BJR
© 2020 The Authors. Published by the British Institute of Radiology
https://doi.org/10.1259/bjr.20190496

Received: 03 June 2019 | Revised: 19 October 2019 | Accepted: 29 October 2019

Cite this article as:
Guha A, Connor S, Anjari M, Naik H, Siddiqui M, Cook G, et al. Radiomic analysis for response assessment in advanced head and neck cancers, a distant dream or an inevitable reality? A systematic review of the current level of evidence. *Br J Radiol* 2020; **93**: 20190496.

SYSTEMATIC REVIEW

Radiomic analysis for response assessment in advanced head and neck cancers, a distant dream or an inevitable reality? A systematic review of the current level of evidence

[1,2]AMRITA GUHA, MD, [3]STEVE CONNOR, [3]MUSTAFA ANJARI, [4]HARISH NAIK, [5]MUSIB SIDDIQUI, [3]GARY COOK and [3]VICKY GOH

[1]School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom
[2]Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, India
[3]Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
[4]Grant Medical college and JJ Group of hospitals, Mumbai, India
[5]School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom

Address correspondence to: Dr Amrita Guha
E-mail: *amritaguha85@gmail.com; amritaguha2006@yahoo.com*

Objective: The recent increase in publications on radiomic analysis as means to produce diagnostic and predictive biomarkers in head and neck cancers (HNCC) reveal complicated and often conflicting results. The objective of this paper is to systematically review the published data, and evaluate the current level of evidence accumulated that would determine clinical application.
Methods: Data sources: Articles in the English language available on the Ovid-MEDLINE and Embase databases were used for the literature search. **Study selection**:Studies which evaluated the role of radiomics as a predictive or prognostic tool for response assessment in HNCC were included in this review.
Study appraisal and synthesis methods: The authors set-out to perform a meta-analysis, however given the small number of studies retrieved that presented adequate data, combined with excessive methodological heterogeneity, we could only perform a structured descriptive systematic review summarizing the key findings. Independent extraction of articles was performed by two authors using predefined data fields and any disagreement was resolved by consensus.
Results: Though most papers concluded that radiomics is an effective predictive and prognostic biomarker in the management of HNCC, significant heterogeneity exists in the study methodology and statistical modelling; thus precluding accurate mathematical comparison or the ability to make clear recommendations going forwards. Moreover, most studies have not been validated and the reproducibility of their results will be a challenge.
Conclusion: Until robust external validation studies on the reproducibility and accuracy of radiomic analysis methods on HNCC are carried out, the current level of evidence remains low, with the authors advising caution against hasty implementation of these tools in the multidisciplinary clinic.
Advances in knowledge: This review is the first attempt to critically analyze the merits and demerits of currently published literature on tumour heterogeneity studies in HNCC, and identifies specific loop holes that need to be addressed by research groups, for a meaningful clinical translation of this potential biomarker.

INTRODUCTION

Head and neck tumours form the seventh leading cancer with respect to incidence, and the eighth with respect to mortality rates.[1] Unfortunately, up to two-thirds of patients with head and neck cancers will present at an advanced stage rather than at an earlier, potentially easily treatable stage.[2] Radiotherapy (RT) with or without chemotherapy forms the mainstay of treatment of advanced head and neck squamous cell cancer (HNSCC) in most subsites.[3]

Understanding the clinical problem

Stratification of patients into response categories that reflect outcome is necessary for treatment optimization of cancers in the head and neck region.[4] Although CT and MRI may be used for post-treatment follow-up of these

patients, anatomical imaging alone to detect residual or recurrent primary disease is limited due to post-treatment tissue distortion.[5] Hermans et al[6] found that follow-up CT scans were definite for local failure in only 41% patients before clinical examination results. King et al[7] reported that residual masses ≥ 1 cm of similar signal to untreated tumor on *T2W* MRI suggest local failure. A recent systematic review by Chung et al[8] found that high pre-treatment apparent diffusion coefficient (ADC) and low rise in ADC with chemoradiation, could be indicators of locoregional failure on diffusion-weighted MRI. Whilst morphological criteria like size, gadolinium enhancement and *T2* signal intensity are routinely used techniques for assessment in the clinic,[9] RECIST criteria are commonly used in trials to monitor response,[10] but are difficult to apply to the complex geometry of primary head and neck tumours. Moreover, standard structural imaging analysis has shown variable diagnostic accuracy, as was demonstrated by Patil et al[11] who found a low correlation between conventional radiological measurements on RECIST with pathological response.

Radiomics is an emerging field that converts imaging data into a high-dimensional mineable feature space using a large number of automatically extracted data-characterization algorithms.[12] In view of the shortcomings of current available imaging approaches, several research groups around the world have initiated research to explore the role of "radiomics" in predicting treatment response to therapy. They hypothesize that rather than using anatomical (*i.e.* CT or MRI) or standard metabolic positron emission tomography (PET) features alone to assess treatment response, tumour characterization and behaviour may be better reflected by quantifying the intratumoural heterogeneity depicted by imaging modalities such as CT, MRI, and PET.

Rationale behind this review

The accumulation of current data suggests there may be potential for radiomics in tumour assessment, risk stratification, and outcome evaluation in head and neck cancer therapy. With a recent increase in the number of research papers published on the prognostic/predictive role of head and neck cancers, it is imperative to perform a critical analysis and systematically review the current available evidence on the translational feasibility of radiomics into clinical practice. The absence of any previously published meta-analysis/systematic review that evaluates the outcomes of these heterogeneously conducted studies, led us to conduct this study in order to determine the current level of evidence.

METHODS

This review was conducted according to the PRISMA guidelines for protocol development (available with author on request) and study reporting[13,14]; and the Centre for Reviews and Dissemination guidelines were used for methodology structuring.[15] The PRISMA checklist is available in the supplementary article.

