HHS Public Access

Author manuscript

Cold Spring Harb Perspect Med. Author manuscript; available in PMC 2021 August 05.

Published in final edited form as:

Cold Spring Harb Perspect Med. ; 11(8): . doi:10.1101/cshperspect.a039537.

Application of Radiomics and AI for Lung Cancer Precision Medicine

Ilke Tunali[1], **Robert J. Gillies**[1], **Matthew B. Schabath**[2]

[1]Department of Cancer Physiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, 33612, USA;

[2]Department of Cancer Epidemiology; H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, 33612, USA;

Abstract

Medical imaging is the standard-of-care for early detection, diagnosis, treatment planning, monitoring, and image-guided interventions of lung cancer patients. Most medical images are stored digitally in a standardized Digital Imaging and Communications in Medicine format that can be readily accessed and used for qualitative and quantitative analysis. Over the several last decades, medical images have been shown to contain complementary and interchangeable data orthogonal to other sources such as pathology, hematology, genomics, and/or proteomics. As such, 'radiomics' has emerged as a field of research that involves the process of converting standard-of-care images into quantitative image-based data that can be merged with other data sources and subsequently analyzed using conventional biostatistics or artificial intelligence (AI) methods. As radiomic features capture biological and pathophysiological information, these quantitative radiomic features have shown to provide rapid and accurate non-invasive biomarkers for lung cancer risk prediction, diagnostics, prognosis, treatment response monitoring and tumor biology. In this chapter, radiomics and emerging AI methods in lung cancer research are highlighted and discussed including advantages, challenges, and pitfalls.

Keywords

Radiomics; Lung Cancer; Quantitative Imaging

Introduction

Medical imaging is the standard-of-care for early detection, diagnosis, treatment planning, monitoring, and image-guided interventions of lung cancer patients. Most medical images are stored digitally in a standardized Digital Imaging and Communications in Medicine (DICOM) format that can be readily accessed and used for qualitative and quantitative

*Corresponding Author: Matthew B. Schabath, Ph.D., H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive MRC-CANCONT, Tampa, FL 33612, Fax: 1-813-745-4150, Matthew.Schabath@Moffitt.org.

Conflict of Interest

R.J. Gillies is an investor and member of the Advisory Board at HealthMyne, Inc., and has Research support form Helix Biopharma. No potential conflicts of interest were disclosed by the other authors.

analysis. Computerized image-based feature extraction from medical images has been an area of research over the last several decades to create computer-aided diagnosis (CAD) systems (Dhawan et al. 1986; Giger et al. 1988; Kawata et al. 1998; Hardie et al. 2008). However, in today's conventional clinical radiology practice, only a few quantitative metrics are used to describe the phenotype of nodules or tumors. This limited list of quantitative metrics with respect to pulmonary lesions includes: 1) computed tomography (CT)-based largest diameter of nodules detected incidentally or in a screening setting following the Fleischner Society (Bueno et al. 2018), or Lung-RADS (ACR) guidelines respectively, 2) CT-based largest diameter of tumors following the Response Evaluation Criteria in Solid Tumors, RECIST (Eisenhauer et al. 2009), 3) standardized uptake value (SUV) derived metrics from positron emission tomography (PET), and 4) percent enhancement of lesions on magnetic resonance imaging (MRI).

Over the last decade, there has been emerging evidence that medical images contain complementary and interchangeable data orthogonal to other sources such as pathology, hematology, genomics, and/or proteomics (Gillies et al. 2016). 'Radiomics' is an emerging field of research that involves the process of converting standard-of-care images into quantitative image-based biomarker that can be merged with these orthogonal clinical and other 'omic-based data and subsequently analyzed using conventional biostatistics or artificial intelligence (AI) methods. As radiomic features capture biological and pathophysiological information from these region-of-interests (ROIs), these quantitative radiomic features have shown to provide rapid and accurate non-invasive biomarkers for lung cancer risk prediction, diagnostics, prognosis, treatment response monitoring, and tumor biology. In this chapter, radiomics and emerging AI methods in lung cancer research are highlighted and discussed including advantages, challenges, and pitfalls.

The Radiomics Pipeline

Medical images can be analyzed by either extracting quantitative features from identified and delineated ROIs or by analyzing an entire image or image series. The former is often referred to as "conventional" radiomics where quantitative image features are extracted from a segmented ROI and analyzed while the latter is often driven by emerging deep learning (DL) methods which do not always depend on segmentation of an ROI. With respect to conventional radiomics, there are five fundamental steps: (1) Image acquisition and digitization, (2) ROI selection and segmentation, (3) Quantitative feature extraction, (4) Biomarker discovery/training, and (5) Test and validation (Figure 1a).

The first step involves image-data acquisition using standard-of-care medical images. Since imaging is widely utilized in oncology care, radiomics has substantial utility with the readily available image-data that exist in every radiology clinical PACS server. However, image acquisition protocols vary widely within and across radiology practices. Though it is advantageous to have homogeneous image acquisition parameters in an analytical cohort, there are downstream image processing steps and analytical methods that can be applied to address heterogeneous image acquisitions, often resulting in more robust predictive models than those obtained with a single homogeneous acquisition protocol (Mu et al. 2019b). For instance, resampling all image voxels on a dataset to a chosen voxel size is a well-known

and applied technique in majority of the radiomic studies to increase robustness of features that are subsequently extracted (Shafiq-Ul-Hassan et al. 2018). When texture features are calculated, a common practice of relative discretization (i.e., clustering of the pixels in the image to a fixed number of bins) or absolute discretization (i.e., clustering of the pixels in the image to a fixed bin size) have also shown to substantially impact the reproducibility of PET (Leijenaar et al. 2015), CT (Larue et al. 2017; Shafiq-Ul-Hassan et al. 2017) and MRI (Duron et al. 2019) derived features. Normalization of radiomic features (e.g., by total number of ROI voxels, a single voxel volume, etc.) that are sensitive to voxel size and image acquisition parameters have also shown to improve the robustness of radiomics across different scanners as well of eliminating the dependence on volume (Shafiq-Ul-Hassan et al. 2017; Shafiq-Ul-Hassan et al. 2018).

The second step involves the ROI selection and segmentation. An ROI may include one or more “index” nodules or tumors, metastatic lesions, or whole organs in 3-dimenational space. As such, it creates an advantage of radiomics over tissue-based biomarkers due to reflection of the entire ROI in 3-dimensional space (e.g., tumor and surrounding parenchyma) compare to just a portion of tumor(s) that was captured on a biopsy. Although there are clear guidelines for “index” lesion selection (Eisenhauer et al. 2009), a radiologist or a very experienced imaging scientist is ideally best suited for ROI selection. Additionally, over-reads of 5 to 10% of all images should be considered and conducted by an independent radiologist and this can result in changing the identity of the “index” lesions. Segmentation of the ROI can be performed using manual, semi-automated, or fully automated approaches. Manual segmentation is laborious and although it can be very accurate when conducted by an experienced radiologist, inter-reader studies show poor reproducibility with Dice coefficients often < 80% (Alilou et al. 2017). Semi-automated (e.g., using an initial parameter such as a single click on the center of a lesion) or fully automated segmentation approaches are faster and more repeatable (Kalpathy-Cramer et al. 2016b; Tunali et al. 2019c) but may still require manual verification and corrections if the segmentations algorithm fails on difficult cases. As some radiomic features are extremely sensitive to segmentation, a segmentation algorithm that produces accurate, reproducible and consistent segmentations in an automated manner and requires minimal user input is a critical requirement in order to increase stability of image-based features (Kalpathy-Cramer et al. 2016a; Tunali et al. 2019c).

The third step is feature extraction where image-based features are calculated from the ROI. The various classes of features have been described in detail elsewhere (Gillies et al. 2016). Briefly, radiomic features are classified into first-, second-, and higher-order features. First-order features include shape- and size-based features and features that describe the distribution of values of individual voxels (i.e., volumetric pixels) without concern for spatial relationships. Shape and size-based features consist of the larger-scale form of the nodule (or tumor) such as sphericity, compactness, surface area, surface to volume ratio, and volume. Nodule/tumor voxel intensities describe the histogram of intensities within the CT image through basic metrics such as mean, median, maximum, minimum, uniformity, or randomness (entropy) of the intensities in the image, as well as the skewness (asymmetry) and kurtosis (flatness) of the histograms of values. Second-order image features are described as “texture” features and they describe statistical interrelationships between voxels

with similar (or dissimilar) contrast values. Texture features consist of image consistency and patterning textures that quantify intra-nodule (or intra-tumor) heterogeneity differences observable within the segmented nodule, tumor, or stromal volume. In practice, there are dozens of methods and multiple variables that can be used to extract texture features, resulting in hundreds of values—far too many to elaborate here. Higher-order features are also described as “texture” features and impose filter grids on the image to extract repetitive or non-repetitive patterns. These include fractal analyses, Minkowski functionals, wavelets, and Laplacian transforms of Gaussian bandpass filters. These generally describe enhanced textural information within the tumor/nodule volume through an iterative process of decomposing the original image into different frequency ranges/scales and then feature extraction occurs from the decomposed images to quantify textural and spatial intensity. In order to enable inter-operability of radiomic signatures (i.e., repeatable of signatures), differences in terminology, algorithms, software implementations, and other methodological facets must be explained distinctly. Recently, an international collaboration initiative called the image biomarker standardisation initiative (IBSI) has worked towards standardizing the extraction of image biomarkers to address part of this inter-operability problem by providing image biomarker nomenclature and definitions, benchmark data sets, and benchmark values to verify image processing and image biomarker calculations, as well as reporting guidelines (Zwanenburg et al. 2018).

