Radiomics for predicting lung cancer outcomes following radiotherapy: a systematic review


Published in:
Clinical oncology (Royal College of Radiologists (Great Britain))

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2021 Elsevier. This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Download date: 15. Sep. 2023
RADIOMICS FOR PREDICTING LUNG CANCER OUTCOMES FOLLOWING RADIOTHERAPY: A SYSTEMATIC REVIEW.

ABSTRACT

Lung cancer’s radiomic phenotype may potentially inform clinical decision-making with respect to radical radiotherapy. At present there are no validated biomarkers available for the individualisation of radical radiotherapy in lung cancer and the mortality rate of this disease remains the highest of all other solid tumours. MEDLINE was searched using the terms ‘radiomics’ and ‘lung cancer’ according to the PRISMA guidance. Radiomics studies were defined as those manuscripts describing the extraction and analysis of at least ten quantifiable imaging features. Only those studies assessing disease control, survival or toxicity outcomes for patients with lung cancer following radical radiotherapy ± chemotherapy were included. Study titles and abstracts were reviewed by two independent reviewers. The Radiomics Quality Score was applied to the full text of included papers. Of 244 returned results, 44 studies met the eligibility criteria for inclusion. Endpoints frequently reported were local (17%), regional (17%) and distant control (31%), overall survival (79%) and pulmonary toxicity (4%). Imaging features strongly associated with clinical outcomes include texture features belonging to the subclasses Gray level run length matrix, Gray level co-occurrence matrix, and kurtosis. The median cohort size for model development was 100 (15–645), and in the 11 studies with external validation in a separate independent population, the median cohort size was 84 (21–295). The median number of imaging features extracted was 184 (10–6538). The median RQS was 11% (0–47). Patient-reported outcomes were not incorporated within any studies identified. No studies externally validated a radiomics signature in a registered prospective study. Imaging-derived indices attained through radiomic analyses could equip thoracic oncologists with biomarkers for treatment response, patterns of failure, normal tissue toxicity and survival in lung cancer. Based on routine scans, their non-invasive nature and cost
effectiveness are major advantages over conventional pathological assessment. Improved tools are required for the appraisal of radiomics studies, as significant barriers to clinical implementation remain such as standardisation of input scan data, quality of reporting, and external validation of signatures in randomised, interventional clinical trials.

KEYWORDS

Radiomics; lung cancer; radiotherapy, biomarker; deep learning.
STATEMENT OF SEARCH STRATEGIES USED AND SOURCES OF INFORMATION

A database search using MEDLINE was conducted on 5th January 2020 using the terms ‘radiomics’ and ‘lung cancer’, their synonyms and Medical Subject Headings from the year 2000 onwards, according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Met-Analyses) guidelines. Radiomics studies were defined as those manuscripts describing the extraction and analysis of at least ten quantifiable imaging features, an arbitrary threshold prospectively agreed by the authors. Only those studies assessing disease control, survival or toxicity outcomes for patients with lung cancer following radical radiotherapy ± chemotherapy were included. Primarily phantom-based investigations were excluded. Studies based on tumour volume, CT number or maximum standard uptake value (SUVmax) alone, or parameters derived from these, were excluded, as it was felt such studies do not fulfil the intended ‘high-throughput’ paradigm of radiomics. Study titles and abstracts were reviewed by two authors independently for inclusion and selected whole papers were graded according to the Radiomics Quality Score (RQS) (1) by GW and SO.

INTRODUCTION

Lung cancer has the highest mortality rate of all neoplasms (2) despite recent advances in screening (3), molecular diagnosis (4), staging (5) and therapeutics (6,7). The biological complexity of lung cancer, related to complicated genetic evolutionary events (8), has thwarted the development of effective therapies (9). The choice of treatment and the monitoring of its response and toxicity are therefore paramount, yet no effective biomarkers exist for personalising radiotherapy (RT), the treatment of choice for ~40% of all stage I-III disease (10,11). Following histological examination, clinicians rely largely on semantic radiological findings for gross assessment of disease behaviour and prognostication (12,13). The utility of the most widely available modality (ie. CT) for clinical decision-making is somewhat limited
owing to the similar radiodensity of tumour, consolidation and treatment scar however (14–17).

Radiomics is defined as “image analysis using an automated high-throughput extraction of large amounts of quantitative features” (18). First reported on in 2012, scans are used for the extraction of non-visible information by bespoke software and deciphered with complex statistical techniques, correlating with clinical endpoints (18,19). Several imaging features have been shown to correspond with specific underpinning biology, such as protein expression (20) and tissue microanatomy (21); thus explaining how histological subtype (22), mutational status (23), disease staging (24) and survival (25) can also be predicted.

Whilst the broad application of radiomics has been covered in recent narrative reviews, from diagnosis, to local treatments, systemic therapy and general prognosis, (26,27), an exhaustive map of radiomics studies in lung cancer radiotherapy is lacking. In employing a comprehensive search strategy exclusively in the radiotherapy space, this comprehensive review should inform Clinical (or Radiation) Oncologists involved in designing radiomics studies, or implementing this technology clinically.

**BASIC PRINCIPLES**

*Image Data*

The concepts of image feature analysis originate from aerial photography studies analysing terrain in the 1970s (28). Modern computational power has permitted the introduction of this approach into medical imaging analysis. The radiological examination of the patient is the first step of the radiomics workflow (**Figure 1**). Standardisation of scanning parameters is vital as acquisition parameters such as X-ray tube settings, magnetic field strength, slice thickness, tracer/contrast parameters and reconstruction algorithms can introduce noise otherwise (29).
To inform radiomics software from where features are to be extracted in a scan dataset, a region of interest must be delineated (ie segmentation), akin to defining the gross tumour volume for RT planning purposes. Radiomic analyses have been shown to be sensitive to the segmentation method employed, with automated methods proving to have greater reproducibility (30).

**Figure 1.** Radiomics workflow schematic. I = Computed tomography images of the lung are acquired; II = regions of interest are delineated; III = Features extracted from region of interest; IV = Redundant and unstable features are removed; V = analysis of remaining features for their prognostic value with machine learning (*Huynh R&O 2016 with permission*)

*Feature Extraction*

Following pre-processing steps, large numbers of imaging features, typically hundreds, are subsequently extracted, ideally according to methodologies harmonised recently by the Imaging Biomarker Standardisation Initiative (31). Features belonging to three different categories (shape (morphology), intensity (‘first order’) and texture (‘second, or higher order’)) are extracted from the original scans. Image filters can also be applied to the scans to further the feature output. All but shape features are known as statistical features.

*Feature Reduction*

Feature reduction is undertaken so that only robust candidate features are analysed. Features that are deemed to not carry useful information by assessment of their variability between different patient groups are usually removed. Features with poor repeatability (between scans) or reproducibility (between scanners) (32), often judged by their inter-class correlation coefficient or concordance correlation coefficients, and those features that are dependent on tumour volume (33) are also among those removed.