Data sources and search strategy

We identified primary studies on the predictive/ prognostic performance of radiomics in head and neck cancers. Ovid-MEDLINE and Embase databases were used for the literature search, which was conducted in February 2018. We combined terms in the following search string to identify relevant studies: "(texture OR heterogeneity OR feature OR radiomics) AND (therapy OR treatment OR response) AND (tumor OR tumour) AND (head and neck) AND (cancer OR malignancy) AND (CT OR tomography OR MRI OR PET)". Further explode and subject heading options were used to include "orophary* OR hypopharyn* OR laryn* OR nasopharyn* OR tongue OR oral cavity OR buccal".

Inclusion criteria

Patients with cancers of the head and neck who received chemoradiotherapy with or without surgery were included in the study. Only studies which analyzed the role of radiomics as a predictive or prognostic tool were included in the review. Articles from 1995 upto January 2018 were included in the study.

Exclusion criteria

Results were then limited to humans (using: AND "humans"[MeSH Terms]), limited to English language (using: AND English[lang]). We used the systematic review filter to identify prior systematic reviews and reviewed the initial search (*i.e.* excluding filters) to identify any articles excluded incorrectly. Studies which evaluated the role of radiomics as a diagnostic tool, or to differentiate tumour types (HPV/p16 status) or to differentiate tumor grading were excluded from our study. This was because these studies did not address the clinical question of predicting residue/recurrence post treatment; and mostly dealt with other facets such as diagnosis for which well validated "gold-standard" techniques are already in place.

Electronic abstracts of identified studies were read and the following exclusion criteria applied. Small cases series (less than five patients), narrative reviews, letters/correspondence and conference abstracts were excluded since these would not contribute sufficient unbiased data to be able to answer our research question.

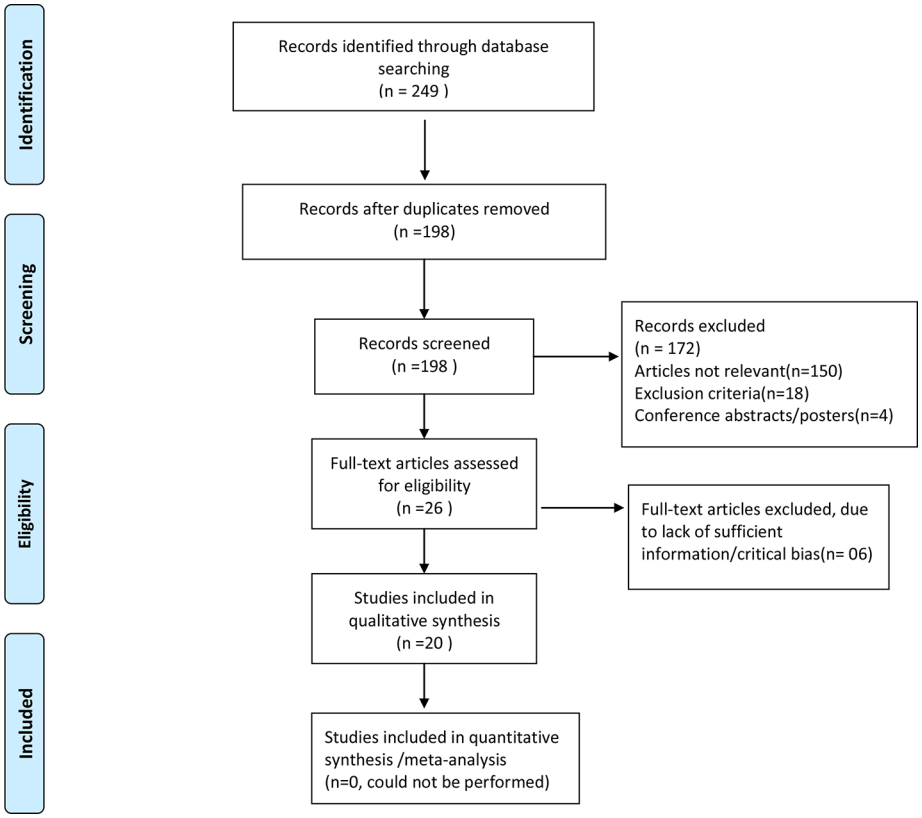
Meta-analysis

At the outset, our intention was to perform a meta-analysis to obtain pooled estimates of survival. However meta-analysis was precluded by the small number of studies retrieved that presented adequate data, combined with excessive methodological heterogeneity. The article was written following PRISMA 2009 guidelines, as has been attached in the supplementary article.

Study selection and data extraction

Following protocol development, a data extraction form was piloted by the lead author in discussion with the other authors. Literature search was performed by two radiologists independently with a special interest in oncological imaging (>5 years' imaging experience) who first screened the titles and abstracts of relevant papers, followed by the main text and any disagreement was resolved by consensus. The ROBINS-1 tool was used for quality assessment of the included studies, and those with a "critical risk of bias" or "no sufficient information" were excluded from the review.

Figure 1. PRISMA flow diagram of the literature search strategy of selected studies.



RESULTS

The PRISMA flow diagram of the literature search strategy of selected studies is shown in Figure 1.The papers published on the role of radiomics imaging for treatment response in head and neck cancer can be broadly divided into those that deal with:

Imaging modality specific radiomics: CT, MRI, PETCT

Methodological standardization (Classifier/feature-related studies)

Validation studies (external validation of findings between two or more centres)

CT radiomics in head and neck cancer

Three papers have been published on the role of CT radiomics in predicting response to therapy in HNSCC (Table 1). Bogowicz et al[16] studied 149 patients with Stage III-IV HNSCC, dividing them into training and testing cohorts. A model comprising of three radiomic features: large size high grey-level emphasis, sum entropy, and difference variance was found to be prognostic for local control. Tumours with greater heterogeneity of CT density distribution were found to have poorer prognosis.