The fourth step is the analytical steps of training or discovery of a model to predict the desired dependent variable (e.g., risk, prognosis, treatment response, etc.) or phenotype (e.g., gene mutation[s], molecular signature, protein expression, etc.) using conventional biostatistics or machine learning (ML) methods (Figure 2). The decision of what analytical method(s) to be deployed is dependent on many different factors including sample size (e.g., ML approaches may overfit smaller sample sizes), study endpoint (e.g., dichotomous vs. continuous or time-dependent), and the statistic or metric of interest (e.g., p-value-driven point estimate vs. performance statistics such as area under curve [AUC)]). There are no set rules and this aspect of radiomics is an area of intense research and debate. Regardless, it is often advantageous that the training model be integrated to include orthogonal information including patient data, clinical data, and other ‘omic information such as driver mutations, immunohistochemistry (IHC) proteomic data, or circulating biomarkers, when available. Image-based features that are non-reproducible can be eliminated (Balagurunathan et al. 2014a; Balagurunathan et al. 2014c; Kalpathy-Cramer et al. 2016a; Tunali et al. 2019c) and principal component analysis or clustering methods could be used to reduce dimensionality of the created models to avoid overfitting (Hosny et al. 2018).

The critical final step is the testing and validation of the model with independent data from internal and/or external datasets. A successful external validation can demonstrate the potential generalizability of the model as it may include different acquisition protocols and patient population(s). Defining a successful test and validated model is dependent on the statistic or metric of interest. For example, if the model was trained to find an integrated model with the highest AUC, then AUC obtained in the test and validation cohorts should not be significantly different from the trained model.

Artificial intelligence in medical imaging

AI is an umbrella term technology that includes ML and DL and enables machines to mimic human intelligence and consequently have downstream effect on transforming industries such as medicine. DL is a subset of ML, while ML is a subset of AI (Figure 2). AI has unprecedented success due to advances in central processing units (CPUs) and graphics processing units (GPUs), and the availability of larger data sets due to increased storages and digitization. AI methods have made substantial advances on interpreting complex and multi-dimensional data in various applications from stock trading to computer-vision and natural language processing to self-driving cars. This advances got its share within health care on various applications such as patient risk management, drug discovery, patient diagnosis and prognosis and medical image interpretations. Specifically, DL methods (e.g., convolutional neural networks) excel at image pattern recognition thus may quantify image information that is subtle to humans in the setting of medical imaging.

Generally, AI in medical image analyses can be applied two ways: i) radiomic features extracted from ROIs can be input into ML algorithms to develop classifiers (i.e., “conventional radiomics”), or ii) an entire image or image series can be an input into DL network to develop classifiers. DL methods are currently being explored in detection, characterizing, and monitoring of cancers. Lung cancer is one of the most prominently researched cancer type by AI due to its medical importance, an abundance of CT or PET/CT images, and the high-contrast high-resolution inherent in CT images (Hosny et al. 2018). Although DL and conventional ML radiomic approaches have similar endpoints in terms of medical image analysis, they differ substantially in terms of training methods. As previously mentioned, “conventional radiomics” requires proper segmentation of the ROI, engineered feature extraction and selection for optimized model creation whereas DL methods do not require ROI annotation but rather uses a single seed point or bounding box that can be identified by an expert human observer or a separately trained AI detection system.

AI methods have already been deployed with great success in radiomics studies. However, to successfully utilize AI methods into clinical setting, several challenges remain. First, large patient and/or image datasets are crucial to the success of any AI approach. This is especially the case with deep learning model, wherein large number of layers are used (Figure 1b). Small datasets are prone to being overfitted, ungeneralizable, and non-reproducible. Since medical imaging is widely deployed in developed countries, access to large imaging datasets should be a real-world reality. However, a critical barrier to access is the time required to annotate and curate such datasets, which is well recognized (Gillies et al. 2016). For DL approaches, data can be “augmented” with a number of approaches (Napel et al. 2018). One of the most widely used approaches is “transfer learning”, where a neural network model is first trained on a similar problem to identify features such as edges, sharpness, etc. and afterwards this network is either reused (partly or completely) or tuned to be adapted for the new task. Another approach is using a large public repository such as The Cancer Imaging Archive (TCIA), or the National Lung Screening Trial (NLST) with annotated medical images to pre-train a neural network and tune it for a specific task. An alternative approach to data augmentation is to artificially inflate the training dataset size by warping, rotating, or inverting the images, which theoretically can be used to overcome the

lack of high sample numbers. A more recent development is to augment image data using Generative Adversarial Networks (GANs) that create new data by generating images with a different modality, such as contrast enhanced to non-enhanced CT, or MRI to CT (Sandfort et al. 2019). However, even with augmentation, data can be insufficient when rare diseases or small data sets are being modelled. An alternative way to tackle this is through a centralized databases or using a distributed learning platform where the 'code' is shared instead of the data (Lambin et al. 2017). One other potential limitation with DL methods is that these systems are a "black-box" in that there is a lack of transparency as how the networks perform the various tasks. However, Chartrand et al. (Chartrand et al. 2017) argues that a highly accurate opaque system is desirable to an inaccurate transparent one, that users may never understand how these networks work, and AI will likely identified patterns those humans cannot interpret. Another limitation is that DL methods are optimized for binary classifications rather than time-dependent endpoints based on a continuous scale, such as survival related outcomes.

Radiomics and AI in early detection/lung cancer screening

Worldwide, lung cancer is the most common diagnosed cancer and leading cause of cancer-related death (Bray et al. 2018). Because of the large number of diagnoses, even incremental improvements in patient outcomes would have profound impact. There is a vast difference in survival outcomes between early stage versus late stage patients where for a localized lung cancer have a 56% 5-year OS and only 5% for distant metastasized diseases (Siegel et al. 2019). Thus, detecting lung cancers early, when they are manageable, or even curable, is a critical need. Asymptomatic lung cancers are detected either incidentally, e.g., when a patient receives an imaging study for another indication, or via a screening program that is designed for individuals who are at high risk. In 2011, results from the National Lung Screening Trial (NLST) demonstrated a 20% relative reduction in lung cancer mortality for individuals screened by LDCT compared with standard chest radiography in a high-risk population of 53,454 current and former smokers ages 55 to 74 years (National Lung Screening Trial Research et al. 2011; Silva et al. 2017; De Koning et al. 2018; Pastorino et al. 2019). More recently, in 2018 the initial results of the Nederlands-Leuvens Longkanker Screenings ONderzo ek (NELSON) trial also showed significant declines in lung cancer mortality (De Koning et al. 2018). Moreover, in 2019 two additional randomized trials conducted in Italy called the Multicentric Italian Lung Detection (MILD) trial (Pastorino et al. 2019) and in Germany called The German Lung cancer Screening Intervention (LUSI) trial (Becker et al. 2019) were published, providing additional validation of lung cancer screening value. As such, the cumulative evidence based on the results of three published trials and anticipated publication of the NELSON trial has demonstrated substantial beneficial mortality reductions associated with LDCT screening.

Despite the mortality benefit associated with lung cancer screening, there are many limitations of early detection by LDCT (De Koning et al. 2018) including detection of large number of indeterminate pulmonary nodules and overdiagnosis of indolent neoplasms that may not otherwise cause clinical symptoms or death (Schabath and Gillies 2015). Additionally, though small pulmonary nodules (< 6 mm) are considered to be "negative" in lung cancer screening, prior analyses have shown that NLST participants who had baseline

negative screens and were diagnosed with incidental lung cancer in follow-up screening had poorer survival rates compared to screening participants who had a baseline positive screen that developed incidence lung cancer (Schabath et al. 2015; Patz et al. 2016).

Developing biomarkers for diagnostic discrimination between malignant tumors and benign nodules is not unique to imaging research and studies have been conducted for decades in this space (Brzakovic et al. 1990; Hadjiiski et al. 1999; Wei et al. 2005). With the emergence of lung cancer screening, there has been a substantial increase in efforts to develop non-invasive image-based classifiers. As such, after the NLST was publicly available, many studies were utilized from patient medical images on this immense cohort. Following is just a small sample of the research that this data set has made possible (Table 1). One of the earliest studies conducted by Hawkins et al. (Hawkins et al. 2016) utilized baseline LDCT scans from the NLST to predict which baseline indeterminate (4–12 mm) pulmonary nodules would subsequently be diagnosed as an incident lung cancer in follow-up screening intervals. The authors developed a machine-learning model of 23 features that yielded a radiomic signature with an under the curve (AUC) of 0.81 for predicting development of cancer in 1 year which was far superior to volume alone (AUC = 0.72) which is the most widely used predictive marker in clinic. Peikert et al. (Peikert et al. 2018) created a radiomics model to compare malignant and benign screen-detected indeterminate lung nodules utilizing the NLST dataset. Using least absolute shrinkage and selection operator (LASSO) multivariable analysis, they reported an AUC of 0.939 could be reached with only 8 non-redundant radiomic features. Huang et al. (Huang et al. 2018) utilizing the NLST to perform a matched case-control study and to identify CT image features to increase the positive predictive value (PPV) and reduce the false positive (FP) rates compared to thoracic radiologist evaluations. Cherezov et al. (Cherezov et al. 2018) utilized images and data from the NLST and improved malignancy prediction accuracy from 74.7% to 81.0% by implementing nodule size-specific models. In their study they used Synthetic Minority Oversampling Technique (SMOTE) to overcome the class imbalances which are inherent in these datasets. Chae et al. (Chae et al. 2014) utilized texture features to differentiate pre-invasive lesions from invasive pulmonary adenocarcinomas that are marked as part-solid ground-glass opacities (GGOs) on chest CT scans. Their artificial neural network (ANN) model showed an excellent performance using five radiomic features with an AUC of 0.981 on 86 part-solid GGOs. Liu et al. (Liu et al. 2017) extracted semantic features (i.e., radiological traits quantified by radiologists) from baseline nodules in the NLST and developed a model that predicted which participants would be diagnosed with lung cancer 1 to 2 years (AUC = 0.80) after the baseline screen. Ardila et al. (Ardila et al. 2019) conducted a DL network to predict the risk of lung cancer utilizing the NLST dataset and their model achieved an AUROC of 94.4% which was validated on an independent validation dataset. Also utilizing images from cancer patients in the NLST, Morales et al. (Morales et al. 2019) identified two radiomic features that stratified patients into three risk-groups (e.g., low, intermediate, and high) with an AUC of 0.878. One of the features was found to be associated with *FOXF2* expression which has been shown to be associated with poor prognosis. Dhara et al. (Dhara et al. 2016) utilized 891 nodules from the Lung Image Database Consortium and Image Database Resource Initiative database and classified malignant versus benign nodules using support vector machine (SVM). Their models