*Feature Selection & Model Building*
Statistical and machine learning procedures are used to analyse features for their predictive capability. To identify features with the most relevance to the clinical trait being studied, supervised or unsupervised machine learning schemes are used (34). Supervised algorithms are prospectively provided with both the radiomic features as inputs, and the observed individual clinical outcomes of interest (‘the ground truth’) as outputs. Unsupervised techniques aim to cluster patients by common features without knowledge of clinical outcomes in advance. Univariate and multivariate analysis may then be used to identify the feature(s) corresponding best with the outcome of interest. The most commonly used measures of the predictive ability include receiver operating characteristic area under the curve (AUC) and concordance index (C-index). AUC refers to the power of a model to discriminate between two outcomes (35) and C-index factors an element of time eg for survival analyses (36). Sensitivity, specificity, and correlation values (Pearson’s or Spearman’s) are frequently reported also.

**Internal and External Validation**

Radiomic feature(s) with predictive potential must then be validated in a separate, smaller proportion of the cohort not used for model training in order to test the model-building procedure ie a test cohort. To avoid splitting a small population into distinct training and test datasets, investigators may employ cross-validation, where the entire dataset is used for training and testing. In cross-validation, near-complete proportions of the total dataset are repeatedly selected and assessed, with the results then combined. Finally, application of the feature(s) in an independent, external cohort in order to test the capability of the model. eg. to confirm the model has sufficient rigour to withstand variability in scanning parameters (37).

**RADIOMICS STUDIES IN LUNG CANCER RADIOTHERAPY**

Of 244 returned results, 44 studies met the eligibility criteria for inclusion (see PRISMA flow diagram (38) **Figure 2**). An additional 24 articles were identified in the citation lists. Due to the
inclusion of less than 10 imaging features, 14 articles were excluded. The title and abstract of several eligible studies did not include the term ‘radiomics’ however this nomenclature is increasingly adopted as shown in Figure 3. Radiomic features were extracted from CT without PET, PET-CT, and MRI in 35/54 (65%), 20/54 (37%) and 1/54 (2%) studies respectively, with several studies including two modalities. AUC was provided to convey predictive ability in 23/54 (43%) studies, and C-index in 16/54 (30%). Endpoints frequently reported included treatment response (defined as shrinkage by the end of treatment) 4/54 (7%), local 9/54 (17%), regional 9/54 (17%) and distant control 17/54 (31%), overall survival 42/54 (79%) and pulmonary toxicity 2/54 (4%). No studies sought radiomic correlates for radiation cardiotoxicity. The median cohort size for model development was 100 (range 15–645), and in the 11 studies with external validation in a separate independent population the median cohort size was 84 (range 21–295). The median number of imaging features extracted was 184 (range 10–6538). The studies identified are narratively summarised below, grouped as conventional and moderately hypofractionated regimes versus stereotactic regimes, given the differences in disease stage, prognosis and toxicity that apply to these two treatment paradigms. The median RQS was 11% (range 0–47). A minority of studies used publicly available scanning protocols (n=8), reported on calibration (n=3) and correlated findings with gene or protein expression (n=1). No studies externally validated a radiomics signature in a registered prospective study or assessed the economical aspects of radiomics. Patient-reported outcomes were not incorporated within any studies identified.

**Figure 2.** PRISMA flow diagram illustrating how studies were identified.

**Figure 3.** The percentage of included studies that include the term ‘radiomics’ in their title or abstract by year of publication.
CONVENTIONAL RADIOTHERAPY

A total of n=30 studies (4,839 patients) testing radiomics for outcome prediction following conventional radiotherapy dose schedules ± chemotherapy, of which 5 were externally validated (476 patients). One study externally validated a previously identified signature in 288 patients (39). Where investigators included clinical variables in their study, these were largely combined with radiomic features (n=12) rather than a head-to-head comparison alone (n=2). Studies correlating imaging features with outcomes including radiotherapy response prediction, distant control, overall survival and radiation pneumonitis.

Table 1. Studies utilising radiomics for prediction of lung cancer outcomes following radical (chemo)RT with conventional dose schedules (SACT Seq = systemic anticancer therapy sequence; CE = contrast enhancement; ind = induction; cc = concurrent; adj = adjuvant; PD = platinum doublet; SA = other single agent; (p) = radiotherapy planning scan; (d) = diagnostic scan; SM = shape metrics; IF = intensity features; TA = texture analysis; FB = filter-based features; DR = ‘delta’ radiomics; CP = clinicopathological; CV = cross-validation; TR = treatment response; pCR = pathological complete response; MPR = major pathological response; GRD = gross residual disease; LC = local control; LRC = loco-regional control; PFS = progression-free survival; DM = distant metastasis; OS = overall survival; LDC = lung density changes; RP = radiation pneumonitis; unspec = unspecified; NR = not recorded)

*study validated a radiomics signature defined previously

are outlined below, with the strongest reported statistic and best predictive capabilities summarised in Table 1.
**Response Prediction**

Tumour shrinkage during the course of radical chemoRT was modelled using features based on 3D planning CT (pCT) scans in patients with locally advanced NSCLC (40,41). Pathological response to neoadjuvant chemoRT could also be predicted from diagnostic CT (dCT) (42) and pCT with tumour and peritumour parenchymal features (43,44). As international guidelines now encourage trimodality therapy for N2 disease as well as Pancoast tumours, radiomics may inform the multidisciplinary management in this setting.

A texture analysis of PET-CT data showed an intensity feature named ‘contrast’ predicted well for treatment response (and overall survival) in a study, where comparisons were made between pre- and post-chemoRT (45). In a small cohort of stage III patients receiving pre-treatment diffusion weighted imaging (DWI)-MRI and PET-CT, features from the former demonstrated better prediction of treatment response (46).

**Local Control**

The numerical value of certain imaging features were noted to evolve during chemoRT by a group investigating temporal changes, ‘delta-radiomics’, with weekly 4D-CT scans (47). Interestingly, texture analysis of images on RT completion could stratify patients by relapse risk categories, but pre-treatment imaging could not. However, predictive models for local (and also distant) failure based on pre-treatment features were not improved by inclusion of delta-radiomics. In a PET-CT study, texture features best predicted local control, compared with SUV-based metrics and intensity features (48), but the opposite trend was observed for locoregional control. A small dynamic MRI-based study confirmed that the ‘maximum enhancement ratio’ was predictive of local control with good accuracy (49), which may be of interest to those groups implementing MRI planning in NSCLC (50).
Distant Control

In an adenocarcinoma-only cohort, pCT-based radiomics were prognostic for the development of distant metastases, and the radiomic model was more powerful when combined with clinical factors such as tumour grade and stage, and patient fitness (51). Only 11% of features in the final model also predicted for OS however, perhaps suggesting that distant disease and patient survival may be influenced by separate elements of the underlying biological phenotype. Incorporation of texture features stable across both 3D- and 4D-CT scans ± contrast in another study improved clinical models of distant control with average classification rates circa 0.89 (52). Although no studies explored features predictive for nodal relapse specifically following conventional RT as a primary focus, a few studies did examine features that corresponded with likelihood of ‘regional failure’, as recorded in Table 1 (47,48,52).

