



On the Fuzziness of Machine Learning, Neural Networks, and Artificial Intelligence in Radiation Oncology

By Issam El Naqa, Associate Editor, Kristy Brock, Associate Editor, Yan Yu, Associate Editor, Katja Langen, Associate Editor, Eric E. Klein, Senior Editor

Received Jun 6, 2017. Accepted for publication Jun 8, 2017

Radiation oncology has been hailed as a potential vanguard for guiding Big Data applications into cancer research, quality assessment, and clinical care. This was emphasized by the joint American Society for Radiation Oncology/National Cancer Institute/American Association of Physicists in Medicine workshop held in Bethesda, Maryland in 2015 (1). Radiation therapy provides a unique combination of clinical patient demographics, physical use of radiation, application of image guidance (“radiomics”), and biological markers (“radiogenomics”) generated over a treatment period that can span a few days to several weeks. The main themes of this inaugural workshop focused on the development of data resources and aggregation. The next step of this process is geared toward data analytics, with machine learning techniques as its computational vehicle.

Driven by this Big Data era, machine learning applications in radiation therapy have witnessed a tremendous surge, ranging from automated contouring, to motion management, to quality assurance, to outcomes modeling (2). This is due to the ability of these algorithms to detect complex patterns in heterogeneous datasets, with superior results when compared with traditional statistical methods. These methods should not be regarded simply as yet another set of toolkits for physicists to toy with but also as powerful means for clinicians to consider for developing decision support systems and designing future clinical trials (3, 4).

This Oncology Scan highlights modern application of machine learning in the challenging areas of (1) predicting radiation pneumonitis (RP) events after stereotactic body radiation therapy (SBRT) using boosting techniques (5), (2) brain image segmentation with deep learning methods (6), and (3) predicting toxicity using computed tomography (CT) radiomics with fuzzy logic (7).

Valdes 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. *Phys Med Biol* 2016. (5)

Summary: This study discusses the use of machine learning methods for predicting RP after early-stage SBRT. The case of RP with early-stage SBRT presents a specific challenge with its very small event rate (<5%). In the Valdes et al study, with 201 consecutive patients treated with SBRT for stage I non-small cell lung cancer and 8 cases of RP, 61 features were assessed, primarily related to known dosimetric and clinical factors. Three methods were considered: the commonly used Decision Trees and Random Forests, in addition to RUSboost (8), which allows handling of imbalanced data as in the case of RP in SBRT; RUSboost provided the relatively best performance, with 2 false-negatives and 28 false-positives. Some of the features that came out of these analyses included DLCO adj%, heart dose to 15 cm³, trachea dose to 4 cm³, and race. Some of these variables have been implicated before in the literature, but others have not been reported before (9), which would require further evaluations in independent datasets.

In general, the article presented a needed approach for better prediction of RP for SBRT patients. However, the study was limited to dosimetric and clinical factors and did not include imaging and biological variables, which have been implicated in several recent studies.

Comments: Several studies have been dedicated to predicting RP using analytical and data-driven approaches based on traditional logistic regression methods and machine learning techniques, primarily in cases of conventional fractionation (9). More recent studies have focused on including molecular biomarkers (10) or imaging (11).

However, the case of RP in early-stage lung cancer patients (in whom RP event rates have been limited) constitutes a specific challenge, probably explaining the variability in reported predictors of RP after SBRT in the literature from typical dose-volume factors, CT changes, to new biomarkers (12). In any case, more in-depth investigation into predictors of SBRT utilizing all available knowledge (clinical, dose-volume, imaging, biomarkers) is still needed. This is particularly true if SBRT is to be expanded into more advanced stages of non-small cell lung cancer patients. Methods based on machine learning, with their high discriminate power and ability to account for data imbalance, are likely to play an important role in predicting radiation therapy outcomes, including RP (3).

Moeskops et al. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans Med Imaging* 2016. (6)

Summary: The challenges of automatic classification of images by machine learning methods have received much attention in the artificial intelligence community. Perhaps most widely known is the ImageNet Large Scale Visual Recognition Challenge, which has served as a public platform for research teams to compete annually for best algorithms that classify and detect objects on the basis of millions of annotated images (13). The success rate of deep learning with convolutional neural network (CNN) image classification is now approaching or exceeding human abilities (14, 15). In the study by Moeskops et al (6), the investigators evaluated a multiscale CNN method for automatic segmentation of magnetic resonance (MR) brain images. In contrast to feature-based approaches, CNN does not require explicit extraction of, for example, intensity, shape, and texture features, but rather applies predefined (or trained) kernels of various patch sizes. In this study, 3 sizes of image patches were used to retain spatial information and to obtain local neighborhood voxels, with correspondingly trained kernel sizes. The weights and biases in the CNN were specifically optimized for each patch size and corresponding kernel size. The method was applied to the segmentation of 5 different sets of images: 3 sets of volumetric-weighted MR brain images of preterm infants and 2 sets of volumetric-weighted MR brain images of adults. The method achieved accurate segmentations in terms of Dice coefficients for all tissue classes with sufficient training data. To illustrate the multiscale problem of anatomic segmentation, it was noted that spatially inconsistent results were obtained for the segmentation of the hippocampus using only the smallest patch size, whereas the largest patch size showed better consistency; yet still, the result with all patches combined showed the most accurate segmentation.

