

Role of radiomics in predicting immunotherapy response

Gargi Kothari^{1,2} 

¹ Department of Radiation Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

² Sir Peter MacCallum Department of Oncology, University of Melbourne, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

G Kothari MBBS, FRANZCR.

Correspondence

Dr Gargi Kothari, Peter MacCallum Cancer Centre, Victorian Comprehensive Cancer Centre Building, 305 Grattan Street, Melbourne Vic. 3000, Australia.
Email: gargi.kothari@petermac.org

Conflict of interest: No conflict of interest.

Submitted 6 March 2022; accepted 2 May 2022.

doi:10.1111/1754-9485.13426

Summary

Immunotherapies have revolutionised cancer management. Despite their success, durable responses are limited to a subset of patients. Prediction of immunotherapy response in patients has proven to be difficult due to a lack of robust biomarkers. Routinely collected imaging may offer an additional information source to personalise patient treatment, with advantages over tissue-based biomarkers. Quantitative image analysis or radiomics, which involves the high-throughput extraction of imaging features, has the potential to non-invasively predict cancer histology, outcomes and prognosis. This review evaluates the value of radiomics in patients undergoing immunotherapy, with a summary provided of the performance of radiomics models in predicting immunotherapy response and toxicity, as well as immune correlates. Much of the literature focussed on clinical endpoints and correlates to tissue biomarkers, particularly in lung cancer, while few studies investigated association with immune-related adverse events. Strengths of the studies included more frequent use of clinical trial datasets, homogenous patient cohorts and high-quality diagnostic scans. Limitations of the studies include heterogeneity in study methodology, lack of well-defined homogenous imaging datasets, limited open publishing of imaging datasets, coding and parameters used for radiomics signature development and limited use of external validation datasets. Future research should address the above limitations, as well as further explore the relationship between radiomics and immune-related adverse effects and less well-studied biological correlates such as tumour mutational burden, and incorporate known clinical prognostic scores into radiomics models.

Key words: biomarkers; immunotherapy; radiomics.

Introduction

For decades, chemotherapy has been the standard of care in patients with advanced cancer. More recently, the advent of immunotherapy including immune checkpoint inhibitors (ICI) has changed the landscape of treatment and clinical outcomes for patients. Immune checkpoints suppress the native anti-tumour immune response. ICI impede these pathways and thereby allow for an effective anti-tumour response by the immune system. The first U.S. Food and Drug Administration (FDA) approved ICI was ipilimumab, which targets the cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) checkpoint protein, and was found to be efficacious in metastatic melanoma.¹ Subsequently, two agents targeting the programmed cell death protein 1/ programmed death ligand 1 (PD-1/PD-L1) immune checkpoint, pembrolizumab and

nivolumab were also approved for metastatic melanoma.^{2,3} Since then, the range of indications for PD-1/PD-L1 inhibitors has expanded rapidly with improvements in survival seen in several different tumour streams in the locally advanced,^{4,5} recurrent⁶ and metastatic setting.^{6–8}

Tissue biomarkers

Despite the success of immunotherapy, the benefits of this treatment are largely seen in a subset of patients, with between 17% and 48% of patients responding to treatment.^{7,9} Some patients also experience significant adverse effects, including colitis, pneumonitis, hepatitis and hypophysitis, which may limit their clinical use and in certain cases can be life-threatening.¹⁰ Improved patient selection and personalisation of treatment is,

therefore, important to identify patients that may respond poorly to immunotherapy alone and potentially may benefit from alternative or combined therapies. Furthermore, it will limit the exposure of patients unnecessarily to possible toxicities associated with immunotherapy treatment. Better patient selection will also improve the cost-effectiveness of immunotherapy, which is expensive compared with other anti-cancer treatments.

Tissue biomarkers may be used to predict response to immunotherapy and select patients most likely to benefit from treatment. PD-L1 is perhaps the most extensively studied and clinically utilised biomarker for immunotherapy response. Notably, in patients with non-small cell lung cancer (NSCLC) with PD-L1 expression $\geq 50\%$, pembrolizumab results in improved overall survival (OS) compared with chemotherapy⁹ and is now routinely tested in the clinic in this patient group. Tumour mutational burden (TMB) is also gaining interest as a predictor of response to immunotherapy, with some studies suggesting a high TMB correlates with better response to immunotherapy.^{11,12} There are, however, a number of disadvantages and challenges related to tissue biomarkers. Tissue biomarkers must be obtained through an invasive procedure, which may be associated with risks such as bleeding or infection. In certain cases, obtaining tissue may not be feasible due to these risks being unacceptably high. Furthermore, due to their invasive nature, tissue biomarkers are typically assessed at a single timepoint only. As a result, assessment of any changes in biomarkers that occur over time and with treatment may not be possible. A biopsy also only allows for sampling of a small part of the total tumour, which may not adequately capture any intra-tumoural or inter-tumoural heterogeneity that exists within a patient.

Imaging biomarkers

Imaging biomarkers have several advantages compared with tissue biomarkers. Imaging is a non-invasive procedure that is performed routinely in patients with cancer. Imaging also allows for visualisation and assessment of the whole tumour as well as any sites of metastatic disease. The ability to capture the entire spectrum of disease within a patient may allow for better assessment of any heterogeneity that exists within or between tumours and which may drive differential prognoses and responses to treatment. Early identification of poor responders, or heterogenous response, may also provide opportunities to escalate or change treatment, or consider multimodality treatment options, such as the addition of stereotactic ablative body radiotherapy (SABR) for oligoprogressive disease.

A widely used system for response assessment to systemic therapies used in clinical trials is 'Response Evaluation Criteria in Solid Tumours' (RECIST).¹³ RECIST measurements and definitions of response are based

upon assessment of the maximum diameter of malignant lesions over a single slice of a CT scan, with total measurement of up to five 'target' lesions. Some patients receiving immunotherapy, however, may experience atypical responses, such as pseudoprogression (PSPD), in which tumours initially enlarge in size, followed by a delayed response and tumour shrinkage. This pattern would be falsely characterised as progressive disease as per RECIST and may lead to premature cessation of treatment. Modified systems have been developed to better account for the differences in response seen following immunotherapy, including immune RECIST (iRECIST), which requires confirmation of progressive disease at a subsequent timepoint.¹⁴ Nevertheless, these systems still rely on a single imaging parameter to determine treatment response, and currently, neither RECIST nor iRECIST consistently provides reliable estimates of OS in patients receiving immunotherapy.¹⁵ As a consequence, our current system for response assessment may result in negative outcomes, including delays in treatment decisions for patients.

Radiomics

Radiomics is a tool that involves the extraction of numerous quantitative features from standard imaging. Radiomics is based upon the assumption that extracted imaging data are the product of mechanisms occurring at a genetic and molecular level linked to the genotypic and phenotypic characteristics of the tissue and may, therefore, be used as imaging biomarkers to predict patient outcomes.¹⁶

The typical workflow for radiomics analysis involves segmentation of a chosen volume of interest (VOI) from imaging datasets. Specialised software is used to perform high-throughput extraction of quantitative image features to characterise the VOI. These features can be broadly grouped into shape, first-order (image intensity) and second-order (texture) features.¹⁷ Analysis of this data and correlation with clinically meaningful endpoints may subsequently be performed using statistical or machine learning techniques, which generate predictions or classifications for specific oncological endpoints. In machine learning, classification is a supervised learning task of inferring a function from labelled training data.¹⁸ Various classification algorithms such as logistic regression, random forests (RF) and support vector machines (SVM) may be used to build predictive models.¹⁹ Unsupervised machine learning techniques (such as hierarchical clustering) may also be used to enhance our ability to analyse images.²⁰

There are numerous advantages of a radiomics approach to image assessment including the ability to gain additional information from a routinely performed procedure, its non-invasive nature and the ability to assess the entire tumour. In addition, tumours may be assessed at multiple timepoints, with assessment of

changes in radiomics features over time termed 'delta-radiomics',²¹ rather than information derived from a single biopsy performed at a single timepoint alone. The utility of radiomics to enhance our understanding of cancer diagnosis, prognosis and response to treatment, as well as better understand tumour heterogeneity and the tumour microenvironment, has been investigated,^{22,23} with studies relating to immunotherapy emerging over recent years. This review aims to summarise the current literature relating to the utility of radiomics in patients with cancer undergoing immunotherapy treatment, including its role in predicting treatment response and toxicity, as well as correlates to relevant tumour biology.

