Radiomics for the Prediction of Treatment Outcome and Survival in Patients With Colorectal Cancer

A Systematic Review

Staal, Femke C.R.; van der Reijd, Denise J.; Taghavi, Marjaneh; Lambregts, Doenja M.J.; Beets-Tan, Regina G.H.; Maas, Monique

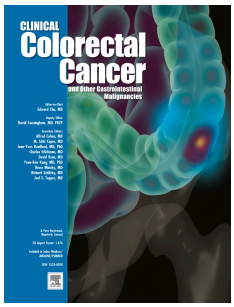
Published in:
Clinical Colorectal Cancer

DOI:
10.1016/j.clcc.2020.11.001

Publication date:
2021

Document version:
Accepted manuscript

Journal Pre-proof



Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review

F.C.R. Staal, D.J. van der Reijd, M. Taghavi, D.M.J. Lambregts, R.G.H. Beets-Tan, M. Maas

PII: S1533-0028(20)30144-4

DOI: https://doi.org/10.1016/j.clcc.2020.11.001

Reference: CLCC 684

To appear in: *Clinical Colorectal Cancer*

Received Date: 2 January 2020

Revised Date: 3 September 2020

Accepted Date: 2 November 2020

Please cite this article as: Staal FCR, van der Reijd DJ, Taghavi M, Lambregts DMJ, Beets-Tan R, Maas M, Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review, *Clinical Colorectal Cancer* (2020), doi: https://doi.org/10.1016/j.clcc.2020.11.001.

Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review

Staal F.C.R.[a,b*], van der Reijd D.J.[a,b*], Taghavi M. [a,b], Lambregts D.M.J. [a], Beets-Tan RGH[a,b,c], Maas M.[a]

[*] *F.C.R Staal and D.J. van der Reijd had an equal contribution to this research and therefore share first authorship.*

[a] Department of Radiology
Antoni van Leeuwenhoek - The Netherlands Cancer Institute
Plesmanlaan 121
1066 CX Amsterdam
The Netherlands

[b] GROW School for Oncology and Developmental Biology
Maastricht University Medical Centre
Universiteitssingel 40
6229 ER Maastricht
The Netherlands

[C] Faculty of Health Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark

Corresponding Author:
Monique Maas
Department of Radiology
The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital
Plesmanlaan 121
1066 CX Amsterdam
The Netherlands
T: +31-20-5121790
F: +31-20-6691101
Email: moniquemaas@live.nl

Abbreviations

ADC: apparent diffusion coefficient

AJCC: American Joint Committee on Cancer

ANN: artificial neural network

AUC: area under the ROC-curve

AUC-CSH: area under the curve of the cumulative SUV-volume histogram

CRC: colorectal cancer

CRLM: colorectal liver metastases

(n)ceCT: (non) contrast enhanced CT

CR: complete response

(n)CRT: (neoadjuvant) chemoradiotherapy

(n)CTx: (neoadjuvant) chemotherapy

DCE: dynamic contrast enhanced

DFS: disease-free survival

DNN: deep neural network

DRE: digital rectal examination

DWI: diffusion weighted imaging

EFS: event free survival

FLR: future liver remnant

GLCM: Gray Level Co-occurrence Matrix

GLDM: Gray Level Dependence Matrix

GLRLM: Gray Level Run Length Matrix

GLZSM: Gray Level Size Zone Matrix

GR: good response

HDFS: hepatic disease free survival

HR: hazard ratio

HU: Hounsfield unit

IBSI: Image Biomarker Standardisation initiative

K_{ep}: exchange between plasma and extravascular extracellular space

K^{trans}: endothelial transfer constant

LASSO: least absolute shrinkage and selection operator

LOOCV: leave-one-out cross-validation

LoG: Laplacian of Gaussian

LRC: locoregional control

MFS: metastases free survival

MRI: magnetic resonance imaging

MwU: Mann Whitney U test

NGTDM: Neighbouring Gray Tone Difference Matrix

OS: overall survival

PET: positron emission tomography

PFS: progression free survival

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

RECIST: Response Evaluation Criteria in Solid Tumours

RF: random forest

RFS: recurrence free survival

ROI: region of interest

ROC: receiver operating curve

RQS: Radiomics Quality Score

(n)RT: (neoadjuvant) radiotherapy

SD: standard deviation

SDD: standard deviation of pixel distribution histogram

SUV: standard uptake volume

SVM: support vector machine

TRG: tumour regression grade

V_e: fraction of tissue volume occupied by extravascular extracellular space

Declaration of Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Background: Prediction of outcome in patients with colorectal cancer (CRC) is challenging, due to lack of a robust biomarker and heterogeneity between and within tumours. The aim of this review was to assess the current possibilities and limitations of radiomics (on CT, MRI and PET) for the prediction of treatment outcome and long-term outcome in CRC.

Methods: MEDLINE/PubMed was searched up to August 2020 for studies that used radiomics for the prediction of response to treatment and survival in patients with CRC (based on pre-treatment imaging). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool and Radiomics Quality Score (RQS) were used for quality assessment.

Results: A total of 76 studies met the inclusion criteria and were included for further analysis. Radiomics analyses were performed on MRI in 41 studies, on CT in 30 studies and on 18F-FDG-PET-CT in 10 studies. Heterogeneous results were reported regarding radiomics methods and included features. High quality studies (n=13), consisting mainly of MRI-based radiomics to predict response in rectal cancer, were able to predict response with a good performance.

Conclusion: Radiomics literature in CRC is highly heterogeneous, nonetheless it holds promise for the prediction of outcome. The most evidence is available for MRI-based radiomics in rectal cancer. Future radiomics research in CRC should focus on independent validation of existing models rather than developing new models.

Keywords: Quantitative imaging analysis, response, survival, artificial intelligence, neoadjuvant chemotherapy, metastasis

Introduction

Colorectal cancer (CRC) is one of the most frequently diagnosed cancers in the Western World, with a 5-year survival rate of 11-65% depending on the initial stage.[1, 2] In the past decade overall survival (OS) has improved, due to the introduction of treatment options such as (new) chemotherapies, (chemo)radiotherapy and immunotherapy.[3, 4] Therefore, management of CRC is now a multimodal approach, but it remains a challenge to know upfront which patients will respond to what type of therapy.[5-7] Currently, we rely on response evaluation after treatment, while prediction of response and long-term outcome remains challenging. Therefore, there is a need for a biomarker that can aid in the selection of the best therapy for each patient before the start of treatment, based on their individual chance for response or good prognosis.

Imaging is routinely used during diagnosis and follow-up and as such it provides an optimal opportunity to identify (non-invasive) biomarkers that could predict response and long-term outcome. Radiomics is a method to optimally use the available imaging data. It refers to extracting quantitative features (i.e. radiomics features) from medical images that provide information about the whole tumour phenotype and microenvironment, which is impossible to be appreciated by visual inspection by a radiologist. Over the past decade radiomics has become a hot topic and an increasing amount of studies in oncology have been published, with promising results in various tumour types, including CRC.[8-12] For CRC, the most interesting areas to explore with radiomics are response and long-term outcome prediction. Because of the complexity, heterogeneity and increasing amount of the literature that has been published in the last decade, it is challenging to interpret the results.[13-15] The purpose of this systematic review is, therefore, to provide an overview of the available literature regarding the use of radiomics for the prediction of treatment outcome and survival in patients with CRC. Furthermore, areas lacking evidence and directions for future research will be identified.

Materials and methods

Radiomics

Radiomics is a technique that can convert medical images (e.g. MRI, CT, and PET) into innumerable quantitative features which describe the relationships between the intensity or density of voxels and position in an image. The analysis involves a few steps as shown in Figure 1 and include: (1) image acquisition (and if necessary, reconstruction), (2) identification and delineation of the region(s) of interest (ROI), (3) extraction of features (Table 1), (4) feature selection and finally, (5) feature classification and data analysis. The steps are explained in more detail in Supplement A1.

Search strategy and selection criteria

This systematic review was conducted in accordance to the PRISMA statement.[16] We conducted a systematic literature search to identify relevant studies published in MEDLINE/PubMed until August 2020, using the following Medical Subject Headings (MeSH) terms: 'Colorectal Neoplasms', 'Colonic Neoplasms' and 'Rectal Neoplasms'. Additionally, the following free search terms were used for the search: 'colorectal neoplasms OR colorectal metastases OR colorectal cancer OR colonic neoplasms OR colon cancer OR rectal neoplasms OR cancer OR colorectal cancer liver metastases OR colorectal liver metastases', 'texture analysis OR textural analysis OR texture parameters OR texture features OR texture or radiomic features OR radiomics OR radiomic* OR radiomics analysis OR quantitative image features OR quantitative image feature analysis'.

Two reviewers [XX and XXXX] independently searched for eligible studies and articles that met the following criteria were included: (1) patients with CRC, (2) radiomics analysis, (3) radiomics of pre-treatment imaging, (4) outcome comprising of response assessment or survival as a reference standard. The whole spectrum of radiomics analyses was considered for inclusion: studies with only (limited) histogram-based features (that do not account for the location of the pixels or spatial interrelationship between pixels) and studies using texture features (e.g. a higher number of features that also take the spatial interrelationship into account).

First, titles and abstracts were screened for potentially eligible studies that met the inclusion criteria. If the references were considered relevant, full text articles were independently reviewed by

both authors to evaluate which studies met the inclusion criteria. Duplicates, reviews, case reports, letters to the editor and comments on studies were excluded. Furthermore, articles that evaluated detection or diagnosis of CRC or CRLM, studies that evaluated imaging after treatment (i.e. not prediction) and studies in another language than English were excluded. Finally, the reference lists of relevant articles were checked manually by both readers, in order to find additional eligible studies. The search and inclusion strategy were supervised by a third reviewer [XX].

The reviewers independently extracted data from the studies. Data extracted from the studies were: (1) study population, (2) study objective, (3) primary tumour, (4) imaging modality, (5) reference standard, (6) classification method, (7) ROI, (8) radiomics workflow, (9) included features, (10) intervention type, (11) most relevant statistical results (p-values, AUC, c-index et cetera). Next, the studies were ordered according to the ROI that was assessed (i.e. primary tumour (colon or rectum) or liver) and what the outcome measure was (i.e. response to treatment or survival). Disagreements were resolved by consensus and if no consensus was reached, a third reviewer was consulted [XX].

Study quality was assessed with the QUADAS-2 checklist.[17] Moreover, the Radiomics Quality Score (RQS), as proposed by Lambin et al. in 2017, was used to specifically assess the radiomics methodology and analysis.[18] The RQS is a score based on sixteen components, with a maximum score of 36 points, where a higher score indicates a higher quality research.

Results

Literature search

The systematic literature search identified a total of 300 relevant studies. All studies were checked for eligibility and based on title and abstract 209 studies were excluded. Full-text analysis of the remaining 91 potentially eligible studies led to exclusion of 19 for the following reasons: thirteen studies did not predict outcome[19-31], three studies used post treatment imaging to predict outcome and pre-treatment data could not be derived separately[32-34], two studies were only available in Chinese[35, 36] and finally one study did not perform radiomics analysis[37], leaving 72 studies to be included. Upon checking the reference lists from the included studies, four additional studies were identified that were

eligible for inclusion.[14, 38-40] So finally, 76 studies were included for analysis as depicted in the PRISMA flowchart (fig 2).

Included studies

The studies were published between 2007 and 2020, with the highest number of publications in the past few years as shown in Figure 3. Of the included studies, seven had a prospective design and twelve were multicentre studies. A total of 8449 patients were included in all studies (range: 8-701). Radiomics analyses were performed on MRI in 41 studies, on CT in 30 studies and on 18F-FDG-PET-CT in 10 studies. The ROI was in the rectum in 49 studies[14, 15, 39-85], in either the colon or rectum in 6 studies[86-91], in the liver in 19 studies[13, 38, 92-108], one study placed the ROI in the mesorectal compartment (both with and without including the actual tumour)[109] and one study analysed all visually identifiable tumour lesions (i.e. primary tumour, all metastases and lymph nodes).[110]

In 30 studies, patients received neoadjuvant chemoradiation ([nCRT] all for rectal cancer) followed by resection. Other therapies included: chemotherapy ([CTx] n=9), resection only (n=7), nCRT without resection (n=3), neoadjuvant (n)CTx followed by resection (n=2), 12 studies had patient cohorts with mixed therapies and finally, in three studies baseline imaging was analysed, but treatment during follow-up was not specified (treatment per study is specified in table 2). A total of 41 studies aimed to predict response to treatment, 19 to predict survival, 10 to predict both response to treatment and survival and seven studies aimed to predict new metastases. One study did not predict response, but only reported feature values pre-treatment in both response groups without doing any statistical analysis.[85] The reference standard for response evaluation was either histopathology or response evaluation criteria in solid tumours (RECIST 1.1)[111], except for 2 studies that used 'size shrinkage' on imaging[102] or downstaging by comparing cTNM with ypTNM.[79] For histopathological response assessment, Tumour Regression Grade (TRG) systems [112-114] were used to define 'good responders' (GR) and 'complete responders' (CR), as specified in table 2. The reference standard for the prediction of new CRLM, was predominantly based on visual assessment of liver metastases on imaging during follow-up. For long term outcome, OS was most frequently studied and other outcome measures for survival prediction are specified in Table 2. A summary of the included studies is

presented in Table 2 and individual study results are available in Supplementary table C1. For the sake of comprehensibility, the results will be primary structured based on the ROI that was analysed.

Quality assessment

The results of the QUADAS-2 assessment are available in the Supplementary table B1. There were no applicability concerns for the included studies and none of the studies were excluded for quality concerns. A total of 44/76 (58%) studies were assessed as low risk on all four domains. The second domain 'index test' is not applicable in radiomics studies. Alternatively, the quality of radiomics was assessed by the RQS score and results are summarized in Table 3. The full results are available in Supplementary table B2. The range of the RQS scores was 0-47%. A majority of studies had a score below 30%, mainly due to lack of (external) validation, the retrospective nature and lack of feature reduction. Noteworthy is that a total of 27/76 (36%) studies scored a quality of 0%. Thirteen studies were considered of high quality based on QUADAS-2 and RQS.[39, 41, 49, 53, 54, 60, 67, 72, 73, 78, 81, 82, 84] These studies performed sufficient feature reduction, compared their results with the golden standard and used an unseen dataset for validation. Eight out of thirteen studies included clinical variables in their analysis. Subgroup analyses of these 13 high quality studies showed moderate to good predictive performance for response, using both logistic regression analysis (AUC 0.69-0.97) and machine learning classifiers (AUC 0.71-0.91). However, no specific features or transformation methods could be identified as most predictive.

Primary tumour radiomics

Rectal cancer: response prediction (CR or GR)

- MRI -

T2W-MRI and diffusion weighted MRI (DWI) were most frequently evaluated (n=23 and n=12 respectively, table 2).[15, 40, 41, 45, 46, 48-56, 60, 62, 64, 66, 69-71, 73, 74, 76-78, 81, 82, 84, 85, 109] For T2W-MRI, entropy was frequently studied and selected in prediction models, but conflicting results were reported.[15, 41, 46, 48, 50, 51, 53, 60, 66, 70, 77, 81, 84, 109] Even, two large and well conducted studies reported opposite findings for entropy.[53, 84] Similar conflicting results were found for energy[15, 46, 48, 54, 56, 66, 70, 77] and kurtosis.[51, 52, 66]

Three studies reported that none of the T2W-MRI radiomics features were significant predictors for response.[41, 48, 70] Multivariable logistic regression models were able to predict response with a moderate AUC (0.63-0.79) based on T2W-MRI.[50, 53, 66, 77, 84] Studies using machine learning classifiers, such as SVM, RF, and Naïve Bayesian network, yielded promising results to predict CR(AUC 0.71-0.87)[41, 54, 71, 82], but performed even better when predicting GR (AUC 0.83-0.90).[54, 73, 74, 82] For DWI, four studies reported no predictive value for ADC and IVIM histogram features for CR[52, 64] and GR[40, 62], while Nie et al. found a lower mean ADC in both CR and GR.[15] GLCM dissimilarity was predictive for GR in two studies[56, 76], but no difference was found in another study.[15] One study reported GLCM IVIM parameters were independent predictors in multivariate analysis for CR (AUC 0.99).[64] Inverse variance was included both in the logistic regression model by van Griethuysen et al. to predict CR (AUC 0.77)[78] and in the random forest model of Yang et al. to predict GR (AUC 0.83).[81] Only one study compared deep learning with a model with handcrafted features and reported the deep learning model to be more accurate to predict GR (AUC 0.73 vs. AUC 0.64, respectively).[55] T1W-MRI based radiomics of rectal cancer yielded moderate results to predict GR, with an Ada boost classifier based model[74] outperforming a logistic regression model[84] (AUC 0.78 vs. 0.63).

Five studies analysed dynamic contrast enhanced (DCE)-MRI.[15, 49, 69, 76, 85] Entropy was selected as a significant predictor for GR (AUC 0.85)[76] and CR (AUC 0.70-0.76).[49, 76] The ANN of Nie et al. was able to predict CR (AUC 0.76) and GR (AUC 0.85), based on DCE only.[15]

Six studies combined MRI sequences, which yielded a high predictive performance for both CR[15, 49, 76, 78] and GR [15, 56, 76, 84] (AUC 0.77-0.94 and AUC 0.72-0.91, respectively). The multi-sequence models outperformed single sequence classifiers for the prediction of response.[49, 76, 78, 84] When comparing individual sequences with one another, DCE[84], ADC[49] and T2W[60, 74] had the best predictive performance. Two studies developed multimodality models: PET/MRI[56] and CT/MRI.[60] The PET/MRI outperformed the MRI-model (AUC 0.86 vs AUC 0.72, respectively), but performance was similar to PET-only (AUC 0.84).[56] The CT/MRI model yielded better performance than CT-only (AUC 0.91 vs AUC 0.78, respectively), but was comparable with the performance of individual MRI sequences (AUC 0.81-0.86).[60]

- CT -

Six studies used CT-based radiomics of the rectum to predict response.[43, 47, 57, 60, 79, 83] On nceCT, Yuan et al. was able to predict GR with logistic regression (accuracy 68%).[83] Their random forest classifier had a good performance to predict CR (accuracy 84%)[83], while the random forest classifier of Hamerla et al. was not able to predict CR (accuracy 50%).[57] On ceCT, multiple histogram features were associated with GR, including a higher kurtosis in two studies.[47, 79] For prediction of GR, a weighted linear model yielded an AUC of 0.70[79] and a logistic regression model had an accuracy of 79%.[60] Only one study used ceCT-based radiomics for the prediction of CR and their SVM model outperformed the DNN while including the same features (AUC 0.72 and AUC 0.62, respectively).[43]

- 18F-FDG-PET-CT -

A total of five studies focused on PET-CT-based radiomics to predict response.[14, 42, 56, 65, 75] Giannini et al. reported that their logistic regression model was able to predict GR (AUC 0.84), with higher GLCM contrast and lower GLCM homogeneity in GR.[56] The random forest classifier of Shen et al. had a better performance to predict CR than their logistic regression model ($p<0.001$).[75] Finally, three studies found no differences in radiomics features between responders and non-responders in multivariable analysis.[14, 42, 65]

Colorectal cancer: survival prediction

Four studies used T2W-MRI-based radiomics of the primary rectal tumour to predict survival .[53, 58, 67, 70] Multiple histogram, GLCM and GLRLM features were correlated with better DFS.[58, 67, 70] One study correlated lower kurtosis with better RFS[58], while another study found no correlation with OS.[53] Dinapoli et al. performed multivariable cox regression to predict OS, but none of the included features remained significant.[53]

Seven studies analysed prediction of survival in primary CRC on CT.[47, 80, 86, 88-91] Dai et al. developed radiomics signatures to predict RFS (AUC 0.74) and OS (AUC 0.77).[90] Conflicting results were found for heterogeneity to predict survival. Some studies reported better OS in heterogeneous primary tumours (i.e. higher entropy and lower uniformity)[88, 89], while other studies reported more

homogeneous tumours to lead to improved DFS or PFS.[47, 91] Combining clinical and radiomics features had a better performance for the prediction of OS than a clinical or radiomics model only (AUC 0.73 [combined] vs. AUC 0.67 [clinical] and AUC 0.66 [radiomics]).[80]

Six studies evaluated 18F-FDG-PET(-CT) to predict survival and reported conflicting results.[14, 42, 59, 63, 65, 87] Moderate C-indexes were achieved when using deep learning to predict OS and recurrence (0.64-0.67).[59, 63]

Liver radiomics (in CRC)

Colorectal liver metastases: response prediction

Six studies used CT-based radiomics to predict response to chemotherapy in CRLM (table 3).[13, 94, 99, 106-108] Lower skewness and narrower SD were predictive for response.[13] Conflicting results were found for uniformity[94, 99, 108] and mean.[13, 94, 99, 107]

Three studies suggest that the predictive value of radiomics features is treatment dependent.[106, 108, 110] Both Ravanelli and Dercle et al. reported that their radiomics signature yielded good performance in the CTx+monoclonal antibody treatment group (AUC 0.80-0.81), however, not in the CTx only group (AUC 0.59).[106, 108]

Only two studies assessed MRI to predict response of CRLM to chemotherapy.[38, 102] Zhang et al. reported higher histogram variance and lower GLCM uniformity on T2W-images in responders. Their classifier was able to predict response with an AUC of 0.81.[102] Liang et al. reported lower mean and 1st-, 10th-, 50th-, 90th- and 99th percentile on ADC in responders. For DCE-MRI none of the histogram parameters were predictive for response.[38]

Colorectal liver metastases: survival prediction

Six studies performed multivariable analysis for prediction of survival based on CRLM.[101, 104, 105, 107, 108, 110] In the study by Shur et al. none of the ceCT or T1W-MRI based radiomics features were predictive for survival. On DCE-MRI, higher histogram pixel minimum (HR 1.66) and lower GLSZM small area emphasis (HR 0.61) were associated with better DFS.[105] On ceCT, more homogeneous tissue of the metastasis was associated with better survival in two studies.[101, 108] Conflicting results

were found for mean and OS.[107, 108] On FDG-PET-CT, two studies reported an association between OS and the AUC of the cumulative SUV-volume histogram (AUC-CSH).[104, 110]

Unaffected liver parenchyma: long-term outcome

A total of eight studies analysed the healthy liver parenchyma on CT in patients with CRC.[93-96, 98, 100, 101, 103] For the prediction of new CRLM, Taghavi et al. developed a combined clinical-radiomics model (random forest, including both histogram and GLDM features) that was able to predict new CRLM up to 24 months (AUC 0.86)[103], while three studies did not find any (histogram) radiomics features predictive of developing metachronous CRLM.[93, 95, 100]

For survival analysis, Beckers et al. reported no radiomics features to be independent predictors for OS in multivariate analysis.[94] Simpson et al. analysed the healthy liver parenchyma that would remain after hepatic surgery and found a lower texture signal (which is a linear combination of energy and entropy) in patient with a better OS (HR 2.19).[101]

Discussion

The aim of this systematic review was to provide an overview of the use of radiomics for the prediction of treatment outcome and survival in patients with CRC. The literature has demonstrated an exponential growth of radiomics studies in the past decade. Initially, radiomics analyses were predominantly bases on “simple” histogram and shape-based features, but as more higher-order features were used, more complex prediction models were developed with the use of machine learning classifiers and deep learning. Many studies have found potential for radiomics analyses in CRC, but the results are difficult to compare. The first issue is the lack of standardized imaging protocols and radiomics workflow, which leads to much variety in imaging sequences and in extraction of different features. Next, consensus in the literature about the mathematical definitions is lacking, sometimes even different names are used for the same mathematical expression, which makes interpretation challenging.[115, 116] A lot of conflicting results were reported and the reason for these differences could be the various sequences and imaging modalities (CT, MRI or PET), outcome definitions (GR, CR, survival) and filter transformations.

To provide more convincing evidence on the value of radiomics in CRC, quality was assessed with both QUADAS-2 and RQS and results of the highest quality studies were evaluated separately. Many studies were of low quality, which is in agreement with a previous study by Sanduleanu et al.. They reported that a majority of their included studies (73%) had a RQS score of less than 30%, comparable with our review.[117] Of the included 76 studies, only 18 were of reasonable quality regarding their radiomics workflow. These high-quality studies were predominantly MRI-based radiomics analysis of the rectum. These robust MRI-based radiomics studies have shown that good predictive performance can be achieved with both logistic regression and machine learning classifiers to predict response in rectal cancer. In all high-quality studies, feature reduction methods were applied, in order to make sure the amount of features was not too high in relation to the number of included patients, which reduces the chance of overfitting and type 1 error. This could indicate that the feature selection/reduction method is more important, rather than the type of classification method. Based on these high quality studies, no specific features, feature groups (e.g. GLCM, GLRLM, etc.) or transformation methods could be identified to predict outcome based on the primary CRC. None of the studies evaluating CRLM or liver parenchyma had a high quality based on both RQS and QUADAS-l. Only three of all liver-based radiomics studies used MRI to assess the predictive value of radiomics in CRLM. Given its high sensitivity for the detection of CRLM compared to CT, MRI-based radiomics might have more potential for outcome prediction.[118, 119]

Many studies evaluated a large amount of features in univariable analysis, in a small sample size or without (external) validation. Although the radiomics workflow of these studies was not desirable, a trend was observed regarding features that measure heterogeneity. Features that measure heterogeneity are frequently included in prediction models. Most studies report a negative association between a heterogeneous tumour and favourable outcome[15, 42, 46, 47, 66, 84, 88, 90] Furthermore, a homogeneous (i.e. features such as uniformity, energy, contrast) tumour was frequently associated with better response and survival.[15, 46, 47, 54, 65, 66, 68, 88] This is in line with reports from biological studies that patients with higher levels of intra-tumoural heterogeneity (i.e. higher entropy) have an inferior response to treatment and/or impaired survival.[6, 7] This heterogeneity is thought to be due to constant complex mutations that occur within a tumour in order to become and remain resistant to

therapy, explaining the poor survival in heterogeneous tumours.[7] Similar to the primary tumour, a more homogeneous tissue is associated with better outcome (response and survival) in CRLM.[101, 102, 104, 110] Combining radiomics features from multiple sequences resulted in a better performance than single-sequence models.[39, 49, 76, 78, 84] One could argue that each sequence provides different and thus complementary information about the tissue texture. Multimodality analysis was only performed in two studies and performance was slightly better compared to single modality models.[56, 60] Considering it is more time-consuming, we would not recommended focusing on multimodality analysis in colorectal tumours based on these results. Moreover, a combination of radiomics and clinical features were selected as relevant variables in multivariable or machine learning models; combining both leads to better performance of the model.[39, 47, 49, 50, 53, 58, 65, 70, 71, 79, 80, 87, 88, 90, 91, 93, 94, 101, 103-105, 109, 110] Also in the high quality studies, clinical features were of additional predictive value.[39, 67, 72, 81, 82, 84] This combination is expected to provide a more holistic model.[18] However, it is important to keep in mind that radiomics should provide additional information, or have additional predictive value over existing biomarkers. Van Griethuysen et al. reported that their radiomics model to predict CR based on MRI was comparable with morphological assessment by expert readers.[78]

The main limitation of this review is that the included studies are heterogeneous with regard to many different factors (e.g. patient cohort, imaging protocols, radiomics workflow), which made it impossible to conduct a meta-analysis. Second, we did not search for unpublished research which leads to an overestimation of results due to publication bias. The quality of the studies was assessed with both QUADAS-2 and RQS. Based on the QUADAS-2 the included studies had reasonable quality, while the RQS was low in most studies. Even though the quality was not always as desired, all studies were included in this review, to provide the most complete overview of the existing literature, including all problems that can arise with radiomics analyses. A subgroup analysis of high quality studies was performed to present more robust evidence on the topics in this study.

Future perspectives

The existing literature has shown that in the majority of studies, a radiomics model performs better than a 'conventional' model with clinical parameters only in both CRC and other tumours.[10, 23, 43, 49, 50,

[53, 58, 65, 67, 93, 101, 120, 121] Furthermore, the recommendation for future research is to combine both clinical and radiomics features into one model, since then the best performance is achieved. Before implementation of radiomics into clinical practice is possible, promising results need to be externally and independently validated. The most critical problem in current radiomics studies is the lack of reproducibility. Each study develops their own model, using different software, which makes it difficult to compare or reproduce the results. A recent study reported that even the use of different software platforms results in different feature values in the same clinical dataset.[116]. Rather than keep developing new models or methods, validation of previously published models should be the first step before prospective trials can be conducted. In order to aid this transition, it is preferable to standardize the radiomics workflow and to use commercially available software and avoid in-house applications (or make the code available).[18] There are several dedicated initiatives (such as the Image Biomarker Standardisation initiative (IBSI)[18, 122] that are trying to establish a standardization of the radiomics workflow between institutions.[18, 122, 123] Finally, before implementation into clinical practice is possible, clinicians such as radiologists, oncologists and surgeons need to acknowledge the potential of radiomics and understand the basics in order to facilitate these trials in oncology. It is also important to note that the goal is not to replace the clinician or to replace the current histopathological and clinical findings, but to complement them with radiomics to further personalize oncological treatment.

Conclusion

Based on this systematic review we can conclude that radiomics in CRC holds promise to predict response to treatment and long-term outcome, especially MRI-based radiomics for rectal cancer has shown the most potential. Future research in CRC should focus on independent validation of existing promising models of focus on developing new models for new research questions.