Zhang et al[17] studied 72 patients with locally advanced HNSCC and found that primary mass entropy and skewness measurements with multiple spatial filters were associated with overall survival (OS); independent of tumor size, N stage, and other clinical variables. In this retrospective study, patients were scanned pre-treatment (chemotherapy only) and no internal validation/boot strapping was performed.

Ou et al[18] matched 120 patients with advanced HNSCC 2:1 into two treatment groups: concurrent chemoradiotherapy (CRT) or bioradiotherapy (BRT). They showed that a 24-feature based radiomic signature significantly predicted for OS and progression-free survival (PFS).

Role of MRI radiomics in head and neck cancers

Six papers have been published on the role of MRI in monitoring treatment response in HNSCC, as summarized in Table 2. Of these, the first four papers in the table are from the same research group,[19–22] all published around the same period and could possibly represent an overlapping/identical patient cohort. This research group retrospectively studied role of radiomics in nasopharyngeal carcinoma with similar basic underlying study design. Patients with Stage III-IV NPC were scanned pre-treatment on 1.5 T MRI (findings reported on CE-T_1W and T_2W sequence) and were divided into training and validation cohorts. Using LASSO as a feature selection technique, they validated radiomics for its role in response prediction (Wang et al[22] ; association with survival/PFS (Ouyang et al[19] ; comparative performace of indivisual CE-T_1W/T_2W *vs* combined CE-T_1W and T_2W MR sequences (Zhang et al[20]) and finally a paper that summarized and included all of the above (Zhang et al[21]). Radiomics consistently performed well in its prognostic and predictive ability in all these papers.

Table 1. CT radiomics in head and neck cancer

Author and year of publication	Sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selectionand number of parametrs derived	Outcome parameter	Result
M.Bogowicz[16] May 2017	*N* = 93 TC *N* = 56 VC Stage III and IV HNSCC (Orp,Hyp,Lar,OC)	Pre-treatment retrospective for TC prospective for VC	Definitive IMRT 70 Gy with cisplatin or cetuximab.	Shape Intensity Texture Wavelet transform 317 features	Grouping: PCA Selection: UVCRA for prognosis (9) For comparison with clinical and combined radiomics-clinical model: MVCRA(3) Split ROC curves at 18 mths	Predict LC: using CI Compare radiomics versus clinical model and a combined clinicoradiomic model for LC	Radiomics signature significantly associated with LC Combined Radiomics + Clinical model performed better than Radiomics model alone in TC, but not VC
H. Zhang et alDec[17] 2013	*N* = 72 Stage III and IV HNSCC (Orp,Hyp,Lar,OC)	Pretreatment Retrospective Median follow up time: 1.9 years Median OS of entire cohort: 2.6 years	All induction TPF chemotherapy.	First-order texture and histogram analysis, Multiple spatial filters applied from fine to coarse. Number of features = NM	MVCRA for primary mass parameters with OS MVA for model with primary size and N stage. No internal validation/ bootstrapping	Predict OS: HR, CI and *p*-value	MVA: Entropy and skewness with multiple filters associated with OS. MVA of clinical and imaging variables: Entropy and skewness with 1.0 spatial filter associated with OS.

(Continued)

Table 1. (Continued)

Author and year of publication	Sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selectionand number of parametrs derived	Outcome parameter	Result
D.Ou et al[18] **June 2017**	*N* = 120 Stage III-IVb HNSCC	Pretreatment Retrospective Median follow-up time: 49.3 months 5 year actuarial OS: 61.2%	3D-CRT or IMRT with Cisplatin or Cetuximab Matched 2:1 into Concurrent Chemoradiotherapy *vs* BioRadiotherapy	Shape Intensity Texture Filter based wavelet 544	Grouping: PCA Feature selection: UVCPHA(24) Data dichotomized: low and high radiomics score Internal validation: dichotomizing data and two sided *p* values	Prognosis and Prediction : OS, PFS AUC at 5 years MVA with HR Test radiomics with combined model using p16	Radiomics model alone, and in combination with p16 predicted OS and PFS. Patients with high signature score significantly benefited more from CRT (*vs* BRT) in terms of OS and PFS, while no benefit difference between CRT and BRT in patients with low signature score.

HR, Hazard Ratios; Hyp, Hypopharynx; IMRT, Intensity Modulated Radiotherapy; LC, Local Control; Lar, Larynx; MVA, Multivariate analysis; MVCRA, Multivariate Cox Regression Analysis; N, Number of patients; NM, Not mentioned; OC, Oral cavity; Orp, Oropharnx; PCA, Principal Component Analysis; ROC, Reciever Operated Curves; TC, Training Cohort; TPF, cisplatin, 5-fluorouracil, and docetaxel; UVCRA, Univariate Cox Regression Analysis; VC, Validation Cohort.