Comments: Deep machine learning algorithms as part of the Big Data framework (16) have for some time entered

our vocabulary in the field of radiation oncology. Yet the exact meaning of such terms is still subject to some interpretation. A simple fact is that fully annotated and accurately segmented very large image training sets of hippocampus or pancreas anatomy are harder to come by than, say, different breeds of dogs. To overcome such challenges, the notion of transfer learning has been used by many investigators as a means to train computer vision algorithms on available generic large databases (eg, ImageNet) using crowdsourcing and then transfer of such learning into domain-specific applications, like segmenting pancreas anatomy. In supervised image classification, practitioners and researchers clearly face a greater challenge than ImageNet, in that expertise in anatomy and radiology techniques is required for image annotation. For example, CNN has been successfully used in computer vision or speech recognition because of its direct adaptation to spectral feature maps (17); when CNN fails, the source of the data detects the error, and a telephone operator can take over. In radiation oncology, however, errors can be fatal; the operator is on standby yet always wanted elsewhere, such as actually seeing patients; and high-quality training data are sparse compared with Google or Facebook. In fact, it can be asserted that radiation oncology is the last frontier for researchers in artificial intelligence, machine learning, computer vision, and all related fields to make a palpable contribution to human health. For instance, one could envision that future generations of radiation oncologists will have increasingly accurate, automatic, and objective classification of critical structures, determination of the extent of the disease and geometric/biological target, spatially optimized dose prescription aided by the latest high-level clinical evidence, and the most robust, optimized dose delivery strategy. Their medical physics colleagues will ensure that this individually optimized, real-time, high-precision medicine is accurately executed. Such is a vision of the marriage of medical art and science.

Pota et al. Early predication of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. *Artif Intell Med* 2017. (7)

Summary: In this work, Pota et al evaluated 74 parotid glands (PGs) from 37 head and neck cancer patients treated with intensity modulated radiation therapy at 2 different institutions, used to gather data classified into 5 areas for all patients, as follows: (i) clinical data, collected before treatment, including age and geometric values associated with PG overlap and patient size, (ii) dosimetric data associated with the patient's individual plan, (iii) radiomics features based on the CT images acquired before RT, including first-order histogram parameters (mean, variance, and global entropy) as well as second-order histogram parameters calculated from the gray level co-occurrence matrix (GLCM) to measure patterns, organization, and textural complexity, (iv) variation in radiomics features

based on CT images acquired at mid-treatment and after radiation therapy using the same acquisition protocol as before treatment, and (v) parotid shrinkage, calculated for all patients and recorded and stratified by the PG volume variation rate. Two classification problems were evaluated; first, data (i)-(iv) were used to predict for (v), and second, data (i)-(v) were used to predict for the grade of late parotid toxicity, rated from no toxicity (0) to severe complication (3), which was available for a subset of 19 patients from 1 institution.

A novel statistical classification technique, likelihood-fuzzy analysis (LFA), was performed, and the results were compared with a well-established method, the naïve Bayes classifier. In summary, the LFA method uses a kernel estimation method to calculate the probability distribution of each variable given each class; each variable is partitioned into a fuzzy membership set, which is then approximated by a linear combination of Membership Functions. The proposed advantages of this approach include no prior class probability distribution assumption; a higher generalization power; interpretability of variable levels and predictor-outcome relations, robustness, and confidence-weighted results; ability to manage heterogeneous variables; and the ability to train and classify data with missing values. Evaluating individual predictors, previously reported predictors of parotid response, early volume variation, and initial volume was confirmed to predict for PG shrinkage. In addition, the half-thickness and early variations of dissimilarity were determined to be the best new predictors. Combined models were then evaluated for predicting PG shrinkage using LFA, leading to results that were greater than 0.8 in accuracy, sensitivity, and specificity. These values exceeded the best results from the naïve Bayes classifier. Similarly, higher levels of accuracy, sensitivity, and specificity were found with LFA compared with the naïve Bayes classifier for predicting 12-month xerostomia.