Table legends

Table 1 Radiomics feature categories and filter/transformation methods

Table 2 Overview of the included studies

Table 3 Details about radiomics workflow of the included studies

Figure legends

Figure 1 Overview of the steps in a radiomics workflow

Figure 2 PRISMA flowchart

Figure 3 Histogram of the included radiomics studies in CRC, sorted by number of publication per year

Supplementary material

Supplement A Stepwise explanation of the radiomics workflow

Supplement B

Supplementary table B1 QUADAS-2 results

Supplementary table B2 Radiomics Quality Score – full results

Supplementary table C1 Complete overview of the individual study results

References

1. Nederlandse Kankerregistratie bdI. Cijfers over kanker. Vol 20182018.

2. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European journal of cancer (Oxford, England : 1990).* 2018;103:356-387.

3. Brouwer NPM, Bos A, Lemmens V, et al. An overview of 25 years of incidence, treatment and outcome of colorectal cancer patients. *International journal of cancer.* 2018;143:2758-2766.

4. Schmoll HJ, Van Cutsem E, Stein A, et al. ESMO Consensus Guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. *Annals of Oncology.* 2012;23:2479-2516.

5. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature.* 2013;501:355-364.

6. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology.* 2017;15:81.

7. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer.* 2012;12:323.

8. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *European journal of radiology.* 2017;94:140-147.

9. Haider MA, Vosough A, Khalvati F, Kiss A, Ganeshan B, Bjarnason GA. CT texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer imaging : the official publication of the International Cancer Imaging Society.* 2017;17:4.

10. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology.* 2015;114:345-350.

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278:563-577.

12. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput Struct Biotechnol J.* 2019;17:995-1008.

13. Ahn SJ, Kim JH, Park SJ, Han JK. Prediction of the therapeutic response after FOLFOX and FOLFIRI treatment for patients with liver metastasis from colorectal cancer using computerized CT texture analysis. *European journal of radiology.* 2016;85:1867-1874.

14. Bundschuh RA, Dinges J, Neumann L, et al. Textural Parameters of Tumor Heterogeneity in (1)(8)F-FDG PET/CT for Therapy Response Assessment and Prognosis in Patients with Locally Advanced Rectal Cancer. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine.* 2014;55:891-897.

15. Nie K, Shi L, Chen Q, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2016;22:5256-5264.

16. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews.* 2015;4:1.

17. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine.* 2011;155:529-536.

18. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology.* 2017;14:749.

19. Li M, Li X, Guo Y, et al. Development and assessment of an individualized nomogram to predict colorectal cancer liver metastases. *Quantitative imaging in medicine and surgery.* 2020;10:397-414.

20. Li Y, Eresen A, Shangguan J, et al. Establishment of a new non-invasive imaging prediction model for liver metastasis in colon cancer. *American journal of cancer research.* 2019;9:2482-2492.

21. Liu M, Ma X, Shen F, Xia Y, Jia Y, Lu J. MRI-based radiomics nomogram to predict synchronous liver metastasis in primary rectal cancer patients. *Cancer medicine.* 2020.

22. Shu Z, Fang S, Ding Z, et al. MRI-based Radiomics nomogram to detect primary rectal cancer with synchronous liver metastases. *Scientific reports.* 2019;9:3374.

23. Chen LD, Liang JY, Wu H, et al. Multiparametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics. *Life sciences.* 2018;208:55-63.

24. Huang YQ, Liang CH, He L, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2016;34:2157-2164.

25. Meng X, Xia W, Xie P, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *European radiology.* 2018.

26. Liu H, Zhang C, Wang L, et al. MRI radiomics analysis for predicting preoperative synchronous distant metastasis in patients with rectal cancer. *European radiology.* 2018.

27. Wagner F, Hakami YA, Warnock G, Fischer G, Huellner MW, Veit-Haibach P. Comparison of Contrast-Enhanced CT and [(18)F]FDG PET/CT Analysis Using Kurtosis and Skewness in Patients with Primary Colorectal Cancer. *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging.* 2017;19:795-803.

28. Huang Y, He L, Dong D, et al. Individualized prediction of perineural invasion in colorectal cancer: development and validation of a radiomics prediction model. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu.* 2018;30:40-50.

29. Hu T, Wang S, Huang L, et al. A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. *European radiology.* 2018.

30. Ganeshan B, Miles KA, Young RC, Chatwin CR. Texture analysis in non-contrast enhanced CT: impact of malignancy on texture in apparently disease-free areas of the liver. *European journal of radiology.* 2009;70:101-110.

31. Liang C, Huang Y, He L, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget.* 2016;7:31401-31412.

32. Liu Z, Zhang XY, Shi YJ, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2017;23:7253-7262.

33. Jeon SH, Song C, Chie EK, et al. Delta-radiomics signature predicts treatment outcomes after preoperative chemoradiotherapy and surgery in rectal cancer. *Radiation oncology (London, England).* 2019;14:43.

34. Li Y, Liu W, Pei Q, et al. Predicting pathological complete response by comparing MRI-based radiomics pre- and postneoadjuvant radiotherapy for locally advanced rectal cancer. *Cancer medicine.* 2019;8:7244-7252.

35. Shu Z, Fang S, Ding Z, Mao D, Pang P, Gong X. [Application value of texture analysis of magnetic resonance images in prediction of neoadjuvant chemoradiotherapy efficacy for rectal cancer]. *Zhonghua wei chang wai ke za zhi = Chinese journal of gastrointestinal surgery.* 2018;21:1051-1058.

36. Zhang J, Zhou Y, Qiu M, Wu B. [Prediction of the therapeutic response after target-combined chemotherapy treatment for patients with liver metastasis from colorectal cancer using computed tomography texture analysis]. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi.* 2018;35:914-920.

37. Kudou M, Nakanishi M, Kuriu Y, et al. Value of intra-tumor heterogeneity evaluated by diffusion-weighted MRI for predicting pathological stages and therapeutic responses to chemoradiotherapy in lower rectal cancer. *Journal of Cancer.* 2020;11:168-176.

38. Liang H-Y, Huang Y-Q, Yang Z-X, Ying-Ding, Zeng M-S, Rao S-X. Potential of MR histogram analyses for prediction of response to chemotherapy in patients with colorectal hepatic metastases. *European radiology.* 2016;26:2009-2018.

39. Tang Z, Zhang XY, Liu Z, et al. Quantitative analysis of diffusion weighted imaging to predict pathological good response to neoadjuvant chemoradiation for locally advanced rectal cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2019;132:100-108.

40. Nougaret S, Vargas HA, Lakhman Y, et al. Intravoxel Incoherent Motion-derived Histogram Metrics for Assessment of Response after Combined Chemotherapy and Radiation Therapy in Rectal Cancer: Initial Experience and Comparison between Single-Section and Volumetric Analyses. *Radiology.* 2016;280:446-454.

41. Antunes JT, Ofshteyn A, Bera K, et al. Radiomic Features of Primary Rectal Cancers on Baseline T(2) -Weighted MRI Are Associated With Pathologic Complete Response to Neoadjuvant Chemoradiation: A Multisite Study. *Journal of magnetic resonance imaging : JMRI.* 2020.

42. Bang JI, Ha S, Kang SB, et al. Prediction of neoadjuvant radiation chemotherapy response and survival using pretreatment [(18)F]FDG PET/CT scans in locally advanced rectal cancer. *European journal of nuclear medicine and molecular imaging.* 2016;43:422-431.

43. Bibault JE, Giraud P, Durdux C, et al. Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Scientific reports.* 2018;8:12611.

44. Boldrini L, Cusumano D, Chiloiro G, et al. Delta radiomics for rectal cancer response prediction with hybrid 0.35 T magnetic resonance-guided radiotherapy (MRgRT): a hypothesis-generating study for an innovative personalized medicine approach. *La Radiologia medica.* 2018.

45. Bulens P, Couwenberg A, Intven M, et al. Predicting the tumor response to chemoradiotherapy for rectal cancer: Model development and external validation using MRI radiomics. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2020;142:246-252.

46. Caruso D, Zerunian M, Ciolina M, et al. Haralick's texture features for the prediction of response to therapy in colorectal cancer: a preliminary study. *La Radiologia medica.* 2018;123:161-167.

47. Chee CG, Kim YH, Lee KH, et al. CT texture analysis in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy: A potential imaging biomarker for treatment response and prognosis. *PloS one.* 2017;12:e0182883.

48. Crimì F, Capelli G, Spolverato G, et al. MRI T2-weighted sequences-based texture analysis (TA) as a predictor of response to neoadjuvant chemo-radiotherapy (nCRT) in patients with locally advanced rectal cancer (LARC). *La Radiologia medica.* 2020.

49. Cui Y, Yang X, Shi Z, et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *European radiology.* 2018.

50. Cusumano D, Dinapoli N, Boldrini L, et al. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. *La Radiologia medica.* 2018;123:286-295.

51. De Cecco CN, Ganeshan B, Ciolina M, et al. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. *Investigative radiology.* 2015;50:239-245.

52. De Cecco CN, Ciolina M, Caruso D, et al. Performance of diffusion-weighted imaging, perfusion imaging, and texture analysis in predicting tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3T MR: initial experience. *Abdominal radiology (New York).* 2016;41:1728-1735.

53. Dinapoli N, Barbaro B, Gatta R, et al. Magnetic Resonance, Vendor-independent, Intensity Histogram Analysis Predicting Pathologic Complete Response After Radiochemotherapy of Rectal Cancer. *International journal of radiation oncology, biology, physics.* 2018.

54. Ferrari R, Mancini-Terracciano C, Voena C, et al. MR-based artificial intelligence model to assess response to therapy in locally advanced rectal cancer. *European journal of radiology.* 2019;118:1-9.

55. Fu J, Zhong X, Li N, et al. Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer. *Physics in medicine and biology.* 2020;65:075001.

56. Giannini V, Mazzetti S, Bertotto I, et al. Predicting locally advanced rectal cancer response to neoadjuvant therapy with (18)F-FDG PET and MRI radiomics features. *European journal of nuclear medicine and molecular imaging.* 2019;46:878-888.

57. Hamerla G, Meyer HJ, Hambsch P, et al. Radiomics Model Based on Non-Contrast CT Shows No Predictive Power for Complete Pathological Response in Locally Advanced Rectal Cancer. *Cancers.* 2019;11.

58. Jalil O, Afaq A, Ganeshan B, et al. Magnetic resonance based texture parameters as potential imaging biomarkers for predicting long-term survival in locally advanced rectal cancer treated by chemoradiotherapy. *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland.* 2017;19:349-362.

59. Li H, Boimel P, Janopaul-Naylor J, et al. DEEP CONVOLUTIONAL NEURAL NETWORKS FOR IMAGING DATA BASED SURVIVAL ANALYSIS OF RECTAL CANCER. *Proceedings. IEEE International Symposium on Biomedical Imaging.* 2019;2019:846-849.

60. Li ZY, Wang XD, Li M, et al. Multi-modal radiomics model to predict treatment response to neoadjuvant chemotherapy for locally advanced rectal cancer. *World journal of gastroenterology.* 2020;26:2388-2402.

61. Liang M, Cai Z, Zhang H, et al. Machine Learning-based Analysis of Rectal Cancer MRI Radiomics for Prediction of Metachronous Liver Metastasis. *Academic radiology.* 2019;26:1495-1504.

62. Liu M, Lv H, Liu LH, Yang ZH, Jin EH, Wang ZC. Locally advanced rectal cancer: predicting non-responders to neoadjuvant chemoradiotherapy using apparent diffusion coefficient textures. *International journal of colorectal disease.* 2017;32:1009-1012.

63. Liu H, Li H, Boimel P, et al. COLLABORATIVE CLUSTERING OF SUBJECTS AND RADIOMIC FEATURES FOR PREDICTING CLINICAL OUTCOMES OF RECTAL CANCER PATIENTS. *Proceedings. IEEE International Symposium on Biomedical Imaging.* 2019;2019:1303-1306.

64. Liu S, Wen L, Hou J, et al. Predicting the pathological response to chemoradiotherapy of non-mucinous rectal cancer using pretreatment texture features based on intravoxel incoherent motion diffusion-weighted imaging. *Abdominal radiology (New York).* 2019;44:2689-2698.

65. Lovinfosse P, Polus M, Van Daele D, et al. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *European journal of nuclear medicine and molecular imaging.* 2018;45:365-375.

66. Meng Y, Zhang C, Zou S, et al. MRI texture analysis in predicting treatment response to neoadjuvant chemoradiotherapy in rectal cancer. *Oncotarget.* 2018;9:11999-12008.

67. Meng Y, Zhang Y, Dong D, et al. Novel radiomic signature as a prognostic biomarker for locally advanced rectal cancer. *Journal of magnetic resonance imaging : JMRI.* 2018.

68. Nardone V, Reginelli A, Scala F, et al. Magnetic-Resonance-Imaging Texture Analysis Predicts Early Progression in Rectal Cancer Patients Undergoing Neoadjuvant Chemoradiation. *Gastroenterology research and practice.* 2019;2019:8505798.

69. Palmisano A, Esposito A, Rancoita PMV, et al. Could perfusion heterogeneity at dynamic contrast-enhanced MRI be used to predict rectal cancer sensitivity to chemoradiotherapy? *Clinical radiology.* 2018;73:911.e911-911.e917.

70. Park H, Kim KA, Jung JH, Rhie J, Choi SY. MRI features and texture analysis for the early prediction of therapeutic response to neoadjuvant chemoradiotherapy and tumor recurrence of locally advanced rectal cancer. *European radiology.* 2020;30:4201-4211.

71. Petkovska I, Tixier F, Ortiz EJ, et al. Clinical utility of radiomics at baseline rectal MRI to predict complete response of rectal cancer after chemoradiation therapy. *Abdominal radiology (New York).* 2020.

72. Petresc B, Lebovici A, Caraiani C, Feier DS, Graur F, Buruian MM. Pre-Treatment T2-WI Based Radiomics Features for Prediction of Locally Advanced Rectal Cancer Non-Response to Neoadjuvant Chemoradiotherapy: A Preliminary Study. *Cancers.* 2020;12.

73. Shayesteh SP, Alikhassi A, Fard Esfahani A, et al. Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB).* 2019;62:111-119.

74. Shayesteh SP, Alikhassi A, Farhan F, et al. Prediction of Response to Neoadjuvant Chemoradiotherapy by MRI-Based Machine Learning Texture Analysis in Rectal Cancer Patients. *Journal of gastrointestinal cancer.* 2020;51:601-609.

75. Shen WC, Chen SW, Wu KC, et al. Predicting pathological complete response in rectal cancer after chemoradiotherapy with a random forest using (18)F-fluorodeoxyglucose positron emission tomography and computed tomography radiomics. *Annals of translational medicine.* 2020;8:207.

76. Shi L, Zhang Y, Nie K, et al. Machine learning for prediction of chemoradiation therapy response in rectal cancer using pre-treatment and mid-radiation multi-parametric MRI. *Magn Reson Imaging.* 2019;61:33-40.

77. Shu Z, Fang S, Ye Q, et al. Prediction of efficacy of neoadjuvant chemoradiotherapy for rectal cancer: the value of texture analysis of magnetic resonance images. *Abdominal radiology (New York).* 2019;44:3775-3784.

78. van Griethuysen JJM, Lambregts DMJ, Trebeschi S, et al. Radiomics performs comparable to morphologic assessment by expert radiologists for prediction of response to neoadjuvant chemoradiotherapy on baseline staging MRI in rectal cancer. *Abdominal radiology (New York).* 2020;45:632-643.

79. Vandendorpe B, Durot C, Lebellec L, et al. Prognostic value of the texture analysis parameters of the initial computed tomographic scan for response to neoadjuvant chemoradiation therapy in patients with locally advanced rectal cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2019;135:153-160.

80. Wang J, Shen L, Zhong H, et al. Radiomics features on radiotherapy treatment planning CT can predict patient survival in locally advanced rectal cancer patients. *Scientific reports.* 2019;9:15346.

81. Yang C, Jiang ZK, Liu LH, Zeng MS. Pre-treatment ADC image-based random forest classifier for identifying resistant rectal adenocarcinoma to neoadjuvant chemoradiotherapy. *International journal of colorectal disease.* 2020;35:101-107.

82. Yi X, Pei Q, Zhang Y, et al. MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Frontiers in oncology.* 2019;9:552.

83. Yuan Z, Frazer M, Zhang GG, et al. CT-based radiomic features to predict pathological response in rectal cancer: A retrospective cohort study. *Journal of medical imaging and radiation oncology.* 2020;64:444-449.

84. Zhou X, Yi Y, Liu Z, et al. Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer. *Annals of surgical oncology.* 2019;26:1676-1684.

85. Zou HH, Yu J, Wei Y, Wu JF, Xu Q. Response to neoadjuvant chemoradiotherapy for locally advanced rectum cancer: Texture analysis of dynamic contrast-enhanced MRI. *Journal of magnetic resonance imaging : JMRI.* 2018.

86. Badic B, Desseroit MC, Hatt M, Visvikis D. Potential Complementary Value of Noncontrast and Contrast Enhanced CT Radiomics in Colorectal Cancers. *Academic radiology.* 2018.

87. Nakajo M, Kajiya Y, Tani A, et al. A pilot study for texture analysis of (18)F-FDG and (18)F-FLT-PET/CT to predict tumor recurrence of patients with colorectal cancer who received surgery. *European journal of nuclear medicine and molecular imaging.* 2017;44:2158-2168.

88. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology.* 2013;266:177-184.

89. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *European journal of radiology.* 2013;82:342-348.

90. Dai W, Mo S, Han L, et al. Prognostic and predictive value of radiomics signatures in stage I-III colon cancer. *Clinical and translational medicine.* 2020;10:288-293.

91. Badic B, Hatt M, Durand S, et al. Radiogenomics-based cancer prognosis in colorectal cancer. *Scientific reports.* 2019;9:9743.

92. Beckers RCJ, Beets-Tan RGH, Schnerr RS, et al. Whole-volume vs. segmental CT texture analysis of the liver to assess metachronous colorectal liver metastases. *Abdominal radiology (New York).* 2017;42:2639-2645.

93. Beckers RCJ, Lambregts DMJ, Schnerr RS, et al. Whole liver CT texture analysis to predict the development of colorectal liver metastases-A multicentre study. *European journal of radiology.* 2017;92:64-71.

94. Beckers RCJ, Trebeschi S, Maas M, et al. CT texture analysis in colorectal liver metastases and the surrounding liver parenchyma and its potential as an imaging biomarker of disease aggressiveness, response and survival. *European journal of radiology.* 2018;102:15-21.

95. Lee SJ, Zea R, Kim DH, Lubner MG, Deming DA, Pickhardt PJ. CT texture features of liver parenchyma for predicting development of metastatic disease and overall survival in patients with colorectal cancer. *European radiology.* 2018;28:1520-1528.

96. Ganeshan B, Miles KA, Young RC, Chatwin CR. Hepatic enhancement in colorectal cancer: texture analysis correlates with hepatic hemodynamics and patient survival. *Academic radiology.* 2007;14:1520-1530.

97. Lubner MG, Stabo N, Lubner SJ, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdominal imaging.* 2015;40:2331-2337.

98. Miles KA, Ganeshan B, Griffiths MR, Young RC, Chatwin CR. Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology.* 2009;250:444-452.

99. Rao SX, Lambregts DM, Schnerr RS, et al. CT texture analysis in colorectal liver metastases: A better way than size and volume measurements to assess response to chemotherapy? *United European gastroenterology journal.* 2016;4:257-263.

100. Rao SX, Lambregts DM, Schnerr RS, et al. Whole-liver CT texture analysis in colorectal cancer: Does the presence of liver metastases affect the texture of the remaining liver? *United European gastroenterology journal.* 2014;2:530-538.

101. Simpson AL, Doussot A, Creasy JM, et al. Computed Tomography Image Texture: A Noninvasive Prognostic Marker of Hepatic Recurrence After Hepatectomy for Metastatic Colorectal Cancer. *Annals of surgical oncology.* 2017;24:2482-2490.

102. Zhang H, Li W, Hu F, Sun Y, Hu T, Tong T. MR texture analysis: potential imaging biomarker for predicting the chemotherapeutic response of patients with colorectal liver metastases. *Abdominal radiology (New York).* 2018.

103. Taghavi M, Trebeschi S, Simões R, et al. Machine learning-based analysis of CT radiomics model for prediction of colorectal metachronous liver metastases. *Abdominal radiology (New York).* 2020.

104. Rahmim A, Bak-Fredslund KP, Ashrafinia S, et al. Prognostic modeling for patients with colorectal liver metastases incorporating FDG PET radiomic features. *European journal of radiology.* 2019;113:101-109.

105. Shur J, Orton M, Connor A, et al. A clinical-radiomic model for improved prognostication of surgical candidates with colorectal liver metastases. *Journal of surgical oncology*. 2019.

106. Dercle L, Lu L, Schwartz LH, et al. Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting EGFR Pathway. *Journal of the National Cancer Institute.* 2020.

107. Dohan A, Gallix B, Guiu B, et al. Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab. *Gut.* 2020;69:531-539.

108. Ravanelli M, Agazzi GM, Tononcelli E, et al. Texture features of colorectal liver metastases on pretreatment contrast-enhanced CT may predict response and prognosis in patients treated with bevacizumab-containing chemotherapy: a pilot study including comparison with standard chemotherapy. *La Radiologia medica.* 2019;124:877-886.

109. Shaish H, Aukerman A, Vanguri R, et al. Radiomics of MRI for pretreatment prediction of pathologic complete response, tumor regression grade, and neoadjuvant rectal score in patients with locally advanced rectal cancer undergoing neoadjuvant chemoradiation: an international multicenter study. *European radiology.* 2020.

110. van Helden EJ, Vacher YJL, van Wieringen WN, et al. Radiomics analysis of pre-treatment [(18)F]FDG PET/CT for patients with metastatic colorectal cancer undergoing palliative systemic treatment. *European journal of nuclear medicine and molecular imaging.* 2018.

111. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European journal of cancer (Oxford, England : 1990).* 2009;45:228-247.

112. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer.* 1994;73:2680-2686.

113. Dworak O, Keilholz L, Hoffmann A. Pathological features of rectal cancer after preoperative radiochemotherapy. *International journal of colorectal disease.* 1997;12:19-23.

114. Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. 2017;67:93-99.

115. Buvat I, Orlhac F, Soussan M. Tumor Texture Analysis in PET: Where Do We Stand? *Journal of nuclear medicine : official publication, Society of Nuclear Medicine.* 2015;56:1642-1644.

116. Fornacon-Wood I, Mistry H, Ackermann CJ, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *European radiology.* 2020.

117. Sanduleanu S, Woodruff HC, de Jong EEC, et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiotherapy and Oncology.* 2018;127:349-360.

118. Schulz A, Viktil E, Godt JC, et al. Diagnostic performance of CT, MRI and PET/CT in patients with suspected colorectal liver metastases: the superiority of MRI. *Acta radiologica (Stockholm, Sweden : 1987).* 2016;57:1040-1048.

119. Vreugdenburg TD, Ma N, Duncan JK, Riitano D, Cameron AL, Maddern GJ. Comparative diagnostic accuracy of hepatocyte-specific gadoxetic acid (Gd-EOB-DTPA) enhanced MR imaging and contrast enhanced CT for the detection of liver metastases: a systematic review and meta-analysis. *International journal of colorectal disease.* 2016;31:1739-1749.

120. Zhai TT, van Dijk LV, Huang BT, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2017;124:256-262.

121. Kickingereder P, Burth S, Wick A, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. 2016;280:880-889.

122. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003.* 2016.

123. Computational Imaging & Bioinformatics Lab - Harvard Medical School. Radiomics. Vol 20192017.

Table 1: Radiomics feature categories and filter/transformation methods

Radiomics feature categories		Description
Shape features		Describe tumour based characteristics such as shape, volume, surface and derivative measurements
Histogram features		Describe the distribution of grey level frequency within a ROI of an image
Texture features	Gray Level Co-occurrence Matrix (GLCM)	Describes how certain pairs of discretized grey levels of neighbouring voxels are distributed along one image direction
	Gray Level Size Zone (GLSZM)	Describes the number of connected voxels that share the same grey level intensity. The more homogeneous the texture, the wider and flatter the matrix
	Gray Level Run Length Matrix (GLRLM)	Describes the distribution of discretized grey levels in terms of run length. A run length is the size of the number of neighbouring pixels in the same grey level along a fixed image direction. A high number represents rough texture, while a low number describes a thin texture
	Neighbouring Gray Tone Difference Matrix (NGTDM)	Describes differences between the sum of discretized grey levels of voxels and the average grey level of neighbouring voxels in adjacent planes
	Gray Level Dependence Matrix (GLDM)	Describes the difference of grey levels between one voxel and its neighbours in 3 dimensions. Captures the coarseness of the overall texture
Filter or transformation method		
	Gabor	Short time Fourier transformation that can compute features via time-frequency analysis with different frequencies and orientations
	Laplacian of Gaussian	Extraction of areas with different texture patterns (more and less course), with different standard deviations of the Gaussian filter
	Law's filters	A basis of 5 vectors that can emphasize micro-structure characteristics of an image such as edge, spot, ripple, wave, undulation and oscillation.
	Wavelet	Decompose the image in both space and frequency domain and can reduce background tissue and noise

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Ahn et al. (2016)[13]	235	66	S/R	CRC	CRLM	RECIST 1.1 (GR)	CT	Response	CTx	Lower skewness (2D), narrower SD (3D) and higher mean attenuation (3D) are independent predictors for a favourable response to chemotherapy
Antunes et al. (2020)[41]	104	77	M/R	Rectum	Primary	Histopathology (AJCC CR: TRG 0)	MRI	Response	nCRT + resection	RF model were able to predict CR (AUC 0.71). None of the intensity-histogram features were predictive for CR
Badic et al. (2018)[86]	61	66	S/R	CRC	Primary	Death	CT	OS	Resection	NGTDM, GLRLM and GLSZM features (CE-CT and NCE-CT) had a significant association with OS. Histogram features were not predictive
Badic et al. (2019)[91]	64	56	S/R	CRC	Primary	Death, imaging	CT	a) OS b) PFS	Resection	In multivariate analysis, higher ratio 3D surface/volume was associated with OS and lower GLCM_entropy was associated with PFS
Bang et al. (2016)[42]	74	73	S/R	Rectum	Primary	a) Histopathology (AJCC, GR: TRG 0-1) b) Death c) Imaging, histopathology	PET	a) Response b) OS c) DFS	Mixed	Lower GrKurtosis was predictive of better DFS (multivariate). Multiple histogram and GLCM features were predictive of GR (univariate)
Beckers, Beets-Tan, et al. (2017)[92]	46	63	S/R	CRC	Liver parenchyma	Histopathology, imaging	CT	New CRLM	Mixed	No radiomics features were predictive of new CRLM
Beckers, Lambregts, et al. (2017)[93]	165	64	M/R	CRC	Liver parenchyma	Imaging	CT	New CRLM	No	Lower uniformity was predictive of early CRLM (≤6 months)
Beckers et al. (2018)[94]	70	71	S/R	CRC	CRLM + liver parenchyma	a) RECIST 1.1 (GR) b) Death	CT	a) Response b) OS	Mixed	Histogram features were not predictive for response. Feature ratios were associated with OS (univariate)
Bibault et al. (2018)[43]	95	52	M/R	Rectum	Primary	Histopathology (CR[i])	CT	Response	nCRT + resection	DNN was able to predict CR with AUC 0.72, compared to AUC 0.62 of the SVM
Boldrini et al. (2018)[44]	16	81	S/R	Rectum	Primary	MRI, DRE, endoscopy (CR)	MRI	Response	nCRT	No pre-treatment radiomics features were able to predict CR
Bulens et al. (2020)[45]	125	71	M/P	Rectum	Primary	Histopathology (Quirke GR: ypT0-1N0), Imaging (wait-and-see)	MRI	Response	nCRT + resection	Radiomics was not able to predict GR (AUC 0.49)
Bundschuh et al. (2014)[14]	27	ns	S/R	Rectum	Primary	a) Histopathology (GR: <10% residual tumour cells) b) Death c) Imaging	PET	a) Response b) OS c) PFS	nCRT + resection	Responders had lower COV. COV >0.37 was associated with better PFS and OS