Overall Survival

Testing just 24 shape and texture features on dCT, investigators found there was predictive potential for select subtypes and stages of NSCLC only eg T2N0 large cell carcinomas (53). A different group performing texture analysis with 30 features was able to predict survival across histopathological subtypes within stage III disease however (54), as per Table 1. In a larger and multicentre dataset using 440 features of all four feature types to undertake consensus clustering, the AUC for prognosis after (chemo)RT was 0.61, and clustering also predicted for disease stage and histology (55). A study comparing different methods of machine learning established an optimal AUC of 0.68 for OS after radical radiotherapy in the publicly available NSCLC-radiomics dataset (56).

Cone-beam CT (CBCT) scans were also reported to be suitable for radiomics investigations, with both static and time-dependent features shown to relate to OS in a small stage III study (57). In contrast to the features in a serial 4D-CT study (47), CBCT features were noted to be generally stable throughout treatment. Another study indicated that some radiomics signatures may be interchanged between standard CT and CBCT (39). A separate group found CT scans
post-chemoRT more predictive for OS than pre-treatment CT using unsupervised methods, and also published the first nomogram for lung cancer survival incorporating a radiomics signature (58).

OS was best predicted on PET-CT using two texture features showing high stability between different scanner models and slice thicknesses in a large multicentre UK study including early stage and locally advanced disease (59). Other PET-CT studies found alternative texture features correlated with survival (60) or both OS and PFS (61) in pure stage III cohorts. Other investigators have improved the C-index for OS in oligometastatic cohorts by combining PET-CT imaging indices from a primary tumour with clinical parameters (62). An externally validated feature combination from both primary and metastatic lymph nodes produced a relatively low C-index for OS though, suggesting the biology of lymph nodes may not match the primary (63).

Serial PET features during chemoRT that are adjusted for distance from tumour edge, have been utilised for survival prognostication at two years (64). This worked better for cases undergoing sequential chemoRT than concurrent chemoRT, potentially owing to regional changes in blood flow related to timing of cytotoxics. In addition the investigators also found ‘delta-radiomics’ out-performed ‘single time-point’ radiomics (47). In a study using the Cancer Imaging Archive PET-CT datasets for training models, OS was predicted better with 2D features than 3D features, indicating that single slice analysis may be adequate to decipher some features (65). Inconsistent resolution at different transverse planes between scanners in this multi-institutional study may explain the inferiority of 3D features observed. Disappointingly, a study attempting to externally validate PET-CT-based radiomic features predicting OS was negative. A further negative PET-CT study seeking to establish a model for OS also attributed its failure to lack of standardisation between different scanners at different centres and heterogenous patient characteristics (66).
In a small heterogenous cohort of chemoRT cases, models combining CT and MRI features were more accurate for predicting OS (and treatment response) compared with each modality in isolation (67).

**Pulmonary Toxicity**

Radiation pneumonitis can be predicted in models combining clinical, dosimetric and 4D-CT-based radiomics variables (68). Integrating spatial information from the dose distribution further improved a predictive model for radiation pneumonitis, so-called ‘dosimics’ (69).

**STEREOTACTIC RADIOTHERAPY**

A total of n=24 studies (2,288 patients) where radiomics was tested for outcome prediction following stereotactic radiotherapy were identified, of which 5 were externally validated (476 patients). Where investigators included clinical variables in their study, these was largely combined with radiomic features (n=5) rather than a comparison alone (n=2). Outcomes including response prediction, local and distant control, overall survival and pulmonary fibrosis are outlined below and summarised in Table 2.

**Table 2.** Studies utilising radiomics for prediction of lung cancer outcomes following stereotactic RT (BH = breath-hold; 3D = three-dimensional; 4D = four-dimensional; CT = computed tomography; PET = positron emission tomography; MRI = magnetic resonance imaging; CE = contrast enhancement; (p) = radiotherapy planning scan; (d) = diagnostic scan; SM = shape metrics; IF = intensity features; TA = texture analysis; FB = filter-based features; CP = clinicopathological; CV = cross-validation; LC = local control; LOC = lobar control; RC = regional control; LRC = loco-regional control; RFS = recurrence-free survival; PFS = progression-free survival; DM = distant metastasis; OS = overall survival; DSS = disease-specific survival; LDC = lung density changes; unspec = unspecified; NR = not
Local Control

In examining modes of failure following SABR using pCT, a study found radiomic features performed better for local failure, not regional or distant (70). Combined PET-CT features led to good performance in French and German studies building radiomics signatures predictive for local control in SABR cohorts (71,72).

Regional Control

In a small cohort, a radiomics-based classifier from free-breathing pCT predicted nodal relapse in its validation cohort with AUC 0.73 (73). A Japanese group investigating 4D-CT radiomics identified features predicting locoregional failure-free survival, as well as relapse-free survival and OS (74). Different features were found to achieve the same goal in an equivalent study of post-treatment dCT by the same group (75). Unsupervised techniques identified two distinct clustered groups for nodal failure based on PET/CT radiomic features, which also discerned OS predictions, in two separate studies (76,77). Another team also obtained PET-CT radiomic features predictive for regional control (and also disease-specific survival and OS, see Table 2) (78).

Distant Failure

Using a PET-CT-based radiomics approach, a Chinese found that the shell feature, derived from features at the outer edge of the tumour, out-performed all other tested imaging features using the entire tumour for distant metastasis development (79). Another group combined an externally validated PET-CT radiomics feature predictive for distant metastasis with the tumour histology status to improve the AUC from 0.71 to 0.81 (80). Combining clinical factors with a combined PET-CT radiomic signature also produced the best predictive results for
distant failure following SABR (81). The type of pCT was found to be of relevance by a group comparing radiomics in static free-breathing 3D-CT and 4D-CT, with average intensity projections from the latter found to be superior in models for distant metastases (82).

