



Radiogenomics of lung cancer

Chi Wah Wong^{1,2}, Ammar Chaudhry¹

¹Department of Applied AI and Data Science, ²Department of Diagnostic and Interventional Radiology, City of Hope National Medical Center, Duarte, CA, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Ammar Chaudhry, MD. 1500 E. Duarte Rd., Duarte, CA 91010, USA. Email: achaudhry@coh.org.

Abstract: Machine learning (ML) and artificial intelligence (AI) are aiding in improving sensitivity and specificity of diagnostic imaging. The rapid adoption of these advanced ML algorithms is transforming imaging analysis; taking us from noninvasive detection of pathology to noninvasive precise diagnosis of the pathology by identifying whether detected abnormality is a secondary to infection, inflammation and/or neoplasm. This is led to the emergence of “Radiobiogenomics”; referring to the concept of identifying biologic (genomic, proteomic) alterations in the detected lesion. Radiobiogenomic involves image segmentation, feature extraction, and ML model to predict underlying tumor genotype and clinical outcomes. Lung cancer is the most common cause of cancer related death worldwide. There are several histologic subtypes of lung cancer, e.g., small cell lung cancer (SCLC), non-small cell lung cancer (NSCLC) (adenocarcinoma, squamous cell carcinoma). These variable histologic subtypes not only appear different at microscopic level, but these also differ at genetic and transcription level. This intrinsic heterogeneity reveals itself as different morphologic appearances on diagnostic imaging, such as CT, PET/CT and MRI. Traditional evaluation of imaging findings of lung cancer is limited to morphologic characteristics, such as lesion size, margins, density. Radiomics takes image analysis a step further by looking at imaging phenotype with higher order statistics in efforts to quantify intralesional heterogeneity. This heterogeneity, in turn, can be potentially used to extract intralesional genomic and proteomic data. This review aims to highlight novel concepts in ML and AI and their potential applications in identifying radiobiogenomics of lung cancer.

Keywords: Machine learning (ML); artificial intelligence (AI); radiomics; radiogenomics; lung cancer

Submitted Mar 31, 2020. Accepted for publication Apr 20, 2020.

doi: 10.21037/jtd-2019-pitd-10

View this article at: <http://dx.doi.org/10.21037/jtd-2019-pitd-10>

Introduction

Artificial intelligence (AI) enables us to interpret large volume of data into a machine learning (ML) or statistical model without an explicit need of a hypothesis as in traditional research. The era of data science and modern inter-disciplinary research have placed an increasing burden on analytical methodologies and we have to ensure any analysis that we perform is rigorous and correct. To design an AI workflow that is scalable and generalizable to new patients within and between hospital sites, it is critical to be aware of data science best practices and abide

by the principles along the way (1-3) to ensure accurate radiogenomic analysis.

Radiogenomic workflow

In a classical sense, a radiomic or radiogenomic pipeline involves image registration and pre-processing, image segmentation and annotation, feature extraction, and searching for an optimal ML or statistical model to correlate with or to predict a variety of genotype and clinical outcomes (4-12). If the models are to be deployed, the pipeline will be tested prospectively and if possible,

also in an external site. As new data come in, the model in production will be adjusted and improved.

Standardizing data and generating radiomic features

If the imaging data are collected consistently using the same scanner and protocol, data quality and image contrast should be relatively stable across patients. However, this is usually not the case in clinical settings, especially when data come from multiple centers or public data repositories. For instance, using a phantom and cohort of NSCLC patients, a study compared the variability of radiomic features derived from CT images from scanners under different vendors (13). They showed that the inter-scanner variability of the radiomic features can be comparable or even larger than inter-patient variability. We need to homogenize the images to reduce unwanted variability such that the radiomic features calculated from standardized image data have sufficient inter-patient variability for model development (14,15). Image pre-processing and standardization methods typically involve intensity normalization, re-sampling and image filtering (16-18).

Computational radiomic features are mostly sub-visual, they can be largely categorized into intensity, shape, and texture. In addition, we can apply spatial filters such as wavelets and Laplacian of Gaussian to extract a variety of derivative and spatial-frequency information. Multiple tools are available to conveniently extract radiomic features (8,19-22). However, features calculated from different radiomic software may affect downstream analysis and possibly the accuracy of clinical outcome prediction (23-25). Additional assessment may be necessary to identify strength and weakness of different radiomic tools when developing a predictive model.

