. Deep learning (DL) is a specific subfield of machine learning that learns representations from data and it is perfectly suited to handle unstructured data like images or text. The "deep" in DL stands for the idea of the application of successive layers of representations (mathematical operations) as shown in Figure 2.

Although we usually refer to them with one term, deep neural network (DNN) encompasses many different algorithms. Convolutional Neural Networks (CNNs), a special kind of DNN, are best known for image-related prediction. Their name stemmed from the successful application of the signal processing operation of convolution. Recurrent neural networks (RNNs) are another variant of DNN specifically useful for learning sequential data, such as voice or text data. The values of DNNs have also been explored in Radiotherapy. A DNN was proposed to predict Quality of life scores for prostate SBRT patients using DVH data [52]. In a study predicting 2-year local failure following SBRT [53], a multi DNN approach analyzing both patient CT simulation scans and clinical risk factors, was shown to outperform logistic regression, RF, and SVM. To retrieve and classify multimodal medical images (MR,

CT, PET) for 24 body organs with different levels of data extraction [54], a CNN framework was developed, obtaining an excellent prediction accuracy performance. Additionally, CNNs were used to predict survival risks for rectal cancer patients based on imaging data showing improvement over other approaches [55]. Also, DNNs were used to explore state transition probabilities for building an autonomous clinical decision support system to adapt patient dose per fraction in a response-adapted treatment setting [56]. Finally, a DNN algorithm was also shown to outperform other approaches in recognizing clinical texts in medical documents [57] or learning radiomic features for survival prediction in lung and head & neck cancers [58].

In general, DL methods appear to be promising for outcome prediction in radiotherapy since it has relatively high prediction performance [58]. However, high quality big data size is the key for successful application of the DL algorithms. For instance, Google brain explicitly recommended to not use DL with less than thousands of data points even when transfer learning, an approach that reuses parameters of a model trained on another task with more available data , is used [59]. These recommendations stem from the fact that DL models usually contain millions of parameters. With many more parameters to tune than other algorithms, overfitting in the cross validation space is more likely to occur. Therefore, the field of radiotherapy should proceed with caution when these algorithms are explored in small datasets [59].

2.3. Interpretability and causal inference

The challenges of creating humanlike intelligence in machines remains greatly underestimated. Current ML systems lack the essence of human intelligence: understanding the situations we experience or being able to grasp their meaning [60]. Most ML algorithms today find correlation and not causation [61]. For instance, an ANN (a non-interpretable algorithm) that was developed to triage patients with pneumonia for hospital discharge was found to inadvertently label asthmatic patients as low risk [62-64]. Colon cancer screening or abnormal breast finding were found to be highly correlated to the risk of having a stroke with no apparent clinical justification [65]. A CNN was found to be using the hospital system, department or the label “portable” to improve their prediction of pneumonia while disregarding the true findings in the image [66]. Besides the issue of data quality, current state-of-the-art methods lack the interpretability that is essential for patient stratification and safe clinical decision making, although machine learning techniques afford better-than-human performance in numerous domains.

Both clinicians and medical physicists should hesitate to apply prediction models based on black box algorithms [67] [68]. Unfortunately, there is a tradeoff between interpretability and accuracy in ML algorithms. DTs, BNs, and logistic regressions [69-71] are among the preferred choices in medicine [72-80] because they create models that are highly interpretable. However, they are consistently outperformed in terms of accuracy by “black box” algorithms used in many other fields, including ensemble methods,

meta algorithms and deep neural networks [81-84]. The tradeoff between accuracy and interpretability represents a long-standing problem in machine learning, and it is an active area of research from both practical and theoretical points of view [64, 85-88]. Additionally, techniques to interpret black boxes have been created but researchers and clinicians should be aware of their limitations. For instance, variable importance (or pixel importance for images) has limited capability to address how specific variables interact in a highly nonlinear function to build predictive model for a specific patient. Moreover, algorithms designed to explain the black box models do not provide a faithful representation of the original model. As such, the medical physics community should follow with attention recent theoretical developments on causality and interpretability of machine learning algorithms before we fully deploy the models mentioned above.