Table 2. MRI Radiomics in head and neck cancer

Author and year of publication	Patient sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection and number of parameters derived	Outcome parameter	Result
Ouyang et al[19] Aug 2017	*N* = 100 TC = 70; VC = 30 NPC Stage III-IVb	Pre-treatment retrospective MRI: T2 and T1c 970 features Median follow up time: 39.5 months	NM	Shape and size, First-order features, Texture, Wavelet 5: T1C GLCMcorrelation, GLCM_IMc T2: GLRLM,GLCM variance and GLCM homogeneity	Feature selection: LASSO (5) Rad score: was used to dichotomise patients into Low or high risk MVCRA to yield HR	Prognosis: PFS Compare clinical model with combined clinical + radiomics model	Radiomics a significant independent predictor of PFS PFS shorter in high risk Rad score patients.
Zhang et al[20] Aug 2017	*N* = 113 TC = 80; VC = 33 NPC Stage III-IVb	Pretreatment Retrospective MRI: T2 and T1c 970 features	NM	Shape and size, First-order features, Texture, Wavelet (4 T1c and 4 T2 features)	Feature Selection: LASSO logistic regression (8) RAD score Data dichotomised: PFS 3 yrs-Yes or No	Predict progression: PFS using AUC Compare T1c, T2 sequences models individually with a radiomics model using both combined	Radiomic model using joint T1c and T2 yielded highest AUC TC and VC (compared to T1 c or T2 alone)
Zhang et al[21] Aug 2017	*N* = 118 TC = 88; VC = 30 NPC Stage III-IVb	Pretreatment Retrospective MRI: T2 and T1c 970	NM	Shape and size, First-order features, Texture, Wavelet	Feature Selection: LASSO logistic regression Nomogram discrimination and calibration: Using C index	Prognosis: PFS Association b/w radiomics and clinical features using heatmaps	Radiomics significantly associated with PFS Radiomics plus clinical data: better in evaluating PFS than clinical data alone. Radiomic model using joint T1c and T2 better than T1c or T2 alone Radiomics plus TNM model outperformed TNM staging alone.
Wang et al[22] Jan 2018	*N* = 120 (NPC stage II,III and IV)	Pretreatment Retrospective MRI: T1, T1c, T2w and T2wFS 591	2 cycles of IC every 3 weeks (Cisplatin, 5FU and Docetaxel)	Histogram, GLCM,GLRL, Gabor and wavelet features Data dichotomized: responder and non-responder to IC Internal validation.	Feature Selection: LASSO regression model five features from T1c; 15 features from combined model Association with response: Mann Whitney U test. ROC curves for discrimatory ability	Association b/w radiomics and response to IC Compared T1c with prediction Then compared model combining T1, T1c, T2w and T2wFS with prediction	T1c and combined sequences' radiomics signature were independant predictors in discriminating response and non-response pretreatment. Combined model of all 4 MR sequences performed better than single T1c sequence.

(Continued)

Table 2. (Continued)

Author and year of publication	Patient sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection and number of parameters derived	Outcome parameter	Result
Liu et al[23] Dec 2015	*N* = 53 TC = 42; VC = 11 NPC	Pretreatment Retrospective 3T MRI: T1,T2 and DWI sequences. 126 features	RT with two cycles CCCT (Cisplatin)	GLCM GLGCM Gabor transform Intensity size zone matrix	Feature Selection: Fischer's coefficient and PCA Supervised learning: two different algorithms used-kNN and ANN.	Evaluate T1,T2 and DWI combined with supervised machine learning algorithms in predicting tumour response to CRT	All three sequences showed predictive value. T1w texture parameters most accurate in differentiating responders *vs* non-responders
Jansen et al[24] 2016	*N* = 19 HNSCC DCEMRI scans at 1.5T	Pre- and intra treatment Retrospective DCE-MRI images, Ktrans and Ve.	CRT	Energy (E) and homogeneity	Forward sequential feature selection algorithm used, followed by logistic regression analysis, to determine the probability of prediction	Merits of texture analysis on parametric maps derived from pharmacokinetic modeling with DCE-MRI	Chemo-radiation treatment in HNSCC significantly reduces the heterogeneity of tumors. E of Ve was significantly higher in intra treatment scans, relative to pretreatment scans

ANN, Artificial neural network; CCCT, Concurrent Chemotherapy; CRT, Chemoradiotherapy; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; HR, Hazard ratio; IC, Induction Chemotherapy; K^{trans}, volume transfer rate; MVCRA, Multivariable Cox regression analysis; NM, Not mentioned; NM, Not mentioned; RAD Score, Radiomics Score (Using linear combination of selected features weighed by relative coefficients); V^e, volume fraction of the extravascular extracellular space; kNN, kNearest neighbors.

Summary: Though on the outset 6 papers with sufficiently large sample sizes showing good performance of radiomics models may look encouraging, the fact that 4 of these appear to be same institution data with possibly overlapping patient cohorts warrant caution regarding the strength of evidence. Again, all papers were retrospective in design and evaluated the "predictive" role of radiomics as a biomarker, except the paper by Jansen et al[20] which was unique in that they compared pretreatment and intratreatment changes in texture analysis derived from DCE-MRI. Their finding of Energy of Ve increasing on treatment is interesting, however limited by the small sample size and lack of any internal or external validation of findings

Table 3. Role of ^{18}F-FDG PETCT radiomics in head and neck cancers

Author and year of publication	Patient sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selectionand number of parametrs derived	Outcome parameter	Result
Bogowicz et al[16] May 2017	TC = 93 ; VC = 56 Stage III and IV HNSCC(Orp,Hyp,Lar,OC)	Pretreatment Retrospective for TC Prospective for VC	Definitive IMRT 70 Gy with cisplatin or cetuximab.	Shape, Intensity, Texture, Wavelet transform 317 features	Grouping: PCA Selection: UVCRA for prognosis (9) For comparison with clinical and combined radiomics-clinical model: MVCRA(3) Split ROC curves at 18 mths	Predict LC: using CI Compare radiomics versus clinical model and a combined clinicoradiomic model for LC	Radiomics signature significantly associated with LC Combined radiomics + clinical model performed better than radiomics model alone in TC, but not VC
Bogowicz et al[28] June 2017	*N* = 172 TC = 121; VC = 51 Stage III and IV HNSCC(Orp,Hyp,Lar,OC)	Pretreatment Retrospective for TC Prospective for VC	Definitive IMRT 70 Gy with cisplatin or cetuximab TC VC = TC+_ consolidation cetuximab	Shape, Intensity, Texture, Wavelet transform 569	Combination of feature selection using PCA and classification using Cox regression with backward selection: chosen for least complicated and best discriminatory. Model validation: CI using Wilcoxon and bootstrap	Compare CT, PET, PETCT radiomics models for prognosis	CT radiomics overestimates probability of tumor control in high risk group. Mostly due to CT artifacts and variable contrast dose. CT (GLSZM,HLH) PET (Spherical disproportion, GLSZM) Combined (CT HLH and PET GLSZM) All showed similar discriminatory CI > 0.7
El Naqa et al[27] June 2009	*N* = 9 Stage and type: NM	Pretreatment Retrospective Median F/U period of 30 months.	Chemoradiotherapy (details not mentioned)	IVH, Shape, Texture, SUV measures 18	RS and AUC for association between extracted features and post-radiotherapy outcomes. Two-metric logistic regression model	Analyzed for endpoint of overall survival rate	Shape-based metrics had the highest categorical prediction power, while commonly used SUV descriptive statistics had the lowest predictive ability
Cheng et al[26] Sept 2013	*N* = 70 T3-4 OPSCC Follow up: 24 mths In-house (Matlab)	Pretreatment Retrospective	Completed platinum-based CCRT, cetuximab-based CBRT, or RT alone with curative intent	SUV histogram, TLG, NGLCM, NGTDM	MVCRA to identify the independent predictors of PFS, DSS, and OS RS to evaluate the associations between textural characteristics, SUV_{max}, MTV, TLG, and the general characteristics of the study participants.	Can textural features provide any additional prognostic information over TLG and clinical staging	Uniformity extracted from the normalized gray-level cooccurrence matrix found to be an independent prognostic predictor