Predicting response to immunotherapy

A summary of selected full-text English radiomics literature relating to immunotherapy across tumour types reporting outcomes of radiomics models is listed in Table 1 following an initial review of the literature using the search terms 'immunotherapy' AND 'radiomics' on PubMed. Eighteen studies included in the table investigated the ability of radiomics features to predict clinical outcomes, including treatment response and survival outcomes. One of the earliest studies published assessing the value of radiomics in predicting immunotherapy response was by Sun *et al.*²⁴ This study trained a radiomics model on a dataset of 135 patients from the prospective MOSCATO trial. This trial included a mixed cohort of patients with advanced solid tumours and aimed to analyse the benefit of high-throughput genomic analyses. All patients within this study had both CT and RNA-sequencing (RNA-seq) data available for review. The radiomics model, which extracted radiomics features from the CT scan of the biopsied lesion, was trained to predict the quantity of tumour-infiltrating CD8 cells, as estimated by RNA-seq. The model included eight features, including *tum_minValue*, four Grey-Level Run Length Matrix (GLRLM) textural features from both the tumour and periphery, the location of the selected VOI and CT peak kilovoltage (kVp). GLRLM features represent the overall heterogeneity or homogeneity of a region, with results from this study showing higher levels of homogeneity correlating with a high CD8 level. Interestingly, the signature also incorporated the location of the lesion, highlighting potential differences in the tumour immune environment and subsequently the radiomic feature results depending upon the site of metastasis. Furthermore, it confirms the importance in considering the definition and details of the specified VOI particularly in radiomics studies of patients with metastatic disease. The developed radiomics signature was subsequently externally validated using CT and RNA-seq data of 119 patients from The Cancer Genome Atlas (AUC = 0.67, 95% CI 0.57–0.77), as well as a randomly selected cohort of patients from Gustave Roussy Cancer Centre that were classified as having either 'immune-inflamed'

or 'immune-desert' phenotypes based upon the primary tumour type and its recognised response to immunotherapy (AUC 0.76, 95% CI 0.66–0.86). The same model was shown to be associated with OS in multivariate analysis (HR 0.52, 95% CI 0.35–0.79, $P = 0.0022$) on 137 patients enrolled in five Phase I immunotherapy trials at the Gustave Roussy Cancer Centre. Some minor limitations of the study included heterogeneity of CT scanners used with lack of information regarding contrast, no testing of feature repeatability or reproducibility and no details regarding cross-validation used during the feature selection process. Some of the significant strengths, however, of this study included the use of an open-access radiomics platform, details provided regarding image pre-processing, as well as explicit reporting of the radiomics model used and features included, allowing this model to be subsequently also externally validated in patients undergoing immunotherapy and SABR combination treatment with results reported in two separate studies.^{25,26} The first study by Sun *et al.* utilised a dataset composed of 94 patients from six independent clinical trials and extracted radiomics features from 100 irradiated and 189 non-irradiated lesions. The median number of fractions and dose by fraction given were 3 fractions of 8 Gy (IQR 6–12). Median time from start of immunotherapy to start of SABR was 21 days (IQR 8–24), with 16 patients starting SABR prior to immunotherapy. Median time to the first follow-up CT scan was 2.8 months. The radiomics model was predictive of per lesion response measured at the time of the first follow-up CT scan with an AUC of 0.63 (95% CI 0.56–0.71). Interestingly, entropy (indicating heterogeneity) of the distribution of the radiomics scores of lesions within a patient was associated with PFS and OS and could discriminate between patients that had uniform vs. mixed responses to treatment. The second study by Korpics *et al.*²⁶ validated the model using an imaging dataset from 68 patients undergoing SABR followed by pembrolizumab on a Phase I trial (NCT02608385).²⁷ Radiomics features were extracted from 139 irradiated lesions. The radiomics model (using a pre-specified, although arbitrary 25% percentile cut-off), was predictive of per lesion response at first follow-up (odds ratio [OR] 10.2; 95% CI 1.76–59.17; $P = 0.012$). The average of the radiomics score for each irradiated lesion was used to determine a radiomics score for each patient. Patients with a high radiomics score had improved PFS (HR 0.47, 95% CI 0.26–0.85; $P = 0.013$) and OS (HR 0.39, 95% CI, 0.20–0.75; $P = 0.005$).

More recently, several studies investigating the value of radiomics in predicting response to immunotherapy have analysed less heterogeneous patient cohorts, most commonly including only patients with advanced NSCLC. A systematic review and meta-analysis of radiomic models predicting immunotherapy response and outcome in patients with NSCLC was recently published.²⁸ The systematic review included 15 studies with datasets ranging

Table 1. Summary of immunotherapy-related radiomics studies

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Mixed cancers Sun et al., 2018 ²⁴	Mixed cancers	Mixed	T: 135 V1: 100 V2: 137 V3: 119	CT	T, V3: Estimation of CD8 cell infiltrate (RNA-seq) V1: Association with tumour immune phenotype V2: OS	LIFEX software (v 3.44)	T: biopsied lesion V1: largest lesion V2: radiologist selected single RECIST target lesion V3: primary lesion VOI for the lesion and peripheral ring (4 mm thickness) created.	78 radiomics features, five lesion locations and peak kilovoltage extracted. Linear elastic-net model used. Eight features used in radiomics model.	T: 0.74 V1: 0.76 V2: HR 0.52, 95% CI 0.35–0.79, P = 0.0022 V3: 0.67
Trebeschi et al., 2019 ⁶⁵	Non-small cell lung cancer and melanoma	Immunotherapy	T: 133 (81 lung; 52 melanoma) V: 70 (42 lung; 28 melanoma)	CT (CE) at baseline and 12 weeks	Per lesion response 1 year OS	Missing	Target lesions (well demarcated at baseline and follow-up and ≥ 5 mm).	A random forest with wrapper feature selection was used.	Per lesion response (NSCLC and melanoma) V: 0.66 NSCLC V: 0.76 Melanoma: V: 0.77
Sun et al., 2020 ²⁵	Advanced solid cancers	Immunotherapy (anti-PD-1/PD-L1 or anti-CTLA-4 monotherapy) + stereotactic radiotherapy	V: 94 (100 irradiated and 189 non-irradiated lesions)	CT (CE)	Per lesion response at first follow-up (RECIST 1.1) Out of field abscopal response PFS and OS	LIFEX software (v 3.44)	Any irradiated and non-irradiated lesions ≥ 5 mm identifiable on baseline and follow-up CT. VOI for the lesion and peripheral ring (4 mm thickness) created.	Details of model development in Sun et al., 2018.	Per lesion response V: 0.63 Abscopal V: 0.70 PFS V: HR 1.67, P = 0.040 OS V: HR = 2.08, P = 0.023
Korpcs et al., 2020 ²⁶	Metastatic solid cancers	Pembrolizumab + stereotactic radiotherapy	V: 68 (139 lesions)	CT (CE)	Per lesion response at first follow-up (RECIST 1.1)	LIFEX software (v 4.60)	All irradiated lesions. VOI for the lesion and peripheral ring (4 mm thickness) created.	Details of model development in Sun et al., 2018. A pre-specified cut-off of the 25th percentile used to divide 'low' vs. 'high' radiomics score.	Per lesion response V: OR 10.2, 95% CI 1.76–59.17, P = 0.012 PFS V: HR 0.47, 95% CI 0.26–0.85, P = 0.013 OS V: HR 0.39, 95% CI 0.20–0.75, P = 0.005