3. Discussions and conclusion

Radiotherapy plays an increasingly dominant role in the comprehensive multidisciplinary management of cancer. About half of all cancer patients will receive radiotherapy either as a part of the initial treatment with curative intent or as palliative treatment [89]. Improved modeling accuracy for normal tissue complications and tumor control probability is highly important and still one of biggest challenges of physicists in radiation therapy today [35, 90]. In this paper, we reviewed a variety of machine learning approaches for radiation outcome modeling based on structured and un-structured data. Considering a relatively large number of features as the input, ML-based methods have the potential to improve radiation outcome modeling compared to analytical methods. Additionally, complex algorithms like RF, GBM, DL can outperform previously used models like DT, BN and logistic regression in terms of accuracy. However, both clinicians and medical physicists should proceed with caution when employing these relative new algorithms due to their lack of interpretability. Accuracy on a test set is not a good representation of true performance because the complex algorithms can take advantage of spurious correlation (e.g asthma status for pneumonia patients). Therefore, we would like to emphasize the need for the development of accurate and interpretable algorithms as decision support tools for clinical radiation treatment decision making. The use of MediBoost in radiation oncology represents pioneering work in this direction [44].

We also need to highlight that limited clinical outcomes, inappropriate toxicity classification, genomics data, variation in dose and fractionation directly affect the accuracy of current models. In fact, most models reported above have accuracy that oscillates around Area Under ROC curve (AUC) of 0.70 which is suboptimal for their clinical use. Therefore, a conscientious effort needs to be done in the field to collect better and bigger datasets but many challenges wait ahead [91, 92]. Possible ways to remove these

obstacles include, but are not limited to, standardization of processes and data structures, information sharing across institutions, distributed learning, transfer learning or synthetic data generation using generative adversarial networks [93]. Imbalanced data is another intrinsic problem of applying ML approaches for radiation outcome prediction. For instance, if only 10% of patients develop certain toxicity, an algorithm that naively predict that nobody develops the toxicity is 90% accurate. Balancing the imbalanced event rate during cross validation and reporting balanced accuracy is necessary, and many techniques can be used to achieve this goal [94-96].

Additionally, imaging information is widely used in radiation oncology treatment planning, from dose calculation to implementation of the treatment. Quantitative features, such as radiomics and radiogenomics, extracted from these images can provide useful information about tumor characteristics habitats or evolution with the treatment. Therefore, these features have a potential to improve radiation outcome prediction for precision and personalized radiotherapy through advanced machine and deep learning techniques. As the number of features increases in the era of Big Data, feature selection and dataset size plays an even important role though.

Summarizing, the overwhelming evidence shows that the quality of reporting of prediction model studies is poor today. Although ML algorithms are a powerful tool, their complexity and higher chances of overfitting can result in poor replicability and generalization of studies. Then model validation is critical for predictive models to be used in the clinical setting. While cross validation and bootstrapping intend to avoid overfitting in developing the predictive models, external validation through reserved testing dataset or other hospital's datasets is necessary to evaluate their prediction performance. A set of recommendation, named Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD), has been developed [97]. We strongly encourage journals and researchers to follow these recommendations and clear specify the level of validation of each study. Moreover, as it current stands, theoretical advances in current machine learning algorithms is also necessary for the wide acceptance of these models into clinical practice. Since most machine learning algorithms find correlation and not causation, careful examination and testing needs to be performed before deploying models into clinical practice. As such, further development on causal modeling and/or interpretability of algorithms is necessary.