**Overall Survival**

Radiomic features from dCT were shown to be superior over conventional radiological parameters for OS prediction in a study of SABR patients (83). pCT was used to achieve this in a separate study specifically using wavelet functions only (84). A novel strategy training with surgical cases and validating with SABR cases, confirmed the prognostic capability of radiomic features for early stage NSCLC, independent of treatment (85). A further novel study correlating imaging and genomic features also validated a radiomics signature for survival, training on pCT and validating on PET-CT (86). In analyses of various feature reduction techniques, the maximum average AUC for survival outcomes (recurrence, disease-free survival and OS) was achieved with an unsupervised approach (87). In a study of PET-CT features, disease-specific survival was predicted by the dissimilarity feature, but no features predicted OS with statistical significance (88).

**Pulmonary Toxicity**

Radiomic features can be used to identify patients at risk of post-SABR local fibrosis from pre-treatment breath-hold pCT (89). In addition, the investigators found features in the same dataset correlating with local failure, DFS and OS. Post-SABR ground glass opacification can be used to predict future relapse versus ongoing benign change (90). A further study demonstrating the ability of imaging features to discern local fibrosis and local tumour recurrence on follow-up CT, also found semi-automated segmentation to be at least as good as manual (91).

**Table 2.** Studies utilising radiomics for prediction of lung cancer outcomes following
stereotactic RT (3D = three-dimensional; 4D = four-dimensional; CT = computed tomography; PET = positron emission tomography; MRI = magnetic resonance imaging; CE = contrast enhancement; ind = induction; (p) = radiotherapy planning scan; (d) = diagnostic scan; SM = shape metrics; IF = intensity features; TA = texture analysis; FB = filter-based features; LC = local control; LOC = lobar control; RC = regional control; LRC = loco-regional control; RFS = recurrence-free survival; PFS = progression-free survival; DM = distant metastasis; OS = overall survival; DSS = disease-specific survival; LDC = lung density changes; unspec = unspecified; NR = not recorded)

*this signature was developed in a surgical cohort and validated in a stereotactic cohort

DISCUSSION

The practice of radiomics has the potential to equip thoracic oncologists with numerical parameters to support clinical decisions (92–94) (see Figure 4). This is of particular importance in lung cancer given the difficulty in discerning tumour from radiation pneumonitis, fibrotic parenchyma and infective consolidation on CT and where there is an ongoing drive to provide personalised radiotherapy (14–17). As outlined in this review, imaging-derived indices hold promise in search for lung cancer RT biomarkers regarding treatment response, patterns of failure, normal tissue toxicity and survival.

Studies identified in this first systematic review of radiomics for lung cancer RT clinical outcomes were weighted towards the prediction of distant metastases and overall survival endpoints, and few studies examined cardiopulmonary toxicity. In the small number of externally validated studies, features frequently associated strongly with clinical outcomes include texture features belonging to the subclasses Gray level run length matrix (GLRLM) (39,63), Gray level co-occurrence matrix (GLCM) (57,65,71,79,85,86), and kurtosis (65,85). This finding is not surprising given the predominance of texture features in related studies
across cancer types and imaging modalities (95–97). External validation was generally achieved by fitting of the model unchanged (51,80,86), or by assessing it's discrimination or calibration (39,57,65).

There were several common flaws in the identified studies, such as heterogenous populations (disease stage, radiation dose fractionation, chemotherapy usage), extraction of very large numbers of features from modest numbers of patients, and absence of an independent cohort for validation. No studies attempted to incorporate patient-reported outcomes and no publications examine the utility of radiomics in a prospective lung cancer RT study. A return to the principles of conventional clinical studies in future studies could therefore improve the quality of lung cancer RT radiomics research, and may aid the clinical acceptance of radiomics. These weaknesses are reflected in the low median RQS value of 11%, which is comparable to most systematic reviews in other tumour sites treated with definitive radiotherapy (published RQS 19–29% (98–101)). It would be prudent for investigators planning radiomics studies to consider the RQS principles in the study phase, perhaps targeting a score ≥30%. There are no ongoing trials recorded on clinicaltrials.gov presently (102).

As a technology based on routine imaging, radiomics’ non-invasive nature and cost effectiveness are major advantages over conventional pathological assessments, bringing significant gains for the lung cancer population (37). Evidence of correlation with immune profile (20,103) means radiomics could have implications for how patients are selected for RT-immunotherapy combinations (104). Pre-RT work-up could be accelerated if a radiomics signature predictive for pulmonary function tests was validated, which is important in lung cancer, where the diagnosis-to-treatment interval is clinically important (105,106). Biology-guided RT is the holy grail of personalised RT planning, and a radiomic target volume, where dose-painting according to spatial radioresistance risk levels within a tumour, may be feasible (107).
Figure 4. An example of how combining accepted models with radiomics can lead to superior prognostic ability, as demonstrated by rising AUC values (Coroller JTO 2017 with permission)

For the full benefit of radiomics to reach patients generally, several barriers remain. A recent phantom-based solution (ComBat) has made progress in re-balancing the variability between large numbers of scanning protocols, improving the robustness of features and actually improving the predictive performance of radiomic signatures (108,109). A high false discovery rate exists due to uncorrected type 1 errors resulting from large numbers of features tested and small sample sizes (110). Similarly, the common dependence of many imaging features on tumour volume must be accounted for. The extremely heterogenous range of methods employed complicates the meta-analysis necessary to ‘stock take’ progress and decide future directions (32). The first such study examined overall survival alone and included 12 studies (111). The published C-index estimate was 0.57 and significant heterogeneity was noted unsurprisingly. A meta-analysis was not attempted in the current study given the vast heterogeneity observed between reports.

The lack of standardisation on key clinical variables for comparison of radiomics performance also restricts the interpretation of the literature, and tumour-specific international consensus statements to this effect would be constructive. Commonly tested clinicopathologic factors in the studies identified include patient fitness (47,51,112), and tumour histology (47,80,86,112), volume (51,80,86) and staging (47,51,112).

As one of the first grading systems designed for the field (1), the RQS has been readily implemented in the assessment the quality of radiomics studies (113–115). It’s adoption has been secured by its rounded interest in the entire radiomics pipeline, and its endorsement by leading figures in the field. However, as a tool primarily designed to guide the critical appraisal of a radiomics publication, rather than judge the quality of the entire research methodology
per se, the RQS is insufficiently incisive to detect many of the common pitfalls of the radiomic process (101). Specific improvements required include reduction of interobserver variation and addressing the imbalance in weighting of certain aspects of the radiomics workflow, in line with the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidance (116).