reached an AUC of 0.951 which outperformed methods that required manual segmentation of a trained radiologist. Other studies have tried addressing the issue of overdiagnosis by utilizing quantitative features. Maldonado et al. (Maldonado et al. 2013) developed a decision algorithm called CANARY that binned pulmonary nodules as aggressive or indolent. Their algorithm had a validation sensitivity of 98.7% and a specificity of 63.6%. Finally, Lu et al. (Lu et al. 2019) identified a model that included features that were extracted from the tumor and the "difference region" (i.e., the part-solid region of the tumor) and yielded an AUC of 0.846 to discriminate aggressive vs. indolent nodules. The summary of these studies can be found in Table 1.

Radiomics and AI in Prognostication

Though pathologic staging remains to be the most important prognostic factor for lung cancer survival (Mirsadraee et al. 2012)., there is marked variability in patient outcomes and survival among patients with the same stage of disease suggesting that other factors contribute to lung cancer survival, progression, and recurrence (Birim et al. 2006; Ries 2007; Kachroo et al. 2008; Dela Cruz et al. 2011; Pao and Girard 2011; Schabath et al. 2014). Accurately classifying the aggressiveness of tumors is critical as it can help physicians under the potential prognosis of a patient beyond just stage alone and provide options to choose between curative and palliative treatments as well as the aggressiveness of the therapy and follow-up. Radiomic studies have shown that image-based classifiers have the potential to complement staging and improve prognostication of lung cancer. Aerts et al. (Aerts et al. 2014) analyzed NSCLC and head and neck patients and validated a CT radiomic signature that had better prognostic performance than TNM staging and volume with a concordance index of 0.65. They found associations between their signature features and gene-expression patterns using gene-set enrichment analysis where the most informative features were correlated with cell cycling pathways. Grove et al. (Grove et al. 2015) developed two CT features, convexity (hazard ratios [HR] = 0.31) and entropy ratio (HR = 2.36), which were significantly associated with OS of patients diagnosed with primary lung adenocarcinoma utilizing two independent cohorts. Tunali et al. (Tunali et al. 2017) assessed the same cohorts and developed novel radiomic features generated from radial gradient (RG) and radial deviation (RD) maps that also predict OS (HR = 0.40). Coroller et al. (Coroller et al. 2015) built a combined model of CT radiomics and clinical predictors to predict the development of distant metastasis (CI = 0.61) and Wu et al. (Wu et al. 2016a) utilized fluorine 18 (18^F) PET/CT-based radiomic features to also predict the development of distant metastasis (CI = 0.71). Huang et al. (Huang et al. 2016) found radiomic signatures that correlated with disease-free survival (HR = 1.77). Several other studies (Huynh et al. 2017; Li et al. 2017; Oikonomou et al. 2018) investigated the prognostic performance of CT radiomic features for distant metastasis and loco-regional recurrence after stereotactic body radiation therapy. Win et al. (Win et al. 2013) showed that heterogeneity on both CT and PET components of PET/CT were significant predictors of survival. Chae et al. (Chae et al. 2014) and She et al. (She et al. 2018) found CT radiomic signatures that differentiate indolent versus invasive lung adenocarcinoma with an AUROC of 0.98 and 0.95, respectively. The summary of these studies can be found in Table 2.

Radiomics and AI in treatment response

Advances in lung cancer treatments (e.g., checkpoint blockade immunotherapy and tyrosine kinase inhibitors [TKIs]) have decreased lung cancer mortality rates and improve patient survival outcomes. Early assessment of a therapeutic efficacy and predicting treatment outcomes would aid decision support for which treatment has the potential to have optimal benefit for the individual patient. This could eliminate unnecessary treatments, reduce toxicities and costs, and increasing patient survival. However, biomarkers that are highly predictive of both positive and negative responses that can be used prior to initiation of therapy are an unmet clinical need. As such, accurate, generalizable, and ideally non-invasive biomarkers are needed as diagnostic companions for different treatments types (Teng et al. 2018).

Immunotherapy that blocks inhibitory checkpoint signals has been shown to yield durable responses. However, a substantial subset of patients do not respond to immunotherapy and, in some instances, patients experience rapid and lethal immunotherapy-induced hyper-progressive disease (HPD). Several studies have utilized CT radiomics to address this recent clinical unmet need of identifying immunotherapy response. Trebeschi et al. (Trebeschi et al. 2019) developed models to predict outcomes of both NSCLC and melanoma patients in the setting of immunotherapy. Their lesion-level performance in NSCLC patients yielded an AUROC of 0.83, whereas the best patient-level prediction models had an AUROC of 0.76. Sun *et al.* (Sun et al. 2018) developed models that assess CD8 cell tumor infiltration and utilized this model to predict immunotherapy response of NSCLC patients. Their radiomic signature of CD8 cells consisted of 8 radiomic features which yielded an AUC of 0.76 to discriminate inflamed tumors from immune-desert tumors. Tunali et al. (Tunali et al. 2019a) developed radiomics-clinical models to predict rapid disease progression, including HPD, among and NSCLC patients treated with immunotherapy found modest-to-high AUCs of 0.81–0.85. Following up on this work, Tunali *et al.* (Tunali et al. 2019c) developed and validated a clinical-radiomic risk model that identified a very-high risk group of patients associated with extremely rapid and poor survival outcomes (HR for OS = 5.35) compare to the low-risk group (HR = 1.00). Lastly, Mu *et al.* (Mu et al. 2019a) developed a PET/CT radiomic signature to predict patients who are likely to achieve durable clinical benefit from immunotherapy that yielded an AUC of 0.81 in the validation cohort.

Similar to immunotherapy, only a subset of patients benefit of treatment with TKIs (Shepherd et al. 2005). As such, studies have been conducted utilizing radiomics to predict patient outcomes and TKI treatment response. Cook *et al.* (Cook et al. 2015) created models using textural features from fluorine 18 (^{18}F) fluorodeoxyglucose (FDG) PET images to predict outcomes among patients treated with Erlotinib (an *EGFR* TKI). Their model discriminated patients into high vs. low overall survival (26.6 months vs 13.1 months, P = 0.006). Ravanelli *et al.* (Ravanelli et al. 2018) identified CT texture features that predicted 6-month progression with an AUROC of 0.80 and 1-year progression with 0.76. Park et al. (Park et al. 2018) identified texture features from pretreatment FDG-PET/CT to predict early EGFR TKI failure. After adjusting for clinical parameters, high GLCM entropy was associated with worse survival (HR = 4.86).

Studies have also investigated the utility of radiomics for treatment responses to chemotherapy or radiation therapy. Coroller et al. (Coroller et al. 2017) utilized CT radiomic features extracted from primary lung tumors and lymph nodes to predict pathological complete response (pCR) and gross residual disease (GRD) after neoadjuvant chemo-radiation before surgery. They identified a pCR radiomics feature (AUC = 0.75) that significantly outperformed total primary tumor and lymph node volume (AUC = 0.58) and a GRD radiomic clinical model that had a AUC of 0.73. Yu et al. (Yu et al. 2018) trained and validated cohorts a CT radiomic model to predict metastasis (HR = 1.27) among NSCLC patients treated with surgery or stereotactic ablative radiation therapy (SABR). Mattonen et al. (Mattonen et al. 2016) identified a machine learning based radiomics model to detect local recurrence after SABR. Their radiomic signature consisted of 5 features which discriminated local recurrence from fibrosis with an AUC of 0.85 and had a significantly lower false negative rate compared to an expert physicians assessment (99% vs 23%). Khorrami et al. (Khorrami et al. 2019) utilized peri- and intratumoral CT radiomic features to predict pemetrexed-based chemotherapy response and showed that peritumoral features were predictive for time-to-progression (AUC = 0.77). Fave et al. (Fave et al. 2017) utilized delta radiomics (i.e., changes in radiomic features in time) and showed that radiomic feature alterations after radiation therapy were associated with tumor response (C-index = 0.558). The summary of these studies can be found in Table 3.