Figure Legends:

Figure 1. A joint pre-treatment BN for LC and RP2 prediction

Figure 2. A neural network structure for deep learning

Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

References

1. Samuel, A.L., *Some Studies in Machine Learning Using the Game of Checkers.* Ibm Journal of Research and Development, 1959. **3**(3): p. 211-&.
2. Valdes, G., et al., *Clinical decision support of radiotherapy treatment planning: A data-driven machine learning strategy for patient-specific dosimetric decision making.* Radiotherapy and Oncology, 2017. **125**(3): p. 392-397.
3. Shiraishi, S. and K.L. Moore, *Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy.* Medical Physics, 2016. **43**(1): p. 378-387.
4. Kearney, V., et al., *An unsupervised convolutional neural network-based algorithm for deformable image registration.* Physics in Medicine and Biology, 2018. **63**(18).
5. Interian, Y., et al., *Deep nets vs expert designed features in medical physics: An IMRT QA case study.* Medical Physics, 2018. **45**(6): p. 2672-2680.
6. Valdes, G., et al., *Use of TrueBeam developer mode for imaging QA.* Journal of Applied Clinical Medical Physics, 2015. **16**(4): p. 322-333.
7. Kearney, V., et al., *Correcting TG 119 confidence limits.* Medical Physics, 2018. **45**(3): p. 1001-1008.
8. Chen, T., et al., *Objected constrained registration and manifold learning: A new patient setup approach in image guided radiation therapy of thoracic cancer.* Medical Physics, 2013. **40**(4).
9. Peng, L.C., et al., *Application of Machine Learning and Radiomics to Distinguish True Progression From Treatment Effect after Stereotactic Radiation Therapy for Brain Metastases.* International Journal of Radiation Oncology Biology Physics, 2018. **102**(3): p. S93-S94.
10. Valdes, G., et al., *Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy.* Physics in Medicine and Biology, 2016. **61**(16): p. 6105-6120.

11. Gennatas, E.D., et al., *Preoperative and postoperative prediction of long-term meningioma outcomes.* Plos One, 2018. **13**(9).
12. Chao, H.H., et al., *Exploratory analysis using machine learning to predict for chest wall pain in patients with stage I non-small-cell lung cancer treated with stereotactic body radiation therapy.* Journal of Applied Clinical Medical Physics, 2018. **19**(5): p. 539-546.
13. Held, D.K., *Radiobiology for the Radiologist*. Radiation Research, ed. J.E. Hall and J.A. Giaccia. Vol. 166. 2006.
14. Lyman, J.T., *Complication Probability as Assessed from Dose Volume Histograms.* Radiation Research, 1985. **104**(2): p. S13-S19.
15. Guyon, I. and A. Elisseeff, *An introduction to variable and feature selection.* Journal of Machine Learning Research, 2003. **3**(Mar): p. 1157-1182.
16. Tang, J., S. Alelyani, and H. Liu, *Feature selection for classification: A review*, in *Data Classification: Algorithms and Applications*. 2014, CRC Press. p. 37-64.
17. El Naqa, I., et al., *Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors.* International Journal of Radiation Oncology Biology Physics, 2006. **64**(4): p. 1275-1286.
18. Deasy, J.O. and I. El Naqa, *Image-based modeling of normal tissue complication probability for radiation therapy.* Cancer Treat Res, 2008. **139**: p. 215-56.
19. Das, S.K., et al., *Predicting lung radiotherapy-induced pneumonitis using a model combining parametric lyman probit with nonparametric decision trees.* International Journal of Radiation Oncology Biology Physics, 2007. **68**(4): p. 1212-1221.
20. Cosma, G., et al., *A survey on computational intelligence approaches for predictive modeling in prostate cancer.* Expert Systems with Applications, 2017. **70**: p. 1-19.
21. Dean, J.A., et al., *Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy.* Radiotherapy and Oncology, 2016. **120**(1): p. 21-27.
22. Eschrich, S., et al., *Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform.* Int J Radiat Oncol Biol Phys, 2009. **75**(2): p. 497-505.
23. Chen, M., et al., *The impact of overall treatment time on outcomes in radiation therapy for non-small cell lung cancer.* Lung Cancer, 2000. **28**(1): p. 11-19.
24. Munley, M.T., et al., *A neural network to predict symptomatic lung injury.* Phys Med Biol, 1999. **44**(9): p. 2241-9.