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Colen <i>et al.</i> , 2018 ⁴⁴	Advanced cancers	Immunotherapy (immune checkpoint inhibitors, cytokines, vaccines, or immunotherapy-based combinations)	32	CT (CE)	Treatment induced pneumonitis	Missing	6 VOI per patient, with three regions in each lung	1,860 features extracted. Maximum relevance and minimum redundancy, anomaly detection algorithm, and leave-one-out cross-validation used.	1.00
Lung cancer Liu <i>et al.</i> , 2021 ²⁹	Advanced non-small cell lung cancer	Immune checkpoint inhibitors	Baseline: T: 137 V: 60 Delta-radiomics T: 112 V: 49	CT	Responders vs. non-responders (progressive disease as per RECIST) at 6 months	In-house software (Analysis Kit, version 3.2.5, GE Healthcare)	Largest RECIST target lesion	402 features extracted. Minimum Redundancy Maximum Relevance and a multivariate LASSO logistic regression analysis with backward elimination method and cross-validation was used.	Baseline Radiomics T: 0.59 V: 0.51 Clinical-radiomics T: 0.65 V: 0.52 Delta-radiomics Radiomics T: 0.81 V: 0.80 Clinical-Radiomics T: 0.83 V: 0.81
Derle <i>et al.</i> , 2020 ³⁰	Stage IIIB – IV non-small cell lung cancer	Nivolumab	T: 72 V: 20	CT at baseline and 8 weeks	Responders vs. non-responders (defined by change in size of the largest measurable lung lesion)	Missing	Largest measurable lung lesion	Delta-radiomics signature developed 1,160 radiomics features were extracted Machine learning was implemented to select up to four features. Four delta-radiomics features included in model.	T: 0.80 V: 0.77
Khorrani <i>et al.</i> , 2020 ³²	Lung cancer	Immune checkpoint inhibitors	T: 50 V1: 62 V2: 27	CT (CE) at baseline and 6–8 weeks	Responders vs. non-responders (progressive disease as per RECIST 1.1) OS	MATLAB 2018b (Mathworks) with an in-house developed toolbox	Lung nodules and a 30-mm perinodular radius, which was divided into 15 annular rings of 2 mm each.	99 delta-radiomics features extracted. A linear discriminant analysis classifier employed. Features stable on a test–retest dataset and predictive on Wilcoxon rank-sum selected.	Response T: 0.88 V1: 0.85 V2: 0.81 OS T: C-index 0.72 V1: C-index 0.69 V2: C-index 0.68
Yang <i>et al.</i> , 2021 ³¹	Stage IIIB – IV non-small cell lung cancer	Immunotherapy (anti-PD-1/IpD-L1 monotherapy)	200	CT	Responders vs. non-responders (progressive disease as per RECIST 1.1) at 90 days	PyRadiomics (Python 3.7.3, PyRadiomics 2.2.0)	Primary tumour	Serial radiomics, laboratory data and baseline clinical data used to develop deep learning models.	0.80

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Tunali et al. 2019 ³⁷	Non-small cell lung cancer	Immune checkpoint inhibitors	T: 228	CT	Hyperprogression	Missing	Largest lung lesion – tumour and border regions	600 radiomics features extracted Non-reproducible features excluded. Features significant on univariable and multivariable analysis chosen. SMOTE used.	T: 0.674 Clinical-radiomics T: 0.865
Vaidya et al. 2020 ³⁸	Non-small cell lung cancer	Immune checkpoint inhibitors	T: 30 V: 79	CT	Hyperprogression	In-house software implemented on a MATLAB release V.2015 platform.	All RECIST target lesions (intra and peri-tumoural regions)	Model included 1 radiomics feature. 198 features extracted. Minimum redundancy maximum relevance feature selection used with cross-validation. Random forest classifier used to build model.	T: 0.85 V: 0.96
Mu et al. 2020 ³⁴	Stage IIIB – IV non-small cell lung cancer	Immune checkpoint inhibitors	T: 99 V1: 47 V2: 48	PET/CT	DCB PFS OS	Missing	Primary tumour	790 features extracted. Feature selection based upon internal stability of features and LASSO. Features combined into a multiparametric radiomic signature. Cross-validation performed.	DCB T: 0.86 V1: 0.83 V2: 0.81 PFS T: C-index 0.74 V1: C-index 0.74 V2: C-index 0.77 OS
Mu et al. 2021 ³³	Stage IIIB – IV non-small cell lung cancer	Immune checkpoint inhibitors	T: 123 V1: 52 V2: 35	PET/CT	Cachexia DCB	MATLAB 2020a	Primary tumour and muscles at the third lumbar vertebra including rectus abdominus, abdominal, psoas, and paraspinous	1,053 features extracted. Features with high inter-rater agreement and significant on two-sample t-test for cachexia selected. Highly correlated features excluded. LASSO logistic regression analysis and minimum mean cross-validated error used.	Cachexia T: 0.77 V1: 0.75 V2: 0.74 DCB T: 0.71 V1: 0.66 V2: 0.70 AUC 0.90
Valentinuzzi et al. 2020 ³⁶	Metastatic non-small cell lung cancer	Pembrolizumab	30	PET/CT	OS	In-house software	Primary tumour	Eight features extracted. Univariate and multivariate logistic regression used with cross-validation.	AUC 0.90

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Nardone et al. 2020 ³⁹	Metastatic non-small cell lung cancer	Nivolumab	T: 35 V: 24	CT (pre and post CE)	PFS OS	LIFEX software	Evaluable target lesion within the lung	14 features extracted. Features with high interobserver variation in contouring excluded. Five features chosen within model.	Median PFS T: 10 vs. 3 months, P = 0.044 V: 15 vs. 6 months, P = 0.048 Median OS T: Not reached vs. 5 months, P = 0.023 V: 26 vs. 6 months, P = 0.032
Polverari et al. 2020 ³⁵	Stage IIIB – IV non-small cell lung cancer	Immunotherapy (anti-PD-1/PD-L1)	57	PET/CT	Responders vs. non-responders (progressive disease as per RECIST 1.1) PD-L1 (<1%, 1%–50%, >50%)	LIFEX software v 5.1	Primary tumour	Univariate analysis with Kruskal–Wallis test was performed.	Response: MTV (P = 0.035), TLG (0.037), asymmetry (P = 0.032) and kurtosis (P = 0.046). PD-L1: Coarseness (P = 0.025), GLZLM_ZLNU (P = 0.035)
Jiang et al. 2020 ⁴⁹	Stage I – IV non-small cell lung cancer	Surgery	T: 266 V: 133	PET/CT	PD-L1	Python 3.7.0 (packages: simpleITK, pydicom, and pywavelet).	Primary tumour	1,744 features from both PET and CT images extracted. Features filtered with automatic relevance determination and minimised with LASSO model with cross-validation used	PD-L1 > 1% V: 0.97 (CT alone) PD-L1 > 50% V: 0.80 (CT alone)
Sun et al. 2020 ⁶⁶	Non – small cell lung cancer	N/A	T: 260 V: 130	CT	PD-L1	In-house texture analysis software	Lung tumour	200 radiomic features extracted. LASSO multivariable logistic regression analysis used. Nine radiomics features in model.	Clinical-radiomics T: 0.829 V: 0.848
Yoon et al. 2020 ⁶⁷	Stage IIIA – IV adenocarcinoma of the lung	N/A	153	CT (CE)	PD-L1 > 50%	Missing	Primary tumour	58 features extracted. Features with high interobserver agreement and statistical significance for PD-L1 selected. LASSO logistic regression model with three-fold cross-validation used. Radiomics model had 4 features.	Radiomics 0.550 Clinical-radiomics 0.667