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Caruso et al. (2018)[46]	8	75	S/R	Rectum	Primary	Histopathology (CR[f])	MRI	Response	nCRT + resection	CR was associated with higher energy, correlation and inverse different moment and lower contrast and entropy
Chee et al. (2017)[47]	95	62	S/R	Rectum	Primary	a) Histopathology (Dworak, GR: TRG 3+4) b) Imaging, death	CT	a) Response b) DFS	nCRT + resection	Homogeneous radiomics features were higher in GR and had better DFS (lower entropy, higher uniformity, lower SD)
Crimi et al. (2020)[48]	62	68	S/R	Rectum	Primary	Histopathology (Mandard CR: TRG 1)	MRI	Response	nCRT + resection	None of the texture features were significantly different in CR group compared to non-CR
Cui et al. (2018)[49]	186	60	S/R	Rectum	Primary	Histopathology (AJCC CR: TRG 1)	MRI	Response	nCRT + resection	Radiomics signature of 3 sequences combined (T2W, T1W, ADC) resulted in the best predictive performance (AUC 0.94)
Cusumano et al. (2018)[50]	198	ns	M/R	Rectum	Primary	Histopathology (Mandard CR: ypT0N0 or ypNx)	MRI	Response	nCRT + resection	Linear LR model (including cT, cN, skewness, entropy and maxFD) could predict CR (AUC 0.79)
Dai et al. (2020)[90]	701	56	S/R	Colon	Primary	Death, imaging	CT	a) OS b) RFS	ns	Clinical-radiomics model was able to predict OS (AUC 0.80) and RFS (AUC 0.79) and outperformed clinical model
de Cecco et al. (2015)[51]	15	60	S/P	Rectum	Primary	Histopathology (AJCC CR: TRG 4, GR[f])	MRI	Response	nCRT + resection	Lower kurtosis in CR
de Cecco et al. (2016)[52]	12	33	S/P	Rectum	Primary	Histopathology (AJCC CR: TRG 4, GR[f])	MRI	Response	nCRT + resection	Lower kurtosis in CR
Dercle et al. (2020)[106]	667	ns	M/R	CRC	CRLM	a/b) Death c) Imaging	CT	a) Response b) OS c) PFS	CTx + anti-EGFR therapy	Radiomics signatures was able to predict response to chemotherapy with cetuximab (AUC 0.80) better than response to chemotherapy alone (AUC 0.59)
Dinapoli et al. (2018)[53]	226	ns.	M/R	Rectum	Primary	a) Histopathology (CR: TRG 1) b-e) Imaging, death	MRI	a) Response b) OS c) MFS d) DFS e) LRC	nCRT + resection	Multivariable LR (including entropy and skewness) could predict CR (AUC 0.75), but was not predictive for survival
Dohan et al. (2020)[107]	270	68	M/R	CRC	CRLM	a) Imaging (RECIST 1.1 and mRECIST15) b) Death	CT	a) Response b) OS	CTx	None of the baseline radiomics features were predictive for response or OS
Ferrari et al. (2019)[54]	55	56	S/R	Rectum	Primary	Histopathology (Dworak-Rodel GR: TRG 3-4, CR: TRG 4)	MRI	Response	nCRT + resection	RF model predicts GR (accuracy 83%). CR was associated with higher GLCM energy and GR with higher histogram SD, min, max and SD of Shannon Entropy
Fu et al. (2020)[55]	43	65	S/R	Rectum	Primary	Histopathology (CR,GR[f]), imaging, colonoscopy	MRI	Response	Mixed	DL-based model to predict GR performed better than the handcrafted classifier (AUC 0.73 vs AUC 0.64)

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Ganeshan et al. (2007)[96]	27	74	S/R	CRC	Liver parenchyma	Death	CT	OS	Resection	Higher entropy and lower uniformity in patients with higher OS
Giannini et al. (2019)[56]	52	68	S/R	Rectum	Primary	Histopathology (Mandard GR: TRG 1-2)	MRI + PET	Response	Mixed	Predicting GR with LR with both PET+MRI features (AUC 0.86) yielded better performance than PET only (AUC 0.84) and MRI only (AUC 0.72)
Hamerla et al. (2019)[57]	169	75	S/R	Rectum	Primary	Histopathology (CR[f])	CT	Response	nCRT + resection	RF classifier was not able to predict CR
Jalil et al. (2017)[58]	56	61	S/R	Rectum	Primary	a) Death b-c) Histopathology, RECIST 1.1	MRI	a) OS b) DFS c) RFS	Mixed	Lower MPP in patients with higher OS/DFS/RFS and lower kurtosis was associated with better RFS
Lee et al. (2018)[95]	89	47	S/R	CRC	Liver parenchyma	a) Death b) Imaging	CT	a) OS b) New CRLM	CTx	None of the features were predictive for neither new CRLM nor OS
Li, Boimel, et al. (2019)[59]	84	61	S/R	Rectum	Primary	Death	PET + CT	RFS	nCRT	The CNN (C-index 0.64), CPH model (C-index 0.60) and RF model (C-index 0.58) were predictive for RFS
Li, Wang, et al. (2020)[60]	118	67	S/R	Rectum	Primary	Histopathology (AJCC GR: TRG 0-2)	MRI + CT	Response	nCRT + resection	CT + MRI combined model (AUC 0.91) outperformed the CT model (AUC 0.77), but was not significantly better than the T2W model (AUC 0.86)
Liang et al. (2016)[38]	53	64	S/R	CRC	CRLM	RECIST 1.1 (CR/GR)	MRI	Response	CTx	Lower mean, 1st, 10th, 50th, 90th and 99th percentile in GR (AUC 0.76-0.82)
Liang et al. (2019)[61]	108	74	S/R	Rectum	Primary	Histopathology, imaging	MRI	New CRLM	Mixed	Prediction of new CRLM with LR (AUC 0.87) and SVM (AUC 0.83) on T2W and DCE
Liu et al. (2017)[62]	26	ns	S/R	Rectum	Primary	Histopathology (GR: T-downstaging or N0)	MRI	Response	nCRT + resection	Lower energy variance and SdGa47 in responders
Liu, Li, et al. (2019)[63]	83	ns	S / R	Rectum	Primary	Death	PET + CT	OS	nCRT	The CPH model with collaborative clustering predicts OS (C-index 0.67)
Liu, Wen, et al. (2019)[64]	41	66	S / P	Rectum	Primary	Histopathology (Dworak CR: TRG 4)	MRI	Response	nCRT + resection	IVIM-DWI parameters GLCM correlation and difference variance were predictive for CR (AUC 0.986)
Lovinfosse et al. (2018)[65]	86	69	S/R	Rectum	Primary	a) Histopathology (Dworak, GR: TRG 3-4) b-e), imaging, death	PET	a) Response b) OS c) DFS d) DSS	nRT, nCTx, nCRT + resection	None of the parameters predicted CR. Homogeneity and coarseness were associated with DFS. Lower SUVmean, higher dissimilarity and lower NGTDM contrast were associated with OS
Lubner et al. (2015)[97]	77	56	S/R	CRC	CRLM	Death	CT	OS	CTx	None of the features was predictive for OS

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Meng, Zhang, Dong, et al. (2018)[67]	108	66	S/R	Rectum	Primary	Imaging, histopathology, death	MRI	DFS	nCRT + resection	Clinical-radiomics model was predictive for DFS (AUC 0.87)
Meng, Zhang, Zou, et al. (2018)[66]	59	66	S/R	Rectum	Primary	Histopathology (CR: TRG 1-2)	MRI	Response	nCRT + resection	Multivariate LR (including uniformity, entropy and energy) was able to predict CR (AUC 0.76).
Miles et al. (2009)[98]	48	67	S/R	CRC	Liver parenchyma	Death	CT	OS	Resection	Higher uniformity ratios were associated with higher OS
Nakajo et al. (2017)[87]	32	63	S/R	CRC	Primary	Imaging, death	PET	PFS	Resection	Higher 18F-FDG-IV, 18F-FDG-SZV and 1-F-FLT-Entropy were associated with PFS
Nardone et al. (2019)[68]	49	69	S / R	Rectum	Primary	Imaging	MRI	New CRLM	nCRT + resection	Lower GLCM_correlation on DWI predicts distant metastases <3 months (AUC 0.83)
Ng, Ganeshan, et al. (2013)[88]	55	48	S/R	CRC	Primary	Death	CT	OS	ns	Higher entropy, kurtosis and SD and lower uniformity, skewness were associated with better OS
Ng, Kozarski, et al. (2013)[89]	55	48	S/R	CRC	Primary	Death	CT	OS	ns	Higher entropy and lower uniformity were associated with better OS
Nie et al. (2016)[15]	48	63	S/R	Rectum	Primary	Histopathology (Ryan CR: TRG 0, GR: TRG 0-1)	MRI	Response	Mixed	Multiple parameters were associated with CR, but none with GR
Nougaret et al. (2016)[40]	31	71	S/R	Rectum	Primary	Histopathology (Dworak GR: TRG 3-4)	MRI	Response	nCRT + resection	None of the ADC histogram features were predictive for GR
Palmisano et al. (2018)[69]	21	86	S/R	Rectum	Primary	Histopathology (Rodel, GR: TRG 3-4)	MRI	Response	nCRT + resection	Ve kurtosis and skewness were predictive for GR
Park et al. (2020)[70]	78	60	S/R	Rectum	Primary	Histopathology (Dworak GR: TRG 2-4, CR: TRG 4), imaging	MRI	a) Response b) DFS	nCRT + resection	GR was associated with higher SD, sphericity and compacity and lower min. Lower GLRLM_LRLGE was associated with better DFS.
Petkovska et al. (2020)[71]	102	59	S/R	Rectum	Primary	Histopathology (AJCC CR: ypT0N0)	MRI	Response	nCRT + resection	The SVM was able to predict CR (accuracy 74%)
Petresc et al. (2020)[72]	67	66	S / R	Rectum	Primary	Histopathology (Ryan GR: TRG 1-2)	MRI	Response	nCRT + resection	Rad-score including first order and GLCM features was able to predict GR (AUC 0.80)
Rahmim et al. (2019)[104]	52	56	S/R	CRC	CRLM	Imaging	PET	a) OS b) PFS c) EFS	Mixed	EFS was associated with histogram uniformity; OS was associated with the AUC of the intensity-volume histogram (multivariable analysis)
Rao et al. (2014)[100]	19	71	S/R	CRC	Liver parenchyma	Imaging, histopathology	CT	New CRLM	No	None of the features were predictive for new metastases

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Rao et al. (2016)[99]	21	68	S/R	CRC	CRLM	Histopathology (Mandard, GR: TRG 1-2)	CT	Response	nCTx + resection	None of the features were predictive for GR
Ravanelli et al. (2019)[108]	43	51	S / R	CRC	CRLM	a/c) Imaging (RECIST 1.1, Choi's criteria) b) Death	CT	a) Response b) OS c) PFS	CTx	In CTx with bevacizumab, lower uniformity was associated with better PFS, OS and GR. No radiomics features were predictive in CTx without bevacizumab.
Shaish et al. (2020)[109]	132	65	M/R	Rectum	Primary + mesorectal compartment	Histopathology (AJCC CR: TRG 0, GR: TRG 0-1)	MRI	Response	nCRT + resection	Tumour- and mesorectal compartment-based radiomics were able to predict GR (AUC 0.80). Best performance for CR was based on the mesorectal compartment only (AUC 0.80)
Shayesteh et al. (2019)[73]	98	64	S/P	Rectum	Primary	Histopathology (AJCC GR: TRG 0-1)	MRI	Response	nCRT + resection	Predicting GR with an ensemble classifier model (AUC 0.94)
Shayesteh et al. (2020)[74]	67	66	S/R	Rectum	Primary	Histopathology (AJCC GR: TRG 0-1)	MRI	Response	nCRT + resection	Predicting GR using a naïve Bayesian network σ=0.5 model (AUC 0.85)
Shen et al. (2020)[75]	169	68	S/R	Rectum	Primary	Histopathology (Dworak GR: TRG 3-4, CR: TRG 4)	PET	Response	nCRT + resection	The RF model was able to predict CR (AUC 0.94).
Shi et al. (2019)[76]	45	69	S/R	Rectum	Primary	Histopathology (Ryan's CR: TRG 0, GR: TRG 0-1)	MRI	Response	nCRT + resection	T2W+ DCE+DWI –based radiomics model was able to predict GR (AUC 0.91) and CR (AUC 0.80).
Shu, Fang, et al. (2019)[77]	136	62	S/P	Rectum	Primary	Histopathology (Dworak CR: TRG 4)	MRI	Response	nCRT + resection	Lower skewness and energy and higher entropy in CR.
Shur et al. (2019)[105]	102	65	S/R	CRC	CRLM	Imaging, histopathology	MRI + CT	DFS	nCTx + resection	Higher GLSZM_small area emphasis and lower pixel minimum on EOB-MRI were associated with better DFS. Radiomics features on T1W and CE-CT were not predictive
Simpson et al. (2017)[101]	198	60	S/R	CRC	CRLM + liver parenchyma	a) Death b) Imaging	CT	a) OS b) HDFS	Mixed	Higher future liver remnant signal (entropy + energy) and higher tumour texture signal (correlation + contrast) were associated with better OS
Taghavi et al. (2020)[103]	91	60	M/R	CRC	Liver parenchyma	Imaging, histopathology	CT	New CRLM	Mixed	Radiomics model (AUC 0.86) outperformed clinical model (AIC 0.71) in prediction of new CRLM
Tang et al. (2019)[39]	222	62	S/R	Rectum	Primary	Histopathology (GR: ypT0-1N0)	MRI	Response	nCRT + resection	3 DWI features were significantly different between GR and non-GR
Van Helden et al. (2018)[110]	99	66	S/R	CRC	All metastatic lesions	a) RECIST b) Death c) RECIST, death	PET	a) Response b) OS c) PFS	CTx (palliative)	Lower entropy in responders in the first line treatment group (AUC 0.74). None of the features were associated with PFS

Reference	#Pt[a]	%M[b]	Type[c]	Primary	ROI[d]	Reference standard	Mod.[e]	Outcome	Intervention	Main results/conclusion
Van Griethuysen et al. (2020)[78]	133	69	M/R	Rectum	Primary	Histopathology (Mandard CR: TRG 1, GR: TRG 1-2), imaging (wait and see)	MRI	Response	nCRT + resection	Radiomics model was able to predict GR, performance was reader dependent (AUC 0.69-0.79)
Vandendorpe et al. (2019)[79]	121	58	M/R	Rectum	Primary	Downstaging (cTNM vs ypTNM)	CT	Response	nCRT + resection	Radiomics model was able to predict GR (AUC 0.67).
Wang et al. (2019)[80]	411	73	S/R	Rectum	Primary	Histopathology (CR[i])	CT	a) OS b) DFS c) RFS d) MFS	nCRT + resection	Clinical-radiomics model outperformed radiomics model alone to predict OS (AUC 0.73 vs 0.66) and RFS (AUC 0.65 vs 0.56).
Yang et al. (2020)[81]	89	67	S/R	Rectum	Primary	Histopathology (AJCC GR: TRG 0-1)	MRI	Response	nCRT + resection	Clinical-radiomics RF classifier was able to predict GR (AUC 0.83).
Yi et al. (2019)[82]	134	60	S/R	Rectum	Primary	Histopathology (Dworak/Rodel GR: TRG 3-4, CR: TRG 4)	MRI	Response	nCRT + resection	Clinical-radiomics SVM was able to predict GR (AUC 0.90) and CR (AUC 0.87)
Yuan et al. (2020)[83]	91	54	S/R	Rectum	Primary	Histopathology (CAP CR: TRG 0, GR: TRG 0-1)	CT	Response	nCRT + resection	RF classifier was able to predict CR (ACC 0.84). LR model was able to predict GR (AUC 0.76)
Zhang et al. (2018)[102]	26	54	S/R	CRC	CRLM	Imaging (size reduction)	MRI	Response	CTx	Higher variance and lower GLCM energy in GR
Zhou et al. (2019)[84]	425	71	S/R	Rectum	Primary	Histopathology (AJCC non-GR: TRG 3)	MRI	Response	nCRT + resection	Multiparametric model was able to predict GR (AUC 0.77). Clinical-radiomics model had a lower performance compared to radiomics model alone
Zou et al. (2018)[85]	83	69	S/R	Rectum	Primary	Histopathology (Mandard, CR: TRG 1, GR: TRG 1+2)	MRI	Response	nCRT + resection	No significant differences between responders and non-responders

Abbreviations: *a* number of patients, *b* percentage of male patients, *c* type of study, *d* region of interest, *e* modality. *ADC* apparent diffusion coefficient, *AJCC* American Joint Committee on Cancer, *AUC* area under the curve, *CE-CT* contrast enhanced computed tomography, *CI* confidence interval, *CNN* convolutional neural network, *COV* coefficient of variation, *CPH* Cox proportional hazard, *CR* complete response, *CRC* colorectal cancer, *CRLM* colorectal liver metastases, *CT* computed tomography, *CTx* chemotherapy, *DCE* dynamic contrast enhanced, *DFS* disease free survival, *DL* deep learning, *DNN* deep neural network, *DRE* digital rectal examination, *DSS* disease specific survival, *DWI* diffusion weighted imaging, *EFS* event free survival, *EGFR* anti-epidermal growth factor receptor, *EOB* gadoxetic acid enhanced, *FD* fractal dimension, *FDG* fludeoxyglucose, *FLT* fluorothymidine, *GCLM* gray level cooccurence matrix, *GLRLM* gray level run length matrix, *GLSZM* gray level size zone matrix, *GR* good response, *HDFS* hepatic disease free survival, *IVIM* intravoxel incoherent motion, *LRLGE* long run low gray level emphasis, *LR* logistic regression, *LRC* locoregional control, *M* multicentre study, *MFS* metastatic free survival, *MPP* mean positive pixels, *MRI* magnetic resonance imaging, *NCE-CT* non contrast enhanced computed tomography, *nCRT* neoadjuvant chemoradiotherapy, *nCTx* neoadjuvant chemotherapy, *NGTDM* neighbouring gray tone difference matrix, *nRT* neoadjuvant radiotherapy, *ns* not specified, *OS* overall survival, *P* prospective study, *PET* positron emission tomography, *PFS* progression free survival, *R* retrospective study, *RECIST* response evaluation criteria in solid tumours, *RF* random forest, *RFS* recurrence free survival, *S* single centre study, *SD* standard deviation, *SUV* standardized uptake value, *SVM* support vector machine, *TNM* tumour node metastases classification, *TRG* tumour regression grade, *Ve* volume of extravascular extracellular space per unit volume of tissue.

Reference	Mod.[a]	Segmentation	#Ob.[b]	Software	Pre.[c]	Filter	[N] Features	Feature selection	Classification	Validation	RQS
Ahn et al. (2016)[13]	CE-CT	Manual (2D/3D)	1	In-house (C++)	--	--	[44] Shape (4), Histogram (13), GLCM (5)	T-test, MwU, X2, Fisher exact	Multivariable LR	Split set	19%
Antunes et al. (2020)[41]	T2W	Manual (2D/3D)	2	MATLAB	Yes	Gabor wavelet, Laws	[764] Histogram, Haralick co-occurrence, Gradient (organization) operators	Wilcoxon, mRMR, LASSO	RF	External	36%
Badic et al. (2018)[86]	CE-/NCE-CT	Semiautomatic (3D)	1	In-house	Yes	--	[99] Histogram (6), GLCM (16), NGTDM, GLRLM, GLSZM (11)	Spearman	Kaplan-Meier	--	8%
Badic et al. (2019)[91]	CE-CT	Semiautomatic (3D)	1	In-house	Yes	--	[33] Histogram (6), GLCM (16), GLSZM (11)	--	Cox regression, Kaplan-Meier	--	0%
Bang et al. (2016)[42]	PET	Manual (3D)	ns	MaZda	--	--	[50] Histogram (9), absolute gradient (6), GLCM(11), GLRLM(5)	Univariable Cox regression	Multivariable Cox regression, Kaplan-Meier	--	0%
Beckers, Beets-Tan, et al. (2017)[92]	CE-CT	Manual (2D)	1	MATLAB	Yes	LoG	[12] Histogram (3)	--	T-test, MwU	--	0%
Beckers, Lambregts, et al. (2017)[93]	CE-CT	Manual (3D)	1	MATLAB	--	LoG	[12] Histogram (3)	Univariable multilevel LR	Multivariable multilevel LR	--	0%
Beckers et al. (2018)[94]	CE-CT	Manual (3D)	1	In-house (Python)	Yes	LoG	[18] Histogram (2)	--	Multivariable Cox regression	--	0%
Bibault et al. (2018)[43]	CE-CT	Manual (3D)	2	IBEX	--	--	[1683] Shape, Intensity Direct, GLCM (28), NID (28), GLRLM (25)	Wilcoxon, ICC	DNN Classifier Custom Estimator, SVM	K-fold cross	11%
Boldrini et al. (2018)[44]	T2*W, T1W	Manual (3D)	2	Moddicom	--	--	[318] Shape, Histogram, GLCM	--	Wilcoxon-Mann-Whitney	--	0%
Bulens et al. (2020)[45]	T2W, DWI, ADC	Manual (3D)	4	Python	Yes	Wavelet, Gabor	[2131] Shape, intensity, texture	Principle component analysis	LR with LASSO	External	22%
Bundschuh et al. (2014)[14]	PET-CT	Manual (3D)	ns	ns	--	--	[3] Histogram	--	ROC curves, Kaplan-Meier	--	0%
Caruso et al. (2018)[46]	T2W	Manual (2D)	2	ns	--	--	[14] GLCM	--	Linear regression	--	0%
Chee et al. (2017)[47]	CE-CT	Manual (2D)	2	In-house	Yes	LoG	[25] Histogram (5)	--	Multivariable Cox regression	--	8%
Crimi et al. (2020)[48]	T2W	Manual (3D)	2	PMOD	--	--	[33] Histogram, GLCM, GLRLM	--	T-test, ROC curve	--	0%

Reference	Mod.[a]	Segmentation	#Ob.[b]	Software	Pre.[c]	Filter	[N] Features	Feature selection	Classification	Validation	RQS
Cui et al. (2018)[49]	T2W, DWI, DCE	Manual (3D)	2	Analysis KIT	Yes	--	[1188] Histogram (42), GLCM (154), GLRLM (180), GLSZM (11), Form factor (9)	Univariable LR, LASSO, Pearson	Multivariable LR	Split set	44%
Cusumano et al. (2018)[50]	T2W	Manual (3D)	4	Moddicom	Yes	LoG	[126] Histogram	Wilcoxon	LR	External	36%
Dai et al. (2020)[90]	NCE-CT	Semiautomatic (2D)	2	ns	--	Yes	[647] Shape, histogram, GLCM, GLRLM	--	LASSO Cox regression	K-fold cross	19%
de Cecco et al. (2015)[51]	T2W	Manual (2D)	1	TexRAD	--	LoG	[20] Histogram (4)	--	MwU	--	0%
de Cecco et al. (2016)[52]	T2W, DWI	Manual (2D)	1	TexRAD	--	Yes	[3] Histogram	--	MwU	--	0%
Dercle et al. (2020)[106]	CE-CT	Semiautomatic (2D/3D)	4	MATLAB	--	Wavelet, Gabor, LoG, Laws	[3499] Shape (16), Histogram (4), Edge frequency features (3), fractal dimension (2), Sigmoid functions (3), NGTDM (5), Run-Length (5), GLCM (16)	Reproducibility analysis, redundancy analysis, feature ranking (RELIEF, Fisher, X2, mRMR, t-test, Wilcoxon score, univariable model)	RF, SVM, K-nearest neighbour, Naive Bayes, Bagging, LASSO, CNN, Cox regression	Split set	47%
Dinapoli et al. (2018)[53]	T2W	Manual (3D)	4	Moddicom	--	LoG	[1200] Histogram (3)	MwU	Multivariable LR, Cox regression	External	42%
Dohan et al. (2020)[107]	CE-CT	Manual (2D)	3	TexRAD	--	LoG	[36] Histogram (6)	Pearson, t-test	Cox regression, linear combination	External	39%
Ferrari et al. (2019)[54]	T2W	Manual (2D/3D)	1	In-house (Python)	Yes	LoG	[855] Histogram (4), Haralick GLCM (5), Shannon Entropy (4)	T-test	RF, ROC curve	Split set	36%
Fu et al. (2020)[55]	DWI	Manual (2D/3D)	1	PyRadio-mics	--	--	[1473] Shape (14), histogram (18), second order (73)	Feature variance	LASSO penalized LR, CNN	K-fold cross	6%
Ganeshan et al. (2007)[96]	CE-CT	Manual (2D)	ns	MATLAB	Yes	LoG	[39] Histogram (3)		Linear regression	--	0%
Giannini et al. (2019)[56]	PET-CT, T2W, DWI	(Semi-) automatic (2D)	1	MATLAB	Yes	--	[78] Histogram (5), GLCM (21)	Pearson	MwU, stepwise LR	--	0%

Reference	Mod.[a]	Segmentation	#Ob.[b]	Software	Pre.[c]	Filter	[N] Features	Feature selection	Classification	Validation	RQS
Hamerla et al. (2019)[57]	NCE-CT	Manual (3D)	2	In-house (Python)	--	LoG, wavelet, Laws	[1819] Shape, histogram, second order	Recursive feature elimination	RF	K-fold cross	6%
Jalil et al. (2017)[58]	T2W	Manual (2D)	1	TexRAD	--	LoG	[30] Histogram (6)	Kaplan-Meier	Multivariable Cox regression	--	0%
Lee et al. (2018)[95]	CE-CT	Manual (2D)	1	TexRAD	--	LoG	[24] Histogram (4)	MwU, LASSO	Cox regression	LOOCV	6%
Li, Boimel, et al. (2019)[59]	CT, PET	Manual (2D/3D)	--	Python	--	Wavelet	ns	Principle component analysis	Deep CNN, RF survival, Cox regression	Split set	0%
Li, Wang, et al. (2020)[60]	CE-CT, T2W, DCE, ADC	Manual (3D)	2	In-house (AIK)	Yes	--	[1584] Shape (9), histogram (42), second order (334), GLZSM (11)	mRMR, LASSO	LR	Split set	36%
Liang et al. (2016)[38]	ADC, DCE	Manual (2D)	1	MaZda	Yes	--	[27] Histogram (9)	--	T-test, MwU	--	11%
Liang et al. (2019)[61]	T2W, DCE	Manual (3D)	3	Anaconda3	--	Wavelet, exponential, Squareroot	[1029] Shape, histogram, GLCM, GLRLM, GLSZM	LASSO	SVM, LR	K-fold cross	11%
Liu et al. (2017)[62]	ADC	Manual (2D)	1	MATLAB	Yes	Gabor-wavelet	[19] Histogram, GLCM, GLGCM, Gabor/wavelet	T-test, MwU	Multivariable LR	--	0%
Liu, Li, et al. (2019)[63]	CT, PET	Manual	>1	R package	Yes	Wavelet	[2503] Shape, texture	Collaborative clustering	RF survival, Cox regression, Cox with LASSO	K-fold cross	0%
Liu, Wen, et al. (2019)[64]	IVIM-DWI	Manual (3D)	1	MaZda	Yes	--	[16] Histogram (5), GLCM (11)	MwU	Multivariable LR	--	0%
Lovinfosse et al. (2018)[65]	PET	Automatic (3D)	1	In-house (Python)	Yes	--	[16] Histogram (6), GLCM (6), NGTDM (2), ISZ (2)	Spearman, univariable Cox regression	Multivariable Cox regression	--	14%
Lubner et al. (2015)[97]	CE-CT	Manual (2D)	1	TexRAD	--	LoG	[30] Histogram (6)	--	Linear regression, Cox regression	--	6%
Meng, Zhang, Dong, et al. (2018)[67]	T2W	Manual (3D)	2	MATLAB	Yes	Wavelet	[970] Shape, intensity, texture	LASSO Cox regression	Multivariable Cox regression	Split set	39%
Meng, Zhang, Zou, et al. (2018)[66]	T2W	Manual (2D)	2	Omni-Kinetics	--	--	[7] Histogram	Wilcoxon signed-rank, MwU	Multivariable LR	--	3%
Miles et al.	CE-CT	Manual (2D)	1	MATLAB	Yes	LoG	[8] Histogram (2)	--	Kaplan-Meier, ROC	--	8%

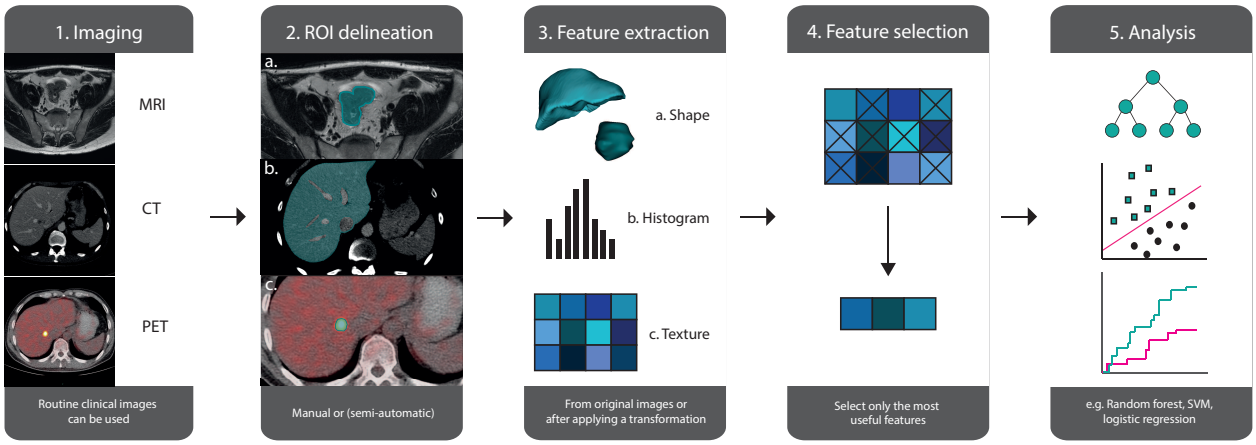
(2009)[98]									curves		
Reference	**Mod.[a]**	**Segmentation**	**#Ob.[b]**	**Software**	**Pre.[c]**	**Filter**	**[N] Features**	**Feature selection**	**Classification**	**Validation**	**RQS**
Nakajo et al. (2017)[87]	PET	Manual (3D)	1	In-house (Python)	Yes	Gaussian	[6] ns	ROC curves	Cox regression	--	0%
Nardone et al. (2019)[68]	DWI, ADC	Manual (3D)	2	LifeX	--	--	[ns] Shape, histogram, GLCM	Pearson, univariable LR	Multivariable LR	Split set	14%
Ng, Ganeshan, et al. (2013)[88]	CE-CT	Semiautomatic (3D)	1	TexRAD	Yes	LoG	[25] Histogram (5)	MwU	Multivariable Cox regression	--	0%
Ng, Kozarski, et al. (2013)[89]	CE-CT	Semiautomatic (2D/3D)	1	TexRAD	Yes	LoG	[10] Histogram (2)	--	ROC curves, Kaplan-Meier	--	3%
Nie et al. (2016)[15]	T1W, T2W, DWI, DCE	Manual (3D)	1	MATLAB	--	--	[103] Shape (5), histogram (12), GLCM (18)	T-test, Wilcoxon-Mann-Whitney	ANN	K-fold cross	11%
Nougaret et al. (2016)[40]	DWI	Manual (2D/3D)	2	Olea Sphere	--	--	[9] Histogram	--	T-test, ROC curves	LOOCV	6%
Palmisano et al. (2018)[69]	MRI	Manual (3D)	1	Olea Sphere	--	--	[7] Histogram	--	MwU	--	6%
Park et al. (2020)[70]	T2W	Manual (3D)	1	LifeX	--	--	[44] Shape (3), histogram (9), GLCM (7), GLRLM 11), NGLDM (3), GLZLM (11)	T-test	Multivariable LR, Cox regression	--	0%
Petkovska et al. (2020)[71]	T2W	Manual (3D)	3	CERR	Yes	LoG, Gabor, Sobel	[108] Shape (7), histogram (20), GLCM (26), GLSZM (15)	Wilcoxon, LASSO	Weighted sum, SVM	Split set	22%
Petresc et al. (2020)[72]	T2W	Manual (3D)	3	PyRadio-mics	Yes	Wavelet, LoG	[960] Shape (14), histogram (18), GLSZM (16), GLRLM (16), GLDM (14)	MwU, Spearman, LASSO	Multivariable LR	Split set	33%
Rahmim et al. (2019)[104]	PET	Automatic (3D)	NA	MATLAB	--	--	[41] PET-specific (5), gETU (10), histogram (19), IVH (7)	Spearman	Cox regression	--	0%
Rao et al. (2014)[100]	CE-CT	Manual (3D)	1	MATLAB	--	LoG	[12] Histogram (3)	--	T-test, MwU	--	0%
Rao et al. (2016)[99]	CE-CT	Manual (3D)	1	MATLAB	--	LoG	[12] Histogram (3)	Univariable LR	Multivariable LR	--	0%
Ravanelli et al. (2019)[108]	CE-CT	Manual (2D)	1	MATLAB	--	LoG	[21] Histogram (3)	T-test, X2	Multivariable LR, Cox regression	LOOCV	0%