Comments: The use of radiomics for predicting radiation toxicity and tumor control, determining tumor boundaries, and evaluating tumor response is an emerging area of research showing great promise. Similar to the study highlighted above, Cunliffe et al (11) evaluated the correlation of radiomics features from thoracic CT with the development of RP. Using 2 diagnostic CT scans (1 before radiation therapy and 1 after radiation therapy) from 115 patients treated for esophageal cancer, 20 texture- and intensity-based features distributed among first-order, fractal, Laws' filter, and GLCM classes were calculated in corresponding regions of interest of all CT scans. Changes in the feature values were found to correspond with increasing radiation dose, 12 of which were found to be related to the development of pneumonitis, leading to a moderate discrimination rate. Studies in this area can aid in quantitative evaluation of toxicity and can potentially aid in the predictive models described in the summary above.

Recently, Lopez et al (18) reported on the association of radiomics and metabolic tumor volumes in the treatment of

glioblastoma multiforme with radiation therapy. Using images from 17 glioblastoma multiforme patients, they built a framework to investigate the association of image-based parameters, clinical target volumes, and metabolic maps of *N*-acetyl aspartate and choline from magnetic resonance spectroscopy imaging. Their results demonstrated that the metabolic information from the *N*-acetyl aspartate was very similar to information gained from standard imaging techniques; however, the choline provided additional information. Studies such as these will be key to determine the value added of additional imaging sequences for defining tumor boundaries and key areas to target.

Another application of radiomics is in the detection of recurrence, such as the work reported by Mattonen et al (19), which compared radiomics assessment with clinicians in the detection of lung cancer recurrence. Posttreatment CT scans of 45 patients were evaluated by 3 radiation oncologists and 3 thoracic radiologists. Twenty-two first-order features, 22 second-order grey-GLCM textures, and 16 size-based and shape-based features were calculated from the consolidative and periconsolidative regions generated semi-automatically. The clinicians demonstrated high sensitivity and moderate specificity, similar to the radiomics assessment; however, the median time to detection of the recurrence was 15.5 months for the clinicians, compared with the less than 6 months for the radiomics assessment. This study highlights the potential clinical impact that radiomics can have in the monitoring of patients using longitudinal imaging.

References

- Benedict SH, El Naqa I, Klein EE. Introduction to big data in radiation oncology: Exploring opportunities for research, quality assessment, and clinical care. *Int J Radiat Oncol Biol Phys* 2016;95:871-872.
- El Naqa I, Li R, Murphy MJ. *Machine Learning in Radiation Oncology: Theory and Application*. 1st ed. Cham, Switzerland: Springer International; 2015;336.
- Kang J, Schwartz R, Flickinger J, et al. Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. *Int J Radiat Oncol Biol Phys* 2015;93:1127-1135.
- Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol* 2013;10:27-40.
- Valdes G, Solberg TD, Heskell M, 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. *Phys Med Biol* 2016;61:6105.
- Moeskops P, Viergever MA, Mendrik AM, et al. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans Med Imaging* 2016;35:1252-1261.
- Pota M, Scalco E, Sanguineti G, et al. Early prediction of radiotherapy-induced parotid shrinkage and toxicity based on CT radiomics and fuzzy classification. *Artif Intell Med* 2017.
- Seiffert C, Khoshgoftaar TM, Hulse JV, et al. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Trans Syst Man Cybern A* 2010;40:185-197.
- Marks LB, Bentzen SM, Deasy JO, et al. Radiation dose-volume effects in the lung. *Int J Radiat Oncol Biol Phys* 2010;76:S70-S76.

10. Luo Y, El Naqa I, McShan DL, et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. *Radiother Oncol* 2017;123:85-92.
11. Cunliffe A, Armato SG 3rd, Castillo R, et al. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *Int J Radiat Oncol Biol Phys* 2015;91:1048-1056.
12. Yamashita H, Takahashi W, Haga A, et al. Radiation pneumonitis after stereotactic radiation therapy for lung cancer. *World J Radiol* 2014;6:708-715.
13. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis* 2015;115:211-252.
14. Roth HR, Farag A, Lu L, et al. Deep convolutional networks for pancreas segmentation in CT imaging. 2015;9413:94131G-94131G-94138.
15. Bar Y, Diamant I, Wolf L, et al. Deep learning with non-medical training used for chest pathology identification. 2015;9414:94140V-94140V-94147.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-444.
17. Abdel-Hamid O, Deng L, Yu D. Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition. Paper presented at the 14th Interspeech Conference; August 25-29, 2013; Lyon, France.
18. Lopez CJ, Nagornaya N, Parra NA, et al. Association of radiomics and metabolic tumor volumes in radiation treatment of glioblastoma multiforme. *Int J Radiat Oncol Biol Phys* 2017;97:586-595.
19. Mattonen SA, Palma DA, Johnson C, et al. Detection of local cancer recurrence after stereotactic ablative radiation therapy for lung cancer: Physician performance versus radiomic assessment. *Int J Radiat Oncol Biol Phys* 2016;94:1121-1128.