Radiomics and AI in Radiogenomics

Radiogenomics is the study of the relationship between imaging features and genomic phenotype(s) (e.g., gene-expression, gene mutations, etc) to inform the potential underlying cellular pathophysiology of a tumor. It should be noted that in some publications, “Radiogenomics” refers to genomic prediction of radiation response patterns, which is not reviewed herein. Cancers are heterogeneous across a wide range of spatial and temporal scales which results in habitat variations in metabolism, vasculature, oxygenation, and gene expression (Gatenby et al. 2013; Yip and Aerts 2016). Genomic heterogeneity, molecular, and microenvironmental events reflect tumor aggressiveness and therapy response. Tumors with same stage and histology still have unique biological underpinning such as driver mutations, proteomic profiling, genomic heterogeneity, and/or microenvironments that reflect and can impact aggressiveness and therapy response. Technical advances allow extensive molecular characterization of tumor cells in each individual patient that enables personalized cancer treatment such as targeted treatments of TKIs. Targeted therapies such as Erlotinib and Gefitinib are used to treat patients with positive *Epidermal Growth Factor Receptor* (*EGFR*) mutations (Riely et al. 2006). However, a single arbitrary sample taken from the tumor using needle biopsy can only represent a small sub-region of the tumor region, which may result in misleading diagnoses. Meanwhile, many radiogenomic studies have shown that radiomic features can capture the link between the cancer genomics and tumor phenotype.

CT imaging is routinely used in the management of lung cancer patients. Thus, CT radiomics could be used to predict mutational status of clinically actionable mutations using non-invasive information. Such an approach would be particularly beneficial as a clinical predictor in patients with unresectable lung cancer, among patients in whom biopsy is

unable to be performed, or to minimize additional biopsy in a molecular test is indeterminate. Additionally, since radiomics can be extracted in immediately and in real-time, data can be captured longitudinally from an ROI to pinpoint potential phenotypic transformations and activation of alternative pathways to recognize potential acquired resistance to therapies earlier. Velazquez et al. (Rios Velazquez et al. 2017) developed a clinical-radiomics signatures to differentiate between *EGFR* and *KRAS* mutations (AUC = 0.70), the most common somatic mutations in lung adenocarcinomas. Gevaert et al. (Gevaert et al. 2017) utilized semantic features to predict *EGFR* and *KRAS* mutations; however their models were only able to predict for *EGFR* mutations accurately (AUC = 0.87). Liu et al. (Liu et al. 2016) utilized CT radiomics to predict *EGFR* mutation status (AUC = 0.709) in an Asian cohort who had surgically-resected peripheral lung adenocarcinomas. Weiss et al. (Weiss et al. 2014) identified CT texture features that discriminated between *KRAS* mutant tumors from pan-wildtype tumors (%89.6 accuracy) and Yamamoto et al. (Yamamoto et al. 2014) combined clinical covariates and CT based features to characterize tumors with anaplastic lymphoma kinase (*ALK+*) rearranged NSCLC. Yoon et al. (Yoon et al. 2015) identified clinical covariates and CT and PET radiomics to predict for *ALK/ROS1/RET* fusion-positive lung adenocarcinoma (Sensitivity = 0.73, Specificity = 0.70) and Zhou et al. (Zhou et al. 2018) combined semantic CT features with next-generation RNA sequencing data and validated 10 metagenes annotated by functional gene enrichment analysis that were significantly associated with semantic features. Wu et al. (Wu et al. 2016b) identified texture CT features were associated with NSCLC tumor histology (AUC = 0.72). The summary of these studies can be found in Table 4.

Limitations and Recommendations

Radiomics have shown promise for providing non-invasive biomarkers to predict diagnosis, prognosis, and treatment response, and for longitudinal monitoring of lung cancer treatment. Despite the compelling results from these studies, there are still many limitations that need to be addressed that result in few reproducible findings and potentially spurious results.

There are many reasons that can be attributed to generating non-reproducible results including heterogeneous image acquisition and segmentation and inappropriate use of statistical methods (e.g., overfit models). Standard-of-care image acquisition parameters have a wide range of parameters that include *inter alia*: pixel spacing and slice thickness, reconstruction kernel, kVp, washout periods on PET scans, administration of contrast agent, echo time and repetition time on MRI scans. Intra- and inter-scanner variabilities affect these parameters which cause radiomic feature distributions to change. To overcome this issue, constant effort needs to be expended to ensure that acquisition and reconstruction protocols are either standardized or correctable. On the other hand, if the radiomic features are extracted from heterogeneous image acquisition parameters, features that are less sensitive to these parameters should be used and decisions should be made to consider eliminating the sensitive features (Shafiq-Ul-Hassan et al. 2017). This does assume that the reproducible features contain as much diagnostic information as those that may be considered for elimination. Meanwhile, computational radiomic feature calculations involve many critical processing steps that include pre-processing, spatial interpolation and intensity discretization. The IBSI (Zwanenburg et al. 2018) provides standardized algorithms for

radiomic feature calculation and give consensus and benchmarks on the most common radiomic features and image processing steps before feature extraction.

Another important factor that affects the reproducibility is the segmentation of the ROI (i.e., tumor parenchyma or peritumoral region). Manual segmentations are particularly time consuming and often leads to intra-observer variations. To overcome this, segmentation of the tumors can be done by semi-automated algorithms which involve minimal user variations such as simple initializations (e.g., seed point), followed by a computer-derived delineation of the ROI. However, to tackle this issue further, segmentation algorithms across institutions have to be standardized to achieve consistent delineations. Nevertheless, many of the features are not reproducible even when acquired within couple minutes using same image acquisition parameters (Balagurunathan et al. 2014b) or when the same segmentation algorithms are being used (Kalpathy-Cramer et al. 2016a). Hence, researchers are encouraged to choose reproducible features by utilizing test re-test datasets such as RIDER, and stable features by utilizing multiple segmentation datasets such as Moist run (Kalpathy-Cramer et al. 2016b; Tunali et al. 2019b) or by performing multiple segmentations on subset of images from their own dataset of interest.

Another issue is poor study design that can result in false positive findings (Yip and Aerts 2016). With the potential wide range of hyperparameters such as number of filters, feature categories, and other adjustable parameters, theoretically there are unlimited numbers of radiomic features available for analysis. Studies often analyze large number of features without accounting for multiple testing errors which leads to selection bias and false positive results. Chalkidou et al. (Chalkidou et al. 2015) suggests using a minimum of 10–15 observations (i.e., patients) per predictor variable (i.e., radiomic feature) to realistically reduce false discovery rates. Another potential application is to correct significant p-values for multiple testing using methods such as the Bonferroni-Holm or Benjamini-Hochberg methods (Holm 1979; Benjamini and Hochberg 1995; Bland and Altman 1995). If estimates of predictive performance are conducted from a cohort of a single institution, multiple-folded repeated cross-validation should be considered to minimize the risk of overfitting. Utilizing one or more validation cohort is the optimal method to validate findings to avoid spurious findings. However, one potential limitation is how to handle differences in demographic and clinical covariates across the training and validations cohorts. As such, another potential approach is to combine patient datasets from different institutions and randomly split all patients into training, test, and validation sets so that the trained model includes heterogeneity. Another suggestion is to assess whether the validated model can be applied on a distinct patient population (e.g., TKI treated patients only) or reflect a *pan-signature* that can be used across multiple patient populations.

Medical image analysis and radiomics have shown to have the ability to characterize phenotypic and biological underpinning of ROIs (Aerts et al. 2014). Linking radiomics to biology is critical to evolve from mere statistical associations to characterizing tumor biology. Moreover, characterizing the biological underpinnings of radiomics will likely ensure the clinical uptake of these models. However, the link between radiomics and the tumor biology are often unknown due to the lack of availability of datasets that include both radiomics and genomic/biological information. Although studies have linked radiomics to

biology (Aerts et al. 2014; Grossmann et al. 2017; Morales et al. 2019; Tunali et al. 2019c), consistent efforts in this domain are needed to expedite the transition of radiomics into clinical practice.

To assess the quality of a radiomics study, a radiomic quality score (RQS) has been proposed by Lambin et al. (Lambin et al. 2017) which evaluates radiomic studies by a series of questions on internal consistency, reproducibility and clinical applicability. The RQS score makes no claims regarding the significance of the study evaluated; rather, the RQS quantifies the proper study design and scientific utility. We highly suggest researchers to evaluate their studies by RQS and try to maximize their score for more repeatable and quality science in the field of radiomics.

Conclusions

Radiomics is a non-invasive tool designed for clinical decision support, for both radiologists and oncologists, designed to use routinely available standard-of-care images of lung cancers. Radiomics have been shown to have utility across the lung cancer care continuum including risk prediction, early detection, diagnosis, prognosis, and treatment response. Despite the aforementioned limitations, radiomics and the radiomic community have evolved, in many ways, faster than other ‘omics fields because the radiomics field was able to leverage from existing disciplines. Though the radiomic community has embraced the rigorous training, testing, and validation of radiomic models, such models have yet to impact clinical practice. However, with the standardization of medical imaging and image-based features and the utilization of emerging technologies such as DL, prospective trials to test the clinical utility of radiomics will be emerging in the near future.

Funding

Funding support came from the National Cancer Institute (NCI) (U01-CA143062 to Drs. Gillies and Schabath) and the NCI Early Detection Research Network (U01-CA200464 to Drs. Gillies and Schabath).

References

ACR. Lung-RADS Assessment Categories, Version 1.0. American College of Radiology. Lung CT Screening Reporting and Data System (Lung-RADS™).

Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D et al. 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5: 4006. [PubMed: 24892406]

Alilou M, Beig N, Orooji M, Rajiah P, Velcheti V, Rakshit S, Reddy N, Yang M, Jacono F, Gilkeson RC et al. 2017. An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung CT. Medical Physics 44: 3556–3569. [PubMed: 28295386]

Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye WX, Corrado G et al. 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography (vol 25, pg 954, 2019). Nat Med 25: 1319–1319.

Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, Kim J, Goldgof DB, Hall LO, Gatenby RA et al. 2014a. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. Transl Oncol 7: 72–87. [PubMed: 24772210]

Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao B et al. 2014b. Test-retest reproducibility analysis of lung CT image features. J Digit Imaging 27: 805–823. [PubMed: 24990346]

Balagurunathan Y, Kumar V, Gu YH, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao BS et al. 2014c. Test-Retest Reproducibility Analysis of Lung CT Image Features. Journal of Digital Imaging 27: 805–823. [PubMed: 24990346]

Becker N, Motsch E, Trotter A, Heussel CP, Dienemann H, Schnabel PA, Kauczor HU, Maldonado SG, Miller AB, Kaaks R et al. 2019. Lung cancer mortality reduction by LDCT screening-Results from the randomized German LUSI trial. Int J Cancer.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Statistical Methodology 57: 289–300.

Birim O, Kappetein AP, van Klaveren RJ, Bogers AJ. 2006. Prognostic factors in non-small cell lung cancer surgery. Eur J Surg Oncol 32: 12–23. [PubMed: 16297591]

Bland JM, Altman DG. 1995. Multiple significance tests: the Bonferroni method. BMJ 310: 170. [PubMed: 7833759]

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394–424. [PubMed: 30207593]

Brzakovic D, Luo XM, Brzakovic P. 1990. An approach to automated detection of tumors in mammograms. IEEE Trans Med Imaging 9: 233–241. [PubMed: 18222769]

Bueno J, Landeras L, Chung JH. 2018. Updated Fleischner Society Guidelines for Managing Incidental Pulmonary Nodules: Common Questions and Challenging Scenarios. Radiographics 38: 1337–1350. [PubMed: 30207935]

Chae HD, Park CM, Park SJ, Lee SM, Kim KG, Goo JM. 2014. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. Radiology 273: 285–293. [PubMed: 25102296]

Chalkidou A, O'Doherty MJ, Marsden PK. 2015. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. PLoS One 10: e0124165. [PubMed: 25938522]

Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, Kadoury S, Tang A. 2017. Deep Learning: A Primer for Radiologists. Radiographics 37: 2113–2131. [PubMed: 29131760]

Cherezov D, Hawkins SH, Goldgof DB, Hall LO, Liu Y, Li Q, Balagurunathan Y, Gillies RJ, Schabath MB. 2018. Delta radiomic features improve prediction for lung cancer incidence: A nested case-control analysis of the National Lung Screening Trial. Cancer Med 7: 6340–6356. [PubMed: 30507033]

Cook GJ, O'Brien ME, Siddique M, Chicklore S, Loi HY, Sharma B, Punwani R, Bassett P, Goh V, Chua S. 2015. Non-Small Cell Lung Cancer Treated with Erlotinib: Heterogeneity of (18)F-FDG Uptake at PET-Association with Treatment Response and Prognosis. Radiology 276: 883–893. [PubMed: 25897473]

Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, Aerts HJWL. 2017. Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC. Journal of Thoracic Oncology 12: 467–476. [PubMed: 27903462]

Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, Lambin P, Haibe-Kains B, Mak RH, Aerts HJ. 2015. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 114: 345–350. [PubMed: 25746350]

De Koning H, Van der Aalst C, Ten Haaf K, Oudkerk M. 2018. Effects of Volume CT Lung Cancer Screening: Mortality Results of the NELSON Randomised-Controlled Population Based Trial. J Thorac Oncol 13: S185–S185.

Dela Cruz CS, Tanoue LT, Matthay RA. 2011. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med 32: 605–644. [PubMed: 22054876]

Dhara AK, Mukhopadhyay S, Dutta A, Garg M, Khandelwal N. 2016. A Combination of Shape and Texture Features for Classification of Pulmonary Nodules in Lung CT Images. J Digit Imaging 29: 466–475. [PubMed: 26738871]

Dhawan AP, Buelloni G, Gordon R. 1986. Enhancement of mammographic features by optimal adaptive neighborhood image processing. IEEE Trans Med Imaging 5: 8–15. [PubMed: 18243977]

Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik JC, Thomassin-Naggara I, Fournier L, Lecler A. 2019. Gray-level discretization impacts reproducible MRI radiomics texture features. PLoS One 14: e0213459. [PubMed: 30845221]

Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M et al. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45: 228–247. [PubMed: 19097774]

Fave X, Zhang LF, Yang JZ, Mackin D, Balter P, Gomez D, Followill D, Jones AK, Stingo F, Liao ZX et al. 2017. Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. Sci Rep-Uk 7.

Gatenby RA, Grove O, Gillies RJ. 2013. Quantitative imaging in cancer evolution and ecology. Radiology 269: 8–15. [PubMed: 24062559]

Gevaert O, Echegaray S, Khuong A, Hoang CD, Shrager JB, Jensen KC, Berry GJ, Guo HH, Lau C, Plevritis SK et al. 2017. Predictive radiogenomics modeling of EGFR mutation status in lung cancer. Sci Rep 7: 41674. [PubMed: 28139704]

Giger ML, Doi K, MacMahon H. 1988. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med Phys 15: 158–166. [PubMed: 3386584]

Gillies RJ, Kinahan PE, Hricak H. 2016. Radiomics: Images Are More than Pictures, They Are Data. Radiology 278: 563–577. [PubMed: 26579733]

Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, Leijenaar RT, Haibe-Kains B, Lambin P, Gillies RJ et al. 2017. Defining the biological basis of radiomic phenotypes in lung cancer. Elife 6.

Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y et al. 2015. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PLoS One 10: e0118261. [PubMed: 25739030]

Hadjiiski L, Sahiner B, Chan HP, Petrick N, Helvie M. 1999. Classification of malignant and benign masses based on hybrid ART2LDA approach. IEEE Trans Med Imaging 18: 1178–1187. [PubMed: 10695530]

Hardie RC, Rogers SK, Wilson T, Rogers A. 2008. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med Image Anal 12: 240–258. [PubMed: 18178123]

Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Krewer H, Li Q, Cherezov D, Gatenby RA, Balagurunathan Y et al. 2016. Predicting Malignant Nodules from Screening CT Scans. J Thorac Oncol 11: 2120–2128. [PubMed: 27422797]

Holm M. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6: 65–70.

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. 2018. Artificial intelligence in radiology. Nat Rev Cancer 18: 500–510. [PubMed: 29777175]

Huang P, Park S, Yan R, Lee J, Chu LC, Lin CT, Hussien A, Rathmell J, Thomas B, Chen C et al. 2018. Added Value of Computer-aided CT Image Features for Early Lung Cancer Diagnosis with Small Pulmonary Nodules: A Matched Case-Control Study. Radiology 286: 286–295. [PubMed: 28872442]

Huang YQ, Liu ZY, He L, Chen X, Pan D, Ma ZL, Liang CS, Tian J, Liang CH. 2016. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology 281: 947–957. [PubMed: 27347764]

Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, Parmar C, Hou Y, Mak RH, Aerts HJWL. 2017. Associations of Radiomic Data Extracted from Static and Respiratory-Gated CT Scans with Disease Recurrence in Lung Cancer Patients Treated with SBRT. Plos One 12.

Kachroo S, Tong L, Spitz MR, Xing Y, Merriman K, Zhu DK, Fueger J, Amos CI, Etzel CJ. 2008. Trends in prevalence of prognostic factors and survival in lung cancer patients from 1985 to 2004 at a tertiary care center. Cancer Detect Prev 32: 101–108. [PubMed: 18639390]

Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P et al. 2016a. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. Tomography 2: 430–437. [PubMed: 28149958]

Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S. 2016b. A Comparison of Lung Nodule Segmentation Algorithms: Methods and Results from a Multi-institutional Study. J Digit Imaging 29: 476–487. [PubMed: 26847203]

Kawata Y, Niki N, Ohmatsu H, Kakinuma R, Eguchi K, Kaneko M, Moriyama N. 1998. Quantitative surface characterization of pulmonary nodules based on thin-section CT images. Ieee T Nucl Sci 45: 2132–2138.

Khorrami M, Khunger M, Zagouras A, Patil P, Thawani R, Bera K, Rajiah P, Fu P, Velcheti V, Madabhushi A. 2019. Combination of Peri- and Intratumoral Radiomic Features on Baseline CT Scans Predicts Response to Chemotherapy in Lung Adenocarcinoma. Radiology: Artificial Intelligence 1.

Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A et al. 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14: 749–762. [PubMed: 28975929]

Larue R, Van De Voorde L, van Timmeren JE, Leijenaar RTH, Berbee M, Sosef MN, Schreurs WMJ, van Elmpt W, Lambin P. 2017. 4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers. Radiother Oncol 125: 147–153. [PubMed: 28797700]

Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, Aerts HJ, Gillies RJ, Lambin P. 2015. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 5: 11075. [PubMed: 26242464]

Li Q, Kim J, Balagurunathan Y, Liu Y, Latifi K, Stringfield O, Garcia A, Moros EG, Dilling TJ, Schabath MB et al. 2017. Imaging features from pretreatment CT scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy. Med Phys 44: 4341–4349. [PubMed: 28464316]

Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, Smith GT, Massion PP, Schabath MB, Gillies RJ. 2017. Radiological Image Traits Predictive of Cancer Status in Pulmonary Nodules. Clin Cancer Res 23: 1442–1449. [PubMed: 27663588]

Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, Ye Z, Gillies RJ. 2016. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. Clin Lung Cancer 17: 441–448 e446. [PubMed: 27017476]

Lu H, Mu W, Balagurunathan Y, Qi J, Abdalah MA, Garcia AL, Ye Z, Gillies RJ, Schabath MB. 2019. Multi-window CT based Radiomic signatures in differentiating indolent versus aggressive lung cancers in the National Lung Screening Trial: a retrospective study. Cancer Imaging 19: 45. [PubMed: 31253194]

Maldonado F, Boland JM, Raghunath S, Aubry MC, Bartholmai BJ, Deandrade M, Hartman TE, Karwoski RA, Rajagopalan S, Sykes AM et al. 2013. Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)--a pilot study. J Thorac Oncol 8: 452–460. [PubMed: 23486265]

Mattonen SA, Palma DA, Johnson C, Louie AV, Landis M, Rodrigues G, Chan I, Etemad-Rezai R, Yeung TP, Senan S et al. 2016. Detection of Local Cancer Recurrence After Stereotactic Ablative Radiation Therapy for Lung Cancer: Physician Performance Versus Radiomic Assessment. Int J Radiat Oncol Biol Phys 94: 1121–1128. [PubMed: 26907916]

Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek E Jr. 2012. The 7th lung cancer TNM classification and staging system: Review of the changes and implications. World J Radiol 4: 128–134. [PubMed: 22590666]

Morales JP, Tunali I, Stringfield O, Eschrich S, Balagurunathan Y, Gillies R, Schabath M. 2019. Peritumoral and Intratumoral Radiomic Features Identify Aggressive Screen-Detected Early-Stage Lung Cancers. Journal of Thoracic Oncology 14: S1130–S1130.

Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. 2019a. Radiomics of (18)F-FDG PET/CT images predicts clinical benefit of advanced NSCLC patients to checkpoint blockade immunotherapy. Eur J Nucl Med Mol Imaging.

Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. 2019b. Radiomics of F-18-FDG PET/CT images predicts clinical benefit of advanced NSCLC patients to checkpoint blockade immunotherapy. Eur J Nucl Med Mol I.

Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. 2018. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. Cancer 124: 4633–4649. [PubMed: 30383900]

National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM et al. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395–409. [PubMed: 21714641]

Oikonomou A, Khalvati F, Tyrrell PN, Haider MA, Tarique U, Jimenez-Juan L, Tjong MC, Poon I, Eilaghi A, Ehrlich L et al. 2018. Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. Sci Rep-Uk 8.

Pao W, Girard N. 2011. New driver mutations in non-small-cell lung cancer. Lancet Oncol 12: 175–180. [PubMed: 21277552]

Park S, Ha S, Lee SH, Paeng JC, Keam B, Kim TM, Kim DW, Heo DS. 2018. Intratumoral heterogeneity characterized by pretreatment PET in non-small cell lung cancer patients predicts progression-free survival on EGFR tyrosine kinase inhibitor. Plos One 13.

Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, Cantarutti A, Sverzellati N, Sozzi G, Corrao G, Marchiano A. 2019. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. Ann Oncol 30: 1162–1169. [PubMed: 30937431]

Patz EF Jr., Greco E, Gatsonis C, Pinsky P, Kramer BS, Aberle DR. 2016. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. Lancet Oncol 17: 590–599. [PubMed: 27009070]

Peikert T, Duan F, Rajagopalan S, Karwoski RA, Clay R, Robb RA, Qin Z, Sicks J, Bartholmai BJ, Maldonado F. 2018. Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the National Lung Screening Trial. PLoS One 13: e0196910. [PubMed: 29758038]

Ravanelli M, Agazzi GM, Ganeshan B, Roca E, Tononcelli E, Bettoni V, Caprioli A, Borghesi A, Berruti A, Maroldi R et al. 2018. CT texture analysis as predictive factor in metastatic lung adenocarcinoma treated with tyrosine kinase inhibitors (TKIs). Eur J Radiol 109: 130–135. [PubMed: 30527295]

Riely GJ, Pao W, Pham D, Li AR, Rizvi N, Venkatraman ES, Zakowski MF, Kris MG, Ladanyi M, Miller VA. 2006. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. Clin Cancer Res 12: 839–844. [PubMed: 16467097]

Ries LAG. 2007. Cancer survival among adults : U.S. SEER program, 1988–2001, patient and tumor characteristics. U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Bethesda, MD.

Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH et al. 2017. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. Cancer Res 77: 3922–3930. [PubMed: 28566328]

Sandfort V, Yan K, Pickhardt PJ, Summers RM. 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep-Uk 9.

Schabath MB, Gillies RJ. 2015. Noninvasive Quantitative Imaging-based Biomarkers and Lung Cancer Screening. Am J Resp Crit Care 192: 654–656.

Schabath MB, Massion PP, Thompson ZJ, Eschrich SA, Balagurunathan Y, Goldof D, Aberle DR, Gillies RJ. 2015. Behavior Differences of Screen-Detected Lung Cancers in the CT Arm of the National Lung Screening Trial (NLST). Journal of Thoracic Oncology 10: S221–S221.

Schabath MB, Thompson ZJ, Gray JE. 2014. Temporal trends in demographics and overall survival of non-small-cell lung cancer patients at Moffitt Cancer Center from 1986 to 2008. Cancer Control 21: 51–56. [PubMed: 24357741]

Shafiq-Ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. 2018. Voxel size and gray level normalization of CT radiomic features in lung cancer. Sci Rep 8: 10545. [PubMed: 30002441]

Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D et al. 2017. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 44: 1050–1062. [PubMed: 28112418]

She YL, Zhang L, Zhu HY, Dai CY, Xie D, Xie HK, Zhang W, Zhao LL, Zou LL, Fei K et al. 2018. The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules. Eur Radiol 28: 5121–5128. [PubMed: 29869172]

Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, Campos D, Maoleekoonpiroj S, Smylie M, Martins R et al. 2005. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 353: 123–132. [PubMed: 16014882]

Siegel RL, Miller KD, Jemal A. 2019. Cancer statistics, 2019. Ca-Cancer J Clin 69: 7–34. [PubMed: 30620402]

Silva M, Pastorino U, Sverzellati N. 2017. Lung cancer screening with low-dose CT in Europe: strength and weakness of diverse independent screening trials. Clin Radiol 72: 389–400. [PubMed: 28168954]

Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao D, Lancia A, Ammari S et al. 2018. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol 19: 1180–1191. [PubMed: 30120041]

Teng F, Meng X, Kong L, Yu J. 2018. Progress and challenges of predictive biomarkers of anti PD-1/PD-L1 immunotherapy: A systematic review. Cancer Lett 414: 166–173. [PubMed: 29155348]

Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Calin AM, Pizzi AD, Lalezari F, Lambregts DMJ, Rohaan M, Parmar C et al. 2019. Predicting Response to Cancer Immunotherapy using Non-invasive Radiomic Biomarkers. Ann Oncol.

Tunali I, Gray JE, Qi J, Abdalah M, Jeong DK, Guvenis A, Gillies RJ, Schabath MB. 2019a. Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: An early report. Lung Cancer 129: 75–79. [PubMed: 30797495]

Tunali I, Hall LO, Napel S, Cherezov D, Guvenis A, Gillies RJ, Schabath MB. 2019b. Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions. Medical Physics.

Tunali I, Stringfield O, Guvenis A, Wang H, Liu Y, Balagurunathan Y, Lambin P, Gillies RJ, Schabath MB. 2017. Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients. Oncotarget 8: 96013–96026. [PubMed: 29221183]

Tunali I, Tan Y, Gray JE, Eschrich S, Guvenis A, Gillies R, Schabath M. 2019c. Clinical-Radiomic Models Predict Overall Survival Among Non-Small Cell Lung Cancer Patients Treated with Immunotherapy. Journal of Thoracic Oncology 14: S1129–S1129.

Wei LY, Yang YY, Nishikawa RM, Jiang YL. 2005. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. Ieee T Med Imaging 24: 371–380.

Weiss GJ, Ganeshan B, Miles KA, Campbell DH, Cheung PY, Frank S, Korn RL. 2014. Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. PLoS One 9: e100244. [PubMed: 24987838]

Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, Meagher M, Shortman RI, Wan S, Kayani I et al. 2013. Tumor Heterogeneity and Permeability as Measured on the CT Component of PET/CT Predict Survival in Patients with Non-Small Cell Lung Cancer. Clinical Cancer Research 19: 3591–3599. [PubMed: 23659970]

Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BW Jr., Diehn M, Li R. 2016a. Early-Stage Non-Small Cell Lung Cancer: Quantitative Imaging Characteristics of (18)F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. Radiology 281: 270–278. [PubMed: 27046074]

Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, Mak R, Aerts HJ. 2016b. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. Front Oncol 6: 71. [PubMed: 27064691]

Yamamoto S, Korn RL, Oklu R, Migdal C, Gotway MB, Weiss GJ, Iafrate AJ, Kim DW, Kuo MD. 2014. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. Radiology 272: 568–576. [PubMed: 24885982]

Yip SS, Aerts HJ. 2016. Applications and limitations of radiomics. Phys Med Biol 61: R150–166. [PubMed: 27269645]

Yoon HJ, Sohn I, Cho JH, Lee HY, Kim JH, Choi YL, Kim H, Lee G, Lee KS, Kim J. 2015. Decoding Tumor Phenotypes for ALK, ROS1, and RET Fusions in Lung Adenocarcinoma Using a Radiomics Approach. Medicine (Baltimore) 94: e1753. [PubMed: 26469915]

Yu W, Tang C, Hobbs BP, Li X, Koay EJ, Wistuba II, Sepesi B, Behrens C, Canales JR, Cuentas ERP et al. 2018. Development and Validation of a Predictive Radiomics Model for Clinical Outcomes in Stage I Non-small Cell Lung Cancer. Int J Radiat Oncol 102: 1090–1097.

Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, Berry GJ, Plevritis SK, Rubin DL, Napel S et al. 2018. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. Radiology 286: 307–315. [PubMed: 28727543]

Zwanenburg A, Leger S, Vallières M, Löck S. 2018. Image biomarker standardisation initiative. https://arxivorg/abs/161207003.

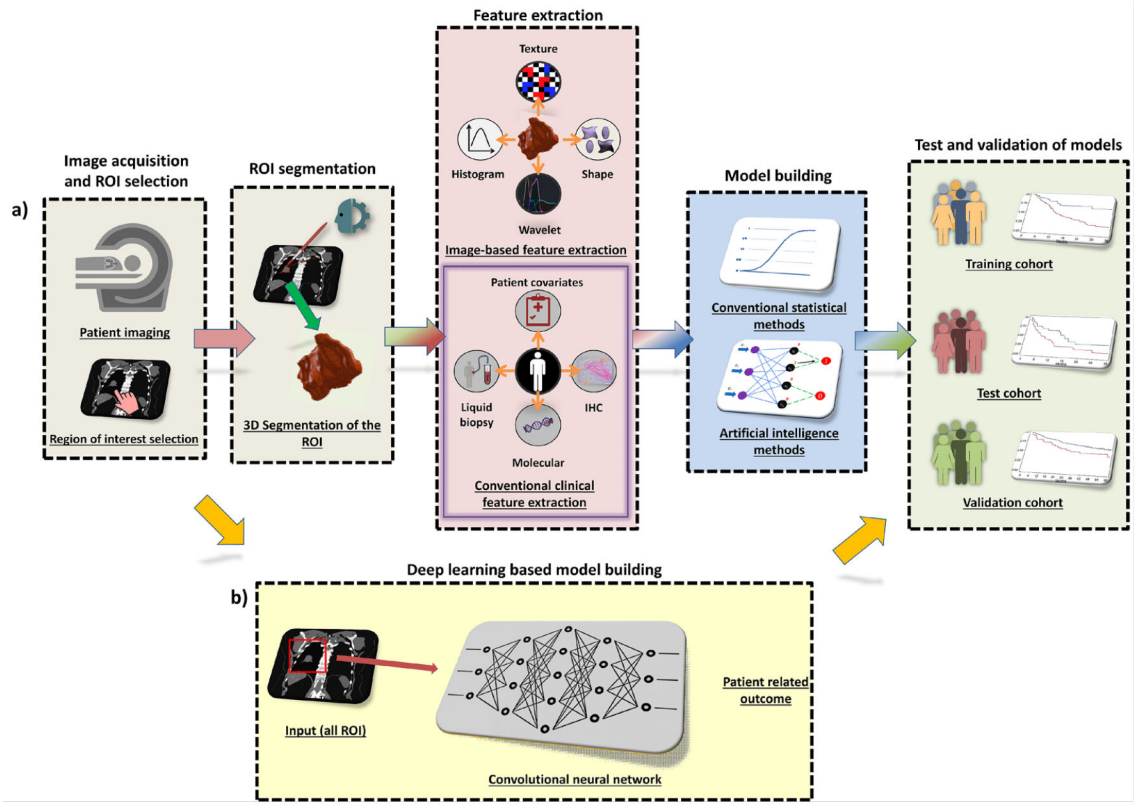

Figure 1: Image-based biomarker model pipelines

a) Conventional radiomics pipeline b) Deep learning pipeline.

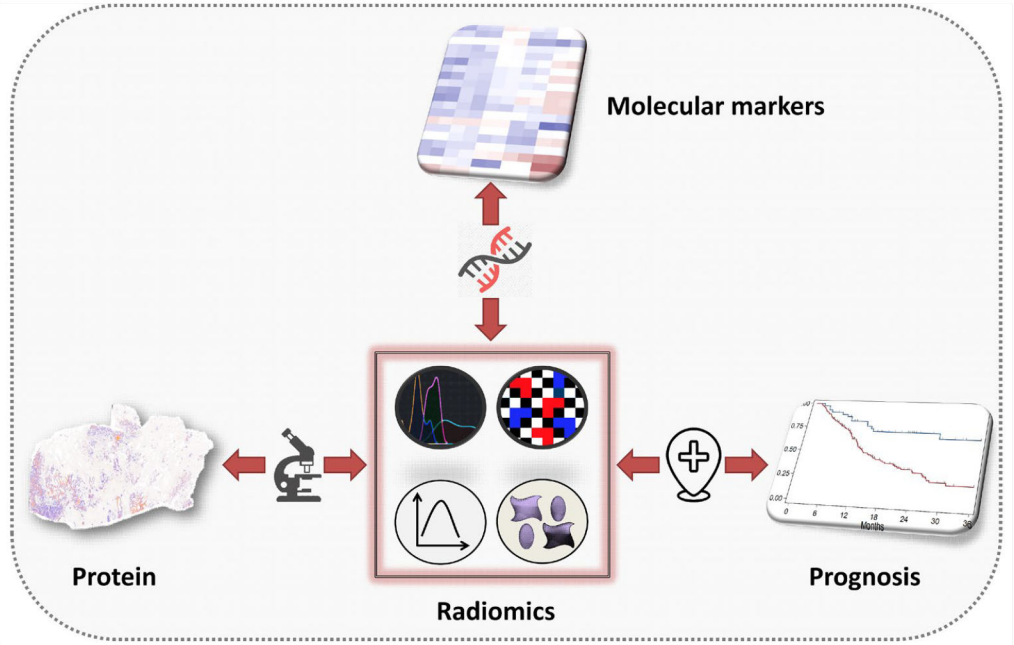

Figure 2: End-points for radiomics modelling.

Radiomics can predict molecular marker and also predict patient outcomes such as overall survival.

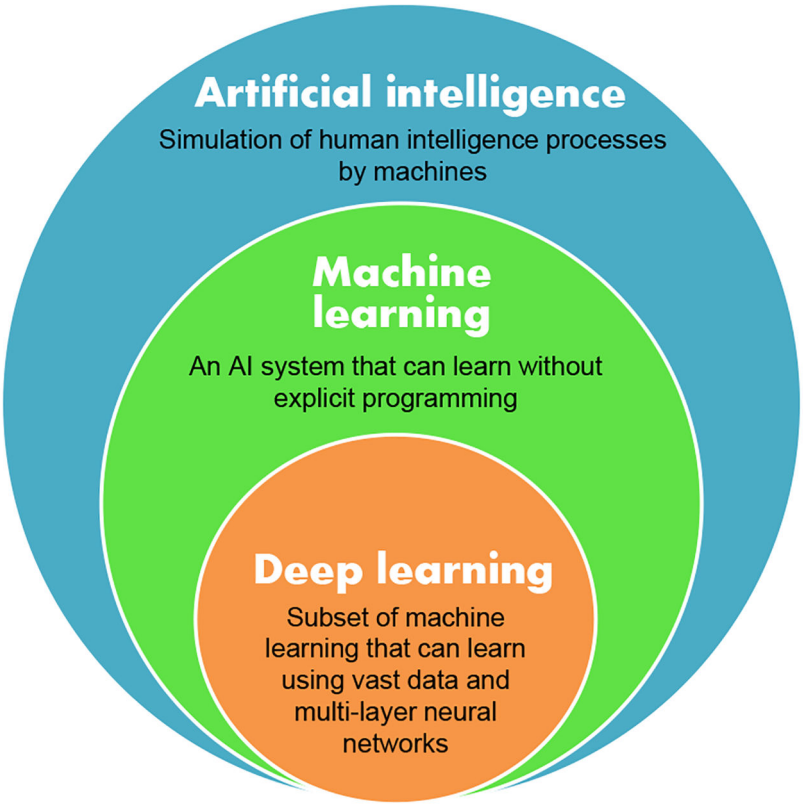

Figure 3:
Artificial intelligence, machine learning, and deep learning.

Table 1.

Radiomics and AI studies in early detection and lung cancer screening.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Hawkins et al. (Hawkins et al. 2016)	CT	Diagnosis	176+152	10	AUC = 0.81	Random forest classifier	Predicting benign vs malignant nodules on a screening cohort	Internal validation with a separate dataset
Peikert et al. (Peikert et al. 2018)	CT	Diagnosis	726	8	AUC = 0.939	LASSO	Differentiating benign vs malignant from indeterminate nodules	Internal validation utilizing bootstrapping
Huang et al. (Huang et al. 2018)	CT	Diagnosis	140+46	5 radiomic + 1 clinical	AUC = 0.9154	Random forest classifier	Differentiating benign vs malignant from small nodules (≤ 20 mm)	Internal validation with a separate dataset
Cherezov et al. (Cherezov et al. 2018)	CT	Diagnosis	255+212	Ranging from 5 to 10	AUC = 0.76 to 0.86	Machine learning	Differentiating benign vs malignant based on nodule size	Internal validation with a separate dataset
Chae et al. (Chae et al. 2014)	CT	Diagnosis	86	2 and 5	AUC = 0.981	Logistic regression + ANN	Differentiating preinvasive lesions from invasive pulmonary adenocarcinomas	No validation
Liu et al. (Liu et al. 2017)	CT	Diagnosis	102+70	4 sematic features	AUC = 0.74 and 0.80	Liner classifier	Differentiating benign vs malignant from incidentally identified nodules	Internal validation with a separate dataset
Dhara et al. (Dhara et al. 2016)	CT	Diagnosis	891	Unknown	AUC = 0.8488 to 0.9505	SVM	Predicting benign vs malignant nodules	No validation
Maldonado et al. (Maldonado et al. 2013)	CT	Tumor behavior	54 + 86	Unknown	Sensitivity = 98.7%	Logistic regression	Categorizing pulmonary nodules as aggressive or indolent	Internal validation with a separate dataset
Lu et al. (Lu et al. 2019)	CT	Tumor behavior	114	5	AUC = 0.846	Logistic regression	Predicting tumor growth (VDT) in lung cancer screening using multi CT windows	No validation
Ardila et al. (Ardila et al. 2019)	CT	Diagnosis	6,716 + 1,139	Deep features	AUC = 0.944	Deep learning	Predicting risk of lung cancer	External validation
Morales et al. (Morales et al. 2019)	CT	Tumor behavior	161 + 73	2	AUC = 0.878	Classification and Regression Tree	Stratifying incident lung cancer patients into risk groups of survival	Internal validation with a separate dataset

Abbreviations: LASSO = least absolute shrinkage and selection operator; ANN = Artificial neural network; SVM = support vector machine; VDT = volume doubling time.