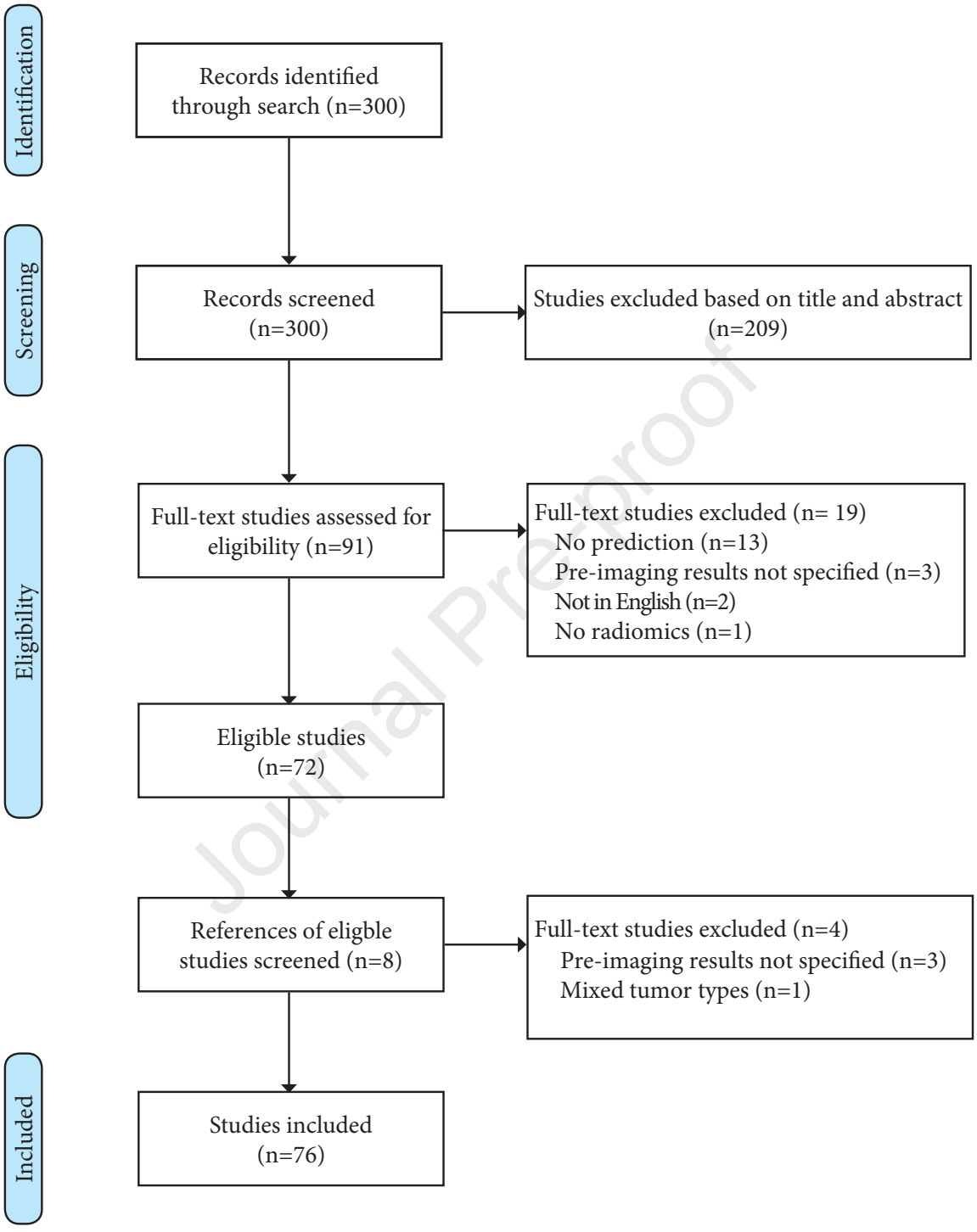
Reference	Mod.[a]	Segmentation	#Ob.[b]	Software	Pre.[c]	Filter	[N] Features	Feature selection	Classification	Validation	RQS
Shaish et al. (2020)[109]	T2W	Manual (3D)	1	PyRadio-mics	--	Wavelet, square, square root, logarithm, exponential, local binary pattern	[1595] Shape (16), histogram (19), GLCM (24), GLRLM (16), GLSZM (5), GLDM (14)	Pearson, LR, qualitative assessing distribution	LR	Split set	0%
Shayesteh et al. (2019)[73]	T2W	Manual (3D)	2	IBEX	Yes	LoG	[160] Shape (17), histogram (9), intensity direct (19), NID (5), COM (19), GLRLM (11)	Pearson, X2	LR, SVM, BN, NN, KNN	Split set	31%
Shayesteh et al. (2020)[74]	T1W, T2W	Manual (3D)	2	IBEX	Yes	LoG	[160] Shape (17), histogram (9), intensity direct (19), NID (5), COM (19), CLRLM (11)	Univariable LR	BN, naïve BN, RF, Ada boost M1, iterative classifier optimizer, logit boost, randomized filtered classifier, random sub space, K logistic model tree	K-fold cross	14%
Shen et al. (2020)[75]	PET	Semiautomatic (3D)	1	In-house	Yes	Gaussian	[68] Histogram (3), GLCM (21), NGLDM (5), GLRLM (11), GLSZM (11)	ROC curve	RF, LR	K-fold cross	11%
Shi et al. (2019)[76]	T2W, DWI, DCE	Manual (3D)	1	MATLAB, TenserFlow	Yes	--	[96] Histogram (12), Haralick's GLCM (18)	ANN, L2 regularization	ANN, LR, SVM, CNN	K-fold cross	6%
Shu, Fang, et al. (2019)[77]	T2W	Manual (2D)	2	AK	--	--	[6] Histogram	T-test	Multifactor LR	--	8%
Shur et al. (2019)[105]	CE-CT, T1W, EOB	Manual (2D)	1	PyRadio-mics	--	--	[104] Shape (12), histogram (18), GLCM (23), GLDM (14), GLRLM (16), GLSZM (16), NGTDM (5)	Pearson	LASSO regularized Cox regression, weighted sum	K-fold cross	22%
Simpson et al. (2017)[101]	CE-CT	Semiautomatic (3D)	ns	In-house	--	--	[5] GLCM	Kaplan-Meier	Multivariable Cox regression	--	6%

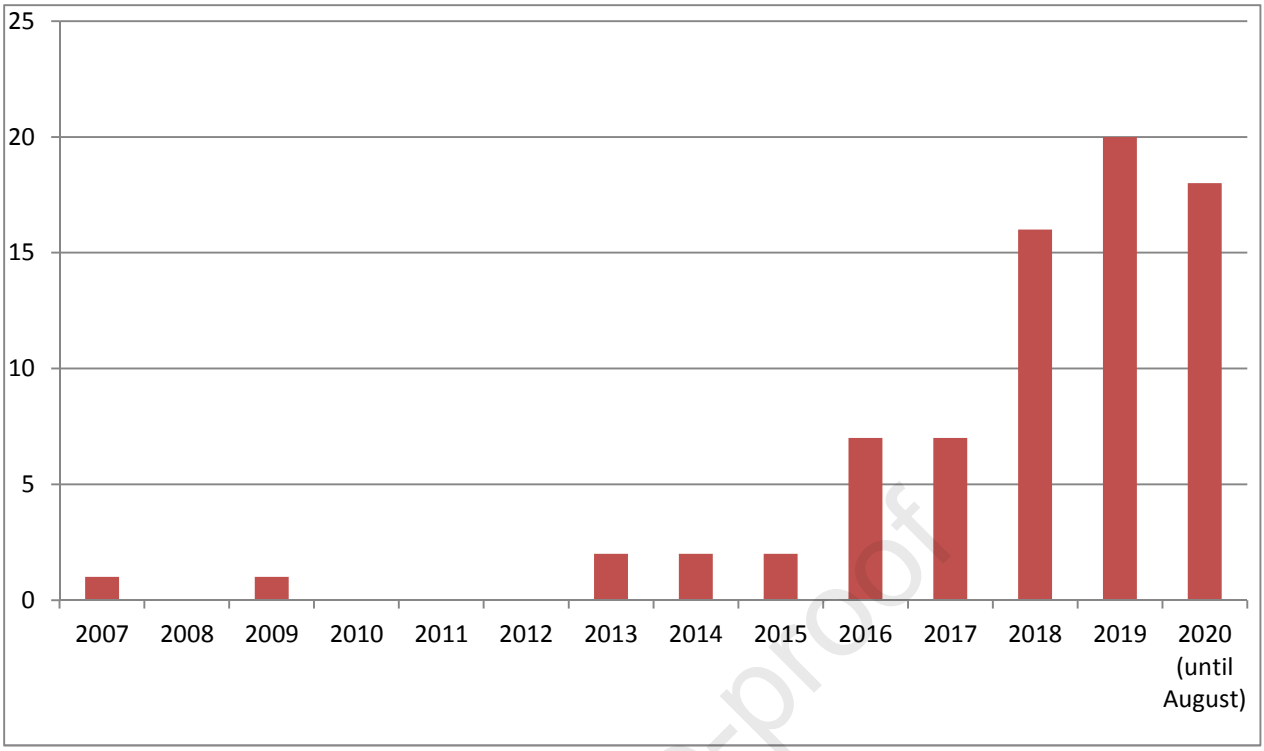
Reference	Mod.[a]	Segmentation	#Ob.[b]	Software	Pre.[c]	Filter	[N] Features	Feature selection	Classification	Validation	RQS
Taghavi et al. (2020)[103]	CE-CT	Automatic + Manual (3D)	2	PyRadio-mics	Yes	LoG, wavelet, exponential, gradient, square, √, logarithm	[1767] Histogram (18), GLCM (24), GLRLM (16), GLSZM (16), NGTDM (5), GLDM (14)	Kruskal-Wallis, Pearson	Wrapper method, RF, Bayesian hyperparameter optimization	Split set	31%
Tang et al. (2019)[39]	T2W, DWI	Manual (ns)	2	MATLAB	--	Wavelet	[1122] Histogram (7), GLCM (9), GLRLM (13), GLSZM (13), NGTDM(5)	DID regression	E-net penalized LR	Split set	47%
Van Helden et al. (2018)[110]	PET	Semiautomatic (3D)	ns	ns	Yes	--	[11] Histogram	Pearson, Spearman	Multivariable Cox regression, Kaplan Meier	--	42%
Van Griethuysen et al. (2020)[78]	T2W, DWI	Semiautomatic + Manual (3D)	4	PyRadio-mics	Yes	Log, exponent, gradient, logarithm	[2505] Shape (13), Histogram (17), GLCM (22), GLRLM (16), GLSZM (16), GLDM (13), NGTDM (5)	ICC, Kruskal-Wallis	mRMR, LR, Wilcoxon rank-sum	External	6%
Vandendorpe et al. (2019)[79]	CE-CT	Manual (2D)	1	TexRAD	--	LoG	[36] Histogram (6)	T-test, penalized LASSO LR, Ridge, ELASTIC-NET	Weighted linear combination, univariable LR	Split set	22%
Wang et al. (2019)[80]	NE-CT	Manual (3D)	2	In-house	--	--	[271] Shape (3), histogram (19), GLCM (20), GLRLM (10), Wavelet (200), FD (18)	Spearman	NMF, multivariable modelling strategy, Cox regression	Split set	31%
Yang et al. (2020)[81]	ADC	Manual (2D)	3	PyRadio-mics	--	Gabor wavelet	[133] Histogram, GLGCM, Intensity-size-zone structure	Spearman, Gini importance	RF	Split set	31%
Yi et al. (2019)[82]	T2W	Manual (2D)	2	MaZda	--	Wavelet	[340] Histogram, gradient, GLRLM, GLCM	LASSO	RF, SVM, linear combination	Split set	31%
Yuan et al. (2020)[83]	NCE-CT	Manual (3D)	2	In-house (C++)	Yes	LoG, wavelet, Laws	[929] Shape, histogram, GLCM, GLSZM, GLRLM, NGTDM, FD	Pearson	LR, RF	Split set	9%
Zhang et al. (2018)[102]	T2W	Manual (2D)	2	MATLAB	--	--	[10] Histogram, GLCM	T-test, MwU	Multivariable LR	--	0%
Zhou et al. (2019)[84]	T1W, T2W, DCE, ADC	Manual (2D)	2	MATLAB	--	LoG	[2424] Histogram (13), GLCM (22)	Wilcoxon rank-sum, Spearman, LASSO	Multivariable LR	Split set	36%

Zou et al. (2018)[85]	DCE	Manual (3D)	1	MATLAB	Yes	--	[7] Histogram	--	T-test, linear combination	--	3%

Abbreviations: *a* modality, *b* number of observers, *c* pre-processing method applied. *ADC* apparent diffusion coefficient, *ANN* artificial neural network, *BN* Bayesian network, *CE-CT* contrast enhanced computed tomography, *CNN* convolutional neural network, *CT* computed tomography, *COM* cooccurence matrix, *DCE* dynamic contrast enhanced, *DID* difference in difference, *DL* deep learning, *DNN* deep neural network, *DWI* diffusion weighted imaging, *EOB* gadoxetic acid enhanced, *FD* fractal dimension, *gETU* generalized effective total uptake, *GLCM* gray level cooccurence matrix, *GLDM* gray level dependence matrix, *GLGCM* gray level gradient cooccurence matrix, *GLRLM* gray level run length matrix, *GLSZM* gray level size zone matrix, *GLZLM* gray level zone length matrix, *ICC* intraclass correlation coefficient, *ISZ* intensity size zone, *IVH* intensity volume histogram, *IVIM* intravoxel incoherent motion, *KNN* k nearest neighbour, *LASSO* least absolute shrinkage and selection operator, *LOOCV* leave one out cross validation, *LRLGE* long run low gray level emphasis, *LR* logistic regression, *LoG* Laplacian of Gaussian, *mRMR* minimum redundancy maximum relevance, *MwU* Mann-Whitney U test, *NCE-CT* non contrast enhanced computed tomography, *NGLDM* neighbourhood gray level difference matrix, *NGTDM* neighbouring gray tone difference matrix, *NID* neighbour intensity direct, *NMF* non-negative matrix factorization, *NN* neural network, *ns* not specified, *PET* positron emission tomography, *RF* random forest, *ROC* receiver operating characteristic, *SVM* support vector machine, *X2* Chi squared test.







Supplementary table C1 Complete overview of the individual study results

Reference	Outcome	Modality/ feature	Parameter	Mean responder - nonresponder	HR (95% CI)	P-value	Outcome
Ahn et al. (2016)[13]	Response (RECIST)	CT	Mean attenuation 2D *	78.59 – 65.98 HU		< 0.001	Higher Mean attenuation 2D in responders
			Skewness 2D *	0.02 – 0.33		0.001	Lower Skewness 2D in responders
			Homogeneity 2D *	0.01 – 0.02		0.004	Lower Homogeneity 2D in responders
			5th percentile 2D *	44.34 – 30.75 HU		< 0.001	Higher 5th percentile 2D in responders
			10th percentile 2D *	50.5 – 37.06 HU		< 0.001	Higher 10th percentile 2D in responders
			25th percentile 2D *	61.98 – 48.58 HU		< 0.001	Higher 25th percentile 2D in responders
			50th percentile 2D *	76.52 – 63.34 HU		< 0.001	Higher 50th percentile 2D in responders
			75th percentile 2D *	92.65 – 80.17 HU		< 0.001	Higher 75th percentile 2D in responders
			90th percentile 2D *	106.78 – 96.20 HU		0.001	Higher 90th percentile 2D in responders
			95th percentile 2D *	114.89 – 105.37 HU		0.003	Higher 95th percentile 2D in responders
			GLCM moments 2D *	0.65 – 0.52		0.001	Higher GLCM moments 2D in responders
			Mean attenuation 3D *	82.94 – 71.76 HU		< 0.001	Higher Mean attenuation 3D in responders
			SD attenuation 3D *	21.69 – 25.06 HU		0.001	Narrower SD attenuation 3D in responders
			Entropy 3D *	4.38 – 4.49		0.017	Lower Entropy 3D in responders
			Homogeneity 3D *	0.01 – 0.02		0.002	Lower Homogeneity 3D in responders
			5th percentile 3D *	47.65 – 33.87 HU		< 0.001	Higher 5th percentile 3D in responders
			10th percentile 3D *	54.38 – 40.98 HU		< 0.001	Higher 10th percentile 3D in responders
			25th percentile 3D *	66.52 – 53.74 HU		< 0.001	Higher 25th percentile 3D in responders
			50th percentile 3D *	81.60 – 69.64 HU		< 0.001	Higher 50th percentile 3D in responders
			75th percentile 3D *	96.96 – 87.08 HU		0.001	Higher 75th percentile 3D in responders
			90th percentile 3D *	110.43 – 102.26 HU		0.005	Higher 90th percentile 3D in responders
			95th percentile 3D *	117.85 – 111.45 HU		0.020	Higher 95th percentile 3D in responders
			GLCM moments 3D *	0.71 – 0.58		0.001	Higher GLCM moments 3D in responders
			Skewness 2D †	0.02 – 0.33		0.025	Lower Skewness 2D in responders
			Mean attenuation 3D †	82.94 – 71.76 HU		0.017	Higher Mean attenuation 3D in responders
			SD attenuation 3D †	21.69 – 25.06 HU		0.002	Narrower SD attenuation 3D in responders
			SD attenuation 2D, Kurtosis 2D, Entropy 2D, Discrete compactness 2D, GLCM ASM 2D, GLCM IDM 2D, GLCM Contrast 2D, GLCM Entropy 2D, Skewness 3D, Kurtosis 3D, Sphericity 3D, Discrete compactness 3D, GLCM ASM 3D, GLCM IDM 3D, GLCM Contrast 3D, GLCM Entropy 3D			ns	
Antunes et al. (2020)[41]	CR	MRI T2W; 2D	Random forest model		AUC 0.71		Accuracy 71% to predict response
		Features included in model	Skewness-Laws-W5R5, Kurtosis-CoLIAGe_informationMetric1_WS3, Skewness-CoLIAGe_Correlation, Kurtosis-Harralick_SumEntropy_WS9				
		MRI T2W; 3D	Random forest model		AUC 0.63		Accuracy 64% to predict response
		Features	Skewness-Laws-W5R5, Kurtosis-				

		included in model	CoLIAGe_informationMetric1_WS3, Skewness-CoLIAGe_Correlation, Kurtosis-Harralick_SumEntropy_WS9			
Badic et al. (2018)[86]	OS	nCE-CT	NCE-CT SZLGEr		0.008	NCE-CT SZLGEr < 0.0004 has better survival
			NCE-CT SZSNUr		0.001	NCE-CT ZSNUr>7698.3732 better survival
			NCE-CT SZHGEl		<0.0006	
			NCE-CT ZSNUr		<0.0006	
			NCE-CT GLNUl		<0.0006	
			NCE-CT GLNUr		<0.0006	
			NCE-CT HGLZEe		<0.0006	
	OS	CE-CT	CE-CT SZLGEr		0.002	CECT SZLGEr <0.0003 has better survival
			CE-CT SZSNUr		0.01	
			CE-CT LGLZEl		<0.0006	
			CE-CT LGREl		<0.0006	
Badic et al. (2019)[91]	OS	Univariate	Ratio 3D Surface/Volume	3.3 (1.2-7.3)	0.022	Lower ratio 3D/Surface → higher survival
			Flatness	0.012 (0.0001-0.6)	0.028	Higher flatness → higher survival
			IDMr	270.1 (1.3-55079)	0.040	Lower IDMr → higher survival
			Inverse difference (IDr)	655.3 (1.1-393927)	0.048	Lower IDr → higher survival
			mean, variance, skewness, kurtosis, energy, HIST_entropy, ASM, GLCM_contrast, correlation, Sum Average, Sum Variance, GLCM_Sum_Entropy (SENT), GLCM_entropy, Difference variance, GLCM_difference entropy, Information correlation, autocorrelation, dissimilarity, Cluster prominence, Joint maximum, Small Zone Emphasis, Large Zone Emphasis, Low grey level zone emphasis, high grey level zone emphasis, small zone low grey level emphasis, small zone high grey level emphasis, large zone low grey level emphasis, large zone high grey emphasis, gray-level non-uniformity, zone size non-uniformity, zon size percentage		ns	
		Multivariable Cox	Ratio 3D Surface/Volume	2.8 (1.0-7.3)	0.04	
			Ratio 3Dsurface/volume, ALDH1A1	8.4 (3.4-20.6)	0.001	
	PFS	Univariate	Flatness	0.0003 (0.0-0.3)	0.023	Higher Flatness → higher PFS
			GLCM_SENT	0.5 (0.31-0.91)	0.021	Higher GLCM_SENT → higher PFS
			GLCM_entropy	0.05 (0.006-0.35)	0.003	Higher GLCM_entropy → higher PFS
			Gray level non-uniformity (GLNU)	0.07 (0.007-0.77)	0.030	Higher GLNU → higher PFS
			mean, variance, skewness, kurtosis, energy, HIST_entropy, ASM, GLCM_contrast, correlation, Sum Average, Sum Variance, Difference variance, GLCM_difference entropy, Information correlation, autocorrelation, dissimilarity, Cluster prominence, Joint maximum, IDM, Inverse Difference, Small		Ns	

			Zone Emphasis, Large Zone Emphasis, Low grey level zone emphasis, high grey level zone emphasis, small zone low grey level emphasis, small zone high grey level emphasis, large zone low grey level emphasis, large zone high grey emphasis, zone size non-uniformity, zon size percentage				
		Multivariable Cox	Stage3, ABCC2, GLCM_entropy		22.8 (3.7-141)	<0.001	
			Stage3, GCLM_entropy		14.2 (2.4-85.3)	0.004	
			ABCC2, GLCM_entropy		14.8 (2.9-76.2)	0.001	
Bang et al. (2016)[42]	Response	Histogram	50% percentile	44.2 - 47.1		0.004	Lower 50% percentile in responders
			90% percentile	54.6 - 57.8		0.018	Lower 90% percentile in responders
			99% percentile	60.5 - 63.3		0.046	Lower 99% percentile in responders
			Mean, variance, skewness, kurtosis, 1% percentile, 10% percentile			ns	
		Absolute gradient	Mean, variance, skewness, kurtosis, percentage of pixels with nonzero gradient			ns	
		Co-occurence matrix	Sum entropy	1.7 - 1.8		0.011	Lower sum entropy in responders
			Entropy	2.5 - 2.6		0.009	Lower entropy in responders
			ASM, contrast, correlation, sum of squares, inverse moment difference, sum average, sum variance, variance difference, entropy difference			ns	
		Run length matrix	Run-length nonuniformity, grey-level nonuniformity, long-run emphasis, short-run emphasis, fraction of image in runs			ns	
	DFS (3y)	GrKurtosis			5.9 (1.4-24.8)	0.015	Higher Grkurtosis → worse 3yDFS
Beckers, Beets-Tan, et al. (2017)[92]	New CRLM (Whole liver)	Histogram Univariate	Entropy0, Uniformity0, Meangreylevelintensity0, entropy0.5, uniformity0.5, meangreylevelintensity0.5, entropy1.5, uniformity1.5, meangreylevelintensity1.5, entropy2.5, uniformity2.5, meangreylevelintensity2.5			ns	
	New CRLM (Liver segments)		Entropy0, Uniformity0, Meangreylevelintensity0, entropy0.5, uniformity0.5, meangreylevelintensity0.5, entropy1.5, uniformity1.5, meangreylevelintensity1.5, entropy2.5, uniformity2.5, meangreylevelintensity2.5			ns	
Beckers, Lambregts, et al. (2017)[93]	New CRLM	Histogram Univariate	Entropy0, Uniformity0, Meangreylevelintensity0, entropy0.5, uniformity0.5, meangreylevelintensity0.5, entropy1.5, uniformity1.5, meangreylevelintensity1.5, entropy2.5, uniformity2.5, meangreylevelintensity2.5			ns	
	New	Multivariable	Uniformity0.5		OR 0.56 (0.31-1.00)	0.05	Lower uniformity0.5 in patients that develop

	CRLM (<6 months from baseline)						early CRLM
Beckers et al. (2018)[94]	Response (RECIST)	Histogram	Liver parenchyma, metastases or metastases/liver ratio: entropy, entropy0.5, entropy2.5, uniformity, uniformity0.5, uniformity2.5			ns	
	OS	Univariate	Emetastases0		2.59 (1.00–6.65)	0.05	Lower entropy → better survival
			Umetastases0		0.90 (0.80−1.00)	0.05	Higher uniformity → better survival
			Emetastases/Eliver0		2.51 (1.44–4.36)	0.001	lower entropy metastases/liver ratio → better survival
			Umetastases/Uliver0		0.41 (0.24–0.72)	0.002	Higher uniformity metastases/liver ratio → better survival
			Emetastases2.5		1.46 (1.13–1.89)	0.004	Lower entropy → better survival
			Umetastases2.5		0.98 (0.96–1.00)	0.012	Higher uniformity → better survival
			Emetastases/Eliver2.5		1.81 (1.06–3.09)	0.03	lower entropy metastases/liver ratio → better survival
			Umetastases/Uliver2.5		0.49 (0.29–0.84)	0.009	higher uniformity metastases/liver ratio → better survival
			Eliver0, Uliver0, Eliver0.5, Uliver0.5, Emetastases0.5, Umetastaes0.5, Emetastases/Eliver0.5, Umetastases/Uliver0.5, Eliver2.5, Uliver2.5			ns	
	OS	Multivariable	Emetastases/Eliver		1.9 (0.95-3.78)	ns	
Bibault et al. (2018)[43]	Response	Deep Neural Network	28 features in GLCM 2D, GLCM 3D and IntensityDirect	not specified	AUC 0.72 (0.65 - 0.87)		
Boldrini et al. (2018)[44]	DRE + MRI + endoscopy		Grey level non-uniformity	not specified		0.027	Grey level non-uniformity different between responders and non-responders
			Short run emphasis, Long run emphasis, Low grey level run emphasis, Short run high grey level emphasis, Long run high grey level emphasis, Run length non uniformity, Normalized run length non-uniformity, Run percentage, Run length variance			ns	
Bulens et al. (2020)[45]	GR	MRI radiomics only	T2W-model		AUC 0.72		
			DWI-model		AUC 0.71		
			ADC-model		AUC 0.70		
			T2W-DWI-model		AUC 0.72		
			T2W-ADC-model		AUC 0.73		
			DWI-ADC-model		AUC 0.72		
			T2W-DWI-AUC-model		AUC 0.70		
		Radiomics + non-radiomics features	Semantic-DWI-ADC-model		Training AUC 0.86 (0.75-0.98) Validation AUC 0.49 (0.24-0.74)		