(Continued)

Table 3. (Continued)

Author and year of publication	Patient sample size and disease stage	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selectionand number of parametrs derived	Outcome parameter	Result
Cheng et al[25] Oct 2014	*N* = 88 T3 or T4 OPSCC In-house (Matlab)	Pretreatment Retrospective	83 patients received CCRT, three received BRT, and the remaining two patients received RT alone with curative intent.	SUV, TLG, GLRLM, GLSZM	UV and MVCRA to identify the independent predictors of PFS and DSS. Kaplan-Meier curves for survival.	Prognostic impact of regional heterogeneity on Progression-free survival (PFS) and disease-specific survival (DSS)	Zone-size nonuniformity (ZSNU) identified as an independent predictor of PFS and DSS. Model combining total lesion glycolysis, uniformity and ZSNU showed a higher predictive value than each variable alone

An independent study by Liu et al[23] on 53 patients of NPC compared the performance of CE-T_1W, T_2W and DWI sequences in predicting response using pre-treatment 3 T imaging. Choosing supervised learning techniques for model construction, they concluded that though all three MR sequences predicted response with high accuracy, CE-T_1W was the single best performer with accuracy >0.9.

Finally, Jansen et al[24] retrospectively evaluated texture analysis on parametric maps derived from 1.5 T dynamic contrast-enhanced MRI (DCE-MRI) performed before and intratreatment in predicting response in 19 patients with head and HNSCC. Though they found no significant changes in the mean and standard deviation for Ktrans (volume transfer rate) and Ve (volume fraction of the extravascular extracellular space) between pre- and intratreatment, texture analysis revealed that the Energy of Ve was significantly higher in intratreatment scans, relative to pre-treatment scans ($p < 0.04$). They concluded that chemoradiation treatment in HNSCC significantly reduces the heterogeneity of tumour.

Role of 18-flu-deoxyglucose PET/CT in head and neck radiomics

Four papers have been published on the role of 18-FDG PET/CT in the prognosis of head and neck cancers(Table 3). Of these, two papers by Cheng et al[25,26] evaluated pre-treatment PET radiomics in Stage T3-4 oropharyngeal squamous cell carcinoma in predicting PFS and disease-specific survival (DSS). Uniformity extracted from the normalized GLCM and Zone-size nonuniformity were identified as independent predictors of PFS and DSS in the two articles, respectively.

El Naqa et al[27] found that a model combining histogram with shape features had highest predictive power for survival, whilst commonly clinically used standardized uptake value descriptive statistics had the lowest predictive ability in a small cohort of nine patients.

Bogowicz et al[16] compared CT, PET and combined 18-FDG PET/CT radiomic models in predicting local tumor control. No significant difference in performance of the models was observed (CI CT = 0.73, CI PET = 0.71, CI PET/CT = 0.73). However, CT radiomics-based model overestimated the probability of tumour control in the poor prognostic group.

Methodological standardization studies

Two groups—one using CT and the other using MRI sought to identify optimal machine-learning methods for their stability and performance in assessment of response in in head and neck cancer (Table 4).

Parmar et al[32] evaluated the performance of 13 feature selection methods and 11 classification methods in predicting OS in a sample of 196 patients with head and neck cancer on CT scan. They observed that three feature selection methods: minimum redundancy maximum relevance, mutual information feature selection, and conditional infomax feature extraction and three

Table 4. Studies on machine learning methods and external validation of radiomics in head and neck cancers.

Author and year of publication	Patient sample	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection	Outcome parameter	Result
Bin Zhang et al[28] June 2017	*N* = 110 TC = 70 VC = 40 Stage III to IVb NPC	Pretreatment Retrospective 1.5 T MRI 3D on T2w and CET1w	Not mentioned.	Shape Intensity Texture Wavelet transform 970	Data dichotomised: No recurrence, local failure and distant failure Quantified AUC and test error of different combination methods for prediction of PFS	Study objective: Which model is best: Compared 54 various permutations & combinations of: six feature selection methods nine machine learning classifiers	RF + RF had highest prognostic value (AUC 0.846) followed by RF + Adaptive Boosting (AUC 0.8204)
C.Parmar et al[25] **Dec 2015**	TC = 101 VC = 95	Pretreatment Retrospective CT images Type and stage: Not mentioned	TC = Either definitive RT alone or concurrent CRT. VC = Definitive RT alone, CRT with or without surgery	Shape Intensity Texture Wavelet transform 440 radiomic	Data dichotomised: Survival at 3 years. Median values of AUC and stability as thresholds to categorize the feature selection and classification methods into low or high performance (stability) groups	Study objective: Which feature classifier model is best Which selection method is best 13 feature selection methods and 11 machine-learning classification methods	Highest prognostic performance and stability was shown by: three feature selection methods: MRMR, MIFS, and CIFE. three classifiers: BY, RF and NN. Analysis investigating performance variability indicated that the choice of classification method is the major factor driving the performance variation