International collaborators have proposed a roadmap for the development of imaging biomarkers (117) and reporting of radiomics-based research (118). Auto-segmentation tools may accelerate radiomic workflows and reduce inter-user variability (119). By rendering imaging datasets publicly available (120,121) and radiomics software 'open source' (122), groups of research teams will be more equipped to focus on their scientific strategy. Such sharing is promoted by the findable, accessible, interoperable, reproducible (FAIR) principles (123). Distributed learning infrastructures can be used to navigate complex data security conflicts (124). Ultimately it may be that commercialisation leads to the routine clinical implementation of radiomics, as evidenced recently as Medicare agreed to cover artificial intelligence for the first time (125). Deep learning, a form of artificial intelligence, involves automated, high-level machine learning (126) and is likely to be adopted in radiomic signature discovery and implementation by the time radiomics is adopted clinically. Convolutional neural networks, one example of the deep learning process, remove the need for the ‘manual’, engineered process outlined above for ‘handcrafting’ radiomic signatures. To fulfil the transparency requirements needed within a health system, the ‘explainability’ of such systems must first be improved however (127).

**CONCLUSIONS**

Given the increasingly imaging-centric nature of radiation oncology, the field of radiomics should be naturally positioned for integration into clinical service workflows and decision-
making. Ultimately, radiomic analysis of the treatment planning, cone-beam and response assessment CTs could potentially contribute to biology-guided RT in lung cancer eg boosting of radioresistant tumour regions, or avoidance of calcific coronary vasculature and highly functional lung regions. Multidisciplinary working groups of thoracic oncologists, radiologists and clinical physicists should be developed locally in those centres wishing to lead on the assimilation of radiomics. Such groups would also contribute to the implementation of related technology such as auto-segmentation and the distributed learning movement (128).

Given the conflicting evidence on dose escalation for NSCLC (129–131), a radiomic biomarker-directed, randomised, interventional study is infeasible currently. A large, prospective, international registry could bridge the gap to clinical trials however, perhaps initially with an emphasis on tailoring radiological surveillance following treatment or predicting cardiopulmonary toxicity. Such prospective registries would be consciously designed using the RQS framework, and include molecular diagnostics which are now widely available for NSCLC (11). Given the abundance of radiomics literature in lung cancer RT, these registries may have the opportunity to focus on select candidate radiomic features with a strong evidence base (eg kurtosis, GLRLM, GLCM), or signatures from studies with a higher RQS (57,60,86). Akin to molecular Oncology, after the high-throughput nature of biomarker discovery, dedicated signature validation studies with clear links to biology (132) may be required for a radiomic parameter/signature to be embraced in national guidelines.

ACKNOWLEDGEMENTS

The authors wish to thank Mrs Gillian McCaughtry, library assistant at the Sir Frank Montgomery Library for Clinical Oncology, Cancer Centre Belfast City Hospital.
ABBREVIATIONS

RT = radiotherapy
SABR = stereotactic ablative radiotherapy
CT = computed tomography
PET = positron emission tomography
MRI = magnetic resonance imaging
PACS = picture archiving and communication systems
AUC = area under the curve
C-index = concordance index
NSCLC = non-small cell lung cancer
PRISMA = preferred reporting items for systematic reviews and met-analyses
SUVmax = maximum standard unit value
RQS = radiomics quality score
3D = three dimensional
pCT = planning computed tomography
dCT = diagnostic computed tomography
DWI = diffusion weighted imaging
CBCT = cone-beam computed tomography
4D = four dimensional
GLRLM = Gray level run length matrix
GLCM = Gray level co-occurrence matrix
FAIR = findable, accessible, interoperable, reproducible
REFERENCES


67. Mahon RN, Hugo GD, Weiss E. Repeatability of texture features derived from


75. Li Q, Kim J, Balagurunathan Y, Qi J, Lie Y, Latifi K, et al. CT imaging features


small cell lung cancer treated with curative radiotherapy. Radiother Oncol. 2020;