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Tian <i>et al.</i> 2021 ⁵⁰	Stage IV non-small cell lung cancer	N/A	T: 750 V: 93 Test: 96	CT	PD-L1 > 50%	Missing	Primary tumour	1,316 features extracted. Radiomics model built using a deep learning approach. The network consisted of three parts, a deep learning feature extraction module based on densenet121, a handcrafted conventional radiomic feature extraction module, and a classifier module based on the fully connected classification layer.	Radiomics: T: 0.71, V: 0.67, Test: 0.75 Deep learning: T: 0.63 V: 0.67 Test: 0.68 Radiomics-deep learning: T: 0.78 V: 0.71 Test: 0.76
Yoon <i>et al.</i> 2020 ⁴⁸	Non-small cell lung cancer	N/A	T: 89 V: 60	CT contrast and non-CE	TILs (type 1 helper T (Th1) cells, type 2 helper T (Th2) cells, and cytotoxic T cells (CTL))	PyRadiomics and in-house code using MATLAB	Missing	239 features extracted. Multiple different machine learning algorithms used.	Th1 (peralised discriminant analysis): T: 0.711 V: 0.564 Th2 (linear discriminant analysis): T: 0.772 V: 0.684 CTL (model used missing): T: 0.681 V: 0.612
He <i>et al.</i> 2020 ⁵³	Non - small cell lung cancer	N/A	T: 236 V: 26 Test: 65	CT	TMB (high vs. low)	Missing	Primary tumour	The feature extraction module of the deep learning model was mainly 3D-densenet. The module contained four blocks and 1,020 deep learning features. The fully connected network was chosen as the classifier.	Radiomics T: 0.75 Test: 0.74 Deep learning T: 0.85 Test: 0.81
Mu <i>et al.</i> 2020 ⁴³	Stage IIIB - IV non-small cell lung cancer	Immune checkpoint inhibitors +/- chemotherapy	T: 97 V1: 49 V2: 48	PET/CT	Immune-related severe adverse events	Missing	Primary tumour	1,092 features extracted. LASSO method and minimum mean cross-validated error used. Five radiomics features in model.	Radiomics: T: 0.88 V1: 0.90 V2: 0.86 Clinical-radiomics T: 0.92 V1: 0.92 V2: 0.88

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Melanoma Basler <i>et al.</i> , 2020 ⁴²	Metastatic melanoma	Immune checkpoint inhibitors	112 (716 lesions)	PET/CT at baseline, 3 and 6 months	Lesion-individual level pseudoprogression (RECIST 1.1)	In-house software 'Z-Rad' written in Python	All metastatic lesions	172 features extracted. Correlated features removed. Features significant on univariate analysis selected. A logistic regression model regularised with elastic net with cross-validation used. PCA and random forest used to develop survival model with 5 variables.	Delta-Radiomics 0.79 Delta-radiomics + blood marker (LDH/5100) 0.82 V: 0.92
Derde <i>et al.</i> , 2022 ⁴¹	Advanced melanoma	Immune checkpoint inhibitors (pembrolizumab/alone) ipilimumab)	T: 252 V: 287 (pembrolizumab/alone)	CT (CE) at baseline and 3 months	OS (6 months post-treatment)	Missing	All measurable lesions at baseline and 3 months		
Bladder cancer Park <i>et al.</i> , 2020 ⁴⁹	Metastatic urothelial carcinoma	Immunotherapy (anti-PD-1/PD-L1)	T: 41 V: 21	CT (CE)	Objective response rate (partial or complete response) Disease control rate (partial or complete response or stable disease) (RECIST 1.1.)	In-house software	Metastatic lesions (up to 2 per organ)	49 features extracted. Robust features selected. Highly correlated features excluded. A LASSO multivariate logistic regression model developed.	Clinical-Radiomics Objective response rate T: 0.87 V: 0.77 Disease control T: 0.87 V: 0.88
Gastrointestinal cancers Chen <i>et al.</i> , 2019 ⁴⁸	Hepatocellular cancer	N/A	T: 150 V: 57	MRI (CE)	Immunoscore (density of CD3+ and CD8+ T cells of tumour core and invasive margin)	Artificial Intelligence Kit software (A.K. software, GE Healthcare)	Primary tumour (intra and peri-tumoural)	1,044 features extracted. Extremely randomised tree method used. 70 radiomics features selected in model.	Radiomics T: 0.904 V: 0.899 Clinical-Radiomics T: 0.926 V: 0.934
Liao <i>et al.</i> , 2019 ⁵⁰	Hepatocellular cancer	N/A	T: 100 V: 42	CT (CE)	CD8+ T cells on immunohistochemistry	Artificial Intelligence Kit software, GE Healthcare)	Primary tumour	385 imaging features extracted. Minimum inhibitory concentration test and interfeature correlation with Elastic-net regularised regression analysis used. Model contains seven radiomics variables.	T: 0.751 V: 0.705
Gao <i>et al.</i> , 2020 ⁷¹	Gastric cancer	N/A	T: 90 V: 45 Test: 30	CT (CE)	Tumour-infiltrating regulatory T (TITreg) cells	PyRadiomics	Primary tumour	859 features extracted. Features robust to interobserver variation in contouring and significant on univariate analysis included. LASSO logistic regression model included six radiomics features.	T: 0.884 V: 0.869 Test: 0.847

Table 1. (continued)

Author, Date	Tumour stream	Treatment	Number of patients	Imaging	Endpoint	Radiomics software	Volume of interest	Radiomics model	Performance of model (AUC)
Wen <i>et al.</i> 2020 ⁵¹	Oesophageal squamous cell carcinoma	N/A	T: 160 V: 60	CT (CE)	PD-L1 and CD8 + TILs	Imaging Biomarker Explorer software (IBEX)	Primary tumour	462 features extracted. LASSO multivariable logistic regression used.	Radiomics PD-L1 T: 0.784 V: 0.750 CD8+ TIL T: 0.764 V: 0.728 Clinical-radiomics PD-L1 T: 0.871 V: 0.692 CD8+ TIL T: 0.832 V: 0.795
Iwatae <i>et al.</i> 2020 ⁷²	Pancreatic ductal adenocarcinoma	N/A	107	CT (CE)	PD-L1	PyRadiomics v2.2.0	Primary tumour +4 mm to include peri-tumoural region	1,037 features extracted. Mann-Whitney U test and recursive feature elimination using random forest used to select features. XGBoost with cross-validation used to develop model.	0.683
Pernicka <i>et al.</i> 2019 ⁵²	Stage II-III colon cancer	N/A	T: 139 V: 59	CT	Microsatellite instability	In-house software code using MATLAB R2015a	Primary tumour	254 features extracted. Wilcoxon rank-sum test used to select features, highly correlated features excluded, random forest used to build model.	Radiomics V: 0.76 Clinical-radiomics T: 0.80 V: 0.79
Li <i>et al.</i> 2021 ⁷³	Central nervous system tumours Low grade glioma	N/A	T: 68 V: 56	MRI (CE)	iMRiskScore (immunophenoscore-derived mRNA risk score)	Pyradiomics (PyTorch module in Python v3.6 used to construct neural network models)	Primary tumour	17,722 features extracted and used to create a neural network-based deep learning model.	T: 0.821 V: 0.708
Yu <i>et al.</i> 2021 ⁷⁴	Breast cancer	N/A	T: 85 V: 36	Mammogram	TILs	Pyradiomics (v2.2.0)	Primary tumour	612 features extracted. Recursive feature elimination and logistic regression used. Model included six radiomics features.	T: 0.83 V: 0.79

AUC, Area under the Receiver Operator Characteristic Curve; CE, contrast-enhanced; CI, confidence interval; C-index, concordance index; CTLA-4, cytotoxic T-lymphocyte-associated antigen 4; DCB, durable clinical benefit; HR, hazard ratio; ICC, intra-class correlation coefficient; iRECIST, immune Response Evaluation Criteria in Solid Tumours; LASSO, least absolute shrinkage and selection operator; N/A, not applicable; OR, odds ratio; OS, overall survival; PCA, principal component analysis; PD-1, programmed cell death protein 1; PD-L1, programmed death ligand 1; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumours; SMOTE, Synthetic Minority Oversampling Technique; T, training; TILs, tumour-infiltrating lymphocytes; TMB, tumour mutational burden; V, validation.