Feature selection

The number of radiomic features is typically in orders of magnitude larger than the number of patients. Fitting all features into a system of equations will lead to overfitting and almost guarantee non-convergence and even if it converges, the model is likely not generalizable to new patient population. To avoid overfitting, we need to control for the degree of freedom consumed by the abundant input features. There have been various strategies for feature selection—namely dimensionality reduction using unsupervised learning, forward or backward feature selection

(wrapper methods), filtering methods such as minimum redundancy maximum relevance (26), and directly embed feature selection in ML models using regularization and dropout (27,28). A pitfall of utilizing feature engineering is that different strategies will likely provide a different set of dominant features. We should not over interpret the significance of individual features in association with the outcome. Alternatively, we can collectively regard the radiomic features set as a radiographic phenotype and relate that to the clinical phenotype and genotype.

Model assessment and external validation

A variety of statistical and ML models can be employed depending on the outcome variable. Using supervised learning, analysis typically falls into either prediction of categorical variables (such as survival groups or gene groups) using classification, and prediction of real-valued variables using regression.

Different assessment options are available for different types of models. In inferential statistics, such as multiple linear regression and survival analysis, P values are often used to depict the significance of association between an input variable and the outcome variable. Associating many radiomic features with every other genomic/molecular/prognostic feature result in multiple statistical analyses. It is a good practice to incorporate multiple comparisons correction to prevent inflation of false positive rates (29).

A common strategy in radiogenomics to predict phenotype or genotype is to cluster the outcome variable into two or more groups and then utilize classification models to make predictions (30-36). The model prediction for classification problems are usually real-valued. Threshold is applied to calculate true positive, true negative, false positive and false negative values for model performance assessment. One of the most commonly used metrics to evaluate classification models is area under the ROC curve (AUROC). The AUROC illustrates the relation of true positive rate and false positive rate when different thresholds are applied to the predicted value of the classification model. Although AUROC is widely accepted, depending on the need, sometimes AUROC alone may not be enough as the only metric and the area under the precision recall curve (AUPRC) could be a useful complement (37), especially when the dataset has imbalanced outcome (38). In addition, since the prediction from the classification model is continuous, we can utilize this score as a proxy for overall survival and then stratify

patients into multiple risk groups with a different number of bins than the original classification model.

An alternative to using classification model is to directly predict survival as a real-valued outcome using ML. Popular routines such as elastic net, support vector machine and tree-based models can be adopted to perform survival analysis (39,40). Metrics such as Harrell's concordance index, or time-dependent cumulative-dynamic AUC can be used for model performance assessment (41,42).

When we develop ML models, an option is to divide a dataset into a training set and a test set. The model is developed using the train set and then applied to the test set for performance evaluation. To be more rigorous, we typically utilize a cross validation approach by dividing the dataset into a cross-validation set and a test set. Within the cross-validation dataset, we can identify the most related set of input features, the best performing ML model with the associated hyperparameters. The commonly used cross-validation strategies are n-fold and leave-one-out. The cross-validated feature set and model are then applied on the test set for a final assessment to see how well the model may be generalized. It is critical that the test set is left completely untouched in the entire feature selection and modeling process and is only used for generalizability testing of the final model. If the best practices are strictly followed, we have more confident that the model is applicable in a prospective manner.

Even if the model is generalizable within site, it does not imply the model will perform well externally. To allow models to be widely adopted, we need to ensure our work are reproducible (4,43,44). In addition, the models should be validated using external data and site-related variability should be assessed and compensated (7,45,46).

Deep learning and autoML

With the advancement of CPU and graphic processing unit (GPU) technologies, deep learning and automated ML ("autoML") approaches are becoming more popular as computationally intensive algorithms can be parallelized and optimized using scalable hardware. Deep learning is an attractive alternative to classical radiomics in that the original images can be directly fed into the neural network configuration with minimal pre-processing (30,47-49). For instance, a clinically validated study using deep learning on low dose CT images achieved 94.4% AUROC on predicting the risk of lung cancer (47). Another multiple timepoint study using convolutional and recurrent neural

networks (CNN and RNN) on CT images acquired on NSCLC patients can predict survival and various cancer-related outcomes including progression, metastases and recurrence with acceptable performance (30). On the other hand, autoML allows scientists to train a large number of ML or deep learning models with little coding (50). The caveat is that the autoML models may not perform better than models designed and trained by researchers with domain and computational science expertise. As autoML is like a black-box model and interpretability may not be trivial, we have to pay extra attention to make sure the data we feed in do not cause discriminatory issues or harm to the patients (50).