Bundschuh et al. (2014)[14]	Response		COV (=SD/mean)	0.44 - 0.53	< 0.05	Lower COV in responders
			Skewness, kurtosis		ns	
	PFS		COV	>0.37	0.03	COV >0.37 → higher PFS
	OS		COV		ns	
Caruso et al. (2018)[46]	CR		Energy 0°, 45°, 90° and 135°	0.88 - 0.07; 0.92 - 0.22; 0.88 - 0.08; 0.92 - 0.06	< 0.05	Higher energy in responders
			Contrast 0°, 45°, 90° and 135°	0.07 - 0.85; 0.07 - 0.74; 0.07 - 0.82; 0.05 - 0.72	< 0.05	Lower contrast in responders
			Correlation 0°, 45°, 90° and 135°	0.89 - 0.49; 0.80 - 0.44; 0.92 - 0.44; 0.87 - 0.42	< 0.05	Higher correlation in responders
			Entropy 0°, 45°, 90° and 135°	0.23 - 1.53; 0.20 - 1.87; 0.22 - 1.50; 0.27 - 1.71	< 0.05	Lower entropy in responders
			IDM 0°, 45°, 90° and 135°	0.77 - 0.26; 0.79 - 0.22; 0.77 - 0.27; 0.82 - 0.22	< 0.05	Higher IDM in responders
			Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure correlation 1, Information measure correlation 2, Maximum correlation		ns	
Chee et al. (2017)[47]	DFS (Imaging)		Entropy0		0.015	Entropy <6.7 → higher survival
			Uniformity0	>0.012	0.025	Uniformity >0.012 → higher survival
			SD0	<28.06	0.038	SD <28.06 → higher survival
		Multivariable Cox prop hazards	Entropy0	3.15(1.23-8.07)	0.017	Lower entropy → higher survival
			Uniformity0	0.33(0.12-0.9)	0.030	Higher uniformity → higher survival
			SD0	2.54(1.06-6.09)	0.036	Lower SD → higher survival
			Filtered features (no further info)		ns	
Crimi et al. (2020)[48]	CR	Histogram	Mean, variance, skewness, excess kurtosis, energy, entropy		ns	
		GLCM	Energy angular second moment uniformity, Contrast inertia variance, Sum of squares variance, Homogeneity inverse different moment, Sum average, Sum variance, Sum entropy, Entropy, Difference variance, Difference entropy, Information correlation, Autocorrelatoin,		ns	

			Dissimilarity, Cluster shade, Cluster prominence, Maximum probability, Inverse difference				
		GLRLM	Short run emphasis, Long run emphasis, High gray-level emphasis, Gray-level non-uniformity, Run-length non-uniformity, Run percentage, Short run low gray-level emphasis, Long run high gray-level emphasis, Short run high gray-level emphasis, Long run low gray-level emphasis			ns	
Cui et al. (2018)[49]	CR	Training dataset	Radiomics score: ADC: Percentile 50, IDM angle 90, GLCM Entropy angle 45, GLCM Entropy angle 135, Intensity Variability, T2w: IDM angle 90, Correlation angle 90, Haralick correlation angle 90, Small Area Emphasis, T1w: Relative Deviation, Kurtosis, Correlation angle 0.	0.71 - -4.91		< 0.0001	Higher rad score in responders
		T2W			AUC 0.77		
		T1W			AUC 0.76		
		ADC			AUC 0.85		
		T2W/T1W/ADC			AUC 0.94		
		Validation dataset	Radiomics score	r -0.25 - -5.91		0.0001	Higher rad score in responders
Cusumano et al. (2018)[50]	CR	Histogram (Multivariable)	Skewness σ=0.48			0.006	Lower skewness in responders
			Max FD (40-100)			0.002	Lower Max FD in responders
			Kurtosis, entropy			ns	
Dai et al. (2020)[90]	OS	CT	Radiomics model cut-off: 0.121		3.1 AUC 0.77	< 0.001	Lower score → better 5-year OS
		CT + clinical features	Combined clinical - radiomics model		AUC 0.80		
			GLSZM_GLNU	coeff 0.15			
			GLSZM_SZSE7	coeff 0.46			
			GLSZM_SZSE6	coeff -0.12			
			GLSZM_GLV5	coeff -0.65			
			GLSZM_SZLGE1	coeff -1.04			
			GLCM_entropy6	coeff 0.40			
			GLCM_inversevariance1	coeff 0.53			
			range1	coeff 0.60			
			minimum1	coeff -0.50			
			skewness	coeff 0.15			
			mean	coeff -0.41			
			kurtosis	coeff -0.53			
	RFS	CT	Radiomics model cut-off: 0.189		2.8 AUC 0.74	< 0.001	Lower score → better 5-year RFS
		CT + clinical features	Combined clinical - radiomics model		AUC 0.79		
			GLSZM_ZSV7	coeff -0.08			
			GLCM_covariance6	coeff -0.004			

			GLCM_correlation6	coeff -0.25		
			GLCM_covariance5	coeff -0.22		
			GLCM_clustershade5	coeff -0.23		
			GLCM_covariance1	coeff -0.30		
			GLCM_clustershade3	coeff -0.11		
			skewness5	coeff -0.02		
			rootmeansquare5	coeff -0.31		
			median5	coeff 0.03		
			mean5	coeff 0.46		
			median3	coeff 0.14		
			mean3	coeff -0.38		
			skewness1	coeff 0.13		
			meanabsolutedeviation1	coeff -0.46		
			kurtosis1	coeff -0.77		
			GLRLM_SRLGLE4	coeff 0.42		
			GLRLM_SRLGLE2	coeff -0.65		
			GLRLM_LGLRE2	coeff 0.33		
			GLSZM_ZSP1	coeff -0.39		
			GLSZM_SZSE1	coeff -0.09		
			GLSZM_SZLGE1	coeff -0.45		
			GLSZM_LZSE1	coeff 0.07		
			GLSZM_LZLGE1	coeff -0.49		
			GLSZM_LGLZE1	coeff 0.04		
			GLSZM_HGLZE1	coeff -0.09		
de Cecco et al. (2015)[51]	Response		SSF4: Kurtosis CR / GR / NR	r -0.45 / 1.23 / 0.46	0.016	Lower kurtosis in CR Higher kurtosis in GR
			SSF5: Kurtosis CR / GR / NR	r -0.62 / 0.82 / 0.16	0.02	Lower kurtosis in CR Higher kurtosis in GR
			Entropy, skewness, MPP		ns	
de Cecco et al. (2016)[52]	CR	T2w	Kurtosis (median)	Median -0.45 - 0.29	< 0.001	Lower kurtosis in responders
		pMRI	Ve	0.28 - 0.57	0.04	Lower Ve (on pMRI) in responders
			DWI:ADC, pMRI: K trans, IAUGC, Kep		ns	
Dercle et al. (2020)[106]	OS (>17.7m)	HQ/CT: Anti-EGFR cohort / RF classifier	Shape SI4, LoG Z Entropy, LoG X Entropy, GDTM Contrast	Training AUC 0.83 (0.75-0.95) Validation AUC 0.80 (0.69-0.94)		RF classifier outperformed KRAS mutation and tumour shrinkage for the prediction of survival
		SQ/CT: Anti-EGFR cohort / RF classifier	RF classifier: Shape SI4, LoG Z Entropy, LoG X Entropy, GDTM Contrast	Training AUC 0.84 (0.76-0.89) Validation AUC 0.72 (0.59-0.83)		
		HQ/CT: Chemo-only cohort	RF classifier: Shape SI4, LoG Z Entropy, LoG X Entropy, GDTM Contrast	Training AUC 0.75 (0.63-0.85) Validation AUC 0.59 (0.44-0.72)		
		SQ/CT: Chemo-only cohort	RF classifier: Shape SI4, LoG Z Entropy, LoG X Entropy, GDTM Contrast	Training AUC 0.75 (0.67-0.82) Validation AUC 0.55 (0.43-0.66)		
Dinapoli et al. (2018)[53]	CR	Multivariable Rad-model	Also including clinical features (n=2)	Training AUC 0.75 (0.61-0.88)		

					Validation 1 AUC 0.75 (0.55-0.94) Validation 2 AUC 0.79 (0.59-0.98)		
			Skewness σ=0.49	Coeff -3.01		0.01	
			Entropy σ=0.34	Coeff 3.61		< 0.03	
			Kurtosis			ns	
	MFS/DFS/ OS	Histogram				ns	
Dohan et al. (2020)[107]	GR	Univariate					Results not specified
		Multivariable	mean, SD, MPP, skewness, kurtosis, entropy			ns	None of the pre-treatment radiomics features were retained in the multivariable model model
Ferrari et al. (2019)[54]	CR	T2W	GLCM Energy (2D)	0.992 – 0.987		0.05	GLCM energy higher in responders
		AI model					N/A → Also included mid/post CRT features
	GR	T2W	Histogram SD	169 - 109		0.0003	Higher histogram SD in responders
			Shannon Entropy mean	0.83 – 0.56		0.002	Higher Shannon entropy mean in responders
			Shannon Entropy sd	0.48 – 0.41		0.004	Higher Shannon entropy SD in responders
			Shannon Entropy max	2.24 – 1.73		0.0002	Higher Shannon entropy max in responders
		AI model					N/A → Also included mid/post CRT features
Fu et al. (2020)[55]	GR	Handcrafted DWI model			AUC 0.64		
		CNN model			AUC 0.73		
Ganeshan et al. (2007)[96]	OS	Histogram	Uniformity1.5			0.033	Lower uniformity → better survival
			Uniformity1.8			0.019	Lower uniformity → better survival
			Entropy2			0.019	Higher entropy → better survival
			Uniformity2			0.011	Lower uniformity → better survival
			Entropy2.5			0.010	Higher entropy → better survival
			Uniformity2.5			0.005	Lower uniformity → better survival
			Entropy2.8			0.007	Higher entropy → better survival
			Uniformity2.8			0.004	Lower uniformity → better survival
			Entropy3			0.006	Higher entropy → better survival
			Uniformity3			0.003	Lower uniformity → better survival
			Entropy3.5			0.004	Higher entropy → better survival
			Uniformity3.5			0.002	Lower uniformity → better survival
			Entropy3.8			0.004	Higher entropy → better survival
			Uniformity3.8			0.002	Lower uniformity → better survival
			Entropy4			0.004	Higher entropy → better survival
			Uniformity4			0.002	Lower uniformity → better survival
			Mean grey level intensity0, Entropy0, Uniformity0.Mean grey level intensity0.5, Entropy0.5, Uniformity0.5, Mean grey level intensity0.8, Entropy0.8, Uniformity0.8 Mean grey level intensity1, Entropy1, Uniformity1, Mean grey level intensity1.5, Entropy1.5, mean grey level			ns	

			intensity1.8, Entropy1.8, Entropy2, mean grey level intensity2.8, mean grey level intensity3, mean grey level intensity3.5, mean grey level intensity3.8, mean grey level intensity4				
Giannini et al. (2019)[56]	GR	Univariate_PET_ only	GLCM_homogeneity	< 0.18	AUC 0.77	<0.001	Lower GLCM_homogeneity in responders
			GLCM_dissimilarity	> 4.30	AUC 0.75	<0.001	Higher GLCM_dissimilarity responders
			GLCM_IDM	< 0.94	AUC 0.74	<0.001	Lower GLCM_IDM in responders
			Glycolytic volume	< 0.000002	AUC 0.74	<0.001	Lower glycolytic volume lower responders
			GLCM_IDMn	< 0.99	AUC 0.74	<0.001	Lower GLCM_IDM normalized in responders
			GLCM_difference variance	> 34.2	AUC 0.74	0.001	Higher GLCM_difference variance in responders
			GLCM_Contrast	> 34	AUC 0.74	0.001	Higher GLCM_Contrast in responders
			GLCM_sum entropy	> 4.67	AUC 0.74	0.001	Higher GLCM_sum entropy in responders
			Metabolic volume	< 486	AUC 0.73	0.001	Higher Metabolic volume in responders
			GLCM_difference entropy	> 2.47	AUC 0.71	0.005	Higher GLCM_difference entropy in responders
			GLCM_energy	< 0.002	AUC 0.70	0.007	Lower GLCM_energy in responders
			GLCM_maximum probability	< 0.008	AUC 0.67	0.027	Lower GLCM_maximum probability in responders
			Mean intensity, median intensity, 10th percentile, 25th percentile, 75th percentile, GLCM_autocorrelation, GLCM_correlation1, GLCM_correlation2, GCLM_clusterprominence, GLCM_ cluster shade, GLCM_entropy, GLCM_variance, GLCM_sum variance, GLCM_sum average, GLCM_information measure of correlation1, GLCM_information measure of correlation2			ns	
	GR	Univariate_MRI only	GLCM_difference entropy	> 1.98	AUC 0.68	0.014	Higher GLCM_difference entropy in responders
			GLCM_homogeneity	< 0.34	AUC 0.68	0.015	Lower GLCM_homogeneity in responders
			GLCM_dissimilarity	> 2.35	AUC 0.68	0.018	Higher GLCM_dissimilarity in responders
			GLCM_IDM	< 0.96	AUC 0.68	0.017	Lower GLCM_IDM in responders
			GLCM_entropy	> 6.35	AUC 0.66	0.045	Higher GLCM_entropy in responders
			Mean intensity, median intensity, 10th percentile, 25th percentile, 75th percentile, metabolic volume, glycolytic volume, GLCM_autocorrelation, GLCM_contrast, GLCM_correlation1, GLCM_correlation2, GCLM_clusterprominence, GLCM_ cluster shade, GLCM_energy, GLCM_maximum probability, GLCM_variance, GLCM_sum variance, GLCM_sum entropy, GLCM_sum average, GLCM_difference variance, GLCM_information measure of correlation1, GLCM_information measure of correlation2, GLCM_IDM normalized			ns	

	GR	Multivariable_PET	GLCM_homogeneity, GLCM_contrast, metabolic volume, glycolytic volume, 10th percentile	AUC 0.84
		Multivariable_MRI	T2W_10th percentile T2W_GLCM_correlation1, ADC_GLCM_cluster prominence, ADC_GLCM_information measure of correlation 2	AUC 0.72
		Multivariable_PET_MRI	PET_GLCM_homogeneity, PET_GLCM_contrast, PET_10th percentile, PET_glycolytic volume, PET_metabolic volume, T2W_GLCM_correlation 1	AUC 0.86
Hamerla et al. (2019)[57]	CR	RF machine learning	Shape_MinorAxisLength,Shape_Maximum2DDiameterSlice, shape_LeastAxisLength 10Percentile, wav_LHH_Median, Wav_LHH_Mean, wav_LHH_Uniformity, Wav_HHH_Skewness, wav_HLH_Kurtosis, wav_LHH_RootMeanSquared, wav_LHH_Variance, wav_HHH_Variance, wav_LLH_Skewness, wav_LHH_MeanAbsoluteDeviation, LoG2.0_Skewness, gradient_Minimum, GLCM_Imc1, wav_LLL_GLCM_Imc2, LoG2.0_GLCM_Autocorrelation, LoG5.0_GLCM_ClusterShade, LoG2.0_GLCM_JointAverage, LoG3.0_GLCM_Idmn, LoG5.0_GLCM_Idmn, wav_HHH_GLRLM_HighGrayLevelRunEmphasis, wav_HHH_GLRLM_LowGrayLevelRunEmphasis, wav_LHH_GLRLM_HighGrayLevelRunEmphasis, wav_LHH_GLRLM_LowGrayLevelRunEmphasis, LoG2.0_GLRLM_RunEntropy, LoG2.0_GLRLM_LongRunHighGrayLevelEmphasis, LoG2.0_GLRLM_ShortRunLowGrayLevelEmphasis, LoG2.0_GLRLM_HighGrayLevelRunEmphasis, LoG2.0_GLRLM_HighGrayLevelRunEmphasis, LoG1.0_GLRLM_RunLengthNonUniformity, LoG5.0_GLRLM_LowGrayLevelRunEmp	Training AUC 0.87 Validation AUC 0.50

			hasis, LoG3.0_GLRLM_HighGrayLevelRunEmphasis, GLSZM_LargeAreaHighGrayLevelEmphasis, GLSZM_ZoneEntropy wav_HHH_GLSZM_SizeZoneNonUniformityNormalized, lbp-3D-k_GLSZM_ZoneEntropy, wav_HHH_GLSZM_ZoneVariance, wav_HHL_GLSZM_ZoneVariance, wav_LHL_GLSZM_ZoneVariance, Wav_LHL_GLSZM_SmallAreaEmphasis, wav_HLL_GLSZM_LargeAreaHighGrayLevelEmphasis, wav_HHL_GLSZM_LargeAreaEmphasis, wav_HHL_GLSZM_LargeAreaHighGrayLevelEmphasis, wav_HHL_GLSZM_SizeZoneNonUniformityNormalized, wav_LHH_GLSZM_ZoneVariance, wav_LLH_GLSZM_SizeZoneNonUniformityNormalized, wav_LLL_GLSZM_LargeAreaHighGrayLevelEmphasis, wav_HHH_GLSZM_SmallAreaLowGrayLevelEmphasis, wav_HHH_GLSZM_SmallAreaHighGrayLevelEmphasis, wav_LHH_GLSZM_LargeAreaEmphasis, LoG2.0_GLSZM_ZoneEntropy, LoG5.0_GLSZM_ZoneEntropy, wav_HHL_gldm_DependenceVariance, wav_LHH_gldm_GrayLevelVariance, LoG2.0_gldm_LargeDependenceHighGrayLevelEmphasis, LoG3.0_gldm_SmallDependenceLowGrayLevelEmphasis, LoG3.0_gldm_LargeDependenceHighGrayLevelEmphasis, LoG3.0_gldm_LowGrayLevelEmphasis, LoG5.0_gldm_LowGrayLevelEmphasis, LoG5.0_gldm_LargeDependenceLowGrayLevelEmphasis, LoG2.0_gldm_HighGrayLevelEmphasis,		
Jalil et al. (2017)[58]	RFS (Imaging)	Univariate	Mean3	0.017	Lower mean → better RFS
			Mean4	0.026	Lower man → better RFS
			SD0	0.032	Lower SD → better RFS
			SD2	0.034	Lower SD → better RFS
			SD4	0.018	Lower SD → better RFS
			SD5	0.017	Lower SD → better RFS
			SD6	0.011	Lower SD → better RFS

			Entropy0		0.034	Lower entropy → better RFS
			Entropy4		0.016	Lower entropy → better RFS
			Entropy5		0.016	Lower entropy → better RFS
			Entropy6		0.011	Lower entropy → better RFS
			MPP2		0.011	Lower MPP → better RFS
			MPP5		0.011	Lower MPP → better RFS
			MPP6		0.019	Lower MPP → better RFS
			Skewness2		0.037	Lower skewness → better RFS
			Kurtosis4		0.047	Lower kurtosis → better RFS
			Mean2, SD3, Entrop1entropy2entropy3		ns	
	RFS (Imaging)	Multivariable	MPP2	8.9 (2.3-33.1)	0.001	Lower MPP → better RFS
			Kurtosis4	7.7 (2.0-29)	0.002	Lower kurtosis4 → better RFS
	DFS	Univariate	Mean2		0.031	Lower mean → better DFS
			Mean3		0.007	Lower mean → better DFS
			Mean4		0.027	Lower mean → better DFS
			Mean6		0.043	Lower mean → better DFS
			MPP2		0.022	Lower MPP → better DFS
			MPP3		0.045	Lower MPP → better DFS
			MPP4		0.022	Lower MPP → better DFS
			MPP5		0.047	Lower MPP → better DFS
			MPP6		0.047	Lower MPP → better DFS
			Skewness2		0.044	Lower skewness → better DFS
			Mean5, kurtosis		ns	
		Multivariable	Mean3	4.5 (1.5-12.9)	0.001	Lower mean → better DFS
			MPP2	3.3 (1.3-8.3)	0.008	Lower MPP → better DFS
	OS	Univariate	Mean3		0.03	Lower mean → better survival
			MPP2		0.008	Lower MPP → better survival
			MPP3		0.029	Lower MPP → better survival
			MPP4		0.019	Lower MPP → better survival
		Multivariable	Mean3	5.7 (1.6-20.2)	0.007	lower mean → better survival
			MPP2	6.9 (2.4-19.5)	<0.001	lower MPP → better survival
Lee et al. (2018)[95]	New CRLM	Histogram	Entropy0, skewness0, SD0, kurtosis 0, entropy2, skewness2, SD2, kurtosis2, entropy3, skewness3, SD3, kurtosis3, entropy4, skewness4, SD4, kurtosis4, entropy5, skewness5, SD5, kurtosis5, entropy6, skewness6, SD6, kurtosis6		ns	
Li, Boimel, et al. (2019)[59]	OS	DL_CT	DL	AUC 0.62		
		DL_PET	DL	AUC 0.60		
		DL _PET&CT	DL	AUC 0.64		
Li, Wang, et al. (2020)[60]	GR	CT-model	Sum average, volumeCC, GLCMEntropy, Percentile5, ShortRunEmphasis, GLCMEntropySD4, Correlation45, Percentile95, MeanDeviation, LRE90, GLCMEntropySD7, InertiaSD1, GLNUSD7, MaxIntensity, Compactness, LongRunLowGreyLevelEmphasisSD7, LowGreyLevelRunEmphasisSD1, ShortRunLowGreyLevelEmphasisSD7	AUC 0.77		

		DCE-T1W	ShortRunEmphasis1, SurfaceArea, Quantile0.025, HaralickCorrelation4, LongRunHighGreyLevelEmphasis, ShortRunEmphasisSD7, GLCMEntropy135, GLCMEntropySD7, ShortRunEmphasisSD1, LRE45, GLCMEntropySD1, MeanDeviation, HaralickCorrelationSD1, LongRunHighGreyLevelEmphasis135, Inertia_AllDirectionSD7, LongRunLowGreyLevelEmphasis90	AUC 0.81		
		T2W	Maximum3DDiameter, Correlation135, stdDeviation, LowGreyLevelRunEmphasisSD4, uniformity, GLCMEntropySD7, GLCMEntropySD4, RunLengthNonuniformitySD4, Correlation135, LRESD4, RelativeDeviation, LRE0, ClusterShade45, ShortRunEmphasis90	AUC 0.86		
		ADC	GLCMEntropySD1, Inertia_a90, uniformity, ClusterShade0, IntensityVariability, MeanValue, Correlation_AllDirectionSD4, RunLengthNonuniformitySD, Correlation45, InverseDifferenceMoment90, LongRunLowGreyLevelEmphasis135, LongRunHighGreyLevelEmphasis0, Maximum3DDiameter, LongRunHighGreyLevelEmphasisSD7, HighGreyLevelRunEmphasisSD7, MeanDeviation, ClusterShadeSD4	AUC 0.82		
		Combined model	ADC_LargeAreaEmphasis, T1_LongRunHighGreyLevelEmphasis135, T1_ShortRunEmphasis90, ADC_Inertia45, T1_ShortRunEmphasis90, T1_RunLengthNonuniformitySD7, ADC_LongRunHighGreyLevelEmphasis90, CT_sumAverage, T1_GLCMEntropy0, ADC_Percentile15" T2_uniformity, ADC_InverseDifferenceMoment90, CT_SmallAreaEmphasis	AUC 0.93		
Liang et al. (2016)[38]	Response (RECIST 1.1)	ADC	ADC mean	104.3 – 150.1	0.001	Lower ADC mean in responders
			ADC 1st percentile	64.5 – 108.8	0.002	Lower ADC 1st percentile in responders
			ADC 10th percentile	78.7 – 124.5	0.002	Lower ADC 10th percentile in responders
			ADC 50th percentile	103.0 – 149.0	< 0.001	Lower ADC 50th percentile in responders
			ADC 90th percentile	130.6 – 177.1	< 0.001	Lower ADC 90th percentile in responders
			ADC 99th percentile	149.7 – 197.8	< 0.001	Lower ADC 99th percentile in responders

			ADC variance, ADC skewness, ADC kurtosis		ns
Liang et al. (2019)[61]	new CRLM	T2W_LR (n=5)	T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLSZM_SizeZoneNonUniformityNormalized, T2_wav_LHH_Maximum, T2_wav_LHL_GLSZM_SmallAreaEmphasis, T2_wav_LHL_GLCM_Imc1	AUC 0.71 (0.61-0.76)	
		T2W_SVM (n=5)	T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLSZM_SizeZoneNonUniformityNormalized, T2_wav_LHH_Maximum, T2_wav_LHL_GLSZM_SmallAreaEmphasis, T2_wav_LHL_GLCM_Imc1	AUC 0.71 (0.60-0.75)	
		VP_LR (n=8)	VP_square_GLRLM_ShortRunLowGrayLevelEmphasis, VP_wav_LLL_GLSZM_SmallAreaLowGrayLevelEmphasis, VP_exponential_10Percentile, VP_wav_HHL_Energy, VP_wav_HLL_Skewness, VP_10Percentile, VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy	AUC 0.74 (0.57-0.75)	
		VP_SVM (n=8)	VP_square_GLRLM_ShortRunLowGrayLevelEmphasis, VP_wav_LLL_GLSZM_SmallAreaLowGrayLevelEmphasis, VP_exponential_10Percentile, VP_wav_HHL_Energy, VP_wav_HLL_Skewness, VP_10Percentile, VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy	AUC 0.68 (0.56-0.72)	
		Combined_LR (T2W+VP; n=13)	T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLSZM_SizeZoneNonUniformityNormalized, T2_wav_LHH_Maximum, T2_wav_LHL_GLSZM_SmallAreaEmphasis, T2_wav_LHL_GLCM_Imc1, VP_square_GLRLM_ShortRunLowGrayLevelEmphasis, VP_wav_LLL_GLSZM_SmallAreaLowGrayLevelEmphasis, VP_exponential_10Percentile, VP_wav_HHL_Energy, VP_wav_HLL_Skewness,	AUC 0.76 (0.65-0.80)	

		VP_10Percentile, VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLSZM_SizeZoneNonUniformityNormalized, T2_wav_LHH_Maximum, T2_wav_LHL_GLSZM_SmallAreaEmphasis, T2_wav_LHL_GLCM_Imc1,	
	Combined_SVM (T2W+VP; n=13)	VP_square_GLRLM_ShortRunLowGrayLevelEmphasis, VP_wav_LLL_GLSZM_SmallAreaLowGrayLevelEmphasis, VP_exponential_10Percentile, VP_wav_HHL_Energy, VP_wav_HLL_Skewness, VP_10Percentile, VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy	AUC 0.74 (0.61-0.75)
	Optimal _LR (T2W+VP; n=22)	T2_exponential_RobustMeanAbsoluteDeviation, T2_exponential_InterquartileRange, T2_wav_HHL_Skewness, T2_wav_HHL_GLRLM_LongRunLowGrayLevelEmphasis, T2_wav_HLH_Mean, T2_wav_HHH_Mean, T2_wav_LLH_GLCM_Correlation, T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLCM_Correlation, T2_wav_LHH_GLCM_Imc1, T2_wav_HHL_GLCM_Correlation, T2_wav_LHL_GLCM_Imc1, T2_wav-LHL_GLSZM_SmallAreaEmphasis, VP_exponential_10Percentile, VP_wav_HHH_Maximum VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy VP_wav_HHL_TotalEnergy VP_wav-LHL_GLSZM_GrayLevelVariance VP_wav-HHL_Energy, VP_wav-LHL_GLCM_Imc2, VP_wav-HLL_Skewness	AUC 0.87 (0.73-0.88)
	Optimal _SVM (T2W+VP; n=22)	T2_exponential_RobustMeanAbsoluteDeviation, T2_exponential_InterquartileRange, T2_wav_HHL_Skewness, T2_wav_HHL_GLRLM_LongRunLowGrayLevelEmphasis, T2_wav_HLH_Mean, T2_wav_HHH_Mean, T2_wav_LLH_GLCM_Correlation,	AUC 0.83 (0.65-0.84)