from 30 to 228 patients, with 10 studies included within the meta-analysis. Imaging parameters were generally well documented within the studies. While most studies utilised a baseline CT scan to extract radiomics features, a few studies also used follow-up CT scans to assess the value of delta-radiomics.^{29–32} Three studies utilised PET/CT scans.^{33–35} The VOI varied between studies and included the primary lung lesion,^{31,34–36} the largest lesion^{29,30,37} and all target lesions.³⁸ A few studies separately considered the peri-tumoural region as a VOI.^{32,37,38} The majority of studies utilised feature dimension reduction and machine learning algorithms to select radiomics features and develop a radiomics signature. Various endpoints were assessed including overall response using RECIST and iRECIST, durable clinical benefit (DCB), hyperprogression, progression-free survival (PFS) and overall survival (OS). The results of the study found radiomics models had a pooled diagnostic odds ratio for predicting immunotherapy response of 14.99 (95% CI 8.66–25.95), and a pooled hazard ratio (HR) for PFS of 2.39 (95% CI 1.69–3.38, $P < 0.001$) and OS of 1.96 (95% CI 1.61–2.40, $P < 0.001$). Three of five studies that added clinical variables to radiomics models showed an improvement in model performance.^{29,31,32,34,37} Despite the promising outcomes of the study, the review also found the quality of the radiomics studies overall to be poor. Only two studies included within this review performed validation of their model on a dataset from an external centre^{32,39} and the overall radiomics quality score (RQS)⁴⁰ was low, with a mean RQS of 29.6% ranging from 0% to 68.1%.

The study with the highest RQS was by Mu *et al.*, which included 194 patients with stage IIIB–IV NSCLC undergoing immunotherapy, who had baseline PET/CT scans available for analysis.³⁴ Radiomics features were extracted from PET, CT and PET/CT for the primary lung lesion and patients divided into a training ($n = 99$) and two validation cohorts (47 retrospective and 48 prospectively collected). An 'in-house' developed software, rather than an open-access platform, was used to extract radiomics features, limiting the ability of external groups to validate study findings. A radiomics model was built using eight radiomics features that was predictive of OS and importantly was also tested and performed well on a validation dataset (C-indices of 0.83 on training, and 0.83 and 0.80 on validation datasets). Another large study, which also performed relatively well on the RQS, is by Liu *et al.*²⁹ This was a retrospective multicentre study that included 197 patients randomly divided into a training and validation dataset using a 7:3 ratio. The authors compared two VOI in this study, with one VOI using all RECIST target lesions or the second VOI using the largest target lesion only. Lesions were contoured at baseline and the first follow-up CT scan. Both baseline radiomics features and delta-radiomics features, which captured the change in radiomics feature value over time, were extracted using 'in-house' software.

Radiomics features were used to predict overall patient response at 6 months. The authors found a delta-radiomics signature using the largest RECIST target lesion performed better compared with a signature using baseline radiomics features alone with an AUC of 0.81 in training and 0.80 in the validation dataset. The delta-radiomics signature using all RECIST target lesions had an AUC of 0.82 in training and 0.81 in the validation dataset. The authors noted that some patients had a heterogenous response to immunotherapy between metastatic lesions; however, no per lesion analysis was performed.

The second most common type of tumour studied in this setting is advanced melanoma. One of the largest studies assessing the value of radiomics models in assessment of efficacy in melanoma is by Derclé *et al.*⁴¹ This study utilised contrast-enhanced CT scans from two randomised multicentre clinical trials of ICI in advanced melanoma, KEYNOTE-002 and KEYNOTE-006, and split patients into a training ($n = 252$) and validation ($n = 287$) cohort. The training cohort consisted of patients treated with ipilimumab and pembrolizumab, while the validation cohort only consisted of patients on pembrolizumab. The radiomics signature trained to predict OS consisted of delta-radiomics features, as well as total tumour volume at baseline, and change in volume at 3 months. The radiomics signature outperformed RECIST in predicting OS (validation cohort AUC 0.92, 95% CI 0.89–0.95 and AUC 0.80, 95% CI, 0.75–0.84, respectively). Despite the strengths of this study, including the use of relatively large datasets from clinical trials, and comparison with RECIST, minimal information was provided regarding the CT acquisition parameters, the platform used to extract radiomics features, any pre-processing performed and details regarding the model itself that would allow for validation of findings by external groups. Additionally, imaging features were not compared or combined with clinical and histopathological features that may have been collected during the trials. Another study in melanoma by Basler *et al.* 2020 aimed to predict pseudoprogression utilising PET/CT-derived radiomics features on 112 patients with 716 metastases.⁴² Of the 716 metastases, 30 were identified as pseudoprogression at 6 months. The study found a model using delta-radiomics features, based upon the percentage change in features from baseline to 3 months, had a higher AUC compared with a model using baseline features alone (AUC 0.79 and 0.69, respectively). The delta-radiomics model incorporated 2 features 'mc_volume' and 'fractal dimension'. The authors found that an increase in 'fractal dimension' representing a change from a homogenous to a heterogenous lesion was associated with true rather than pseudoprogressive disease. The performance of the model was marginally improved to AUC 0.82 through the addition of blood markers (LDH/S100). Limitations of this study include the small sample size, lack of validation

cohort and use of 'in-house' software to extract radiomics features.

Prediction of immune-related adverse events

Three studies included in Table 1 investigated the ability of radiomics to predict immune-related adverse events. Mu *et al.* were able to train and validate a radiomics model derived from PET/CT features of the primary tumour in patients with NSCLC undergoing ICI treatment to predict the risk of immune-related severe adverse events.⁴³ The model also incorporated clinical features including type and dose of immunotherapy. Strengths of the study include the use of a test and prospective validation cohort and publication of the five features and calculation formula used for the radiomics score. Limitations, however, include lack of detail regarding the platform used to extract radiomics features, the relatively small sample size and low event rate with only 14 out of 97 patients experiencing adverse events in the development cohort. Another study, analysing a mixed cohort of patients, developed a radiomics model based upon CT-derived features from the normal lung to predict the risk of pneumonitis with 100% accuracy, although the number of patients on the study was very small, with only two out of 32 patients experiencing pneumonitis, with no validation cohort.⁴⁴

Correlation of radiomics features with tumour biology

Cellular immunity, and in particular lymphocytes, plays an important role in the body's anti-tumour immune response. Tumour-infiltrating lymphocytes (TIL) are lymphocytes found directly within or surrounding a tumour and have been associated with better outcomes in a range of primary tumour types^{45,46} and also found to be predictive of response to ICI.⁴⁷ Lymphocytes can be broadly divided into T and B lymphocytes, and natural killer cells. T lymphocytes in particular play a leading role in the cellular immune response. T lymphocytes can be further divided into CD4+ helper T lymphocytes (CD3+ CD4+) and CD8+ cytotoxic T lymphocytes (CD3+ CD8+) according to their surface markers. Both the tumour and the peri-tumoural region may be important in providing information regarding immune infiltration and stromal inflammation and thereby response to immunotherapy agents, which may be captured on imaging and changes in imaging over time and with treatment.

Sixteen studies included in Table 1 assessed the correlation of radiomics features with tissue biomarkers. One of the largest series assessing the ability of radiomics features to predict TIL status of a tumour is by Chen *et al.*⁴⁸ This study in patients with hepatocellular carcinoma (HCC) extracted MRI radiomics features from the tumour and peri-tumoural regions to predict the patients' Immunoscore, which assessed the density of CD3+ and

CD8+ T cells within the tumour core and invasive margin. The study found that a radiomics model which incorporated both intra- and peri-tumoural radiomics features had a higher AUC in the validation cohort compared with a model with intra-tumoural features alone (0.89, 95% CI 0.80–0.99 vs. 0.63, 95% CI 0.47–0.80). The model was further improved through the addition of clinical features of AFP, GGT and AST (AUC 0.93, 95% CI 0.86–1.00) although the results were not statistically significant. Strengths of this study include the homogenous patient population, use of both intra-tumoural and peri-tumoural VOIs and reporting of features used in the radiomics model. While the model performed well on a non-random split sample validation cohort, it could be further strengthened by testing on an external validation cohort and potentially further reduction in the number of features utilised within the model to reduce the risk of overfitting the model. Another study, discussed earlier by Sun *et al.*, that included a mixed cohort of 135 patients with advanced solid malignancies from the MOSCATO trial, generated a CT-based radiomics signature to predict for CD8 cell tumour infiltration assessed using RNA-seq genomic data from patient biopsies.²⁴ The authors subsequently validated the model on patients from the Cancer Genome Atlas dataset (AUC 0.67; 95% CI 0.57–0.77). Interestingly, three of the eight features within this model were textural features derived also from the peri-tumoural region.