<table>
<thead>
<tr>
<th>Study</th>
<th>Stage</th>
<th>Dose &amp; Frac</th>
<th>SACT Seq (drugs)</th>
<th>Imaging Modality (phase where applicable)</th>
<th>CE</th>
<th>Total No of Features</th>
<th>Feature Categories</th>
<th>Model Development</th>
<th>Independent Validation Cohort</th>
<th>CP Factors</th>
<th>Segmentation Method</th>
<th>Outcomes Tested</th>
<th>Best Statistic Reported &amp; Outcome</th>
<th>RQS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramella (40) 2018</td>
<td>III</td>
<td>45.75Gy / 25-40Gy</td>
<td>ind or cc (multiple)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>242</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>91</td>
<td>None</td>
<td>Yes</td>
<td>Manual</td>
<td>TR</td>
<td>AUC for TR 0.82</td>
</tr>
<tr>
<td>Hunter (41) 2016</td>
<td>II-IIIB</td>
<td>74Gy / NR #</td>
<td>cc (NR)</td>
<td>4D-CT(p) (end exhale)</td>
<td>No</td>
<td>153</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>66</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>TR</td>
<td>Pearson’s coefficient for TR 0.81</td>
</tr>
<tr>
<td>Khorrami (42) 2019</td>
<td>III</td>
<td>30-60Gy / NR #</td>
<td>cc (PD)</td>
<td>3D-CT(d)</td>
<td>NR</td>
<td>1542</td>
<td>SM, IF, TA</td>
<td>56</td>
<td>45</td>
<td>90</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>MPR, DFS, OS</td>
</tr>
<tr>
<td>Coroller (43) 2016</td>
<td>II-III</td>
<td>45-70Gy / NR #</td>
<td>ind or cc or adj (NR)</td>
<td>CT(p)</td>
<td>NR</td>
<td>1603</td>
<td>SM, IF, TA, FB</td>
<td>102</td>
<td>25</td>
<td>127</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>pCR, LC, DM, OS</td>
</tr>
<tr>
<td>Coroller (44) 2017</td>
<td>III</td>
<td>45-70Gy / NR #</td>
<td>ind +/- cc or adj (NR)</td>
<td>3D-CT (p)</td>
<td>NR</td>
<td>635</td>
<td>SM, IF, TA</td>
<td>60</td>
<td>25</td>
<td>85</td>
<td>None</td>
<td>No</td>
<td>NR</td>
<td>pCR, GRD, LC, DM, OS</td>
</tr>
<tr>
<td>Dong (45) 2016</td>
<td>III</td>
<td>62.4-68Gy / 36-40Gy</td>
<td>cc +/- adj (PD)</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>20</td>
<td>SM, IF, TA, IF</td>
<td>None</td>
<td>58</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>TR, PFS, OS</td>
<td>AUC for TR 0.80</td>
</tr>
<tr>
<td>Fave (47) 2017</td>
<td>II-IV</td>
<td>66-74Gy / NR #</td>
<td>cc (NR)</td>
<td>4D-CT(d) (end exhale)</td>
<td>No</td>
<td>212</td>
<td>SM, IF, TA, DR</td>
<td>CV</td>
<td>137</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>LRC, DM, OS</td>
<td>C-index for OS 0.69</td>
</tr>
<tr>
<td>Vaidya (48) 2012</td>
<td>I-IV</td>
<td>32.5-74Gy / NR #</td>
<td>unspec (NR)</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>32</td>
<td>IF, TA</td>
<td>CV</td>
<td>27</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>LC, LRC</td>
<td>Spearman’s coefficient for LC 0.59</td>
</tr>
<tr>
<td>Coroller (51) 2015</td>
<td>II-III</td>
<td>54-60Gy / NR #</td>
<td>ind or cc or adj (NR)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>635</td>
<td>SM, IF, TA, FB</td>
<td>CV</td>
<td>98</td>
<td>84</td>
<td>Combined</td>
<td>Manual</td>
<td>DM, OS</td>
<td>C-index for DM 0.61</td>
</tr>
<tr>
<td>Fried (52) 2014</td>
<td>III</td>
<td>60-74Gy / 53-42Gy</td>
<td>cc (PD)</td>
<td>4D-CT(p) (ave + end exhale) &amp; 3D-CT(d)</td>
<td>Yes</td>
<td>198</td>
<td>TA, IF</td>
<td>CV</td>
<td>91</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>LRC, DM, OS</td>
<td>Classification for DM 79%</td>
</tr>
<tr>
<td>Chaddad (53) 2017</td>
<td>I-IV</td>
<td>51-66Gy / 33-42Gy</td>
<td>ind +/- cc (PD)</td>
<td>3D-CT(d)</td>
<td>Yes</td>
<td>24</td>
<td>SM, TA</td>
<td>CV</td>
<td>315</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>OS</td>
<td>Spearman’s coefficient for OS 0.76</td>
</tr>
<tr>
<td>Bianconi (54) 2018</td>
<td>NR</td>
<td>51-79.2Gy / 33-44</td>
<td>ind +/- cc (PD)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>30</td>
<td>SM, IF, TA</td>
<td>None</td>
<td>248</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
<td>HR 0.76 for OS</td>
</tr>
<tr>
<td>Parmar (55) 2015</td>
<td>I-IIIB</td>
<td>51-79.2Gy / 33-44</td>
<td>ind +/- cc (PD)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>440</td>
<td>SM, IF, TA, FB</td>
<td>422</td>
<td>225</td>
<td>645</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Sun (56) 2018</td>
<td>I-IIIB</td>
<td>51-79.2Gy / 33-44</td>
<td>ind +/- cc (PD)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>339</td>
<td>SM, IF, TA, FB</td>
<td>188</td>
<td>95</td>
<td>283</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Shi (133) 2019</td>
<td>III</td>
<td>&gt;50Gy in 2Gy #</td>
<td>cc (PD)</td>
<td>3D-CT(p) &amp; CBCT</td>
<td>NR</td>
<td>658</td>
<td>SM, IF, TA, FB, DR</td>
<td>None</td>
<td>23</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
<td>HR for OS 0.21</td>