			T2_wav_LHH_GLCM_MaximumProbability, T2_wav_LHH_GLCM_Correlation, T2_wav_LHH_GLCM_Imc1, T2_wav_HHL_GLCM_Correlation, T2_wav_LHL_GLCM_Imc1, T2_wav-LHL_GLSZM_SmallAreaEmphasis, VP_exponential_10Percentile, VP_wav_HHH_Maximum VP_squareroot_GLSZM_LargeAreaHighGrayLevelEmphasis, VP_wav_HHL_GLRLM_RunEntropy VP_wav_HHL_TotalEnergy VP_wav-LHL_GLSZM_GrayLevelVariance VP_wav-HHL_Energy, VP_wav-LHL_GLCM_Imc2, VP_wav-HLL_Skewness			
Liu et al. (2017)[62]	Response (TRG)		Energy	0.097 - 0.17	0.006	Lower energy in responders
			Energy variance	0.01 - 0.02	0.006	Lower energy variance in responders
			Correlation	r -0.14 - -0.6	0.024	Lower correlation in responders
			Shade cluster variance	17.05 - 36.74	0.008	Lower shade cluster variance in responders
			Gabor-04	1.6 - 8.43	0.022	Lower Gabor-04 in responders
			Gabor-11	0.38 - 2.03	0.019	Lower Gabor-11 in responders
			Gabor-26	9.56 - 15.91	0.023	Lower Gabor-26 in responders
			Gabor-47	0.38 - 2.07	0.019	Lower Gabor-47 in responders
			SdGa-00	46.65 - 35.29	0.038	Higher SdGa-00 in responders
			SdGa-04	59.59 - 102.89	0.016	Lower SdGa-04 in responders
			SdGa-11	29.55 - 50.48	0.016	Lower SdGa-11 in responders
			SdGa-37	79.13 - 61.93	0.02	Higher SdGa-37 in responders
			SdGa-44	72.59 - 57.96	0.038	Higher SdGa-44 in responders
			SdGa-47	29.55 - 51.31	0.016	Lower SdGa-47 in responders
			High-intensity small area emphasis	9.75 - 4.59	0.043	Higher HISAE in responders
			Histogram energy	0.01 - 0.02	0.033	Lower energy in responders
			Mean ADC, SdGa-43		ns	
Liu, Li, et al. (2019)[63]	OS	PET+CT features				
		Cox regression		c-index 0.63		
		Cox with LASSO		c-index 0.60		
		RF		c-index 0.54		
		PCA + Cox		c-index 0.63		
		PCA + Cox LASSO		c-index 0.58		
		PCA + RF		c-index 0.50		
		CC + Cox		c-index 0.67		
		CC + Cox with LASSO		c-index 0.67		
		CC + RF		c-index 0.55		
Liu, Wen, et al. (2019)[64]	CR	ADC-, D-, D*-, *f*-map	Histogram: kurtosis, mean, skewness, variance, median		ns	
		ADC-map -	GLCM_Contrast	119.0 - 220.9	0.025	Lower GLCM contrast in responders

		univariate				
			GLCM_Difference_variance	59.1 - 93.1	0.017	Lower GLCM difference variance in responders
			GLCM_Sum_average	78.0 - 67.4	0.049	Higher GLCM sum average in responders
			Histogram: kurtosis, mean, skewness, variance, median. GLCM: ASM, Correlation, Difference_entropy, Entropy, IDM, Sum_entropy, Sum_of_squares, Sum_variance		ns	
		D*-map - univariate	GLCM_Contrast	151.3 - 300.3	0.002	Lower GLCM contrast in responders
			GLCM_Correlation	0.6 - 0.1	0.001	Higher GLCM correlat in responders
			GLCM_Difference_variance	65.5 - 96.5	0.024	Lower GLCM difference variance in responders
			Histogram: kurtosis, mean, skewness, variance, median. GLCM: ASM, Difference_entropy, Entropy, IDM, Sum_average, Sum_entropy, Sum_of_squares, Sum_variance		ns	
		D-map - univariate	GLCM_Contrast	64.8 - 248.3	0.001	Lower GLCM contrast in responders
			GLCM_Correlation	-0.1 - 0.2	0.039	Lower GLCM correlat in responders
			GLCM_Difference_variance	16.5 - 76.1	0.001	Lower GLCM difference variance in responders
			GLCM_Sum_average	74.3 - 64.2	0.027	Higher GLCM sum average in responders
			Histogram: kurtosis, mean, skewness, variance, median. GLCM: ASM, Difference_entropy, Entropy, IDM, Sum_entropy, Sum_of_squares, Sum_variance		ns	
		f-map - univariate	GLCM_Correlation	0.8 - 0.06	0.002	Higher GLCM correlat in responders
			Histogram: kurtosis, mean, skewness, variance, median. GLCM: ASM, Contrast, Difference_entropy, Difference_variance, Entropy, IDM, Sum_average, Sum_entropy, Sum_of_squares, Sum_variance		ns	
		GLCM - multivariable	D-map_Difference _variance, D*-map_Correlate, ADC-map_Difference_variance		AUC 0.99	
Lovinfosse et al. (2018)[65]	Response	Univariate				
			Skewness		OR 0.3 (0.11-0.89)	0.029	Lower skewness in responders
			Excess kurtosis		OR 0.2 (0.1-0.6)	0.003	Lower excess kurtosis in responders
			Contrast GLCM		OR 2.8 (1.0-7.9)	0.047	Higher contrast GLCM in responders
			Homogeneity		OR 0.3 (0.1-0.8)	0.014	Lower homogeneity in responders
			Dissimilarity		OR 3.8 (1.3-10.7)	0.014	Higher dissimilarity in responders
			Coarseness		OR 0.3 (0.1-0.8)	0.023	Lower coarseness in responders
			NGDTM Contrast		OR 3.8 (1.3-10.7)	0.014	Higher NGDTM contrast in responders
			Busyness		OR 0.3 (0.1-0.9)	0.029	Lower busyness in responders
			Intensity Variability		OR 0.4 (0.1-1.0)	0.047	Lower intensity variability in responders

		Multivariable	Skewness, Excess kurtosis, Contrast GLCM, Homogeneity, Dissimilarity, Coarseness, NGDTM Contrast, Busyness, Intensity Variability		Ns	
	Response	Univariate	Skewness, Excess kurtosis, Contrast GLCM, Homogeneity, Dissimilarity, Coarseness, NGDTM Contrast, Busyness, Intensity Variability		Ns	
	DSS	Univariate	ASM	0.2 (1.7-25.1)	0.011	higher ASM → better DSS
			Contrast(GLCM)	0.1 (0.01-0.8)	0.027	higher contrast (GLSM) --> better DSS
			Entropy	7.9 (2.1-30.2)	0.002	lower entropy → better DSS
			Correlation	13.2 (1.6-106)	0.015	Lower correlation → better DSS
			Dissimilarity	0.1 (0.01-0.8)	0.033	Higher dissimilarity → better DSS
			Coarseness	10.2 (2.1-49.2)	0.004	Lower Coarseness → better DSS
			Contrast (NGTDM)	0.1 (0.01-0.7)	0.024	Higher contrast (NGTDM) → better DSS
			Busyness	5.6 (1.2-26.4)	0.031	Lower busyness → better DSS
			Intensity variability	14.1 (1.7-117)	0.015	Lower intensity variability → better DSS
			Size zone variability	0.2 (0.07-0.9)	0.036	Higher size zone variability → better DSS
		Multivariable	Coarseness	7.1 (1.3-38.7)	0.024	Lower Coarseness → better DSS
	DFS	Univariate	Excess kurtosis	0.3 (0.1-0.8)	0.020	Higher excess kurtosis → better DFS
			Homogeneity	0.2 (0.1-0.6)	0.005	Higher homogeneity → better DFS
			Coarseness	3.8 (1.4-10.3)	0.009	Lower Coarseness → better DFS
		Multivariable	Homogeneity	0.06 (0.01-0.3)	0.000	Higher homogeneity → better DFS
			Coarseness	11.2 (3.1-40)	0.000	Lower coarseness → better DFS
	OS (death)	Multivariable	SUV mean	0.2	0.005	Higher SUVmean → better survival
			Dissimilarity	53.6	0.000	Lower dissimilarity → better survival
			Contrast (NGTDM)	0.01	<0.0001	Higher contrast (NGTDM) → better survival
Lubner et al. (2015)[97]	OS (death)		Entropy5	0.6 (0.4-0.95)	0.02	Higher entropy → better survival
Meng, Zhang, Dong, et al. (2018)[67]	Response	TRG 1-2 vs TRG 3-5	Kurtosis	3.57 - 3.24	0.02	Higher kurtosis in responders
			Mean, SD, skewness, uniformity, energy, entropy		ns	
		CR vs non-CR	Uniformity	0.82 - 0.79	0.03	Higher uniformity in responders
			Energy	0.95 - 0.50	<0.01	Higher energy in responders
			Entropy	0.22 - 1.39	0.01	Lower entropy in responders
			Mean, SD, skewness, kurtosis		ns	
Meng, Zhang, Zou, et al. (2018)[66]	DFS (3y)		Radiomic signature (training)	6.8 (3.7-12.8)	<0.001	Lower radiomics signature → Better DFS
			Radiomic signature (validation)	2.9 (1.9-4.5)	<0.001	Lower radiomics signature → Better DFS
		Features included in signature	X1__energy, X1__maximum, X6__kurtosis, X6_GLCM_correlation, X1_GLRLM_LRE, X1_GLRLM_SRLGLE, X1_GLRLM_LRLGLE, X7_GLRLM_mean			
Miles et al.	OS (death)	Univariate	Uniformity1.5_2.5		0.027	Higher uniformity → better DFS

(2009)[98]			Uniformity2.0_2.5		0.005	Higher uniformity → better DFS
			Meanintensity0.5_2.5, Meanintensity1.0_2.5, Meanintensity1.5_2.5, Meanintensity2.0_2.5, Uniformity0.5_2.5, Uniformity1.0_2.5		ns	
Nakajo et al. (2017)[87]	PFS	Kaplan Meier	IV		0.015	Lower IV → better PFS
			SZV		0.013	Lower SZV → better PFS
			SUVmax, SUVmean, TLG, COV, entropy, homogeneity, dissimilarity, zone percentage		ns	
		Univariate Cox	Entropy	7. 9(0.9-72.1)	0.066	
			IV	1.1 (1.1-1.2)	<0.0001	Lower IV → better PFS
			SZV	1.0 (1.0-1.01)	0.002	Lower SZV → better PFS
			SUVmax, SUVmean, TLG, COV, homogeneity, dissimilarity, zone percentage		ns	
Nardone et al. (2019)[68]	New CRLM	Univariate	DWI_GLCM_contrast	OR 2.4	0.009	Higher DWI_GLCM_contrast in patients with early PD
			DWI_GLCM_correlation	OR 0.2	<0.001	Lower DWI_GLCM_correlation in PD
			ADC_GLCM_correlation	OR 0.32	0.001	Lower ADC_GLCM_correlation in PD
		Multivariable	DWI_GLCM_correlation	OR 0.002 (0.001-0.18)		
				Training AUC 0.85 (0.73-0.98) Validation AUC 0.83 (0.619-1.000)	0.001	Lower DWI_GLCM_correlation in PD
Ng, Ganeshan, et al. (2013)[88]	OS (death)	Histogram, univariate	Skewness0		0.004	Lower skewness → better survival
			Entropy1.0		0.001	Higher entropy → better survival
			Uniformity1.0		0.18	Lower uniformity → better survival
			Kurtosis1.0		0.03	Higher kurtosis → better survival
			Skewness1.0		0.008	Lower skewness → better survival
			SDD1.0		0.001	Higher SDD → better survival
			Entropy1.5		0.01	Higher entropy → better survival
			Skewness1.5		0.03	Lower skewness → better survival
			SDD1.5		0.003	Higher SDD → better survival
			Entropy2.0		0.049	Higher entropy →better survival
			SDD2.5		0.01	Higher SDD → better survival
			SDD2.5		0.03	Higher SDD → better survival
			Entropy0, uniformity0, kurtosis0, SD0, Uniformity1.5, kurtosis1.5, uniformity2.0, kurtosis2.0, skewness2.0, entropy2.5, uniformity2.5, kurtosis2.5, skewness2.5		ns	
Ng, Kozarski, et al.	OS (death), largest	Histogram, univariate	Entropy0, Uniformity0, Entropy1.0, uniformity1.0, entropy1.5, uniformity1.5, entropy2.0, uniformity2.0, entropy2.5,		ns	

(2013)[89]	cross section		uniformity2.5			
	Whole tumour volume	Histogram, univariate	Entropy1.0		0.002	Higher entropy → better survival
			Uniformity1.0		0.02	Lower uniformity → better survival
			Entropy1.5		0.01	Higher entropy → better survival
			Entropy2.0		0.049	Higher entropy → better survival
Nie et al. (2016)[15]	Response	T1w	GLCM Autocorrelation	1.4 - 1.5	0.03	Lower autocorrelation in responders
			GLCM Cluster prominence	231 - 298	0.03	Lower cluster prominence in responders
			GLCM Cluster shade	16.7 - 21.6	0.03	Lower cluster shade in responders
			GLCM Contrast	0.02 - 0.03	0.03	Lower contrast in responders
			GLCM Dissimilarity	0.003 - 0.004	0.03	Lower dissimilarity in responders
			GLCM Energy	0.99 - 0.98	0.03	Higher energy in responders
			GLCM Entropy	0.04 - 0.05	0.03	Lower entropy in responders
			GLCM Homogeneity 1	0.9995 - 0.9995	0.03	
			GLCM Homogeneity 2	0.9995 - 0.9994	0.03	Higher homogeneity 2 in responders
			GLCM Max Probability	0.993 - 0.991	0.03	Higher max probability in responders
			GLCM Sum Average	2.09 - 2.1	0.03	Lower sum average in responders
			GLCM Sum Variance	5.43 - 5.87	0.03	Lower sum variance in responders
			GLCM Sum Entropy	0.04 - 0.05	0.03	Lower sum entropy in responders
			GLCM Difference Variance	0.03 - 0.03	0.03	
			GLCM Difference Entropy	0.004 - 0.005	0.03	Lower difference entropy in responders
			GLCM IMC 2	0.25 - 0.28	0.03	Lower IMC2 in responders
			GLCM correlation, CLCM IMC 1		ns	
	Response	T2w	GLCM Autocorrelation	1.4 - 1.5	0.03	Lower autocorrelation in responders
			GLCM Cluster prominence	225 - 293	0.03	Lower cluster prominence in responders
			GLCM Cluster shade	16.3 - 21.3	0.03	Lower cluster shade in responders
			GLCM Contrast	0.03 - 0.03	0.03	
			GLCM Dissimilarity	0.004 -0.004	0.05	
			GLCM Energy	0.987 - 0.982	0.03	Higher energy in responders
			GLCM Entropy	0.04 - 0.05	0.03	Lower entropy in responders
			GLCM Homogeneity 1	0.999 - 0.9995	0.05	
			GLCM Homogeneity 2	0.9995 - 0.9994	0.05	
			GLCM Max Probability	0.993 - 0.991	0.03	Higher max probability in responders
			GLCM Sum Average	2.1 - 2.1	0.03	
			GLCM Sum Variance	5.4 - 5.8	0.03	Lower sum variance in responders
			GLCM Sum Entropy	0.04 - 0.05	0.03	Lower sum entropy in responders
			GLCM Difference Variance	0.026 - 0.03	0.05	Lower difference variance in responders
			GLCM Difference Entropy	0.004 - 0.005	0.04	Lower difference entropy in responders
			GLCM Information Measure Correlation 1	r -0.90 - -0.90	0.04	
			GLCM Information Measure Correlation 2	0.25 - 0.27	0.04	Lower IMC2 in responders
			GLCM Correlation	0.95 - 0.95	ns	
	Response	DCE	Kurtosis	3.09 - 2.29	0.05	
			Skewness	0.41 - 1.49	0.02	Lower skewness in responders
			GLCM Autocorrelation	1.13 - 1.21	0.008	Lower GLCM autocorrelation in responders
			GLCM Cluster prominence	73 - 109	0.02	Lower GLCM cluster prominence in responders
			GLCM Cluster shade	5.29 - 7.93	0.02	Lower GLCM cluster shade in responders
			GLCM Energy	0.995 - 0.993	0.01	Higher GLCM energy in responders
			GLCM Entropy	0.02 - 0.03	0.007	Lower GLCM entropy in responders

		GLCM Max Probability	0.998 - 0.997	0.01	Higher GLCM max probability in responders
		GLCM Sum Average	2.03 - 2.04	0.02	Lower GLCM sum average in responders
		GLCM Sum Variance	4.4 - 4.7	0.008	Lower GLCM sum variance in responders
		GLCM Sum Entropy	0.017 - 0.024	0.01	Lower GLCM sum entropy in responders
		GLCM Difference Variance	0.025 - 0.032	0.03	Lower GLCM difference variance in responders
		GLCM Difference Entropy	0.004 - 0.005	0.03	Lower GLCM difference entropy in responders
		GLCM Information Measure Correlation 1	r -0.7 - -0.8	0.03	Higher GLCM IMC1 in responders
		GLCM Information Measure Correlation 2	0.14 - 0.17	0.01	Lower GLCM IMC2 in responders
		10% percentile, 20% percentile, 30% percentile, 40% percentile, 50% percentile, 60% percentile, 70% percentile, 80% percentile, 90% percentile, Max, GLCM contrast, GLCM correlation, GLCM dissimilarity, GLCM homogeneity 1, GLCM homogeneity 2		ns	
CR/non-CR	DWI	Mean ADC	0.90 - 0.97	0.005	Lower mean ADC in responders
		50% percentile	0.90 - 0.92	0.002	Lower 50% percentile in responders
		60% percentile	0.98 - 1.06	0.005	Lower 60% percentile in responders
		70% percentile	1.06 - 1.19	0.01	Lower 70% percentile in responders
		80% percentile	1.13 - 1.31	0.02	Lower 80% percentile in responders
		GLCM Autocorrelation	1.42 - 1.49	0.008	Lower CLGM autocorrelation in responders
		GLCM Cluster prominence	239 – 272	0.02	Lower cluster prominence in responders
		GLCM Cluster shade	17.6 - 20.2	0.02	Lower cluster shade in responders
		GLCM Energy	0.98 - 0.98	0.01	
		GLCM Entropy	0.06 - 0.06	0.007	
		GLCM Max Probability	0.99 - 0.991	0.01	Lower max probability in responders
		GLCM Sum Average	2.10 - 2.11	0.02	Lower sum average in responders
		GLCM Sum Variance	5.53 - 5.77	0.008	Lower sum variance in responders
		GLCM Sum Entropy	0.05 - 0.06	0.01	Lower sum entropy in responders
		GLCM Difference Variance	0.06 - 0.07	0.03	Lower difference variance in responders
		GLCM Difference Entropy	0.01 - 0.02	0.03	Lower difference entropy in responders
		GLCM Information Measure Correlation 1	r -0.79 - -0.78	0.03	Lower IMC1 in responders
		GLCM Information Measure Correlation 2	0.26 - 0.27	0.01	Lower IMC2 in responders
		10% percentile, 20% percentile, 30% percentile, 40% percentile, 90% percentile, Max, kurtosis, skewness, GLCM contrast, GLCM correlation, GLCM dissimilarity, GLCM homogeneity 1, GLCM homogeneity 2		ns	
Response (GR / non-GR)	T1w	GLCM Autocorrelation, GLCM Cluster prominence, GLCM Cluster shade, GLCM Contrast, GLCM Correlation, GLCM Dissimilarity, GLCM Energy, GLCM Entropy, GLCM Homogeneity 1, GLCM Homogeneity 2, GLCM Max Probability, GLCM Sum Average, GLCM Sum Variance, GLCM Sum Entropy, GLCM Difference Variance, GLCM Difference Entropy, GLCM Information Measure		ns	

		Correlation 1, GLCM Information Measure Correlation 2			
Response (GR / non-GR)	T2w	GLCM Autocorrelation	1.44 - 1.53	0.050	Lower autocorrelation in responders
		GLCM Cluster prominence	259 – 309	0.050	Lower cluster prominence in responders
		GLCM Cluster shade	18.8 - 22.4	0.050	Lower cluster shade in responders
		GLCM Dissimilarity	0.004 - 0.005	0.047	Lower dissimilarity in responders
		GLCM Energy	0.99 - 0.98	0.050	Higher energy in responders
		GLCM Max Probability	0.993 - 0.991	0.050	Higher max probability in responders
		GLCM Sum Average	2.1 - 2.1	0.050	
		GLCM Sum Variance	5.6 - 6.0	0.050	Lower sum variance in responders
		GLCM Contrast, GLCM Correlation, GLCM Entropy, GLCM Homogeneity 1, GLCM Homogeneity 2, GLCM Sum Entropy, GLCM Difference Variance, GLCM Difference Entropy, GLCM Information Measure Correlation 1, GLCM Information Measure Correlation 2			
Response (GR / non-GR)	DCE	Kurtosis	3.14 - 2.21	0.030	Higher kurtosis in responders
		GLCM Autocorrelation	1.14 - 1.22	0.001	Lower autocorrelation in responders
		GLCM Cluster prominence	83 – 130	0.001	Lower cluster prominence in responders
		GLCM Cluster shade	6.1 - 9.4	0.001	Lower cluster shade in responders
		GLCM Correlation	0.80 - 0.89	0.030	Lower correlation in responders
		GLCM Energy	0.995 - 0.99	0.001	Higher energy in responders
		GLCM Entropy	0.02 - 0.03	0.001	Lower entropy in responders
		GLCM Max Probability	0.997 - 0.996	0.001	Higher max probability in responders
		GLCM Sum Average	2.0 - 2.1	0.001	Lower sum average in responders
		GLCM Sum Variance	4.5 - 4.8	0.001	Lower sum variance in responders
		GLCM Sum Entropy	0.02 - 0.03	0.001	Lower sum entropy in responders
		GLCM Information Measure Correlation 1	r -0.73 - -0.82	0.010	Higher IMC1 in responders
		GLCM Information Measure Correlation 2	0.15 - 0.19	0.001	Lower IMC2 in responders
		10% percentile, 20% percentile, 30% percentile, 40% percentile, 50% percentile, 60% percentile, 70% percentile, 80% percentile, 90% percentile, Max, skewness, GLCM contrast, GLCM dissimilarity, GLCM homogeneity 1, GLCM homogeneity 2, GLCM Difference Variance, GLCM Difference Entropy		ns	
Response (GR / non-GR)	DWI	Mean ADC	0.90 - 0.97	0.03	Lower mean ADC in responders
		50% percentile	0.90 - 0.93	0.002	Lower 50% percentile in responders
		60% percentile	0.98 - 1.03	0.005	Lower 60% percentile in responders
		70% percentile	1.06 - 1.19	0.01	Lower 70% percentile in responders
		80% percentile	1.14 - 1.31	0.02	Lower 80% percentile in responders
		Kurtosis	6.98 - 4.21	0.03	Higher kurtosis in responders
		GLCM Entropy	0.06 - 0.07	0.01	Lower entropy in responders
		10% percentile, 20% percentile, 30%		ns	

			percentile, 40% percentile, 90% percentile, Max, skewness, GLCM autocorrelation, GLCM Cluster prominence, GLCM Cluster shade, GLCM contrast, GLCM dissimilarity, GLCM energy, GLCM homogeneity 1, GLCM homogeneity 2, GLCM Max Probability, GLCM Sum Average, GLCM Sum Variance, GLCM Sum Entropy, GLCM Difference Variance, GLCM Difference Entropy, GLCM Information Measure Correlation 1, GLCM Information Measure Correlation 2				
Nougaret et al. (2016)[40]	GR	ADC	Minimum, 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile, Maximum, Skewness, Kurtosis			ns	
Palmisano et al. (2018)[69]	Response	Ve	Kurtosis	6.27 - 36.40		0.011	Lower kurtosis in responders
			Skewness	4.89 - 1.81		0.003	Higher skewness in responders
			25th percentile, 50th percentile, 75th percentile			ns	
		Ktrans	Mean, SD, kurtosis, skewness, 25th percentile, 50th percentile, 75th percentile			ns	
Park et al. (2020)[70]	GR	Multivariable LR	Min	cutoff: 45.0	OR 0.3 (0.1-0.8)	0.013	Lower min in responders
			SD	cutoff: 128.9	OR 3.2 (1.2-8.1)	0.016	Higher SD in responders
			Sphericity	cutoff: 0.4	OR 3.2 (1.3-8.0)	0.015	Higher sphericity in responders
			Compacity	cutoff: 3.2	OR 2.8 (1.1-7.1)	0.032	Higher compacity in responders
	DFS	Cox	GLRLM: LRLGE	cutoff: 0.006	2.3 (1.0-5.1)	0.046	Lower GLCM_LRLGE → better survival
	GR/CR/DFS	Shape	Volume(mL), volume (#vx), sphericity, compacity			ns	
		Histogram	Min, mean, std, max, skewness, kurtosis, entropy_log10, entropy_log2, energy			ns	
		GLCM	Homogeneity, energy, contrast, correlation, entropy_log10, entropy_log2, dissimilarity			ns	
		GLRLM	SRE, LRE, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, GLNU, RLNUU, RP			ns	
		NGLDM	Coarseness, contrast, busyness			ns	
		GLZLM	SZE, LZE, LGZE, HGZE, SZLGE, SZHGE, LZLGE, LZHGE, GLNU, ZLNU, ZP			ns	
Petkovska et al. (2020)[71]	CR	MRI T2W	Rad score; cutoff 1.39		AUC 0.75 (0.63–0.87)	0.001	Lower rad-score was associated with CR
		Features included in model	Surface area	coeff 0.28			Lower surface area in responders
			Compactness	coeff -0.02			Higher compactness in responders
			GLCM difference variance	coeff -0.0005			Higher GLCM difference variance in

						responders
			GLSZM low-gray level emphasis	coeff 0.10		Lower GLSZM low-gray level emphasis in responders
			SD Gabor σ=2, θ=30°	coeff 0.11		Lower SD Gabor σ=2, θ=30° in responders
			Kurtosis Gabor σ=2√2, θ=30°	coeff 0.05		Lower Kurtosis Gabor σ=2√2, θ=30° in responders
Petresc et al. (2020)[72]	non-response		Log5.0_GLSZM_SmallAreaEmphasis	AUC 0.80		
			wav-LHL_GLCM_Correlation	AUC 0.74		
			wav-LHL_10Percentile	AUC 0.71		
			wav-HHL_GLCM_MCC	AUC 0.69		
			wav-HHL_GLCM_Imc1	AUC 0.75		
			wav-HHL_Kurtosis	AUC 0.69		
			wav-HHL_GLSZM_SmallAreaHighGrayLevelEmphasis	AUC 0.71		
		Rad score	Log5.0_GLSZM_SmallAreaEmphasis, wav-LHL_GLCM_Correlation, wav-LHL_10Percentile, wav-HHL_GLCM_MCC, wav-HHL_GLCM_Imc1, wav-HHL_Kurtosis, wav-HHL_GLSZM_SmallAreaHighGrayLevelEmphasis	AUC 0.94 (0.82-0.99)	<0.001	Lower Rad-Score in responders
		Combined (Radscore+clinical)	Radscore, tumor length, tumor differentiation, grade (poor), MRF positive	OR 6.52 (1.87-22.72) AUC 0.80 (0.58-0.94)	0.003	Lower Rad-Score in responders
Rahmim et al. (2019)[104]	OS	PET - univariate	MTV; cutoff 9.3	2.6 (1.4-5.0)	0.003	Lower MTV → better OS
			TLG; cutoff 58.3	2.6 (1.4-5.0)	0.003	Lower TLG → better OS
			SUVpeak; cutoff 6.8	2.1 (1.09-3.9)	0.027	Lower SUVpeak → better OS
			SUVmean, SUVmax		ns	
		PET - multivariable	AUC-IVH, Number of CRLM, Liver-therapy-3mon-prior,	4.3 (2.2-8.6)	<0.001	
	PFS	PET - univariate	MTV, TLG, SUVpeak, SUVmean, SUVmax		ns	
		PET - multivariable	SUVmax, Number of CRLM	4.0 (1.7-9.7)	0.002	
	EFS	PET - univariate	MTV; cutoff 9.3	2.3 (1.2-2.2)	0.009	Lower MTV → better EFS
			TLG; cutoff 58.3	2.3 (1.2-4.2)	0.009	Lower TLG → better EFS
			SUVpeak, SUVmean, SUVmax		ns	
		PET - multivariable	MTV, uniformity, Number CRLM	3.2 (1.7-5.9)	<0.001	
Rao et al. (2014)[100]	New CRLM	Histogram, univariate	Entropy0, Uniformity0, meangreylevelintensity0, entropy0.5, uniformity0.5, meangreylevelintensity0.5, entropy1.5, uniformity1.5, meangreylevelintenstiy1.5, entropy2.5, uniformity2.5, meangreylevelintensity2.5		ns	

Rao et al. (2016)[99]	Response	Histogram, univariate	Mean intensity, mean intensity0.5, mean intensity1.5, mean intensity2.5, entropy, entropy0.5, entropy 1.5, entropy2.5, uniformity, uniformity0.5, uniformity1.5, uniformity2.5			ns	
Ravanelli et al. (2019)[108]	GR RECIST	Univariate - CTx+b	Uniformity_1	0.37 - 0.46		<0.001	Lower uniformity in responders
			Uniformity_1.5	0.54 - 0.64		<0.001	Lower uniformity in responders
			Uniformity_2	0.61 - 0.75		<0.001	Lower uniformity in responders
			Uniformity_2.5	0.67 - 0.82		<0.001	Lower uniformity in responders
			Uniformity_3	0.70 - 0.86		<0.001	Lower uniformity in responders
			Uniformity_3.5	0.73 - 0.87		<0.001	Lower uniformity in responders
			Uniformity, CT density			ns	
	GR RECIST	Univariate – CTx only	CT density	60.5-19.4		0.02	Higher CT density in responders
			Uniformity, Uniformity_1, Uniformity_1.5, Uniformity_2, Uniformity_2.5, Uniformity_3, Uniformity_3.5, CT density			ns	
	OS	Multivariable - CTx+b	CT density; cutoff > 53 HU		RR 3.7	0.028	Higher density → Better OS
			Uniformity; cutoff > 0.42		RR 6.9	0.005	Lower uniformity → Better OS
	PFS	Multivariable - CTx+b	Uniformity; cutoff > 0.42		RR 5.1	0.004	Lower uniformity → Better PFS
	GR RECIST	Multivariable - CTx+b	Uniformity; cutoff > 0.42		OR 20.0	0.01	Lower uniformity in responders
	GR Choi	Multivariable - CTx+b	Uniformity σ=1		Training AUC 0.89 Validation AUC 0.81		Lower uniformity in responders
Shaish et al. (2020)[109]	GR	MRI T2W: features included in model	Com_exp_firstorder_10Percentile, Com_exp_GLSZM_GrayLevelNonUniformity, Com_exp_GLSZM_ZoneVariance, Com_log_firstorder_Minimum, Com_orig_shape_Elongation, Com_wav-HHH_firstorder_Kurtosis, Com_wav-HHH_GLSZM_GrayLevelNonUniformityNormalized, Com_wav-HHH_GLSZM_HighGrayLevelZoneEmphasis, Com_wav-HHH_GLSZM_SizeZoneNonUniformityNormalized, Com_wav-HHL_GLSZM_HighGrayLevelZoneEmphasis, Com_wav-HLH_firstorder_Entropy, Com_wav-HLH_GLSZM_GrayLevelNonUniformityNormalized, Com_wav-HLH_GLSZM_HighGrayLevelZoneEmphasis, Com_wav-HLL_firstorder_Mean, Com_wav-HLL_GLSZM_GrayLevelNonUniformity		AUC 0.80		