Many studies have explored the association of radiomics features with PD-L1 status. One of the first studies to assess the ability of radiomics features to predict PD-L1 status was by Jiang *et al.*⁴⁹ This was a retrospective single-centre study that extracted radiomics features from both the CT and PET components of a pre-operative PET/CT in patients with lung cancer undergoing surgery. The PD-L1 status of patients was assessed on post-operative samples via immunohistochemistry testing. The authors found that the performance of the CT-derived radiomics features was excellent (AUC 0.97, 95% CI 0.93–1.0) and significantly better than PET-derived features. The performance of the model was excellent when tested using both the Ventana PD-L1 test kit (PD-L1 \geq 1% AUC 0.97; PD-L1 \geq 50% AUC 0.80) and the pharmDex PD-L1 test kit (PD-L1 \geq 1% AUC 0.86; PD-L1 \geq 50% AUC 0.91). Strengths of this study include the relatively large sample size and use of post-operative samples to detect PD-L1 status, as well as relatively homogenous imaging parameters collected from a single centre. While the radiomics features included within the radiomics model are provided, explicit details regarding any pre-processing performed and details of the radiomics signature to allow for validation by external groups were not provided. Another study considering a more advanced cohort of patients with Stage IV NSCLC found a radiomics-based model predictive of PD-L1 $>$ 50% on biopsy samples had moderate performance (AUC 0.67 validation, AUC 0.75 test cohort) and was slightly

improved through the addition of deep learning features (AUC 0.71 validation, AUC 0.76 test cohort).⁵⁰ This study included a very large cohort of patients to develop the model ($n = 750$), relevant particularly given the use of a deep learning model. Minimal details, however, were provided regarding the platform used for radiomics feature extraction and the features included within the radiomics model. Another study in oesophageal cancer using mixed biopsy and surgical samples found that the addition of clinical features (grade and stage) to a radiomics model resulted in a higher AUC compared with a model with radiomics or clinical features alone (AUC 0.817, 0.750 and 0.692, respectively).⁵¹

Other tissue biomarkers that are gaining interest include microsatellite instability (MSI) and TMB. One study was able to validate a radiomics model incorporating 40 radiomics features to predict microsatellite instability in colon cancer with an AUC of 0.76.⁵² Another study in patients with NSCLC was able to develop both a CT-based radiomics and deep learning model to predict TMB (AUC 0.74, 95% CI 0.69 to 0.79 and AUC 0.85, 95% CI 0.84 to 0.87, respectively).⁵³ The overall number of patients, to some extent for developing a deep learning model, and in particular within the validation cohort, however, in this study were relatively small. CT images were obtained using a large range of manufacturers with varying slice thicknesses; however, detailed information regarding CT acquisition parameters was not provided. Information regarding the platform used to extract radiomics features was also missing.

Limitations

Compared with the broader radiomics literature, immunotherapy-related studies to date have a number of strengths including more frequent use of prospectively collected datasets, use of relatively homogenous patient groups and treatment and use of high-quality diagnostic CT scans for extraction of radiomics features. Many studies additionally investigated the value of delta-radiomics and peri-tumoural radiomics features. The studies, however, also highlight a number of challenges and limitations that exist. Despite many studies reporting good to excellent performance of radiomics-based models, there was heterogeneity in the number and type of features extracted from imaging, the pre-processing of features, the feature selection methods and machine learning algorithms used to develop the radiomics model. Differences in methodology may affect model performance^{54,55} and complicate the task of comparing results across radiomics studies. While there is little consensus within the literature as to which feature selection processes and machine learning models should be used, this should certainly be carefully considered in any radiomics analysis.⁵⁶ The overall quality of studies could also be improved, particularly through the adherence to guidelines such as the Image Biomarker Standardisation

Initiative (IBSI),⁵⁷ the radiomics quality score (RQS)⁴⁰ and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.^{58,59}

The IBSI guidelines aimed to standardise feature definitions and provide guidance on the pre-processing methods used for features to allow comparison of features across studies. Unfortunately, a number of radiomics studies did not report which software package was used to extract radiomics features or utilised 'in-house' software to extract features. Few studies did not report the exact features used within radiomics models. Even when these were reported, sufficient information was not routinely provided regarding the pre-processing methods or mathematical definitions of features, potentially resulting in features with the same name across studies, measuring different image characteristics, while features with different names may be measuring the same characteristics.⁶⁰ The lack of use of standardised image processing and image biomarker nomenclature and definitions within the literature severely impedes any comparison across studies, collation and meta-analysis of findings across studies and contributes to the replication crisis within the field, undermining both the credibility of findings and the ability to translate results to a clinical setting.

The RQS is a radiomics-specific quality assessment tool that identifies 16 key components, each assigned a number of points, to clearly evaluate the validity of a radiomics study and any bias or limitations that exist. Studies within the radiomics literature performed relatively poorly on the use of prospective studies that were pre-registered in a trial database, use of calibration statistics, use of multiple imaging timepoints, cost-effectiveness analysis and open publishing of the imaging, VOI and codes used for analysis within the study. In particular, upfront inclusion of radiomics analysis into prospective trials affords studies the opportunity for more homogenous image datasets through prospectively defining and recording image acquisition parameters and reconstruction algorithms used, as well as better reporting details such as contrast protocols, which is often not well documented within studies and may lead to variation in outcomes. Open publishing of the codes used for analysis and presentation of the full model developed, including all regression coefficients, would allow for replication of findings and validation of models by independent groups on external datasets.

The importance of validation, and particularly external validation of models, is highlighted in the TRIPOD statement. This is particularly hampered by the limited availability of publicly accessible annotated imaging datasets in patients undergoing immunotherapy. Several radiomics studies, therefore, included relatively small numbers of patients, limiting their ability to both develop and validate models. While cross-validation and random split sample validation methods were commonly utilised as

internal validation techniques, most studies lacked external validation cohorts. Lack of external validation may not only result in over-estimation of the performance of models but also limit their generalisability.

Future directions

Despite the limitations and challenges highlighted by the current radiomics literature, a number of areas for future research are also emerging. Studies investigating the predictive value of radiomics models in patients undergoing immunotherapy in rarer cancer types and association with less well-studied tissue biomarkers, such as TMB, are required. There is limited research on the ability of radiomics models to predict immune-related adverse effects. Studies did not routinely assess clinical features and the added value of imaging biomarkers to known clinically utilised prognostic and predictive indicators in patients undergoing immunotherapy, such as the Gustave Roussy Immune Score.⁶¹

Unlike much of the prior radiomics work which assessed early-stage or locally advanced disease, assessment of immunotherapy response often required interrogation of datasets of patients with metastatic disease and thereby multiple potential VOIs. The appropriate VOI in these patients is uncertain. In these patients, the primary tumour may not always be easily defined or previously treated and, therefore, not available for analysis. Contouring all sites of metastatic disease can be a laborious process and introduces further uncertainties. Not all tumours are easily and consistently defined on imaging due to various factors such as associated collapse or consolidation surrounding the tumour. Among the current studies, VOI definitions varied widely from the primary tumour, to the largest lesion, to all RECIST target lesions or all sites of metastatic disease, with limited and inconsistent findings as to the most reliable and valid VOI in this setting. Further complicating the issue is the finding that the peri-tumoural region may also hold pertinent information and radiomics features extracted from this region often featured within radiomics models used. However, the definition and size of the peri-tumoural region also varied between studies. Further research is, therefore, required as to the optimal VOI in this setting. The introduction of multiple VOIs per patient also raises the issue of the ideal way to aggregate radiomic results to utilise in per patient endpoints such as overall response and survival. In addition to determining the optimal VOI, further research is required as to whether 'semantic' imaging features, such as those qualitatively described by radiologists, can be quantitatively defined to create novel medical expert-defined features that could be used for image analysis. Further research assessing per lesion responses, to investigate not only hyperprogression and pseudoprogression, but also patients with mixed responses and potentially oligoprogressive disease, is also required.