</tr>
<tr>
<td>Van Timmeren (39) 2017*</td>
<td>I-IV</td>
<td>44-79.2Gy / NR #</td>
<td>cc (NR)</td>
<td>3D or 4D-CT(p) (50% exhale) &amp; CBCT</td>
<td>Yes</td>
<td>1119</td>
<td>SM, IF, TA, FB</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>288</td>
<td>Combined</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Van Timmeren (57) 2019</td>
<td>I-IV</td>
<td>55-75.6Gy / 33-44</td>
<td>ind or cc (NR)</td>
<td>3D or 4D-CT(p) (50% exhale) &amp; CBCT</td>
<td>Yes</td>
<td>2317</td>
<td>SM, IF, TA, FB</td>
<td>CV</td>
<td>141</td>
<td>196</td>
<td>Compared only</td>
<td>Semi-automated</td>
<td>LRC, OS</td>
<td>C-index for OS 0.63</td>
</tr>
<tr>
<td>Wang (58) 2019</td>
<td>I-III</td>
<td>NR</td>
<td>cc (PD)</td>
<td>3D-CT(d)</td>
<td>Yes</td>
<td>1045</td>
<td>SM, IF, TA, FB</td>
<td>CV</td>
<td>118</td>
<td>None</td>
<td>No</td>
<td>Combined</td>
<td>Semi-automated</td>
<td>OS</td>
</tr>
<tr>
<td>Ashad (59) 2019</td>
<td>I-III</td>
<td>NR</td>
<td>unspec (NR)</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>665</td>
<td>SM, IF, TA</td>
<td>133</td>
<td>204</td>
<td>337</td>
<td>21</td>
<td>Combined</td>
<td>Semi-automated</td>
<td>OS</td>
</tr>
<tr>
<td>Ohi (60) 2016</td>
<td>III</td>
<td>&gt;60Gy / NR #</td>
<td>cc +/- adj (PD)</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>45</td>
<td>TB, TA</td>
<td>CV</td>
<td>121</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>OS</td>
<td>AUC for OS 0.63</td>
</tr>
<tr>
<td>Liu (61) 2018</td>
<td>IIIA-IIIB</td>
<td>40-80Gy / 20-39Gy</td>
<td>unspec (PD or SA)</td>
<td>PET-CT(p)</td>
<td>NR</td>
<td>72</td>
<td>TA</td>
<td>None</td>
<td>35</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>OS, PFS</td>
<td>p=0.005 for OS in high/tow groups</td>
</tr>
<tr>
<td>Study (Year)</td>
<td>Phase</td>
<td>Dose (Gy)</td>
<td>Seq</td>
<td>Ind/CC/Adj</td>
<td>Image Modality</td>
<td>TR</td>
<td>N</td>
<td>Methodology</td>
<td>CV</td>
<td>DR</td>
<td>Combined</td>
<td>AUC</td>
<td>RP</td>
<td>AUC</td>
</tr>
<tr>
<td>-------------</td>
<td>-------</td>
<td>------------</td>
<td>-----</td>
<td>-------------</td>
<td>----------------</td>
<td>----</td>
<td>---</td>
<td>-------------</td>
<td>----</td>
<td>-----</td>
<td>-----------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Fried (62) 2016</td>
<td>III</td>
<td>NR</td>
<td>cc (PD)</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>27</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>195</td>
<td>None</td>
<td>Combined</td>
<td>Semi-automated</td>
<td>OS</td>
<td>C-index for OS 0.62</td>
</tr>
<tr>
<td>Carvalho (63) 2016</td>
<td>I-III</td>
<td>45-100 GY / NR</td>
<td>unspec (NR)</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>118</td>
<td>SM, TA, IF</td>
<td>CV</td>
<td>262</td>
<td>50</td>
<td>Compared</td>
<td>Manual</td>
<td>OS</td>
<td>C-index for OS 0.59</td>
</tr>
<tr>
<td>Buizza (64) 2013</td>
<td>I-IV</td>
<td>NR</td>
<td>ind or cc (PD)</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>290</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>30</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>OS</td>
<td>AUC for OS 0.87</td>
</tr>
<tr>
<td>Shen (65) 2017</td>
<td>I-IV</td>
<td>NR</td>
<td>PET-CT(p)</td>
<td>Yes</td>
<td>1014</td>
<td>IF, TA, FB</td>
<td>CV</td>
<td>483</td>
<td>125</td>
<td>No</td>
<td>Semi-automated</td>
<td>OS</td>
<td>AUC for OS 0.66</td>
<td>4 (11)</td>
</tr>
<tr>
<td>Astaraki (26) 2019</td>
<td>I-IV</td>
<td>&lt;79.2 GY / NR</td>
<td>ind or cc (PD)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>451</td>
<td>SM, IF, TA, FB</td>
<td>None</td>
<td>30</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>OS</td>
<td>AUC for OS 0.96</td>
</tr>
<tr>
<td>Van Timmeren (66) 2019</td>
<td>II-IV</td>
<td>45-74 GY / NR</td>
<td>ind/CC (NR)</td>
<td>PET-CT(p)</td>
<td>NR</td>
<td>2695</td>
<td>SM, IF, TA, FB</td>
<td>103</td>
<td>35</td>
<td>138</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Mahon (67) 2019</td>
<td>II-IV</td>
<td>60 GY / 30#</td>
<td>cc (NR)</td>
<td>MRI + 4D-CT(d) (0% &amp; 50%)</td>
<td>No</td>
<td>59</td>
<td>IF, DR</td>
<td>CV</td>
<td>15</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>TR, OS</td>
<td>AUC for OS 0.88</td>
</tr>
<tr>
<td>Krafft (68) 2019</td>
<td>NR</td>
<td>59-64 GY / 30-40#</td>
<td>ind, CC or Adj (NR)</td>
<td>4D-CT(p) (50%)</td>
<td>NR</td>
<td>6538</td>
<td>IF, TA, FB</td>
<td>CV</td>
<td>192</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>RP</td>
<td>AUC for RP 0.68</td>
</tr>
<tr>
<td>Liang (69) 2019</td>
<td>I-IV</td>
<td>NR</td>
<td>cc (NR)</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>129</td>
<td>IF</td>
<td>CV</td>
<td>70</td>
<td>None</td>
<td>Compared</td>
<td>Manual</td>
<td>RP</td>
<td>AUC for RP 0.78</td>
</tr>
</tbody>
</table>