Table 2.

Radiomics and AI studies in prognostication.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Aerts et al. (Aerts et al. 2014)	CT	OS	422+ 225	4	Concordance index = 0.65	Cox proportional hazards regression	Prognostic power of radiomic features and the underlying gene-expression patterns.	External validation
Grove et al. (Grove et al. 2015)	CT	OS	61+47	1	HR = 0.31 and 2.36 P value : 0.008	Cox proportional hazards regression	Prognostic power of newly developed radiomic features	External validation
Tunali et al. (Tunali et al. 2017)	CT	OS	61+47	1	HR = 0.40 P value : 0.014	Cox proportional hazards regression	Prognostic power of newly developed radiomic features	External validation
Coroller et al. (Coroller et al. 2015)	CT	OS & Distant metastasis	98+84	3	Concordance index = 0.61	Cox proportional hazards regression	Predicting distant metastasis	External validation
Wu et al. (Wu et al. 2016a)	PET	Distant metastasis	70+31	2	Concordance index = 0.71	Cox proportional hazards regression	Predicting distant metastasis	Internal validation with a separate dataset
(Huang et al. 2016)	CT	DFS	141+141	5	HR: 1.77 Concordance index = 0.691	LASSO	Predicting DFS in early stage patients	Internal validation with a separate dataset
Huynh et al.(Huynh et al. 2017)	CT	OS	131	13	AUC = 0.667	Correlation analysis (Spearman's correlation coefficient)	Early stage disease recurrence prediction of patients treated with SBRT	No validation
Li et al. (Li et al. 2017)	CT	OS	92	2 radiomic + 1 clinical + 1 semantic	Log-rank p-value = 0.0002	Cox proportional hazards regression	Early stage disease survival prediction of patients treated with SBRT	No validation
Oikionomou et al. (Oikonomou et al. 2018)	CT + PET	OS	150	7	Log-rank p-value = 0.002	Cox proportional hazards regression	Predict clinical outcome in lung cancer patients treated with SBRT	No validation
Win et al. (Win et al. 2013)	FDG-PET/CT	OS	56 + 66	2 radiomic + 1 clinical	Cox regression p-value < 0.001	Cox proportional hazards regression	Predicting OS	External validation
Chae et al. (Chae et al. 2014)	CT	Prognostic	86	2	AUC = 0.981	ANN	Differentiate preinvasive lesions from invasive pulmonary adenocarcinomas	No validation
She et al. (She et al. 2018)	CT	Prognostic	207 + 195	5	AUC = 0.95	Logistic regression	Differentiate indolent from invasive pulmonary adenocarcinomas	Internal validation with a separate dataset

Abbreviations: LASSO = least absolute shrinkage and selection operator; ANN = Artificial neural network; OS = overall survival; HR = hazard ratio; FPR; False positive rate; FNR = False negative rate; OS = overall survival

Table 3.

Radiomics and AI studies in treatment response.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Trebeschi et al. (Trebeschi et al. 2019)	CT	Response to immunotherapy	133+70	2	AUC = 0.83, $p < 0.001$	Machine learning	Identifying radiomic biomarkers for immunotherapy response	Internal validation with a separate dataset
Sun et al. (Sun et al. 2018)	CT	Response to immunotherapy	135 + 256	8	AUC = 0.76	Machine learning	Identifying radiomic markers to assess tumor infiltrating CD8 cells and immunotherapy response	External validation
Tunali et al. (Tunali et al. 2019a)	CT	Rapid disease progression	228	4 radiomic + 4 clinical	AUC = 0.804	Logistic regression	Identifying rapid disease progression phenotypes in patients treated with immunotherapy	Internal validation with a bootstrapping
Tunali et al. (Tunali et al. 2019c)	CT	Response to immunotherapy	180 + 90 + 58	1 radiomic + 2 clinical	Log-rank $p < 0.001$	Classification and Regression Tree	Identifying survival risk groups for patients treated with immunotherapy	External validation
Mu et al. (Mu et al. 2019a)	FDG-PET/CT	Response to immunotherapy	194 + 47 + 48	8	AUC = 0.81	LASSO	Predicting durable clinical benefit from immunotherapy	Internal validation with a separate datasets
(Cook et al. 2015)	FDG-PET	Response to Erlotinib	47	1 delta radiomic *	$p = 0.01$	Wilcoxon signed-rank test	Identifying baseline radiomic biomarkers and delta-radiomic biomarkers of treatment response and OS of patient treated with *EGFR* TKI	No validation
(Ravanelli et al. 2018)	CT	Response to Erlotinib	50	5	AUC = 0.8	LASSO	Determining high and low risk groups of patients treated with *EGFR* TKI	Cross-validation and bootstrapping
(Park et al. 2018)	FDG-PET/CT	Response to Erlotinib	161 + 21	8 univariable features	C-index range = 0.630–0.669	Harrell's C-index	Determining overall survival risk groups of patients treated with *EGFR* TKI	Internal validation with a separate dataset
Coroller et al. (Coroller et al. 2017)	CT	pCR	85	1	AUC = 0.75, $p = 0.01$	Random forest	Predicting pathological response after neoadjuvant chemoradiation	No validation
Yu et al. (Yu et al. 2018)	CT	Response to surgery or SABR	147 + 295	2	HR = 1.27; $p < 2.\ 10^{-16}$	Random survival forests	Predicting OS of Stage I NSCLC patients	Internal validation with a separate dataset
Mattonen et al. (Mattonen et al. 2016)	CT	Local recurrence	45	5	FPR = 24.0%, FNR = 23.1%	Machine learning	Assess physician ability to detect timely local recurrence and to compare physician performance with a radiomics tool	No validation

Author	Imaging modality	Major end-point	N (training + test cohorts)	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Khorrami et al. (Khorrami et al. 2019)	CT	OS and TTP	72 + 53	7	AUC = 0.77	Minimum redundancy maximum relevance	Discriminative ability of radiomic features on response to chemotherapy	Internal validation with a separate dataset
Fave et al. (Fave et al. 2017)	CT	Local recurrence	107	1 delta radiomic *	Log-rank p-value = 0.269 C-index = 0.558	Cox proportional hazards regression	Assessing radiation therapy response by utilizing delta radiomics	No validation

Abbreviations: ANN = artificial neural network; FPR; false positive rate; FNR = false negative rate; TTP = time-to-progression; AUC = area under curve; pCR = pathological complete response; SABR = stereotactic ablative radiation therapy; LASSO = least absolute shrinkage and selection operator;

* Delta-radiomics is the difference in radiomic features between multiple scans.

Table 4.

Radiomics and AI studies in radiogenomics.

Author	Imaging modality	Major end-point	N (training + test cohorts)	Number of features in the model	Evaluation metric	Analytical method used	Study description	Validation Type
Wu et al. (Wu et al. 2016b)	CT	Tumor histology	350	5	AUC = 0.72; $p = 2.3 \times 10^{-7}$	Machine learning	Finding association between radiomic features and the tumor histologic subtypes	No validation
Velazquez et al. (Rios Velazquez et al. 2017)	CT	*EGFR* mutation	353+352	21	AUC = 0.70	Random forest	Discriminating between $EGFR^+$ and $EGFR^-$ cases	External validation
Gevaert et al. (Gevaert et al. 2017)	CT	*EGFR* mutation	186	5 semantic features	AUC = 0.87	Multivariate decision tree	Discriminating between $EGFR^+$ and $EGFR^-$ cases	No validation
Liu et al. (Liu et al. 2016)	CT	*EGFR* mutation	298	4 radiomic +2 clinical	AUC = 0.709	Logistic regression	Discriminating between $EGFR^+$ and $EGFR^-$ cases	No validation
Weiss et al. (Weiss et al. 2014)	CT	*KRAS* mutation	48	2	Accuracy = 89.6%	Recursive decision tree	Differentiation between *KRAS* mutation and pan-wildtype	No validation
Yamamoto et al. (Yamamoto et al. 2014)	CT	*ALK* mutation	59 +113	3 radiomic +1 clinical	Accuracy = 78.8%	Random forest	Discriminating between ALK^+ and ALK^- cases	External validation
Yoon et al. (Yoon et al. 2015)	PET/CT	*ALK*/*ROS1*/ *RET* mutations	539	4 radiomics + 1 qualitative image features + 2 clinical	Sensitivity = 0.73, Specificity = 0.70	Chi-squared test and Student t test	Predictors of tumors with *ALK*, *ROS1*, or *RET* fusions	Cross-validation

Abbreviations: AUC = area under curve;