25. Su, M., et al., *An artificial neural network for predicting the incidence of radiation pneumonitis.* Med Phys, 2005. **32**(2): p. 318-25.
26. Chen, S.F., et al., *A neural network model to predict lung radiation-induced pneumonitis.* Medical Physics, 2007. **34**(9): p. 3420-3427.
27. El Naqa, I., J.D. Bradley, and J. Deasy, *Nonlinear Kernel-based Approaches for Predicting Normal Tissue Toxicities.* Seventh International Conference on Machine Learning and Applications, Proceedings, 2008: p. 539-544.
28. Smith, W.P., et al., *Personalized treatment planning with a model of radiation therapy outcomes for use in multiobjective optimization of IMRT plans for prostate cancer.* Radiation Oncology, 2016. **11**.
29. Oh, J.H. and I. El Naqa, *Bayesian network learning for detecting reliable interactions of dose-volume related parameters in radiation pneumonitis.* Eighth International Conference on Machine Learning and Applications, Proceedings, 2009: p. 484-488.
30. Luo, Y., et al., *Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis.* Radiotherapy and Oncology, 2017. **123**(1): p. 85-92.
31. Luo, Y., et al., *Development of a Fully Cross-Validated Bayesian Network Approach for Local Control Prediction in Lung Cancer.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2018.
32. Luo, Y., et al., *A multiobjective Bayesian networks approach for joint prediction of tumor local control and radiation pneumonitis in nonsmall-cell lung cancer (NSCLC) for response-adapted radiotherapy.* Medical Physics, 2018. **45**(8): p. 3980-3995.
33. Jochems, A., et al., *Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital - A real life proof of concept.* Radiotherapy and Oncology, 2016. **121**(3): p. 459-467.
34. Palma, D.A., et al., *Predicting Radiation Pneumonitis After Chemoradiation Therapy for Lung Cancer: An International Individual Patient Data Meta-analysis.* International Journal of Radiation Oncology Biology Physics, 2013. **85**(2): p. 444-450.
35. Valdes, G., et al., *Salvage HDR Brachytherapy: Multiple Hypothesis Testing Versus Machine Learning Analysis.* International Journal of Radiation Oncology Biology Physics, 2018. **101**(3): p. 694-703.

36. Ospina, J.D., et al., *Random Forests to Predict Rectal Toxicity Following Prostate Cancer Radiation Therapy.* International Journal of Radiation Oncology Biology Physics, 2014. **89**(5): p. 1024-1031.
37. Soares, I., et al., *Predicting xerostomia after IMRT treatments: a data mining approach.* Health and Technology, 2018. **8**(1-2): p. 159-168.
38. Ishwaran, H., et al., *Random Survival Forests.* Annals of Applied Statistics, 2008. **2**(3): p. 841-860.
39. Ishwaran, H. and U.B. Kogalur, *Consistency of random survival forests.* Statistics & Probability Letters, 2010. **80**(13-14): p. 1056-1064.
40. Jain, R., et al., *Outcome Prediction in Patients with Glioblastoma by Using Imaging, Clinical, and Genomic Biomarkers: Focus on the Nonenhancing Component of the Tumor.* Radiology, 2014. **272**(2): p. 484-493.
41. Chollet, F. and J.J. Allaire, *Deep Learning with R*. 2018: Manning Publications.
42. Wolpert, D.H., *The lack of A priori distinctions between learning algorithms.* Neural Computation, 1996. **8**(7): p. 1341-1390.
43. Caruana, R. and A. Niculescu-Mizil. *An empirical comparison of supervised learning algorithms*. in *ICML '06 Proceedings of the 23rd international conference on Machine learning*. 2006. Pittsburgh, Pennsylvania, USA: ACM New York, NY, USA
44. Valdes, G., et al., *MediBoost: a Patient Stratification Tool for Interpretable Decision Making in the Era of Precision Medicine.* Scientific Reports, 2016. **6**.
45. El Naqa, I., R. Li, and M.J. Murphy, *Machine Learning in Radiation Oncology: Theory and Applications*. 2015, Switzerland: Springer.
46. Deist, T.M., et al., *Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers.* Medical Physics, 2018. **45**(7): p. 3449-3459.
47. Ylijoki1, O. and J. Porras1, *Perspectives to Definition of Big Data: A Mapping Study and Discussion.* Journal of Innovation Management, 2016. **4**(1): p. 69-91.
48. Birkhead, G.S., M. Klompas, and N.R. Shah, *Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health.* Annual Review of Public Health, Vol 36, 2015. **36**: p. 345-359.
49. Thompson, R.F., et al., *Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?* Radiother Oncol, 2018.
50. Feng, M., et al., *Machine Learning in Radiation Oncology: Opportunities, Requirements, and Needs.* Front Oncol, 2018. **8**: p. 110.