			Normalized, Com_wav-LHH_GLSZM_HighGrayLevelZoneEmphasis, Com_wav-LLH_firstorder_Entropy, Com_wav-LLH_GLRLM_GrayLevelNonUniformity Normalized Tum_grad_firstorder_Kurtosis, Tum_lbp-3D-m1_firstorder_10Percentile, Com_wav-HLL_firstorder_Mean, Tum_logarithm_ngtdm_Busyness, Tum_orig_shape_Elongation, Tum_wav-HHH_GLSZM_GrayLevelNonUniformity Normalized, Tum_wav-HHH_GLSZM_HighGrayLevelZoneEmphasis, Tum_wav-HHH_GLSZM_SizeZoneNonUniformityNormalized, Tum_wav-HHL_firstorder_Mean, Tum_wav-HHL_firstorder_Skewness, Tum_wav-HHL_GLSZM_SizeZoneNonUniformityNormalized, Tum_wav-LHH_GLSZM_HighGrayLevelZoneEmphasis, LVI Stage, Sequence Institute, NEX, Bandwidth TE, Field strength, Tum_wav-LHL_GLSZM_GrayLevelNonUniformity Normalized, Tum_wav-LLL_GLSZM_SmallAreaEmphasis Age Com_exponential_GLSZM_ZoneVariance Com_wav-HHH_GLSZM_ SizeZoneNonUniformityNormalized, Com_wav-HHL_firstorder_Mean, Com_wav-HLH_GLSZM_ GrayLevelNonUniformityNormalized, Com_wav-LHH_firstorder_Mean, Com_wav-LHH_firstorder_Median Tum_lbp-3D-m1_firstorder_10Percentile, Tum_lbp-3D-m2_firstorder_90Percentile, Tum_wav-HHH_firstorder_Mean, Tum_wav-LHH_GLSZM_HighGrayLevelZoneEmphasis, Stage, Sequence			
	CR	MRI T2W: features included in model		AUC 0.80		
Shayesteh et al. (2019)[73]	GR	Univariate	GLCLM: Dissimilarity, Sum Average, Inter Quartile Range, Cluster Tendency, Variance, Cluster prominence		ns	
		Ensemble classifier model	SVM.NN.BN.KNN σ=0	AUC 0.90		Accuracy 89%
			SVM.NN.BN.KNN σ=0.5	AUC 0.95		Accuracy 90%
			SVM.NN.BN.KNN σ=1.5	AUC 0.94		Accuracy 91%
Shayesteh et al. (2020)[74]	GR	MRI T1W - univariate	Max probability	AUC 0.61 (0.55-0.66)	<0.001	
			Percentile	AUC 0.60 (0.54-	0.001	

					.066)		
			Long-run high gray-level		AUC 0.59 (0.53-0.61)	0.003	
			IDMn, Percentile area			ns	
		MRI T2W univariate	Dissimilarity		AUC 0.65 (0.58-0.71)	<0.001	
			Sum average		AUC 0.64 (0.58-0.70)	<0.001	
			Inter-quartile range		AUC 0.63 (0.55-0.68)	0.002	
			Cluster tendency, Variance, Cluster Prominence			ns	
		T1W	Bayesian network σ=0		AUC 0.64 (0.57-0.68)	<0.001	
			Random sub space σ=0.5		AUC 0.59 (0.52–0.63)	0.001	
			Random sub space σ=1.0		AUC 0.79 (0.75–0.83)	0.002	
		T2W	Naïve Bayesian network σ=0.5		AUC 0.85 (0.77–0.89)	<0.001	
			Naïve Bayesian network σ=1.5		AUC 0.81 (0.77–0.84)	<0.001	
			Iterative classifier optimizer σ=0		AUC 0.72 ((0.61-0.77)	<0.001	
			Randomizable filtered classifier σ=1.5		AUC 0.68 (0.62-0.74)	<0.001	
		T1W/T2W	Bayesian network σ 0.5-1.0-1.5, Naïve Bayesian network σ=0-1.0, Ada boost M1 σ=0-0.5-1.0-1.5, Iterative classifier optimizer σ=0.5-1.0-1.5, Logit Boost σ=0-0.5-1.0-1.5, Randomizable filtered classifier σ=0-0.5-1.0, Random sub space σ=0-1.5, Random forest σ=0-0.5-1.0-1.5, K logistic model tree σ=0-0.5-1.0-1.5			ns	
Shen et al. (2020)[75]	CR	PET / RF model	Contrast-N, LRE, Diversity-D4, LGRE, Information measure of correlation , Diversity-D2, Contrast-G, LRLGE, Sum average, SZE, Sum variance, LZE, LZLGE, Difference variance, ZLNU, LGHRE, Coarseness, Sum of squares variance, Information measure of correlation 2, RLNUU, Busyness, MTV, Strength, Complexity, Autocorrelation, SRHGE, GLNUr	Cutoff: 0.30	Training AUC 0.94 Validation ns		Training: Accuracy 95% Validation: Accuracy 89%
	CR	PET/ Logistic regression Univariable	short-run high gray-level emphasis (SRHGE)		OR 1.0 (1.0-1.1) AUC 0.68		RF model performed better than logistic regression model (p<0.001)
			MTV		AUC 0.32	0.007	
			GLCM_Contrastg		AUC 0.65	0.026	
			GLCM_Dissimilarity		AUC 0.65	0.022	
			GLCM_homogeneitym		AUC 0.31	0.005	
			GLCM_homogeneityp		AUC 0.32	0.007	
			Sum of squares variance		AUC 0.68	0.007	

			Sum average	AUC 0.68	0.006
			Sum variance	AUC 0.68	0.006
			Information measure of correlation1	AUC 0.30	0.002
			Information measure of correlation 2	AUC 0.68	0.006
			Inverse difference normalized	AUC 0.32	0.005
			IDM normalized	AUC 0.33	0.011
			GLRLM_SRE	AUC 0.69	0.005
			GLRLM_LRE	AUC 0.32	0.006
			GLRLM_GLNUr	AUC 0.29	0.002
			GLRLM_RP	AUC 0.36	0.030
			GLRLM_RLNUU	AUC 0.32	0.008
			GLRLM_LGRE	AUC 0.31	0.005
			GLRLM_HGRE	AUC 0.68	0.008
			GLRLM_SRLGE	AUC 0.32	0.005
			GLRLM_SRHGE	AUC 0.68	0.006
			GLRLM_LRLGE	AUC 0.29	0.001
			GLRLM_LRHGE	AUC 0.65	0.028
			NGLDM_Coarseness	AUC 0.71	0.002
			NGLDM_$Contrast_N$	AUC 0.69	0.005
			NGLDM_Busyness	AUC 0.29	0.001
			NGLDM_Complexity	AUC 0.68	0.006
			NGLDM_Strength	AUC 0.66	0.017
			GLSZM_SZE	AUC 0.64	0.038
			GLSZM_LZE	AUC 0.31	0.003
			GLSZM_GLNUz	AUC 0.32	0.006
			GLSZM_ZLNU	AUC 0.36	0.03
			GLSZM_HGZE	AUC 0.65	0.022
			GLSZM_LZHGE	AUC 0.26	<0.001
			SUVmax, Mean, Median, Variance, Std. Dev., Skewness, Kurtosis, 25th percentile, 75th percentile, Peak, TLGmax, TLGmean, TLGpeak, Total, Entropy, Energy, DiversityD2, DiversityD3, DiversityD4, GLCM_Autocorrelation, , GLCM_Correlation, GLCM_Cluster prominence, GLCM_Cluster shade, GLCM_Energy, GLCM_Entropy, GLCM_Sum entropy, GLCM_Difference variance, GLCM_Difference entropy, GLSZM_ZP, GLSZM_LGZE, GLSZM_SZLGE, GLSZM_SZHGE, GLSZM_LZHGE		ns
Shi et al. (2019)[76]	CR	T1W+T2W	T1 correlation (AUC=0.62), T2 homogeneity1 (AUC=0.65), T1 dissimilarity (AUC=0.0.69), T2 information measure2 (AUC=0.70), T1 homogeneity1 (AUC=0.70)	AUC 0.72	
		DWI (ADC)	Dissimilarity (AUC=0.65), sum variance (AUC=0.69), kurtosis (AUC=0.72), entropy (AUC=0.73), 50% (AUC=0.75)	AUC 0.75	
		DCE (LAVA2)	Information measure1 (AUC=0.57), 60% (AUC=0.67), 70% (AUC=0.70),	AUC 0.76	

			correlation (AUC=0.73), dissimilarity (AUC=0.75), entropy (AUC=0.76)				
		All radiomics	DCE 60% (AUC=0.60), T2 correlation (AUC=0.65), DWI sum variance (AUC=0.71), DWI 60% (AUC=0.74), DCE information measure1 (AUC=0.76)		AUC 0.78		
		Radiomics + ADC/DCE enhancements	T2 correlation (AUC=0.62), DWI sum variance (AUC=0.69), ADC (AUC=0.74), DCE 60% (AUC=0.77), DCE information measure1 (AUC=0.78), DWI 50% (AUC=0.79)		AUC 0.80		
		CNN			AUC 0.59		
	GR	T1W+T2W	T2 correlation (AUC=0.62), T1 information measure1 (AUC=0.69), T1 entropy (AUC=0.71), T1 contrast (AUC=0.72), T2 information measure1 (AUC=0.72)		AUC 0.74		
		DWI (ADC)	homogeneity2 (AUC=0.59), dissimilarity (AUC=0.64), 40% (AUC=0.69), information measure1 (AUC=0.74), contrast (AUC=0.76)		AUC 0.76		
		DCE (LAVA2)	dissimilarity (AUC=0.72), entropy (AUC=0.79), kurtosis (AUC=0.83), sum variance (AUC=0.84), 60% (AUC=0.85)		AUC 0.85		
		All radiomics	DCE dissimilarity (AUC=0.72), DWI 40% (AUC=0.81), DCE kurtosis (AUC=0.85), T2 information measure1 (AUC=0.87), T1 entropy (AUC=0.87)		AUC 0.88		
		Radiomics +ADC/DCE enhancements	DCE dissimilarity (AUC=0.72), Volume (AUC=0.83), ADC (AUC=0.87), DCE kurtosis (AUC=0.89), T2 information measure1 (AUC=0.90), T1 entropy (AUC=0.90)		AUC 0.91		
		CNN			AUC 0.47		
Shu, Fang, et al. (2019)[77]	CR	Histogram - univariate	Energy	0.008 - 0.012		0.016	Lower energy in responders
			Skewness	0.2 - 0.7		0.043	Lower skewness in responders
			Entropy	7.1 - 6.8		0.03	Higher entropy in responders
			Kurtosis, variance, uniformity			ns	
		Histogram - multivariable	Energy (cutoff > 0.0092)		OR 11.1 (1.8–68.4) AUC 0.75	0.01	Lower energy in responders
			skewness, entropy			ns	
Shur et al. (2019)[105]	DFS	MRI-T2W	Radiomics model; cutoff -0.71			ns	
			Combined clinical - radiomics model; cutoff -0.28			0.0007	Lower score → better DFS
		Features included in model	Pixel minimum		1.7 (1.3-2.2)	<0.001	Lower pixel minimum → better DFS
			GLSZM_small area emphasis		0.6 (0.5-0.8)	0.001	Higher GLDZM small area emphasis → better DFS

Simpson et al. (2017)[101]	HDFS	Histogram, univariate	FLR homogeneity		70.2 (3.6-1384)	0.005	lower FLR homogeneity → better HDFS
			FLR contrast		0.5 (0.3-1.0)	0.05	Higher FLR contrast → better HDFS
			FLR energy		6.4 (1.7-23.3)	0.005	Lower FLR energy → better HDFS
			FLR entropy		0.2 (0.04-0.7)	0.017	Higher FLR entropy → better HDFS
			Tumor correlation		6.0 (1.5-24.0)	0.011	Lower tumor correlation → better HDFS
			FLRcorrelation, tumor energy, tumor entropy			ns	
		Histogram, multivariable	FLR texture signal		2.2 (1.2-4.0)	0.01	Lower texture signal → better HDFS
			Tumor texture signal			ns	
	OS (death	Histogram, univariate	FLR homogeneity		18.7 (1.3-271.1)	0.032	lower FLR homogeneity → better survival
			FLR contrast		0.6 (0.3-1.0)	0.055	higher FLR contrast → better survival
			FLR energy		3.2 (0.9-11.0)	0.065	lower FLR energy → better survival
			FLR entropy		0.2 (0.1-0.7)	0.014	higher FLR entropy → better survival
			tumor correlation		5.2 (1.6-17.4)	0.007	lower tumor correlation → better survival
			tumor homogeneity		8.5 (1.1-63.1)	0.037	Lower tumor homogeneity → better survival
			tumor contrast		0.7 (0.6-0.9)	0.009	Higher tumor contrast → better survival
			FLRcorrelation, tumor energy, tumor entropy			ns	
		Histogram, multivariable	FLR texture signal		2.2 (1.1-4.3)	0.029	Lower texture signal → better survival
			Tumor texture signal		2.4 (1.2-4.6)	0.013	Lower texture signal → better survival
Taghavi et al. (2020)[103]	New CRLM	CT: Radiomics RF model	median_LoGσ=1, median_LoGσ=2.5, median_WavLHH, GLDM_SmallDependenceHighGrayLevel Emphasis_Exponential		AUC 0.86 (0.85-0.87)		
		CT: Combined RF model	CEA level, adjuvant chemotherapy, neoadjuvant chemotherapy, median_LoGσ=1, median_LoGσ=2.5, GLDM_SmallDependenceHighGrayLevel Emphasis_Exponential, GLDM_SmallDependenceHighGrayLevel Emphasis_WavHLL		AUC 0.86 (0.85-0.87)		
Tang et al. (2019)[39]	GR	DWI	WavD8_GLRLM_LGRE			0.055	
			WavF1_NGTDM_contrast			<0.01	
			NGTDM_coarseness			<0.01	
			GLSZM_SZHGE			<0.01	
Van Helden et al. (2018)[110]	Response (RECIST 1.1)	1st line treatment	Mean entropy	5.27 - 5.38		0.04	Entropy lower in responders
			Mean SUV max, mean SUVpeak, mean SUVmean, Mean compactness, mean sphericity, Mean AUC-CSH, mean entropy FXD, mean MATV, SUM MATV, mean TLG, SUM TLG			ns	
		3rd line	Mean SUV max, mean SUVpeak, mean			ns	

	treatment	SUVmean, Mean compactness, mean sphericity, Mean AUC-CSH, mean entropy, mean entropy FXD, mean MATV, SUM MATV, mean TLG, SUM TLG			
PFS	1st line treatment, univariate	Mean AUC-CSH	0.9 (0.8-0.97)	0.02	Higher mean AUC-CSH → better PFS
		Mean MATV	1.2 (1.0-1.3)	0.01	Lower mean MATV → better PFS
		SUM MATV	1.1 (1.0-1.3)	0.02	Lower SUM MATV → better PFS
		Mean TLG	1.2 (1.0-1.3)	0.02	Lower mean TLG → better PFS
		SUM TLG	1.1 (1.0-1.3)	0.05	Lower Sum TLG → better PFS
		Mean SUVmax, Mean SUVpeak, Mean SUVmean, Mean compactness, mean sphericity, Mean entropy, Mean entropy FXD		ns	
	Multivariable			ns	none of the features remained significant predictors
OS (death)	1st line treatment, univariate	Mean MATV	1.2 (1.1-1.4)	<0.01	Lower mean MATV → better survival
		SUM MATV	1.2 (1.0-1.4)	0.01	Lower SUM MATV → better survival
		Mean TLG	1.2 (1.1-1.4)	0.01	Lower mean TLG → better survival
		SUM TLG	1.2 (1.0-1.4)	0.02	Lower sum TLG → better survival
		Mean AUC-CSH	0.8 (0.7-0.9)	<0.01	Higher mean AUC-CSH → better survival
		Mean SUVmax, Mean SUVpeak, Mean SUVmean, Mean compactness, mean sphericity, Mean entropy, Mean entropy FXD		ns	
	Multivariable	Mean AUC-CSH	0.6 (0.5-0.9)	0.016	Higher mean AUC-CSH → better survival
		SUM MATV	2.6 (1.0-6.9)	0.048	Lower SUM MATV → better survival
PFS	3rd line treatment, univariate	Mean MATV	1.3 (1.1-15)	0.02	Lower mean MATV → better PFS
		SUM MATV	1.4 (1.1-1.7)	0.01	Lower SUM MATV → better PFS
		Mean TLG	1.3 (1.1-1.6)	0.01	Lower mean TLG → better PFS
		SUM TLG	1.3 (1.1-1.5)	0.01	Lower Sum TLG → better PFS
		Mean SUVmax, Mean SUVpeak, Mean SUVmean, Mean compactness, mean sphericity, Mean AUC-CSH, Mean entropy, Mean entropy FXD		ns	
	multivariable	mean MATV	1.4 (1.0-1.8)	0.03	lower mean MATV → better PFS
		SUM MATV	1.4 (1.1-1.9)	0.01	lower SUM MATV → better PFS
		Mean TLG	1.4 (1.0-1.8)	0.016	lower mean TLG → better PFS
		SUM TLG	1.4 (1.0-1.8)	0.03	lower Sum TLG → better PFS
OS (death)	3rd line treatment, univariate	Mean SUVmax	1.2 (1.0-1.4)	0.03	Lower mean SUVmax → better survival
		Mean SUVpeak	1.2 (1.0-1.5)	0.04	Lower mean SUVpeak → better survival
		SUM MATV	2.0 (1.4-3.1)	<0.01	Lower SUM MATV → better survival
		Mean TLG	1.5 (1.2-2.1)	<0.01	Lower mean TLG → better survival
		SUM TLG	1.8 (1.2-2.6)	<0.01	lower SUM TLG → better survival

			Mean SUVmean, Mean compactness, mean sphericity, Mean AUC-CSH, Mean entropy, Mean entropy FXD		ns	
		multivariable	mean MATV	2.4 (1.4-4.3)	<0.01	lower mean MATV → better PFS
			SUM MATV	2.5 (1.5-4.2)	<0.01	lower SUM MATV → better PFS
			Mean TLG	1.7 (1.2-2.5)	<0.01	lower mean TLG → better PFS
			SUM TLG	1.7 (1.2-2.5)	<0.01	lower Sum TLG → better PFS
Van Griethuysen et al. (2020)[78]	GR	MRI –T2W/DWI (expert_segm1)	DWIb1000_gradient_NGTDM_busyness, DWIb0_LoG1_Root mean squared, ADC_LoG3_GLSZM_gray level non uniformity normalized, DWIb1000_logarithm_energy, DWIb0_exp_skewness, T2W_kurtosis, ADC_logarithm_GLRLM_short run emphasis, DWIb1000_robust mean absolute deviation, ADC_LoG5_NGTDM_strength DWIb0_grad_GLCM_cluster shade, DWIb0_LoG1_GLRLM_run length non uniformity normalized,	AUC 0.79 (0.66–0.93)	0.001	
		MRI –T2W/DWI (expert_segm2)	T2W_LoG1_kurtosis, DWIb0_exp_median, ADC_LoG3_GLSZM_gray level non uniformity, T2W_LoG5_10th percentile, ADC_logarithm_GLDM_dependence variance, DWIb1000_grad_entropy, T2W_logarithm_interquartile range DWIb0_grad_NGTDM_complexity, DWIb1000_LoG3_GLDM_dependence non uniformity, T2W_kurtosis,	AUC 0.69 (0.52–0.86)	0.028	
		MRI –T2W/DWI (non expert_segm1)	DWIb1000_LoG1_skewness, T2W_LoG5_kurtosis, DWIb0_exp_GLRLM_gray level non uniformity normalized, DWIb1000_grad_GLCM_inverse difference normalized, DWIb1000_grad_entropy, ADC_logarithm_GLRLM_short run emphasis ADC_LoG3_NGTDM_complexity, ADC_grad_GLSZM_large area emphasis, ADC_exp_GLDM_dependence non uniformity normalized, T2W_kurtosis,	AUC 0.78 (0.64–0.92)	0.001	
		MRI –T2W/DWI (non expert_segm2)	T2W_LoG5_10th percentile, ADC_logarithm_GLDM_Large dependence emphasis, DWIb1000_LoG3_NGTDM_busyness, T2W_logarithm_rubust mean absolute deviation, DWIb1000_LoG5_NGTDM_complexity	AUC 0.70 (0.53–0.86)	0.024	
		MRI –T2W/DWI (semi-	DWIb0_LoG5_GLCM_correlation, DWIb0_gradient_NGTDM_strength,	AUC 0.60 (0.43–0.78)	ns	

	auto_segm)	DWIb0_LoG1_maximum, DWIb1000_GLDM_gray level non uniformity, T2W_kurtosis, DWIb0_exp_GLRLM_Low gray level run emphasis, ADC_logarithm_GLSZM_size zone non uniformity normalized, DWIb1000_LoG5_variance, T2W_logarithm_robust mean absolute deviation		
CR	MRI –T2W/DWI (expert_segm1)	ADC_exp_GLDM_dependence, nonuniformity normalized, DWIb100_exp_GLSZM_gray level nonuniformity, ADC_LoG3_GLSZM_large area low gray level emphasis, DWIb0_exp_GLRLM_low gray level run emphasis, DWIb0_exp_NGTDM_busyness, ADC_Root mean squared, ADC_LoG3_GLCM_inverse difference normalized, DWIb1000_LoG3_GLCM_inverse variance, DWIb0_LoG1_root mean squared	AUC 0.77 (0.58–0.96)	0.01
	MRI –T2W/DWI (expert_segm2)	ADC_Logarithm_Skewness, DWIb1000_LoG3_GLCM_clustershade, T2W_LoG1_kurtosis, DWIb1000_GLDM_gray level non uniformity, T2W_Logarithm_interquartile range, DWIb0_exp_median, DWIb0_LoG1_GLSZM_gray level variance, DWIb0_LoG1_Mean, DWIb1000_LoG5_NGTDM_busyness	AUC 0.69 (0.47–0.91)	ns
	MRI –T2W/DWI (non-expert_segm1)	ADC_exp_GLDM_dependence, nonuniformity normalized, DWIb1000_LoG3_robust mean absolute deviation, DWIb0_exp_uniformity, DWIb0_Log1_GLSZM_size zone non uniformity normalized, DWIb1000_exp_GLSZM_gray level non uniformity, DWIb0_exp_NGTDM_busyness, DWIb0_grad_GLCM_cluster prominence, DWIb1000_grad_NGTDM_busyness, T2W_logarithm_GLSZM_zone entropy	AUC 0.73 (0.51–0.94)	0.03
	MRI –T2W/DWI (non-expert_segm2)	T2W_LoG5_10th percentile, ADC_exp_minimum, DWIb1000_grad_GLCM_inverse difference normalzed, DWIb1000_GLDM_gray level non uniformity, DWIb0_exp_GLCM_joint energy, DWIb0_grad_GLCM_cluster shade, DWIb0_LoG3_GLCM_correlation,	AUC 0.66 (0.42–0.89)	ns

		MRI –T2W/DWI (semi-auto_segm)	DWIb1000_grad_90th percentile, DWIb0_LoG1_GLCM_inverse variance ADC_exp_GLDM_dependence non uniformity normalized, DWIb1000_LoG5_GLRLM_Long run high gray level emphasis, ADC_LoG3_GLRLM_long run emphasis, DWIb1000_LoG1_GLSZM_gray level non uniformity, T2W_logarithm_GLSZM_zone entropy, ADC_LoG5_GLCM_joint average, DWIb0_grad_GLCM_cluster shade, DWIb0_LoG5_kurtosis, DWIb0_exp_GLRLM_Gray level non uniformity normalized	AUC 0.63 (0.42–0.84)	ns	
Vandendorpe et al. (2019)[79]	GR	CE-CT	Combined clinical - radiomics model (validation)	AUC 0.70		Higher score is associated with responders
	GR	CE-CT	Radiomics model (training)	OR 13.3 AUC 0.86 (0.73–0.98)		
		Features included in model	Mean_SSF0	coeff 0.02		Higher mean in responders (no filter)
			MPP_SSF0	coeff 0.02		Higher mean positive pixel in responders (no filter)
			Mean_SSF2	coeff 0.03		Higher mean in responders (fine filter)
			Entropy_SSF4	coeff 1.7		Higher entropy in responders (medium filter)
			Kurtosis_SSF5	coeff 0.5		Higher kurtosis in responders (medium filter)
			Mean_SSF6	coeff -.00003		Lower mean in responders (coarse filter)
	GR	Univariate	Mean_SSF0	65.4 - 57.4	0.014	Mean is higher in responders
			MPP_SSF0	67.6 - 60.0	0.012	Mean positive pixel is higher in responders
			Entropy_SSF2	5.2 - 5.0	0.022	Entropy is higher in responders (fine filter)
			Entropy_SSF3	5.0 - 4.8	0.017	Entropy is higher in responders (medium filter)
			Kurtosis_SSF3	0.5 - -0.1	0.012	Kurtosis is higher in responders (medium filter)
			Entropy_SSF4	4.9 - 4.7	0.002	Entropy is higher in responders (medium filter)
			Kurtosis_SSF4	0.3 - -0.3	0.006	Kurtosis is higher in responders (medium filter)
			Kurtosis_SSF5	0.1 - -0.5	0.002	Kurtosis is higher in responders (medium filter)
			Entropy_SSF6	4.6 - 4.4	0.016	Entropy is higher in responders (coarse filter)
			Kurtosis_SSF6	-0.1 - -0.5	0.004	Kurtosis is higher in responders (coarse filter)
			SD_SSF0, Entropy_SSF0, Skewness_SSF0, Kurtosis_SSF0, Mean_SSF2, SD_SSF2, MPP_SSF2, Skewness_SSF2, Kurtosis_SSF2, Mean_SSF3, SD_SSF3, MPP_SSF3, Skewness_SSF3, Mean_SSF4, SD_SSF4, MPP_SSF4, Skewness_SSF4,		ns	

			Mean_SSF5, SD_SSF5, Entropy_SSF5, MPP_SSF5, Mean_SSF6, SD_SSF6, MPP_SSF6, Skewness_SSF6	
Wang et al. (2019)[80]	RFS (=local control)	CT-radiomics	absolute_median, GLRLM_SRE, GLRLM_LRE, GLRLM_GLN, GLRLM_RLNU, GLRLM_RP, LL_GLCM_Contrast, LL_GLCM_Difference_variance, HH_GLCM_Information_measures_I, LL_GLRLM_RLNU, Minkowsky.Dimension_2, LH_GLRLM_RLNU, LH_GLRLM_RP, HL_GLRLM_RLNU, HL_GLRLM_RP, HH_GLRLM_GLN, HH_GLRLM_RLNU, HH_GLRLM_RP, LL_absolute_mean, LL_absolute_median, HL_absolute_median	AUC 0.56 (0.47 0.66)
		CT-combined	absolute_median, GLRLM_SRE, GLRLM_LRE, GLRLM_GLN, GLRLM_RLNU, GLRLM_RP, LL_GLCM_Contrast, LL_GLCM_Difference_variance, HH_GLCM_Information_measures_I, LL_GLRLM_RLNU, Minkowsky.Dimension_2, LH_GLRLM_RLNU, LH_GLRLM_RP, HL_GLRLM_RLNU, HL_GLRLM_RP, HH_GLRLM_GLN, HH_GLRLM_RLNU, HH_GLRLM_RP, LL_absolute_mean, LL_absolute_median, HL_absolute_median	AUC 0.65 (0.55 0.75)
	OS	CT-radiomics	absolute_median, GLRLM_SRE, GLRLM_LRE, GLRLM_GLN, GLRLM_RLNU, GLRLM_RP, LL_GLCM_Contrast, LL_GLCM_Difference_variance, HH_GLCM_Information_measures_I, LL_GLRLM_RLNU, Minkowsky.Dimension_2, LH_GLRLM_RLNU, LH_GLRLM_RP, HL_GLRLM_RLNU, HL_GLRLM_RP, HH_GLRLM_GLN, HH_GLRLM_RLNU, HH_GLRLM_RP, LL_absolute_mean, LL_absolute_median, HL_absolute_median	AUC 0.66 (0.89-0.72)
		CT-combined	absolute_median, GLRLM_SRE, GLRLM_LRE, GLRLM_GLN, GLRLM_RLNU, GLRLM_RP, LL_GLCM_Contrast, LL_GLCM_Difference_variance, HH_GLCM_Information_measures_I, LL_GLRLM_RLNUU,	AUC 0.73 (0.66 0.80)