Various machine learning approaches have been utilised to create radiomics models, with further research required as to the optimal approach in the immunotherapy setting. Deep learning approaches are increasingly being incorporated into image analysis studies and into all parts of the workflow including segmentation, generation of imaging features and model development. Convolutional neural networks (CNN) are a commonly employed class of deep learning, in which pre-defined features (as used in radiomics) are not required. In contrast, CNN layers are used to automate the process of selection and quantification of imaging features that are important for certain tasks or outputs such as prognostication. Additionally, while radiomics feature extraction is highly dependent upon the accurate segmentation of a VOI, deep learning processes can handle less well-defined regions of interest or entire image datasets and independently determine relevant and significant regions within an image. Deep learning approaches may, therefore, mitigate the limitations of using pre-defined features and manually contoured VOI, which inherently introduces bias into the image analysis process. The ability to substantially automate the process of image analysis through deep learning, thereby facilitating the analysis of larger datasets combined with promising initial outcomes using deep learning in immunotherapy⁶² and non-immunotherapy settings,⁶³ makes this an exciting area for further research. Challenges of a deep learning approach, however, include the lack of interpretability and transparency of deep learning models, raising ethical and trust issues and limiting their clinical use, although solutions to partly overcome these issues are being investigated.⁶⁴

Conclusions

The incorporation of immunotherapy into the arsenal for cancer therapies has led to an exciting phase of change, with improvements in patient outcomes and substantial progress made in our understanding of the biology of the immune response. The challenge of delivering personalised care for patients in this setting remains, with better biomarkers needed to augment clinician decision making in this increasingly complex field.

Radiomics and image analysis provide an underutilised avenue for further exploration, potentially in combination with currently employed clinical and tissue biomarkers. Overall studies evaluating the predictive power of radiomics-based models based upon both baseline and change in imaging features over time in patients undergoing immunotherapy are promising, with many studies reporting good model performance with respect to prediction of immunotherapy response and outcome. Furthermore, there is increasing evidence that tumour immune states may be characterised by radiomic features.

However, despite the burgeoning number of studies in this field, current radiomic studies remain immature with room for improvement in quality, with heterogeneous results and lack of thorough model validation. Integration of radiomics models into a clinical setting will require prospective studies with externally validated models, integration of other relevant biological and clinical data to develop superior composite models and resolution of technical barriers required for model implementation in routine clinical care.

Acknowledgement

Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

- Hodi FS, O'Day SJ, McDermott DF *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010; **363**: 711–23.
- Minor WJDAS. Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* 2015; **16**: 375–84.
- Dummer R, Daud A, Puzanov I *et al.* A randomized controlled comparison of pembrolizumab and chemotherapy in patients with ipilimumab-refractory melanoma. *J Transl Med* 2015; **13**: 1–2.
- Antonia SJ, Villegas A, Daniel D *et al.* Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med* 2017; **377**: 1919–29.
- Finn RS, Qin S, Ikeda M *et al.* Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *N Engl J Med* 2020; **382**: 1894–905.
- Burtneß B, Harrington KJ, Greil R *et al.* Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet* 2019; **394**: 1915–28.
- Gandhi L, Rodríguez-Abreu D, Gadgeel S *et al.* Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* 2018; **378**: 2078–92.
- Fradet Y, Bellmunt J, Vaughn DJ *et al.* Randomized phase III KEYNOTE-045 trial of pembrolizumab versus paclitaxel, docetaxel, or vinflunine in recurrent advanced urothelial cancer: results of >2 years of follow-up. *Ann Oncol* 2019; **30**: 970–6.
- Reck M, Rodríguez-Abreu D, Robinson AG *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016; **375**: 1823–33.
- Hussaini S, Chehade R, Boldt RG *et al.* Association between immune-related side effects and efficacy and benefit of immune checkpoint inhibitors – a systematic review and meta-analysis. *Cancer Treat Rev* 2021; **92**: 102134.
- Marabelle A, Fakih M, Lopez J *et al.* Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol* 2020; **21**: 1353–65.
- Goodman AM, Kato S, Bazhenova L *et al.* Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther* 2017; **16**: 2598–608.
- Eisenhauer EA, Therasse P, Bogaerts J *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45**: 228–47.
- Seymour L, Bogaerts J, Perrone A *et al.* iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017; **18**: e143–52.
- Mulkey F, Theoret MR, Keegan P, Pazdur R, Sridhara R. Comparison of iRECIST versus RECIST V.1.1 in patients treated with an anti-PD-1 or PD-L1 antibody: pooled FDA analysis. *J Immunother Cancer* 2020; **8** (1): e000146.
- Aerts HJ, Velazquez ER, Leijenaar RT *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**: 4006.
- Lambin P, Rios-Velazquez E, Leijenaar R *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**: 441–6.
- Soofi AA, Awan A. Classification techniques in machine learning: applications and issues. *J Basic Appl Sci* 2017; **13**: 459–65.
- Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion* 2019; **50**: 71–91.
- Olaode A, Naghdy G, Todd C. Unsupervised classification of images: a review. *Int J Image Process* 2014; **8**: 325–42.
- Carvalho S, Leijenaar R, Troost E *et al.* Early variation of FDG-PET radiomics features in NSCLC is related to overall survival-the “delta radiomics” concept. *Radiother Oncol* 2016; **118**: S20–1.
- Kothari G, Korte J, Lehrer EJ *et al.* A systematic review and meta-analysis of the prognostic value of radiomics based models in non-small cell lung cancer treated with curative radiotherapy. *Radiother Oncol* 2021; **155**: 188–203.
- Peikert T, Bartholmai BJ, Maldonado F. Radiomics-based management of indeterminate lung nodules? Are we there yet? *Am J Respir Crit Care Med* 2020; **202**: 165–7.