51. Morin, O., et al., *A Deep Look Into the Future of Quantitative Imaging in Oncology: A Statement of Working Principles and Proposal for Change.* International Journal of Radiation Oncology Biology Physics, 2018. **102**(4): p. 1074-1082.
52. Qi, X., J. Neylon, and A. Santhanam, *Dosimetric Predictors for Quality of Life After Prostate Stereotactic Body Radiation Therapy via Deep Learning Network.* International Journal of Radiation Oncology Biology Physics, 2017. **99**(2): p. S167-S167.
53. Aneja, S., et al., *Deep Neural Network to Predict Local Failure Following Stereotactic Body Radiation Therapy: Integrating Imaging and Clinical Data to Predict Outcomes.* International Journal of Radiation Oncology Biology Physics, 2017. **99**(2): p. S47-S47.
54. Qayyum, A., et al., *Medical image retrieval using deep convolutional neural network.* Neurocomputing, 2017. **266**: p. 8-20.
55. Li, H., et al., *Deep Convolutional Neural Networks for Imaging Based Survival Analysis of Rectal Cancer Patients.* International Journal of Radiation Oncology Biology Physics, 2017. **99**(2): p. S183-S183.
56. Tseng, H.H., et al., *Deep reinforcement learning for automated radiation adaptation in lung cancer.* Medical Physics, 2017. **44**(12): p. 6690-6705.
57. Wu, Y., et al., *Named Entity Recognition in Chinese Clinical Text Using Deep Neural Network.* Stud Health Technol Inform, 2015. **216**: p. 624-8.
58. Jochems, A., et al., *Developing and Validating a Survival Prediction Model for NSCLC Patients Through Distributed Learning Across 3 Countries.* International Journal of Radiation Oncology Biology Physics, 2017. **99**(2): p. 344-352.
59. Valdes, G. and Y. Interian, *Comment on 'Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study'.* Physics in Medicine and Biology, 2018. **63**(6).
60. Mitchell, M. *Artificial Intelligence Hits the Barrier of Meaning*. 2018; Available from: https://www.nytimes.com/2018/11/05/opinion/artificial-intelligence-machine-learning.html.
61. Meyer, P., et al., *Survey on deep learning for radiotherapy.* Computers in Biology and Medicine, 2018. **98**: p. 126-146.
62. Cooper, G.F., et al., *An evaluation of machine-learning methods for predicting pneumonia mortality.* Artificial Intelligence in Medicine, 1997. **9**(2): p. 107-138.
63. Cooper, G.F., et al., *Predicting dire outcomes of patients with community acquired pneumonia.* Journal of Biomedical Informatics, 2005. **38**(5): p. 347-366.

64. Caruana, R., et al., *Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission*, in *the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015: Sydney, Australia.

65. Mullainathan, S. and Z. Obermeyer, *Does Machine Learning Automate Moral Hazard and Error?* American Economic Review, 2017. **107**(5): p. 476-480.

66. Zech, J.R., et al., *Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study.* Plos Medicine, 2018. **15**(11).

67. Kang, J., et al., *Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective.* International Journal of Radiation Oncology Biology Physics, 2015. **93**(5): p. 1127-1135.

68. Lynch, C.M., V.H. van Berkel, and H.B. Frieboes, *Application of unsupervised analysis techniques to lung cancer patient data.* PLoS One, 2017. **12**(9): p. e0184370.