			Minkowsky.Dimension_2, LH_GLRLM_RLNU, LH_GLRLM_RP, HL_GLRLM_RLNU, HL_GLRLM_RP, HH_GLRLM_GLN, HH_GLRLM_RLNU, HH_GLRLM_RP, LL_absolute_mean, LL_absolute_median, HL_absolute_median			
Yang et al. (2020)[81]	GR	MRI ADC / RF model	Entropy-mean, inverse variance, energy-mean, small area emphasis, ADC-min, ADC-mean, sdGa02, small gradient emphasis, age, size	Training AUC 0.84 Validation AUC 0.83		Accuracy 91.3% to predict response
Yi et al. (2019)[82]	GR	MRI T2W / SVM model	Mean (coeff 1.8), Perc_01 (coeff 3.2)	AUC 0.90 (0.80-0.99)		
			Clinicopathological features included in model: OB-test, cN-stage, age, distance_to_anus, pathological_type, red_cell_counts, platelet_counts, neutrophil_counts, monocyte_counts, globuline			
	CR	MRI T2W / SVM model	Mean (coeff 2.5), Variance (coeff -5.6), Perc_01 (coeff3.1), Perc_10 (coeff 9.2), Vert1_RLNUU (coeff -2.4), x135dr_RLNUU (coeff -4.5), WvEnLL_s1 (coeff 3.6), WavEnLL_s3 (coeff -1.2)	AUC 0.87 (0.75-0.99)		
			Clinicopathological features included : OB-test, cN-stage, age, distance_to_anus, pathological_type, red_cell_counts, platelet_counts, neutrophil_counts, monocyte_counts, globuline			
Yuan et al. (2020)[83]	GR	neCT-LR	Short (coeff 16.4), Log3norm_contrast (coeff 17.9), FD_SD (coeff -2.1), GTD-Cnorm TextureStrength (coeff -161.9), GRD-Coarseness (coeff 637.0), SZ_LIE (coeff -0.04), RL_LRHGE (coeff -1.0), CoCmax probability (coeff -3.2), ISLECoeff_Var (coeff 6.5), ILSLentropy (coeff 8.5)	Training AUC 0.90 Validation AUC 0.76	0.02	Accuracy 68% to predict GR (validation)
	CR	neCT-RF	Features included: n=8 (not specified)	AUC not provided		Accuracy 84% to predict CR (validation)
Zhang et al. (2018)[102]	Response (size changes)	Histogram	Variance	446.1 - 210.2	<0.001	Variance higher in responders
			Entropy1	5.69 - 5.47	0.008	Entropy higher in responders
		GLCM	GLCM ASM	0.96 - 0.98	<0.001	ASM lower in responders
			GLCM Entropy2	0.14 - 0.07	<0.001	Entropy higher in responders
			GLCM Contrast	0.20 - 0.08	<0.001	Contrast higher in responders
			GLCM Correlation	4.13 - 8.56	0.001	Correlation lower in responders
			GLCM IDM	0.96 - 0.98	<0.001	IDM lower in responders
		Histogram	Mean, skewness, kurtosis		ns	
Zhou et al. (2019)[84]	non-GR	Single sequence MRI	T1W LR	Validation AUC 0.63		
			T2W LR	Validation AUC		

			CE-T1W LR			0.63 Validation AUC	
			ADC LR			0.76 Validation AUC	
		Multiparametric MRI	T1W-T2W-CET1W-ADC-model			0.68 Training AUC 0.82 (0.75-0.89) Validation AUC 0.77 (0.608–0.937)	Accuracy 77% to predict non-GR (validation)
		T1W	GLCM_entropy_135	coeff 1.7	OR 3.28	0.006	
			LoG_1.5-GLCM_correlation1_0, LoG_3-GLCM_correlation1_90, LoG_3-_mean			ns	
		T2W	LoG_3-GLCM_dentropy_90			ns	
		CE-T1W	LoG_1-GLCM_correlation1_135	coeff 1.6	OR 1.8	0.050	
			LoG_1-GLCM_saverage_90	coeff 2.7	OR 3.0	0.001	
			GLCM_IMC1_0			ns	
		ADC	LoG_2-GLCM_mprobability_45, LoG_2-GLCM_mprobability_0, GLCM_homogeneity2_90, GLCM_dissimilarity_90, LoG_1-GLCM_correlation1_90, GLCM_cshade_45, GLCM_entropy_135, LoG_1_median, skewness			ns	
Zou et al. (2018)[85]	Response	3d Ktrans map CR non CR	Variance, skewness, kurtosis, uniformity, energy, entropy, correlation			ns	
		3d Ktrans map GR non GR	Variance, skewness, kurtosis, uniformity, energy, entropy, correlation			ns	

Abbreviations appendix B1

ADC apparent diffusion coefficient, *ASM* angular second moment, *AUC* area under the curve, *AUC-CSH* area under the curve of the cumulative SUV volume, *BN* Bayesian network, *CC* collaborative clustering, *CE* contrast enhanced, *CI* confidence interval, *coeff* coefficient, *CNN* convolutional neural network, *Com* mesorectal compartment, *COV* coefficient of variation, *CR* complete response, *CRLM* colorectal liver metastases, *CT* computed tomography, *CTx+b* chemotherapy combined with bevacizumab, *D-map* pure diffusion coefficient, *D*-map* pseudo-diffusion coefficient, *DCE* dynamic contrast enhanced, *DFS* disease free survival, *DL* deep learning, *DWI* diffusion weighted imaging, *EGFR* anti-epidermal growth factor, *exp* exponential, *f-map* perfusion fraction, *FD* fractal dimension, *FLR* future liver remnant, *GLCM* gray level co-occurrence matrix, *GLNU* gray level nonuniformity, *GLRLM* gray level run length matrix, *GLSZM* gray level size zone matrix, *GLZLM* gray level zone length matrix, *GR* good response, *grad* Gradient, *GTDC* gray tone difference coarseness, *HDFS* hepatic disease free survival, *HGLZE/HGZE* high gray level zone emphasis, *HGRE* high gray level run emphasis, *HHH/HLH/LHH/LLH/LLL* 3D wavelet transform band, *HR* hazard ratio, *HU* Hounsfield units, *HQ* high quality, *IAUGC* initial area under the gadolinium curve, *IDM* inverse difference moment, *IDMn* inverse difference moment normalized, *IMC* information measure correlation, *IV* intensity variability, *IVH* intensity volume histogram, *Kep* rate constant form extracellular extravascular space back into blood plasma, *KNN* K nearest neighbour, *Ktrans* transfer constant from the blood plasma into the extracellular extravascular space, *LASSO* least absolute shrinkage and selection operator, *LGLRE/LGRE* low gray level run emphasis, *LGLZE/LGZE* low gray level zone emphasis, *LR* logistic regression, *LRE* long run emphasis, *LRHGE* long-run high gray level emphasis, *LRLGE* long-run low gray level emphasis, *LZE* long zone emphasis, *LZHGE* long zone high gray level emphasis, *LZLGE* large zone low gray level emphasis, *LZSE* large zone size emphasis, *MATV* metabolic active tumour volume, *MCC* maximum correlation coefficient, *MPP* mean positive pixels, *MRF* mesorectal fascia, *MRI* magnetic resonance imaging, *MTV* metabolic tumour volume, *nCE* non contrast enhanced, *NGLDM* neighbouring gray level dependence matrix, *NGTDM* neighbouring gray tone difference matrix, *NN* neural network, *ns* not specified, *orig* original, *OR* odds ratio, *OS* overall survival, *PCA* principal component analysis, *Perc* percentile, *PET* positron emission tomography, *PFS* progression free survival, *pMRI* perfusion MRI, *RECIST* response evaluation criteria in solid tumours, *RF* random forest, *RFS* recurrence free survival, *RLNU* run length non uniformity, *RP* run percentage, *RR* relative risk, *SD* standard deviation, *segm* segmentation, *SQ* standard quality, *SRE* short run emphasis, *SRHGE* short run high gray level emphasis, *SRLGE*/*SRLGLE* short run low gray level emphasis, *SUV* standard uptake value, *SVM* support vector machine, *SZE* short zone emphasis, *SZHGE* small zone high gray emphasis, *SZLGE* small zone low gray emphasis, *SZSNU* small zone size nonuniformity, *SZV* size zone variability, *T1W* T1 weighted, *T2W* T2 weighted, *TLG* total lesion glycolysis, *Tum* Tumour, *Ve* volume of extravascular extracellular space per unit volume of tissue, *VP* venous phase sequence, *wav* Wavelet, *ZLNU* zone length nonuniformity, *ZP* zone percentage, *ZSNU* zone size nonuniformity, *ZSP* zone size percentage, *ZSV* zone size variance

Supplement A Stepwise explanation of the radiomics workflow

1. Image acquisition

A radiomics analysis can be performed on standard of care images and no additional imaging is necessary. However, it is important to keep in mind that imaging protocols are not always standardized, which results in variability between the images, even within institutions and patients. Therefore, the Image Biomarker Standardisation initiative (IBSI) has made recommendations for the pre-processing of the images before analysis which can reduce this variability.[1]

2. Identification and delineation of the ROI

Identification of the ROI can be done either manually, semi-automatically or automatically (by validated software).[2] The analysis can be performed both on tumour and on normal tissue, but the ROI has to be derived from the same tissue across the analysed cohort.[3]

3. Extraction of features

After segmentation of the ROI, the features can be extracted. These features are essentially values obtained from the images by mathematical algorithms and can be extracted by different tools, such as open source software (e.g. Texrad[4], MaZda[5], PyRadiomics[6], IBEX[7], CERR[8], CGITA[9]) or an in-house tool can be developed. Features can either be directly extracted from the 'original' medical images or after applying a filter or transformation method. In general, imaging filtration allows for better enhancement of edges, it can remove noise or provide information about the spatial location of image features. Commonly used filters are Gabor filters, Law's filters, Laplacian transform of Gaussian (LoG) filters, Wavelet and Fourier.[10] Different subgroups of features are defined which are described in short below (and in table 1).[3, 11, 12]

a) Shape features are structural features that describe the 3D shape and size of the ROI, for instance shape, volume, diameter and surface.
b) Histogram-based features (or first order features) are derived from the histogram of an image and describe the distribution of voxel intensities or densities within a ROI. They do not account for the location of the pixels or spatial interrelationship between pixels.

c) Texture features describe the spatial relationship between neighbouring voxels in a ROI.[13, 14] (In the medical literature, the term texture analysis is also used to describe histogram-based features that do not take the spatial interrelationship into account. For the purpose of this article, the term 'texture features' will only be used to describe features that take the spatial interrelationship into account). Various (different) groups of texture features are used in current scientific research to predict outcome such as: Grey Level Co-occurrence Matrix (GLCM), Gray Level Size Zone (GLSZM), Gray Level Run Length Matrix (GLRLM), Neighbouring Gray Tone Difference Matrix (NGTDM), Gray Level Dependence Matrix (GLDM)(table 1).[15, 16]

4. Feature selection

Feature selection is necessary to select only the most useful features in the dataset. In radiomics analyses the number of variables often exceeds the number of patients, while not all features might be useful and some features can be highly correlated with each other and therefore need to be excluded. Features selection methods can be divided into univariate or multivariate. Univariate methods select features based only on their association with a specific outcome. Multivariate methods investigate the interaction between features and can be done either supervised or unsupervised. Supervised (multivariate) methods try to find a subset of features that has the best discriminatory value of outcome. Unsupervised methods are based on feature reduction only and do not take outcome into account.[17]

5. Feature classification and data analysis

The last step is to combine the best performing features into a predictive model. Different approaches can be used to generate a predictive model such as logistic regression models or more complex (machine learning) models such as Bayesian statistics, support vector machine (SVM), random forest (RF) or deep neural network (DNN)). Then, the model needs to be validated on an independent dataset to prove its generalizability, which should ideally be performed on an external dataset.[3, 10]

References

1. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*. 2016.
2. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Seminars in Radiation Oncology*. 2019;29:185-197.
3. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*. 2017;14:749.
4. TexRAD. 2019, August 5.
5. MaZda. 2019, August 5.
6. Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. 2017;77:e104-e107.
7. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. ibex: An open infrastructure software platform to facilitate collaborative work in radiomics. 2015;42:1341-1353.
8. Apte AP, Iyer A, Crispin-Ortuzar M, et al. Technical Note: Extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research. *Medical physics*. 2018.
9. Fang Y-HD, Lin C-Y, Shih M-J, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. 2014;2014.
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278:563-577.
11. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clinical radiology*. 2004;59:1061-1069.
12. Larue RTHM, Defraene G, Ruysscher DD, Lambin P, Elmpt Wv. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. 2017;90:20160665.
13. Ergen B, Baykara M. Texture based feature extraction methods for content based medical image retrieval systems. *Biomedical materials and engineering*. 2014;24:3055-3062.
14. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*. 1973;SMC-3:610-621.
15. Forghani R, Savadjiev P, Chatterjee A, Muthukrishnan N, Reinhold C, Forghani B. Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Comput Struct Biotechnol J*. 2019;17:995-1008.
16. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: The promise of radiomics. *Physica Medica*. 2017;38:122-139.
17. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning methods for Quantitative Radiomic Biomarkers. *Scientific reports*. 2015;5:13087.

Supplement B

Supplementary table B1 QUADAS-2 results

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	Patient selection	Index test	Reference standard	Flow & timing	Patient selection	Index test	Reference standard
Ahn et al. (2016)[13]	?	n/a	?	?	?	n/a	?
Antunes et al. (2020)[41]	?	n/a	?	?	?	n/a	?
Badic et al. (2018)[86]	?	n/a	?	?	?	n/a	?
Badic et al. (2019)[91]	?	n/a	?	?	?	n/a	?
Bang et al. (2016)[42]	?	n/a	?	?	?	n/a	?
Beckers, Beets-Tan, et al. (2017)[92]	?	n/a	?	?	?	n/a	?
Beckers, Lambregts, et al. (2017)[93]	?	n/a	?	?	?	n/a	?
Beckers et al. (2018)[94]	?	n/a	?	?	?	n/a	?
Bibault et al. (2018)[43]	?	n/a	?	?	?	n/a	?
Boldrini et al. (2018)[44]	?	n/a	?	?	?	n/a	?
Bulens et al. (2020)[45]	?	n/a	?	?	?	n/a	?
Bundschuh et al. (2014)[14]	?	n/a	?	?	?	n/a	?
Caruso et al. (2018)[46]	?	n/a	?	?	?	n/a	?
Chee et al. (2017)[47]	?	n/a	?	?	?	n/a	?
Crimi et al. (2020)[48]	?	n/a	?	?	?	n/a	?
Cui et al. (2018)[49]	?	n/a	?	?	?	n/a	?
Cusumano et al. (2018)[50]	?	n/a	?	?	?	n/a	?
Dai et al. (2020)[90]	?	n/a	?	?	?	n/a	?
de Cecco et al. (2015)[51]	?	n/a	?	?	?	n/a	?
de Cecco et al. (2016)[52]	?	n/a	?	?	?	n/a	?
Dercle et al. (2020)[106]	?	n/a	?	?	?	n/a	?
Dinapoli et al. (2018)[53]	?	n/a	?	?	?	n/a	?
Dohan et al. (2020)[107]	?	n/a	?	?	?	n/a	?
Ferrari et al. (2019)[54]	?	n/a	?	?	?	n/a	?
Fu et al. (2020)[55]	?	n/a	?	?	?	n/a	?
Ganeshan et al. (2007)[96]	?	n/a	?	?	?	n/a	?
Giannini et al. (2019)[56]	?	n/a	?	?	?	n/a	?
Hamerla et al. (2019)[57]	?	n/a	?	?	?	n/a	?
Jalil et al. (2017)[58]	?	n/a	?	?	?	n/a	?
Lee et al. (2018)[95]	?	n/a	?	?	?	n/a	?
Li, Boimel, et al. (2019)[59]	?	n/a	?	?	?	n/a	?
Li, Wang, et al. (2020)[60]	?	n/a	?	?	?	n/a	?
Liang et al. (2016)[38]	?	n/a	?	?	?	n/a	?
Liang et al. (2019)[61]	?	n/a	?	?	?	n/a	?
Liu et al. (2017)[62]	?	n/a	?	?	?	n/a	?
Liu, Li, et al. (2019)[63]	?	n/a	?	?	?	n/a	?
Liu, Wen, et al. (2019)[64]	?	n/a	?	?	?	n/a	?

Lovinfosse et al. (2018)[65]	+	n/a	+	+	+	n/a	+
Lubner et al. (2015)[97]	+	n/a	+	?	+	n/a	+
Meng, Zhang, Dong, et al. (2018)[67]	+	n/a	+	+	+	n/a	+
Meng, Zhang, Zou, et al. (2018)[66]	+	n/a	+	+	+	n/a	+
Miles et al. (2009)[98]	+	n/a	+	+	+	n/a	+
Nakajo et al. (2017)[87]	+	n/a	+	?	+	n/a	+
Nardone et al. (2019)[68]	+	n/a	+	+	+	n/a	+
Ng, Ganeshan, et al. (2013)[88]	+	n/a	+	+	+	n/a	+
Ng, Kozarski, et al. (2013)[89]	+	n/a	+	+	+	n/a	+
Nie et al. (2016)[15]	?	n/a	+	+	+	n/a	+
Nougaret et al. (2016)[40]	+	n/a	+	+	+	n/a	+
Palmisano et al. (2018)[69]	+	n/a	+	+	+	n/a	+
Park et al. (2020)[70]	+	n/a	+	+	+	n/a	+
Petkovska et al. (2020)[71]	+	n/a	+	+	+	n/a	+
Petresc et al. (2020)[72]	+	n/a	+	+	+	n/a	+
Rahmim et al. (2019)[104]	+	n/a	+	+	+	n/a	+
Rao et al. (2014)[100]	+	n/a	+	-	+	n/a	+
Rao et al. (2016)[99]	+	n/a	+	-	+	n/a	+
Ravanelli et al. (2019)[108]	-	n/a	+	+	+	n/a	+
Shaish et al. (2020)[109]	+	n/a	+	+	+	n/a	+
Shayesteh et al. (2019)[73]	+	n/a	+	+	+	n/a	+
Shayesteh et al. (2020)[74]	+	n/a	+	+	+	n/a	+
Shen et al. (2020)[75]	+	n/a	+	+	+	n/a	+
Shi et al. (2019)[76]	?	n/a	+	+	?	n/a	+
Shu, Fang, et al. (2019)[77]	+	n/a	+	+	+	n/a	+
Shur et al. (2019)[105]	+	n/a	+	?	+	n/a	+
Simpson et al. (2017)[101]	+	n/a	+	+	+	n/a	+
Taghavi et al. (2020)[103]	+	n/a	+	-	+	n/a	+
Tang et al. (2019)[39]	+	n/a	+	+	+	n/a	+
Van Helden et al. (2018)[110]	+	n/a	+	+	+	n/a	+
Van Griethuysen et al. (2020)[78]	-	n/a	+	+	+	n/a	+
Vandendorpe et al. (2019)[79]	+	n/a	+	+	+	n/a	+
Wang et al. (2019)[80]	+	n/a	?	?	+	n/a	?
Yang et al. (2020)[81]	+	n/a	+	+	+	n/a	+
Yi et al. (2019)[82]	+	n/a	+	+	+	n/a	+
Yuan et al. (2020)[83]	+	n/a	+	+	+	n/a	+
Zhang et al. (2018)[102]	+	n/a	+	+	+	n/a	+
Zhou et al. (2019)[84]	+	n/a	+	+	+	n/a	+
Zou et al. (2018)[85]	+	n/a	+	+	+	n/a	+

Supplementary table B2 RQS results

Reference	1 Image protocol	2 Multiple segmentations	3 Phantom study	4 Multiple time points	5 Feature reduction for multiple testing	6 Multivariable analysis with non radiomics features	7 Biological correlates	8 Cut-off analysis	9 Discriminations statistics	10 Calibration statistics	11 Prospective	12 Validation	13 Gold standard	14 Clinical utility	15 Cost-effectiveness	16 Open data	Total	RQS
Score range	0 – 2	0 – 1	0 – 1	0 – 1	-3 – 3	0 – 1	0 – 1	0 – 1	0 – 2	0 – 2	0 – 7	-5 – 5	0 – 2	0 – 2	0 – 1	0 – 4	-12 – 36	0% – 100%
Ahn et al. (2016)[13]	1	0	0	0	3	0	0	1	1	0	0	2	2	0	0	0	10	28%
Antunes et al. (2020)[41]	1	1	0	0	3	0	0	0	2	0	0	4	2	0	0	0	13	36%
Badic et al. (2018)[86]	1	0	1	0	3	0	1	1	1	0	0	-5	0	0	0	0	3	8%
Badic et al. (2019)[91]	1	0	0	0	-3	1	1	0	1	0	0	-5	0	0	0	1	-3	0%
Bang et al. (2016)[42]	1	0	0	0	-3	0	1	0	1	0	0	-5	0	0	0	0	-5	0%
Beckers, Beets-Tan, et al. (2017)[92]	1	0	0	0	-3	1	1	0	1	0	0	-5	0	0	0	0	-4	0%
Beckers, Lambregts, et al. (2017)[93]	1	0	0	0	-3	0	1	0	0	0	0	-5	0	0	0	0	-6	0%
Beckers et al. (2018)[94]	1	0	0	0	3	1	0	0	0	0	0	-5	0	0	0	0	0	0%
Bibault et al. (2018)[43]	1	1	0	0	3	0	0	0	2	0	0	-5	2	0	0	0	4	11%
Boldrini et al. (2018)[44]	0	1	0	0	-3	0	0	0	0	0	0	-5	0	0	0	1	-6	0%
Bulens et al. (2020)[45]	1	1	0	0	3	0	0	0	2	0	0	0	0	0	0	1	8	22%
Bundschuh et al. (2014)[14]	1	0	0	0	-3	0	1	0	1	0	0	-5	0	0	0	0	-5	0%
Caruso et al. (2018)[46]	1	1	0	0	3	0	0	0	0	0	0	-5	0	0	0	0	0	0%
Chee et al. (2017)[47]	1	1	0	0	3	1	1	1	0	0	0	-5	0	0	0	0	3	8%
Crimi et al. (2020)[48]	1	1	0	0	-3	0	1	0	0	0	0	-5	0	0	0	0	-5	0%
Cui et al. (2018)[49]	1	1	0	0	3	1	1	0	2	1	0	2	2	2	0	0	16	44%
Cusumano et al. (2018)[50]	1	1	0	0	3	1	1	0	0	0	0	3	2	0	0	1	13	36%
Dai et al. (2020)[90]	1	0	0	0	3	1	0	1	2	0	0	-5	2	2	0	0	7	19%
de Cecco et al. (2015)[51]	1	0	0	0	-3	0	0	1	1	0	0	-5	0	0	0	0	-5	0%
de Cecco et al. (2016)[52]	1	0	0	0	-3	0	1	1	1	0	0	-5	0	0	0	0	-4	0%
Dercle et al. (2020)[106]	1	1	1	0	3	0	1	1	2	0	0	5	2	0	0	0	17	47%
Dinapoli et al. (2018)[53]	1	1	0	0	3	1	0	0	2	1	0	4	2	0	0	0	15	42%
Dohan et al. (2020)[107]	1	1	0	0	3	0	1	1	2	1	0	3	0	0	0	1	14	39%
Ferrari et al. (2019)[54]	1	0	0	0	3	1	0	0	1	0	0	2	2	2	0	1	13	36%
Fu et al. (2020)[55]	1	0	0	0	3	0	0	0	2	0	0	-5	0	0	0	1	2	6%
Ganeshan et al. (2007)[96]	0	0	0	0	-3	0	1	0	0	0	0	-5	0	0	0	0	-7	0%
Giannini et al. (2019)[56]	1	0	0	0	-3	0	0	1	1	0	0	-5	2	0	0	0	-3	0%
Hamerla et al. (2019)[57]	0	1	0	0	3	0	0	0	2	0	0	-5	0	0	0	1	2	6%

Jalil et al. (2017)[58]	0	1	0	0	-3	1	1	1	1	0	0	-5	0	0	0	1	-2	0%
Lee et al. (2018)[95]	1	1	0	0	3	0	0	0	1	0	0	-5	0	0	0	1	2	6%
Li, Boimel, et al. (2019)[59]	0	0	0	0	3	0	0	0	1	0	0	-5	0	0	0	0	-1	0%
Li, Wang, et al. (2020)[60]	1	1	0	0	3	1	0	0	2	1	0	2	2	0	0	0	13	36%
Liang et al. (2016)[38]	1	0	0	0	3	0	1	1	1	0	0	-5	2	0	0	0	4	11%
Liang et al. (2019)[61]	1	1	0	0	3	0	0	0	2	0	0	-5	2	0	0	0	4	11%
Liu et al. (2017)[62]	1	0	0	0	-3	0	0	0	1	0	0	-5	0	0	0	0	-6	0%
Liu, Li, et al. (2019)[63]	0	0	0	0	3	0	0	0	1	0	0	-5	0	0	0	0	-1	0%
Liu, Wen, et al. (2019)[64]	1	0	0	0	-3	0	1	0	1	0	0	-5	2	0	0	0	-3	0%
Lovinfosse et al. (2018)[65]	2	0	0	0	3	1	1	1	2	0	0	-5	0	0	0	0	5	14%
Lubner et al. (2015)[97]	0	1	0	0	3	0	1	0	1	0	0	-5	0	0	0	1	2	6%
Meng, Zhang, Dong, et al. (2018)[67]	1	1	0	0	3	0	0	1	0	0	0	-5	0	0	0	0	1	3%
Meng, Zhang, Zou, et al. (2018)[66]	1	1	0	0	3	1	1	1	2	0	0	2	2	0	0	0	14	39%
Miles et al. (2009)[98]	1	0	1	0	3	0	1	1	1	0	0	-5	0	0	0	0	3	8%
Nakajo et al. (2017)[87]	1	1	0	0	-3	1	0	1	0	0	0	-5	0	0	0	0	-4	0%
Nardone et al. (2019)[68]	0	1	0	0	-3	1	1	1	1	1	0	0	2	0	0	0	5	14%
Ng, Ganeshan, et al. (2013)[88]	1	0	0	0	-3	1	1	1	1	0	0	-5	0	0	0	0	-3	0%
Ng, Kozarski, et al. (2013)[89]	1	0	0	0	3	0	0	1	1	0	0	-5	0	0	0	0	1	3%
Nie et al. (2016)[15]	1	0	0	0	3	0	1	0	2	0	0	-5	2	0	0	0	4	11%
Nougaret et al. (2016)[40]	1	1	0	0	3	0	0	0	2	0	0	-5	0	0	0	0	2	6%
Palmisano et al. (2018)[69]	1	0	0	0	3	0	1	1	1	0	0	-5	0	0	0	0	2	6%
Park et al. (2020)[70]	1	0	0	0	-3	1	0	0	1	0	0	-5	0	0	0	0	-5	0%
Petkovska et al. (2020)[71]	1	1	0	0	3	1	1	1	2	0	0	-5	2	0	0	1	8	22%
Petresc et al. (2020)[72]	1	1	0	0	3	1	0	1	2	0	0	0	2	0	0	1	12	33%
Rahmim et al. (2019)[104]	0	0	0	0	3	1	0	0	1	0	0	-5	0	0	0	0	0	0%
Rao et al. (2014)[100]	1	0	0	0	-3	0	1	0	1	0	0	-5	0	0	0	0	-5	0%
Rao et al. (2016)[99]	0	0	0	0	-3	0	1	0	0	0	0	-5	0	0	0	0	-7	0%
Ravanelli et al. (2019)[108]	1	0	0	0	-3	0	1	1	2	0	0	-5	2	0	0	0	-1	0%
Shaish et al. (2020)[109]	1	1	0	0	3	0	0	0	1	0	0	2	2	0	0	1	11	31%
Shayesteh et al. (2019)[73]	1	1	0	0	3	0	0	0	2	0	0	-5	2	0	0	1	5	14%
Shayesteh et al. (2020)[74]	0	0	0	0	-3	1	0	0	2	0	0	-5	2	0	0	0	-3	0%
Shen et al. (2020)[75]	1	0	0	0	3	0	0	1	2	0	0	-5	2	0	0	0	4	11%
Shi et al. (2019)[76]	0	0	0	0	3	0	0	0	1	0	0	-5	2	0	0	1	2	6%
Shu, Fang, et al. (2019)[77]	1	1	0	0	3	0	0	0	1	0	0	-5	2	0	0	0	3	8%
Shur et al. (2019)[105]	1	1	0	0	3	1	1	1	2	0	0	-5	2	0	0	1	8	22%
Simpson et al. (2017)[101]	1	0	0	0	3	1	1	0	1	0	0	-5	0	0	0	0	2	6%

Taghavi et al. (2020)[103]	1	1	0	0	3	1	0	0	2	0	0	0	2	0	0	1	11	31%
Tang et al. (2019)[39]	1	1	0	0	3	1	1	0	2	2	0	2	2	2	0	0	17	47%
Van Helden et al. (2018)[110]	1	1	0	0	3	0	0	0	2	0	0	3	2	2	0	1	15	42%
Van Griethuysen et al. (2020)[78]	2	0	0	0	3	1	0	0	1	0	0	-5	0	0	0	0	2	6%
Vandendorpe et al. (2019)[79]	0	0	0	0	3	1	1	0	2	1	0	0	0	0	0	0	8	22%
Wang et al. (2019)[80]	1	1	0	1	3	1	0	0	2	0	0	2	0	0	0	0	11	31%
Yang et al. (2020)[81]	1	1	0	0	3	1	0	0	2	0	0	0	2	0	0	1	11	31%
Yi et al. (2019)[82]	0	1	0	0	3	1	1	0	1	0	0	2	2	0	0	0	11	31%
Yuan et al. (2020)[83]	0	1	0	0	3	0	0	0	1	0	0	2	2	0	0	0	9	9%
Zhang et al. (2018)[102]	1	1	0	0	-3	0	1	1	1	0	0	-5	0	0	0	0	-3	0%
Zhou et al. (2019)[84]	1	1	0	0	3	1	1	0	1	1	0	2	2	0	0	0	13	36%
Zou et al. (2018)[85]	1	0	0	0	3	0	1	1	0	0	0	-5	0	0	0	0	1	3%
Mean±SD	0.8 ±0.4	0.5 ±0.5	0.04 ±0.2	0.01 ±0.1	1.2 ±2.8	0.4 ±0.5	0.5 ±0.5	0.4 ±0.5	1.2 ±0.7	0.1 ±0.4	0 ±0	-2.9 ±3.3	0.9 ±1.0	0.1 ±0.5	0 ±0	0.3 ±0.4	3.5 ±6.8	13%