24. Sun R, Limkin EJ, Vakalopoulou M *et al.* 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* 2018; **19**: 1180–91.
25. Sun R, Sundahl N, Hecht M *et al.* Radiomics to predict outcomes and abscopal response of patients with cancer treated with immunotherapy combined with radiotherapy using a validated signature of CD8 cells. *J Immunother Cancer* 2020; **8**: e001429.
26. Korpics MC, Polley M-Y, Bhawe SR *et al.* A validated T cell radiomics score is associated with clinical outcomes following multisite SBRT and pembrolizumab. *Int J Radiat Oncol Biol Phys* 2020; **108**: 189–95.
27. Luke JJ, Lemons JM, Karrison TG *et al.* Safety and clinical activity of pembrolizumab and multisite stereotactic body radiotherapy in patients with advanced solid tumors. *J Clin Oncol* 2018; **36**: 1611–8.
28. Chen Q, Zhang L, Mo X *et al.* Current status and quality of radiomic studies for predicting immunotherapy response and outcome in patients with non-small cell lung cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 2021; **49**: 345–60.
29. Liu Y, Wu M, Zhang Y *et al.* Imaging biomarkers to predict and evaluate the effectiveness of immunotherapy in advanced non-small-cell lung cancer. *Front Oncol* 2021; **11**: 657615.
30. Dercle L, Fronheiser M, Lu L *et al.* Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clin Cancer Res* 2020; **26**: 2151–62.
31. Yang Y, Yang J, Shen L *et al.* A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-PD-1/PD-L1 immunotherapy in advanced stage non-small-cell lung cancer. *Am J Transl Res* 2021; **13**: 743.
32. Khorrani M, Prasanna P, Gupta A *et al.* Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer Immunol Res* 2020; **8**: 108–19.
33. Mu W, Katsoulakis E, Whelan CJ, Gage KL, Schabath MB, Gillies RJ. Radiomics predicts risk of cachexia in advanced NSCLC patients treated with immune checkpoint inhibitors. *Br J Cancer* 2021; **125**: 229–39.
34. Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. 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* 2020; **47**: 1168–82.
35. Polverari G, Ceci F, Bertaglia V *et al.* (18)F-FDG pet parameters and radiomics features analysis in advanced nsclc treated with immunotherapy as predictors of therapy response and survival. *Cancer* 2020; **12**: 1163.
36. Valentinuzzi D, Vrankar M, Boc N *et al.* [18F] FDG PET immunotherapy radiomics signature (iRADIOMICS) predicts response of non-small-cell lung cancer patients treated with pembrolizumab. *Radiol Oncol* 2020; **54**: 285.
37. Tunali I, Gray JE, Qi J *et al.* Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: an early report. *Lung Cancer* 2019; **129**: 75–9.
38. Vaidya P, Bera K, Patil PD *et al.* Novel, non-invasive imaging approach to identify patients with advanced non-small cell lung cancer at risk of hyperprogressive disease with immune checkpoint blockade. *J Immunother Cancer* 2020; **8**: e001343.
39. Nardone V, Tini P, Pastina P *et al.* Radiomics predicts survival of patients with advanced non-small cell lung cancer undergoing PD-1 blockade using Nivolumab. *Oncol Lett* 2020; **19**: 1559–66.
40. Lambin P, Leijenaar RT, Deist TM *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749.
41. Dercle L, Zhao B, Gönen M *et al.* Early readout on overall survival of patients with melanoma treated with immunotherapy using a novel imaging analysis. *JAMA Oncol* 2022; **8**: 385–92.
42. Basler L, Gabrys HS, Hogan SA *et al.* Radiomics, tumor volume, and blood biomarkers for early prediction of pseudoprogression in patients with metastatic melanoma treated with immune checkpoint inhibition. *Clin Cancer Res* 2020; **26**: 4414–25.
43. Mu W, Tunali I, Qi J, Schabath MB, Gillies RJ. Radiomics of 18F fluorodeoxyglucose PET/CT images predicts severe immune-related adverse events in patients with NSCLC. *Radiol Artif Intell* 2020; **2**: e190063.
44. Colen RR, Fujii T, Bilen MA *et al.* Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Investig New Drugs* 2018; **36**: 601–7.
45. Hurkmans DP, Kuipers ME, Smit J *et al.* Tumor mutational load, CD8+ T cells, expression of PD-L1 and HLA class I to guide immunotherapy decisions in NSCLC patients. *Cancer Immunol Immunother* 2020; **69**: 771–7.
46. Emens LA, Cruz C, Eder JP *et al.* Long-term clinical outcomes and biomarker analyses of atezolizumab therapy for patients with metastatic triple-negative breast cancer: a phase 1 study. *JAMA Oncol* 2019; **5**: 74–82.
47. Paijens ST, Vledder A, de Bruyn M, Nijman HW. Tumor-infiltrating lymphocytes in the immunotherapy era. *Cell Mol Immunol* 2021; **18**: 842–59.
48. Chen S, Feng S, Wei J *et al.* Pretreatment prediction of immunoscore in hepatocellular cancer: a radiomics-based clinical model based on Gd-EOB-DTPA-enhanced MRI imaging. *Eur Radiol* 2019; **29**: 4177–87.
49. Jiang M, Sun D, Guo Y *et al.* Assessing PD-L1 expression level by radiomic features from PET/CT in nonsmall cell lung cancer patients: an initial result. *Acad Radiol* 2020; **27**: 171–9.
50. Tian P, He B, Mu W *et al.* Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics* 2021; **11**: 2098–107.

51. Wen Q, Yang Z, Zhu J *et al.* Pretreatment CT-based radiomics signature as a potential imaging biomarker for predicting the expression of PD-L1 and CD8+ TILs in ESCC. *Onco Targets Ther* 2020; **13**: 12003.
52. Pernicka JSG, Gagniere J, Chakraborty J *et al.* Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol* 2019; **44**: 3755–63.
53. He B, Dong D, She Y *et al.* Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* 2020; **8**: e000550.
54. Wang L, Dong T, Xin B *et al.* Integrative nomogram of CT imaging, clinical, and hematological features for survival prediction of patients with locally advanced non-small cell lung cancer. *Eur Radiol* 2019; **29**: 2958–67.
55. Sun W, Jiang M, Dang J, Chang P, Yin FF. Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis. *Radiat Oncol* 2018; **13**: 197.
56. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging* 2020; **11**: 91.
57. Zwanenburg A, Vallières M, Abdalah MA *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; **295**: 328–38.
58. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation* 2015; **131**: 211–9.
59. Park JE, Kim D, Kim HS *et al.* Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol* 2020; **30**: 523–36.
60. Korte JC, Cardenas C, Hardcastle N *et al.* Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer. *Sci Rep* 2021; **11**: 17633.
61. Bigot F, Castanon E, Baldini C *et al.* Prospective validation of a prognostic score for patients in immunotherapy phase I trials: the Gustave Roussy Immune Score (GRIm-Score). *Eur J Cancer* 2017; **84**: 212–8.
62. Mu W, Jiang L, Shi Y *et al.* Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer* 2021; **9**: e002118.
63. Hosny A, Parmar C, Coroller TP *et al.* Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS Med* 2018; **15**: e1002711.
64. Singh A, Sengupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imaging* 2020; **6**: 52.
65. Trebeschi S, Drago SG, Birkbak NJ *et al.* Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 2019; **30**: 998–1004.
66. Sun Z, Hu S, Ge Y *et al.* Radiomics study for predicting the expression of PD-L1 in non-small cell lung cancer based on CT images and clinicopathologic features. *J Xray Sci Technol* 2020; **28**: 449–59.
67. Yoon J, Suh YJ, Han K *et al.* Utility of CT radiomics for prediction of PD-L1 expression in advanced lung adenocarcinomas. *Thorac Cancer* 2020; **11**: 993–1004.
68. Yoon HJ, Kang J, Park H, Sohn I, Lee S-H, Lee HY. Deciphering the tumor microenvironment through radiomics in non-small cell lung cancer: correlation with immune profiles. *PLoS One* 2020; **15**: e0231227.
69. Park KJ, Lee JL, Yoon SK, Heo C, Park BW, Kim JK. Radiomics-based prediction model for outcomes of PD-1/PD-L1 immunotherapy in metastatic urothelial carcinoma. *Eur Radiol* 2020; **30**: 5392–403.
70. Liao H, Zhang Z, Chen J *et al.* Preoperative radiomic approach to evaluate tumor-infiltrating CD8+ T cells in hepatocellular carcinoma patients using contrast-enhanced computed tomography. *Ann Surg Oncol* 2019; **26**: 4537–47.
71. Gao X, Ma T, Bai S *et al.* A CT-based radiomics signature for evaluating tumor infiltrating Treg cells and outcome prediction of gastric cancer. *Ann Transl Med* 2020; **8**: 469.
72. Iwatate Y, Hoshino I, Yokota H *et al.* Radiogenomics for predicting p53 status, PD-L1 expression, and prognosis with machine learning in pancreatic cancer. *Br J Cancer* 2020; **123**: 1253–61.
73. Li Z-z, Liu P-f, An T-t, Yang H-c, Zhang W, Wang J-x. Construction of a prognostic immune signature for lower grade glioma that can be recognized by MRI radiomics features to predict survival in LGG patients. *Transl Oncol* 2021; **14**: 101065.
74. Yu H, Meng X, Chen H *et al.* Predicting the level of tumor-infiltrating lymphocytes in patients with breast cancer: usefulness of mammographic radiomics features. *Front Oncol* 2021; **11**: 393.