69. Quinlan, J.R., *C4.5: Programs for Machine Learning*. 1993: San Mateo: Morgan Kaufmann.

70. Breiman, L., et al., *Classification and Regression Trees*. 1984: Chapman and Hall/CRC.

71. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*. 1988, San Mateo: Morgan Kaufmann.

72. Cain, K.P., et al., *An Algorithm for Tuberculosis Screening and Diagnosis in People with HIV.* New England Journal of Medicine, 2010. **362**(8): p. 707-716.

73. Berlowitz, D.R., et al., *Inadequate management of blood pressure in a hypertensive population.* New England Journal of Medicine, 1998. **339**(27): p. 1957-1963.

74. Haydel, M.J., et al., *Indications for computed tomography in patients with minor head injury.* New England Journal of Medicine, 2000. **343**(2): p. 100-105.

75. Chen, H.Y., et al., *A five-gene signature and clinical outcome in non-small-cell lung cancer.* New England Journal of Medicine, 2007. **356**(1): p. 11-20.

76. Gilbert, M.R., et al., *A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma.* New England Journal of Medicine, 2014. **370**(8): p. 699-708.

77. Lionetti, E., et al., *Introduction of Gluten, HLA Status, and the Risk of Celiac Disease in Children.* New England Journal of Medicine, 2014. **371**(14): p. 1295-1303.

78. Spiegelhalter, D.J., *Probabilistic Expert Systems in Medicine: Practical Issues in Handling Uncertainty.* Statistical Science, 1987. **2**(1): p. 3-44.

79. Shortliffe, E.H., *The Adolescence of Ai in Medicine - Will the Field Come of Age in the 90s.* Artificial Intelligence in Medicine, 1993. **5**(2): p. 93-106.

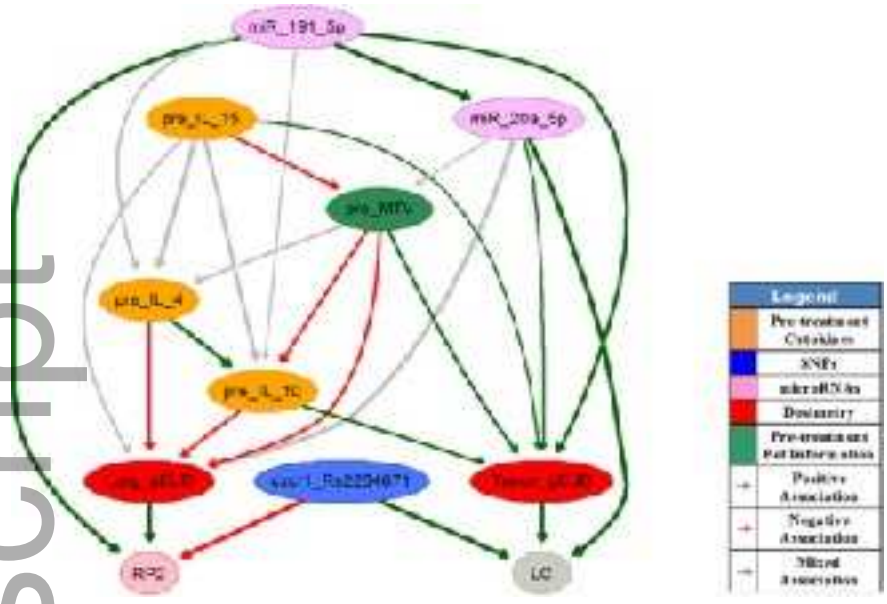
80. Warner, H.R., D.K. Sorenson, and O. Bouhaddou, *Knowledge Engineering in Health Informatics*. 1997, Berlin, New York: Springer Verlag.
81. Breiman, L., *Random forests.* Machine Learning, 2001. **45**(1): p. 5-32.
82. Freund, Y. and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting.* Journal of Computer and System Sciences, 1997. **55**(1): p. 119-139.
83. Friedman, J., T. Hastie, and R. Tibshirani, *Additive logistic regression: A statistical view of boosting - Rejoinder.* Annals of Statistics, 2000. **28**(2): p. 400-407.
84. Friedman, J.H., *Greedy function approximation: A gradient boosting machine.* Annals of Statistics, 2001. **29**(5): p. 1189-1232.
85. Cabitza, F., R. Rasoini, and G.F. Gensini, *Unintended Consequences of Machine Learning in Medicine.* JAMA, 2017. **318**(6): p. 517-518.
86. Castelvecchi, D., *Can we open the black box of AI?* Nature News, 2016. **538**.
87. Ribeiro, M.T., S. Singh, and C. Guestrin, *Why should i trust you?: Explaining the predictions of any classifier*, in *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, ACM: San Francisco, CA, USA.
88. Kim, B., J. Shah, and F. Doshi-Velez, *Mind the Gap: A Generative Approach to Interpretable Feature Selection and Extraction.* Advances in Neural Information Processing Systems 28 (Nips 2015), 2015. **28**.
89. Grau, C. and M. Hoyer, *High-precision Radiotherapy.* Business Briefing: European Oncology Review, 2005.
90. Bortfeld, T. and R. Jeraj, *The physical basis and future of radiation therapy.* British Journal of Radiology, 2011. **84**(1002): p. 485-498.
91. El Naqa, I., et al., *Machine learning and modeling: Data, validation, communication challenges.* Medical Physics, 2018. **45**(10): p. E834-E840.
92. Lambin, P., et al., *Predicting outcomes in radiation oncology--multifactorial decision support systems.* Nat Rev Clin Oncol, 2013. **10**(1): p. 27-40.
93. Goodfellow, I.J., et al., *Generative Adversarial Nets.* Advances in Neural Information Processing Systems 27 (Nips 2014), 2014. **27**.
94. Blagus, R. and L. Lusa, *SMOTE for high-dimensional class-imbalanced data.* Bmc Bioinformatics, 2013. **14**.
95. Parker, B.J., S. Guenter, and J. Bedo, *Stratification bias in low signal microarray studies.* Bmc Bioinformatics, 2007. **8**.