Table 1. Studies utilising radiomics for prediction of lung cancer outcomes following radical (chemo)RT with conventional dose schedules (SACT Seq = systemic anticancer therapy sequence; CE = contrast enhancement; ind = induction; cc = concurrent; adj = adjuvant; PD = platinum doublet; SA = other single agent; (p) = radiotherapy planning scan; (d) = diagnostic scan; SM = shape metrics; IF = intensity features; TA = texture analysis; FB = filter-based features; DR = ‘delta’ radiomics; CP = clinicopathological; CV = cross-validation; TR = treatment response; pCR = pathological complete response; MPR = major pathological response; GRD = gross residual disease; LC = local control; LRC = loco-regional control; PFS = progression-free survival; DM = distant metastasis; OS = overall survival; LDC = lung density changes; RP = radiation pneumonitis; unspec = unspecified; NR = not recorded)

*study validated a radiomics signature defined previously
<table>
<thead>
<tr>
<th>Study</th>
<th>Stage</th>
<th>Dose &amp; Frac.</th>
<th>Imaging Modality (phase where applicable)</th>
<th>CE</th>
<th>Total No of Features</th>
<th>Features Categories</th>
<th>Model Development</th>
<th>CP Factors</th>
<th>Segmentation Method</th>
<th>Outcomes Tested (All)</th>
<th>Best Statistic Compared</th>
<th>Reported and Outcome</th>
<th>RQS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lafata (70) 2018</td>
<td>I</td>
<td>Mean 51 Gy / NR #</td>
<td>CT(p)</td>
<td>NR</td>
<td>43</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>70</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>LC, DM, PFS</td>
<td>AUC for LC 0.83</td>
</tr>
<tr>
<td>Pyka (72) 2015 I-IIA</td>
<td>24-45 Gy / 3-5 #</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>10</td>
<td>SM, IF, TA</td>
<td>None</td>
<td>45</td>
<td>None</td>
<td>No</td>
<td>Automated</td>
<td>LC, RC, DM, DSS, OS</td>
<td>AUC for LC 0.87</td>
<td>-4 (0)</td>
</tr>
<tr>
<td>Dussault (71) 2018</td>
<td>I-II</td>
<td>48-60 Gy / 3-5 #</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>184</td>
<td>SM, IF, TA</td>
<td>None</td>
<td>64</td>
<td>23</td>
<td>No</td>
<td>Semi-automated</td>
<td>LC, DM, RFS, DSS, OS</td>
<td>Accuracy for LC 0.98</td>
</tr>
<tr>
<td>Franceschini (73) 2019</td>
<td>I-IIA</td>
<td>48 Gy / 4 #</td>
<td>3D-CT(p)</td>
<td>No</td>
<td>41</td>
<td>SM, IF, TA</td>
<td>70</td>
<td>32</td>
<td>102</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>RC, PFS, DSS, OS</td>
</tr>
<tr>
<td>Li (74) 2017</td>
<td>I-IIA</td>
<td>48-60 Gy / 4-8 #</td>
<td>4D-CT(p) (Ave)</td>
<td>No</td>
<td>219</td>
<td>SM, TA</td>
<td>CV</td>
<td>92</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>LRC, RFS, OS</td>
<td>C-index for LRC 0.66</td>
</tr>
<tr>
<td>Li (75) 2018</td>
<td>I-IIA</td>
<td>48-60 Gy / 4-8 #</td>
<td>3D-CT(d)</td>
<td>Yes</td>
<td>219</td>
<td>IF, TA, FB</td>
<td>CV</td>
<td>59</td>
<td>None</td>
<td>No</td>
<td>Semi-automated</td>
<td>LRC, RFS, OS</td>
<td>C-index for LRC 0.77</td>
</tr>
<tr>
<td>Davatzikos (76) 2018</td>
<td>I</td>
<td>50 Gy / 4-5 #</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>343</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>80</td>
<td>None</td>
<td>No</td>
<td>NR</td>
<td>RC, OS</td>
<td>p&lt;0.001 for RC in 2 unsupervised clusters</td>
</tr>
<tr>
<td>Li (77) 2018</td>
<td>I-IIA</td>
<td>50 Gy / 4-5 #</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>722</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>100</td>
<td>None</td>
<td>Combined</td>
<td>Semi-automated</td>
<td>RC, OS</td>
<td>C-index for RC 0.66</td>
</tr>
<tr>
<td>Oikonomou (78) 2018</td>
<td>I-II</td>
<td>48-56 Gy / 4-5 #</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>42</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>150</td>
<td>None</td>
<td>Combined</td>
<td>Manual</td>
<td>RC, DM, RFS, DSS, OS</td>
<td>HR for RC 0.51</td>
</tr>
<tr>
<td>Hao (79) 2018</td>
<td>I</td>
<td>NR</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>34</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>71</td>
<td>23</td>
<td>No</td>
<td>Automated</td>
<td>DM</td>
<td>AUC for DM 0.79</td>
</tr>
<tr>
<td>Wu (80) 2016</td>
<td>I</td>
<td>25-60 Gy / 1-5 #</td>
<td>PET-CT(p)</td>
<td>NR</td>
<td>70</td>
<td>SM, IF, TA, FB</td>
<td>CV</td>
<td>70</td>
<td>31</td>
<td>Combined</td>
<td>Manual</td>
<td>DM</td>
<td>AUC for DM 0.81</td>
</tr>
<tr>
<td>Zhou (81) 2017</td>
<td>I</td>
<td>NR</td>
<td>PET-CT(d)</td>
<td>NR</td>
<td>30</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>52</td>
<td>None</td>
<td>Combined</td>
<td>Semi-automated</td>
<td>DM</td>
<td>AUC for DM 0.87</td>
</tr>
<tr>
<td>Huyynh (82) 2017</td>
<td>I-IIA</td>
<td>48-60 Gy / 10-16 #</td>
<td>3D-CT(p) &amp; 4D-CT(p) (Ave)</td>
<td>NR</td>
<td>644</td>
<td>SM, IF, TA,</td>
<td>CV</td>
<td>119</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>LRC, DM</td>
<td>C-index for DM 0.67</td>
</tr>
<tr>
<td>Huyynh (83) 2016</td>
<td>I-II</td>
<td>54-60 Gy / 3-5 #</td>
<td>3D-CT(d)</td>
<td>Yes</td>
<td>1605</td>
<td>SM, IF, TA, FB</td>
<td>90</td>
<td>23</td>
<td>113</td>
<td>None</td>
<td>Compared only</td>
<td>Manual</td>
<td>LC, LOC, RC, LRC, DM, CSS, OS</td>
</tr>
<tr>
<td>Starkov (84) 2019</td>
<td>I</td>
<td>25-60 Gy / 1-5 #</td>
<td>3D-CT(p)</td>
<td>No</td>
<td>30</td>
<td>FB</td>
<td>CV</td>
<td>116</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>LC, RC, DM, PFS, OS</td>
<td>p&lt;0.05 for OS</td>
</tr>
<tr>
<td>Yu (85) 2017*</td>
<td>I</td>
<td>50-70 Gy / 4-10 #</td>
<td>3D-CT(d)</td>
<td>Yes</td>
<td>12</td>
<td>IF, TA</td>
<td>CV</td>
<td>147</td>
<td>295</td>
<td>No</td>
<td>Manual</td>
<td>RC, DM, OS</td>
<td>C-index for OS 0.64</td>
</tr>
<tr>
<td>Lee (86) 2018</td>
<td>I</td>
<td>NR</td>
<td>3D-CT(p) &amp; PET-CT(p)</td>
<td>NR</td>
<td>39</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>117</td>
<td>88</td>
<td>Combined</td>
<td>Manual</td>
<td>OS</td>
<td>C-index for OS 0.64</td>
</tr>
<tr>
<td>Zhang (87) 2017</td>
<td>NR</td>
<td>NR</td>
<td>3D-CT(NR)</td>
<td>NR</td>
<td>30</td>
<td>SM, IF, TA</td>
<td>CV</td>
<td>112</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>LRC, DM, RFS, OS</td>
<td>AUC for OS 0.79</td>
</tr>
<tr>
<td>Soufi (134) 2018</td>
<td>IIIB</td>
<td>51-9.2 Gy / 33-44 #</td>
<td>3D-CT(p)</td>
<td>NR</td>
<td>47</td>
<td>IF, TA</td>
<td>11</td>
<td>221</td>
<td>232</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Du (135) 2019</td>
<td>I-II</td>
<td>48-60 Gy / 4-6 #</td>
<td>3D-CT &amp; 4D-CT(p) (all phases)</td>
<td>NR</td>
<td>841</td>
<td>SM, IF, TA, FB</td>
<td>20</td>
<td>140</td>
<td>160</td>
<td>None</td>
<td>No</td>
<td>Manual</td>
<td>OS</td>
</tr>
<tr>
<td>Lovinfosse (88) 2016</td>
<td>I-IIA</td>
<td>45-60 Gy / 3-5 #</td>
<td>PET-CT(d)</td>
<td>No</td>
<td>17</td>
<td>SM, IF, TA</td>
<td>None</td>
<td>63</td>
<td>None</td>
<td>Compared only</td>
<td>Semi-automated</td>
<td>RFS, DSS, OS</td>
<td>HR for DSS 0.82</td>
</tr>
</tbody>
</table>
Table 2. Studies utilising radiomics for prediction of lung cancer outcomes following stereotactic RT (BH = breath-hold; 3D = three-dimensional; 4D = four-dimensional; CT = computed tomography; PET = positron emission tomography; MRI = magnetic resonance imaging; CE = contrast enhancement; (p) = radiotherapy planning scan; (d) = diagnostic scan; SM = shape metrics; IF = intensity features; TA = texture analysis; FB = filter-based features; CP = clinicopathological; CV = cross-validation; LC = local control; LOC = lobar control; RC = regional control; LRC = loco-regional control; RFS = recurrence-free survival; PFS = progression-free survival; DM = distant metastasis; OS = overall survival; DSS = disease-specific survival; LDC = lung density changes; unspec = unspecified; NR = not recorded)

Bousabarah (89) 2019  
I-IIA & IV  
25-60Gy / 1-8#  
BH-CT(p)  
Yes  
803  
SM, IF, TA  
CV  
110  
None  
No  
Manual  
LC, RFS, OS, LDC  
C-index for LDC 0.72  
5 (14)

Mattonen (90) 2014  
I-IIA  
55-60Gy / 3-8#  
3D-CT(d)  
Yes  
10  
SM, IF, TA  
CV  
22  
None  
No  
Manual  
LDC  
AUC for LDC 0.875  
-6 (0)

Mattonen (91) 2015  
I-IIA  
54-60Gy / 3-8#  
4D-CT(p) (max inhale)  
NR  
104  
SM, IF, TA  
CV  
22  
None  
No  
Semi-automated  
LDC  
AUC for LDC 0.77  
1 (3)

*this signature was developed in a surgical cohort and validated in a stereotactic cohort