(Continued)

Table 4. (Continued)

Author and year of publication	Patient sample	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection	Outcome parameter	Result
H.*H. Aerts*[12] **June 2014**	*N* = 1019 NSCLC and Stage I to IVb HNSCC	Pretreatment Retrospective TC = Lung1 n=422 Maastro NSCLC **VC (four cohorts):** *Lung2 n* = 225 Radboud NSCLC; *H&N1 n* = 136 Maastro HNSCC; *H&N2 n* = 95 VU Amst HNSCC; *Lung3 n* = 89 Maastro NSCLC.	Definitive radiotherapy alone or chemoradiation with (*n* = 36) or without surgery	Shape Intensity Texture Filter based wavelet 440	Unsupervised clustering Single best performer on Stability ranks (4) MVCPHA	Validation of TC on the VC. Compared Radiomics with clinical parameters Prognosis: CI; Kaplan Meir Survival curves	Radiomic signature of TC had good performance on the VC and could be transferred from lung to head-and-neck cancer. Combined radiomics with TNM staging showed better performance compared to TNM staging alone. Radiomics preserved its prognostic performance in all treatment groups (RT or CTRT), for both Lung and H&N cancer patients Significant association with survival; primary T stage and overall stage

(Continued)

Table 4. (Continued)

Author and year of publication	Patient sample	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection	Outcome parameter	Result
C.Parmar[29] **June 2015**	*N* = 878 NSCLC and Stage I to IVb HNSCC	Pretreatment Retrospective **TC (two cohorts):** *Lung1 n* = 422 Maastro NSCLC; *H&N1 n* = 136 Maastro HNSCC **VC (two cohorts):** *Lung2 n* = 225 Radboud NSCLC; *H&N2 n* = 95 VU Amst HNSCC.	Definitive radiotherapy alone or chemoradiation	Shape Intensity Texture Filter based wavelet 440	Feature extraction: Consensus clustering (11 lung & 13 HNSCC) Cluster validation: Rand Statistic to assess agreement between TC and VC Independent external validation: MVCPHA With both: mean CI and mean AUC	Comparison of the prognostic performance of radiomic features in Lung and H&N cancer Association b/w radiomic feature clusters and patient survival: CI Association b/w feature cluster and a categorical clinical parameter : AUC	11 Lung and 6 HNSCC clusters had a significant prognostic association with patient survival. Both common as well as cancer-specific clustering and clinical associations of radiomic features. Strongest HNSCC associations : Prognosis (CI = 0.68±0.01) ; and stage (AUC = 0.77±0.02) Although five cluster pairs had substantial overlap between Lung and HNSCC, radiomic features also possess cancer-specific prognostic ability since signatures performed better in validation cohorts of the same cancer type.
R.***R. Leijenaar***[30] **Aug 2015**	*N* = 542 OP SCC	Pretreatment Retrospective **TC:** *n* = 422 *Lung1* Maastro NSCLC **VC:** *N* = 542 PMH OP SCC	Radiotherapy (IMRT) or CCRT	First order statistics Shape Gray level run length Wavelet	Signature model fit in a Cox regression and assessed model discrimination with Harrell's c-index. Kaplan-Meier survival curves between high and low signature predictions were compared with a log-rank test.	Prognostic index (PI) of the radiomic signature Effect of CT artifacts on radiomics signature	Radiomics validated well, demonstrating a good model fit and preservation of discrimination. Although CT artifacts were of influence, the signature had significant prognostic power regardless if patients with CT artifacts were included.

(Continued)

Table 4. (Continued)

Author and year of publication	Patient sample	Time of imaging and study type	Therapy	Number of parameters derived	Method of feature grouping/selection	Outcome parameter	Result
BogowiczNov[31] 2017	TC = 128 VC = 50 HNSCC (Stage I-IV) (Orp,Hyp,Lar,OC)	Retrospective PETCT scan, 3 months post CRT	Definitive RCT	Shape Intensity Texture Wavelet transform 649 features	2 independent models studied: Feature selection: PCA Classifier: LOSSY Cox and logistic regression models; CI MAASTRO indicator: Histogram range USZ: GLCM difference entropy	Association of Post CRT PET radiomics with local tumor control Compare reproducibility of 2 different software programs: USZ and MAASTRO	Independantly each software based model was prognostic for local tumor control. However 88% features were not reproducible in the two groups

AUC, Area under the ROC curve; CBRT, Concurrent Bioradiotherapy; CCCT, Concurrent Chemotherapy; Hyp, Hypopharynx; IVH, Intensity Volume histogram; LC, Local Control; Lar, Larynx; MVCRA, Multivariate Cox Regression Analysis; N, Number of patients; NM, Not mentioned; OC, Oral cavity; Orp, Oropharnx; RS, Spearman's rank correlation; TC, Training Cohort; UVCRA, Univariate Cox Regression Analysis; VC, Validation Cohort.
Summary: There is sufficient diversity in the scope ofarticles evaluating the role of 18-FDG PET/CTradiomics. Whilst 4 of these 5 papers are essentially from 2 groups, all groupsclaim good "predictive" power for overall survival or local tumour control. Anadditional paper by Bogowicz et al[24] comparing the performance of CT,PET and combined 18-FDG PET/CT found that CT radiomics had a tendency tooverestimate chance of tumor control in poor responders. They advocated designof local control models on PET scans, rather than CT.

classification methods: Bayesian, Random Forest and Nearest neighbour showed highest prognostic performance and stability.

Similarly, Zhang et al[28]evaluated 6 feature selection methods and 9 classification methods in 110 patients with advanced NPC on MRI in predicting local failure and distant failure. Their results showed that the combination methods Random Forest (RF) + RF had the highest prognostic performance, followed by RF + Adaptive Boosting and Sure Independence Screening + Linear Support Vector Machines.