Lung cancer and radiogenomics

Lung cancer is the most common cause of cancer related death worldwide (51). Lung cancer is usually diagnosed on medical imaging [radiographs or computed tomography (CT)] with imaging findings usually describing presence of a space occupying lesion within the lung parenchyma and its relationship to surrounding tissues (pleural, ribs, hilum, etc.). Detected lesions are usually biopsied to confirm diagnosis of cancer as well as for histologic characterization of the tumor [small cell lung cancer (SCLC), non-small cell lung cancer (NSCLC), etc.] (51). Although tissue-based diagnosis is the current gold standard, it is important to be mindful of its limitations; e.g., sampling error, biologic heterogeneity between primary neoplasm and metastatic site, etc. (51). Radiogenomics allows for structural characterization of the lesions as well as provide functional information about the tumor including tumor biology (51,52). Additionally, radiogenomic map of the lesion can help characterize intralesional heterogeneity (52). Radiogenomics can be performed using multimodal (X-rays, CT, PET, MRI) and/or multiparametric (multiple MRI sequences, e.g., diffusion MRI, perfusion MRI, etc.) techniques (52). Furthermore, recent studies have shown that in addition to increasing diagnostic specificity, radiogenomics can aid in treatment selection and prognostication (51,52). With emergence of targeted therapeutics [e.g., anti-programmed death receptor 1 (anti-PD1), anti-programmed death ligand 1 (anti-PDL1), etc.], radiogenomics powered by ML and AI can potentially help identify targeted therapeutics patients can benefit from. Recent studies have shown radiomics of NSCLC can help in early cancer detection, evaluate treatment efficacy, and predict treatment-related outcomes (51-55). Additionally,

radiomic approach can be used to evaluate outcomes in patients treated with radiotherapy. Zhou *et al.* (52) recently showed efficacy of radiogenomics as noninvasive diagnostic and prognostic biomarkers in lung cancer. For example, radiogenomics can be used to identify epidermal growth factor receptor (EGFR) expression in lung cancer as well as other transcriptional factors (52). Zhou *et al.* also demonstrated that lung cancer CT Hounsfield attenuation and lesion margins correlated with cell-cycle genes. Zhou *et al.* noted that presence of irregular borders and ground glass opacities in the lesion correlated with EGFR expression in lung cancer. More recently, Forghani *et al.* showed radiomics and AI platform can be utilized to help non-invasively differentiate between adenocarcinoma and squamous cell carcinoma of the lung (56).

Conclusions

Radiogenomic analysis of NSCLC showed multiple associations between semantic image features and metagenes that represented canonical molecular pathways, and it can result in noninvasive identification of molecular properties of NSCLC.

Acknowledgments

The authors would like to take this opportunity to thank the Editorial Board of *Journal of Thoracic Disease* for this unique opportunity. Also, the authors would like to thank the staff of *Journal of Thoracic Disease* for the endless support in organizing this series.

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editor (Ammar Chaudhry) for the series “Role of Precision Imaging in Thoracic Disease. The article was sent for external peer review organized by the Guest Editor and the editorial office.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at: <http://dx.doi.org/10.21037/jtd-2019-pitd-10>). The authors have no conflicts of interest to declare. The series “Role of Precision Imaging in Thoracic Disease” was commissioned by the editorial office without any funding or sponsorship. AC served as the unpaid Guest Editor of the series.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Blei DM, Smyth P. Science and data science. *Proc Natl Acad Sci U S A* 2017;114:8689-92.
2. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science* 2015;349:255-60.
3. Kang J, Rancati T, Lee S, et al. Machine Learning and Radiogenomics: Lessons Learned and Future Directions. *Front Oncol* 2018;8:228.
4. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016;6:23428.
5. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-62.
6. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278:563-77.
7. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
8. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017;77:e104-7.
9. Aerts HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* 2016;2:1636-42.
10. Rudie JD, Rauschecker AM, Bryan RN, et al. Emerging Applications of Artificial Intelligence in Neuro-Oncology. *Radiology* 2019;290:607-18.
11. Langlotz CP, Allen B, Erickson BJ, et al. A Roadmap for Foundational Research on Artificial Intelligence in

- Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019;291:781-91.
12. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012;30:1234-48.
 13. Mackin D, Fave X, Zhang L, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol* 2015;50:757-65.
 14. Lambin P. Radiomics Digital Phantom. *CancerData*, 2016.
 15. Zhao B, Tan Y, Tsai WY, et al. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl Oncol* 2014;7:88-93.
 16. Shafiq-Ul-Hassan M, Latifi K, Zhang G, et al. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep* 2018;8:10545.
 17. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS One* 2017;12:e0178524.
 18. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys* 2017;44:1050-62.
 19. Pfähler E, Zwanenburg A, de Jong JR, et al. RaCaT: An open source and easy to use radiomics calculator tool. *PLoS One* 2019;14:e0212223.
 20. Nioche C, Orhac F, Boughdad S, et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res* 2018;78:4786-9.
 21. Dinapoli N, Alitto AR, Vallati M, et al. Moddicom: a complete and easily accessible library for prognostic evaluations relying on image features. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:771-4.
 22. Gotz M, Nolden M, Maier-Hein K. MITK Phenotyping: An open-source toolchain for image-based personalized medicine with radiomics. *Radiother Oncol* 2019;131:108-11.
 23. Liang ZG, Tan HQ, Zhang F, et al. Comparison of radiomics tools for image analyses and clinical prediction in nasopharynx cancer. *Br J Radiol* 2019:20190271.
 24. Caramella C, Allorant A, Orhac F, et al. Can we trust the calculation of texture indices of CT images? A phantom study. *Med Phys* 2018;45:1529-36.
 25. Berenguer R, Pastor-Juan MDR, Canales-Vazquez J, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* 2018;288:407-15.
 26. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226-38.
 27. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157-82.
 28. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23:2507-17.
 29. Nair VS, Gevaert O, Davidzon G, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res* 2012;72:3725-34.
 30. Xu Y, Hosny A, Zeleznik R, et al. Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. *Clin Cancer Res* 2019;25:3266-75.
 31. Tu W, Sun G, Fan L, et al. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer* 2019;132:28-35.
 32. Jia TY, Xiong JF, Li XY, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol* 2019;29:4742-50.
 33. Gu Q, Feng Z, Liang Q, et al. Machine learning-based radiomics strategy for prediction of cell proliferation in non-small cell lung cancer. *Eur J Radiol* 2019;118:32-7.
 34. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559-67.
 35. Emaminejad N, Qian W, Guan Y, et al. Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients. *IEEE Trans Biomed Eng* 2016;63:1034-43.
 36. Yamamoto S, Korn RL, Oklu R, et al. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. *Radiology* 2014;272:568-76.
 37. Avati A, Jung K, Harman S, et al. Improving palliative care with deep learning. *BMC Med Inform Decis Mak* 2018;18:122.
 38. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
 39. Pölsterl S, Navab N, Katouzian A. Fast Training of Support Vector Machines for Survival Analysis. In: Appice A, Rodrigues PP, Costa VS, et al. editors. *Machine Learning and Knowledge Discovery in Databases*. Cham: Springer International Publishing. 2015:243-59.
 40. Pölsterl S, Gupta P, Wang L, et al. Heterogeneous

- ensembles for predicting survival of metastatic, castrate-resistant prostate cancer patients. *F1000Res* 2016;5:2676.
41. Uno H, Cai T, Tian L, et al. Evaluating Prediction Rules for t-Year Survivors with Censored Regression Models. *J Am Stat Assoc* 2007;102:527-37.
 42. Hung H, Chiang CT. Estimation methods for time-dependent AUC models with survival data. *Can J Stat* 2010;38:8-26.
 43. Peng RD. Reproducible research in computational science. *Science* 2011;334:1226-7.
 44. Mesirov JP. Computer science. Accessible reproducible research. *Science* 2010;327:415-6.
 45. Orlhac F, Frouin F, Nioche C, et al. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. *Radiology* 2019;291:53-9.
 46. Orlhac F, Boughdad S, Philippe C, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. *J Nucl Med* 2018;59:1321-8.
 47. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;25:954-61.
 48. Lakshmanaprabu S, Mohanty SN, Shankar K, et al. Optimal deep learning model for classification of lung cancer on CT images. *Future Gener Comput Syst* 2019;92:374-82.
 49. Ravi D, Wong C, Deligianni F, et al. Deep Learning for Health Informatics. *IEEE J Biomed Health Inform* 2017;21:4-21.
 50. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digital Health* 2019;1:e232-42.
 51. Shi L, He Y, Yuan Z, et al. Radiomics for Response and Outcome Assessment for Non-Small Cell Lung Cancer. *Technol Cancer Res Treat* 2018;17:1533033818782788.
 52. Zhou M, Leung A, Echegaray S, et al. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. *Radiology* 2018;286:307-15.
 53. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. *Br J Radiol* 2017;90:20160642.
 54. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev* 2016;1:207-26.
 55. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol* 2017;86:297-307.
 56. Forghani R, Savadjiev P, Chatterjee A, et al. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput Struct Biotechnol J* 2019;17:995-1008.

Cite this article as: Wong CW, Chaudhry A. Radiogenomics of lung cancer. *J Thorac Dis* 2020;12(9):5104-5109. doi: 10.21037/jtd-2019-pitd-10