96. Schiller, T.W., et al., *Modeling radiation-induced lung injury risk with an ensemble of support vector machines.* Neurocomputing, 2010. **73**(10-12): p. 1861-1867.

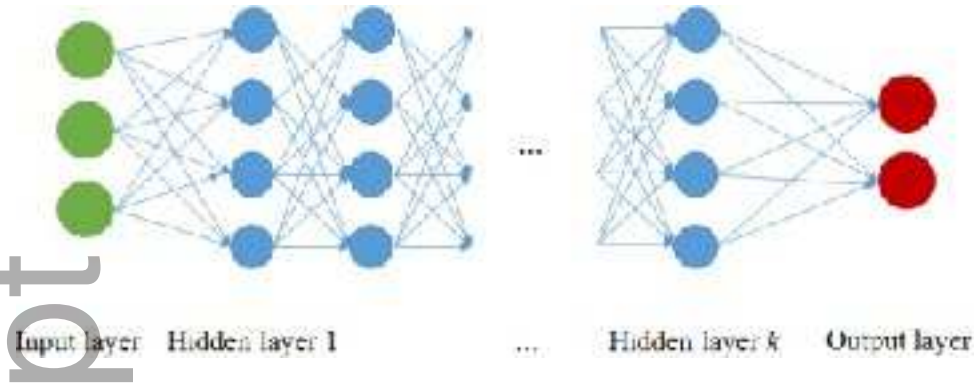
97. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) RESPONSE.* Annals of Internal Medicine, 2015. **162**(10): p. 735-736.

Table 1 Patients' input features and radiation treatment outcome

Patients' ID	Local Control	Tumor Dose	Age	Chemo	…
Patient #1	1 (yes)	75.60	78	0 (no)	…
Patient #2	0 (no)	71.39	70	1 (yes)	…
Patient #3	1 (yes)	66.00	56	1 (yes)	…
…	…	…	…	…	…



mp_13570_f1.tif



mp_13570_f2.tif