Validation studies

Lastly, there are four papers published on external validation of the performance of radiomics as a prognostic and predictive tool (Table 4). However, most of these have focussed on CT radiomics. In fact, none of these papers evaluated MRI radiomics, which is more widely used to monitor treatment response in head and neck cancer.

Spearheading efforts in this direction was the landmark paper by Aerts et al,[12] who conducted a radiomic analysis of 440 features extracted from a pre-treatment CT database of 1019 patients with either lung or head and neck cancer across different institutions. They demonstrated a transferable capability of radiomics across two cancer types indicating that radiomics quantifies a general prognostic cancer phenotype that can broadly be applied to other cancer types.

Parmar et al[29] from the same group, on the other hand demonstrated that cancer-specific prognostic ability of radiomics signatures performed better in validation cohorts of a particular cancer type than common clustering across diverse cancer types.

Leijenaar et al[30] demonstrated external validation on an independent cohort of 542 oropharyngeal squamous cell carcinoma patients. They undertook to test if the radiomics study in Netherlands performed by Aerts et al,[12] could be validated in a large and independent cohort of North American patients. Their results demonstrated that the radiomics signature validated well, demonstrating good model fit and preservation of discrimination.

Bogowicz et al[31] also published their findings on reproducibility of radiomics for predicting tumor control on post radiochemotherapy 18-FDG PET/CT scans. They compared their in-house USZ (University hospital of Zurich) software with that of MAASTRO (Netherlands). However, though they found that both models were prognostic for tumour control independently in advanced head and neck cancers, 88% of the features were not reproducible between the implementations. Moreover, this study only looked at post-treatment scans of patients.

DISCUSSION AND CRITICAL APPRAISAL OF ARTICLES REVIEWED

With regards to this review's applicability (NICE[33]) most of the included studies had a well-defined study population and considered an appropriate, relevant intervention, irrespective of the study design. However, the use of advanced imaging varied

across the studies in the following respects: the use of different imaging techniques (*e.g.* MRI, CT, PET); the different types of scanners used within each imaging technique (*i.e.* different scanners might have different imaging settings and levels of accuracy); and particularly, the different types of radiomics model design along with varying treatment strategies. Some studies incorporated an internal validation of their analysis, by stratifying patients into a training and validation cohort, some performed external validation, while some did not perform any validation at all. This makes direct comparison between the studies extremely difficult and affects the studies' generalizability across research groups.

Inherent ambuiguity in the very process of performing radiomic analysis makes it very difficult to expect uniformity in research publications. There are a wide array of feature reduction/selection and classification methods available, there being no single "correct" or "incorrect" method. However, the choice of methodology would definitely affect reproducibility and this would need further research. Moreover, various researchers use in-house proprietary software for performing their radiomic extraction, as well as varying statistical and bioinformatics approaches for data analysis and interpretation which adds to the complexity.

Most studies evaluated the role of radiomics "pre-treatment," except Bogowicz et al[31] who studied the role of PET radiomics 3 months post-treatment. Also, most authors only looked at imaging findings at a single pre- or post-treatment time point rather than monitoring changes over time, except Jansen et al[24] who monitored significance of radiomics changes with treatment. The vast data-mine of serial imaging, since it is standard practice in the clinic to monitor response to therapy on MRI/PET remains virtually untapped. Comparing the relative performance of serial anatomical imaging with temporal changes in heterogeneity of both the primary tumour and node would definitely aid in answering the clinical dilemma of monitoring treatment response.

Finally, the greatest concern of the authors here is the "translational potential". Until such time that radiomics analysis become more widely accessible and standardized to a minimum data set, most radiomics-related studies are only being conducted by a handful of niche research groups worldwide.

LIMITATIONS

Though the authors set-out to perform a meta-analysis, the limited number of papers published along with extreme methodological heterogeneity and reporting meant we could only perform a descriptive systematic review and though unavoidable, this is a major limitation of our study. Another limitation is that we excluded grey literature and papers published in non-English journals.

CONCLUSIONS

Though most individual papers claim radiomics to be a good performer as a "predictive" tool in head and neck cancer, the current level of evidence remains low given the lack of validation and reproducibility studies. Moreover, quality assurance and quality control parameters should be agreed upon by researchers, such that a specific imaging acquisition protocol matched with a specific radiomic and analytic protocol can be validated to perform within an estimated error range as a predictive, prognostic and evaluative tool. For radiomics to make it through the "translational gap" and gain traction in clinical practice, prospective randomized controlled trials will be required to demonstrate consistency, reproducibility, efficacy, cost effectiveness and prognostic impact, else this would be another potential imaging biomarker that got "lost in translation".

REFERENCES

1. Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893–917. doi: https://doi.org/10.1002/ijc.25516
2. Brouha XDR, Tromp DM, De Leeuw JRJ, Hordijk GJ, Winnubst JAM. Increasing incidence of advanced stage head and neck tumours. *Clin Otolaryngol Allied Sci* 2003; **28**: 231–4. doi: https://doi.org/10.1046/j.1365-2273.2003.00696.x
3. National Comprehensive Clinical Guidelines: Clinical Practice Guidelines in Oncology. *Head and Neck cancers. Version 1. 2016 [Website]* 2016;.
4. Birchard KR, Hoang JK, Herndon JE, Patz EF. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. *Cancer* 2009; **115**: 581–6. doi: https://doi.org/10.1002/cncr.24060
5. El-Khodary M, Tabashy R, Omar W, Mousa A, Mostafa A. The role of PET/CT in the management of head and neck squamous cell carcinoma. *The Egyptian Journal of Radiology and Nuclear Medicine* 2011; **42**: 157–67. doi: https://doi.org/10.1016/j.ejrnm.2011.05.006
6. Hermans R, Pameijer FA, Mancuso AA, Parsons JT, Mendenhall WM. Laryngeal or hypopharyngeal squamous cell carcinoma: can follow-up CT after definitive radiation therapy be used to detect local failure earlier than clinical examination alone? *Radiology* 2000; **214**: 683–7. doi: https://doi.org/10.1148/radiology.214.3.r00fe13683
7. King AD, Keung CK, Yu K-H, Mo FKF, Bhatia KS, Yeung DKW, et al. T2-Weighted MR imaging early after chemoradiotherapy to evaluate treatment response in head and neck squamous cell carcinoma. *AJNR Am J Neuroradiol* 2013; **34**: 1237–41. doi: https://doi.org/10.3174/ajnr.A3378
8. Chung SR, Choi YJ, Suh CH, Lee JH, Baek JH. Diffusion-Weighted magnetic resonance imaging for predicting response to chemoradiation therapy for head and neck squamous cell carcinoma: a systematic review. *Korean Journal of Radiology* 2019; **20**: 649–61. doi: https://doi.org/10.3348/kjr.2018.0446
9. King AD. Mri assessment of treatment response. *Cancer Imaging* 2015; **15**: O26. doi: https://doi.org/10.1186/1470-7330-15-S1-O26
10. de Bree R, Wolf GT, de Keizer B, Nixon IJ, Hartl DM, Forastiere AA, et al. Response assessment after induction chemotherapy for head and neck squamous cell carcinoma: from physical examination to modern

imaging techniques and beyond. *Head Neck* 2017; **39**: 2329–49. doi: https://doi.org/10.1002/hed.24883
11. Patil V, Noronha V, Joshi A, Muddu Krishna V, Juvekar S, Pantvaidya G, et al. Is there a limitation of RECIST criteria in prediction of pathological response, in head and neck cancers, to postinduction chemotherapy? *ISRN Oncol* 2013; **2013**: 1–6. doi: https://doi.org/10.1155/2013/259154
12. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**: 4006. doi: https://doi.org/10.1038/ncomms5006
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Alessandro Liberati DGA, Clarke M, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; **6**: e1000100. doi: https://doi.org/10.1371/journal.pmed.1000100
14. David Moher LS, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P.Lesley A Stewart & PRISMA-P Group Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, (2015; **2015**Article number: 1.
15. Systematic Reviews. Centre for Reviews and Dissemination. 2009. Available from: https://www.york.ac.uk/crd/guidance/.
16. Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed tomography Radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 2017; **99**: 921–8. doi: https://doi.org/10.1016/j.ijrobp.2017.06.002
17. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, et al. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology* 2013; **269**: 801–9. doi: https://doi.org/10.1148/radiol.13130110
18. Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to human papillomavirus status. *Oral Oncol* 2017; **71**: 150–5. doi: https://doi.org/10.1016/j.oraloncology.2017.06.015
19. Zhang B, Ouyang F, Gu D, Dong Y, Zhang L, Mo X, et al. Advanced nasopharyngeal carcinoma: pre-treatment prediction of progression based on multi-parametric MRI radiomics. *Oncotarget* 2017; **8**: 72457–65. doi: https://doi.org/10.18632/oncotarget.19799
20. Ouyang F-S, Guo B-L, Zhang B, Dong Y-H, Zhang L, Mo X-K, et al. Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma. *Oncotarget* 2017; **8**: 74869–79. doi: https://doi.org/10.18632/oncotarget.20423
21. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clinical Cancer Research* 2017; **23**: 4259–69. doi: https://doi.org/10.1158/1078-0432.CCR-16-2910
22. Wang G, He L, Yuan C, Huang Y, Liu Z, Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. *Eur J Radiol* 2018; **98**: 100–6. doi: https://doi.org/10.1016/j.ejrad.2017.11.007
23. Liu J, Mao Y, Li Z, Zhang D, Zhang Z, Hao S, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. *Journal of Magnetic Resonance Imaging* 2016; **44**: 445–55. doi: https://doi.org/10.1002/jmri.25156
24. Jansen JFAet al. Texture analysis on parametric maps derived from dynamic contrast-enhanced magnetic resonance imaging in head and neck cancer. *World J Radiol* 2016; **8**: 90–7. doi: https://doi.org/10.4329/wjr.v8.i1.90
25. Cheng N-M, Fang Y-HD, Lee L-yu, Chang JT-C, Tsan D-L, Ng S-H, L-y L, et al. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur J Nucl Med Mol Imaging* 2015; **42**: 419–28. doi: https://doi.org/10.1007/s00259-014-2933-1
26. Cheng N-M, Dean Fang Y-H, Tung-Chieh Chang J, Huang C-G, Tsan D-L, Ng S-H, et al. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-Stage oropharyngeal squamous cell carcinoma. *Journal of Nuclear Medicine* 2013; **54**: 1703–9. doi: https://doi.org/10.2967/jnumed.112.119289
27. El Naqa I, Grigsby PW, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit* 2009; **42**: 1162–71. doi: https://doi.org/10.1016/j.patcog.2008.08.011
28. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett* 2017; **403**: 21–7. doi: https://doi.org/10.1016/j.canlet.2017.06.004
29. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep* 2015; **5**: 13087. doi: https://doi.org/10.1038/srep13087
30. Leijenaar RTH, Carvalho S, Hoebers FJP, Aerts HJWL, van Elmpt WJC, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol* 2015; **54**: 1423–9. doi: https://doi.org/10.3109/0284186X.2015.1061214
31. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, et al. Post-radiochemotherapy PET radiomics in head and neck cancer - The influence of radiomics implementation on the reproducibility of local control tumor models. *Radiother Oncol* 2017;: 0606..
32. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic Machine-Learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol* 2015; **5**: 272. doi: https://doi.org/10.3389/fonc.2015.00272
33. National Institute for Health and Care Excellence(NICE): The guidelines manual. Process and methods [PMG6]. 2012. Available from: https://www.nice.org.uk/process/pmg6/chapter/developing-review-questions-and-planning-the-systematic-review.