



SAPIENZA
UNIVERSITA' DI ROMA

DOTTORATO DI RICERCA IN MEDICINA SPERIMENTALE
XXXII CICLO

“Performance of texture analysis in
predicting tumoral response to neoadjuvant
chemoradiotherapy in rectal cancer patients
studied with 3T MR”

DOTTORANDO

Dott. Davide Bellini

DOCENTE GUIDA

Prof. A. Laghi

COORDINATORE DEL DOTTORATO
Prof. Maurizio Sorice

ANNO ACCADEMICO 2018-2019

Table of Contents

Introduction

Role of MRI in Rectal Cancer	5
Texture Analysis	17
Artificial Intelligence in Radiology	27
References	41

Experimental Study

Abstract	50
Background	51
Material and Methods	53
Results	58
Discussion	67
References	71

Introduction

Role of MRI in Rectal Cancer

Rectal cancer – defined as cancer (usually adenocarcinoma) occurring in the distal 15 cm of the intestinal tract as measured to or from the anal verge – is one of the major causes of cancer-related mortality worldwide. Although imaging can be suggestive of the diagnosis of rectal cancer, particularly when obstruction hinders endoscopic access or biopsy fails, the primary role of imaging is to assist in treatment triage of histologically diagnosed tumors. Surgical resection with negative margins (i.e., no tumor extension within 1 mm of the resected margins on histology) is the only standard locally curative therapy for rectal cancer. Failure to attain negative margins (i.e., positive postoperative margins) often results in tumor recurrence and the possibility of incurable disease, a poor quality of life, and reduced disease-free survival.

The initial local staging is performed to determine which patients require preoperative chemoradiation therapy (CRT) or to plan surgery in those not requiring CRT with the intent to obtain a negative margin. For tumors in the upper two thirds of the rectum, the standard procedure is low anterior resection (LAR) with total mesorectal excision where the rectum (except the distal portion) and the surrounding mesorectum are removed. For tumors in the distal one third of the rectum, depending on local extension, sphincter-sparing surgeries (e.g., ultra-LAR or intersphincteric resection) or abdominoperineal resection (APR) surgeries are attempted. Post-CRT staging aims at assessing treatment response; knowing the post-CRT stage is important for selecting further treatment such as surgical resection or extended CRT.

Transrectal ultrasound (TRUS) is an accurate imaging modality for differentiating T1 from T2 tumors and is similar to MRI in differentiating T2 from T3 tumors. However, at higher disease stages, MRI is better than TRUS in the assessment of the tumoral border and mesorectal fascia (MRF), surrounding viscera, and pelvic nodes. In addition, dynamic contrast-enhanced (DCE) MRI can provide functional information that may predict response to treatment or help detect recurrent

disease. CT is not suitable for T staging of rectal cancer because of its lower contrast resolution, but it is the preferred modality for detecting distant metastasis, especially when combined with PET. Currently, MRI is the imaging modality of choice for the local evaluation of rectal cancer.

Rectal MRI Techniques

MRI performed at a higher field strength benefits from faster image acquisition, higher spatial resolution, and higher signal-to-noise ratio (SNR), which may improve the visibility of the rectal wall.

New pelvic phased-array multichannel coils provide high spatial resolution, high SNR, and larger-FOV imaging for visualization of the lateral pelvic structures and lymph nodes. Endorectal coils are not recommended. Bowel preparation is generally not necessary before the examination, but antispasmodic agents are useful for decreasing bowel peristalsis and resultant motion artifacts. Filling of the rectal lumen with gel or contrast material probably facilitates the detection of small tumors. However, compression of the mesorectal fat due to rectal distention may critically alter the staging because it leads to underestimation of the distance of the tumor to MRF and possibly to nonvisualization of the mesorectal nodes; therefore, the routine use of endorectal filling is not recommended.

The rectal MRI includes multiplanar conventional and high-resolution oblique T2-weighted and axial T1-weighted pulse sequences and multiparametric MRI sequences including diffusion-weighted imaging (DWI) and contrast-enhanced MRI. The mandatory part of this protocol is T2-weighted imaging; the other sequences are optional when the MRI examination must be shortened. A sidewall- to-sidewall sagittal T2-weighted sequence provides the initial images for localizing the tumor, and axial and coronal T2-weighted imaging should be performed in the same manner. Then, high-resolution oblique T2- weighted images with thin (3 mm) slices and a large matrix size (e.g., 320 × 320) should be obtained perpendicular to the tumoral axis in the sagittal view in one or more planes depending on the size and shape of the tumor. High-resolution oblique imaging provides the optimal anatomic information for an improved assessment of the depth of invasion and of tumoral

relationships especially anteriorly and in relation to the sphincter complex and levator muscle in patients with low rectal tumors. The use of T1-weighted imaging for rectal cancer staging is recommended mainly for the evaluation of coincidental findings and the pelvic bones. DWI may help in the assessment of response to CRT and may improve the accuracy of MRI for the detection of rectal cancers and involved pelvic nodes.

Currently there is no agreement regarding the role of gadolinium-enhanced MRI in patients with rectal cancer. However, it may improve the detection of tumors and malignant lymph nodes, increase the accuracy of MRI for diagnosing T3 tumors and locoregional extensions, and help in the assessment of treatment response after CRT. In addition, the DCE-MRI-derived quantitative parameters that represent the tumor microcirculation may help in the prediction of the status of the circumferential resection margin (CRM), the presence of metastases, and response to CRT. A recent meta-analysis showed that multiparametric MRI had a promising role in restaging of rectal cancer after preoperative CRT through a more accurate diagnosis of nodal disease and in predicting and detecting good treatment response. Contrast-enhanced MRI may also facilitate the assessment of extramural vascular invasion (EMVI) and T4 tumors and characterization of coexistent pelvic abnormalities.

MRI for Initial Staging of Rectal Cancer

MRI staging of rectal cancer comprises the assessment of tumor location and relationship to MRF and sphincters, tumor size, extent of extramural spread (T stage), peritoneal reflection, EMVI, lymph nodes, and bony metastasis.

For tumor localization, the distance of the lowest portion of the tumor from the anal verge is measured. A rectal tumor is characterized as low, middle, or high when its most caudal border is less than 5 cm from the anal verge, 5–10 cm from the anal verge, or more than 10 cm from the anal verge, respectively. In the lower parts of the rectum, the mesorectal fat surrounding the rectum is circumferentially bound by MRF. However, in higher portions, the peritoneum starts covering the anterior part of mesorectal fat to a point called the “anterior peritoneal reflection.” Upward from the

anterior peritoneal reflection, the peritoneum gradually extends posteriorly and finally encircles the rectosigmoid. The peritoneal reflection appears as a thin (1–2 mm) hypointense line on T2-weighted imaging that attaches the anterior aspect of the rectum and should be assessed in both the axial and sagittal planes. On sagittal images, the peritoneal reflection may be depicted above the tip of the seminal vesicles in men and at the uterocervical angle in women. The relationship of tumor to and invasion of the peritoneal reflection should be carefully reported. Low rectal cancers should be differentiated from anal squamous cell carcinomas before MRI interpretation; this distinction is possible only by histopathologic results and not by location and is important because the staging, behavior, and management of these entities are completely different.

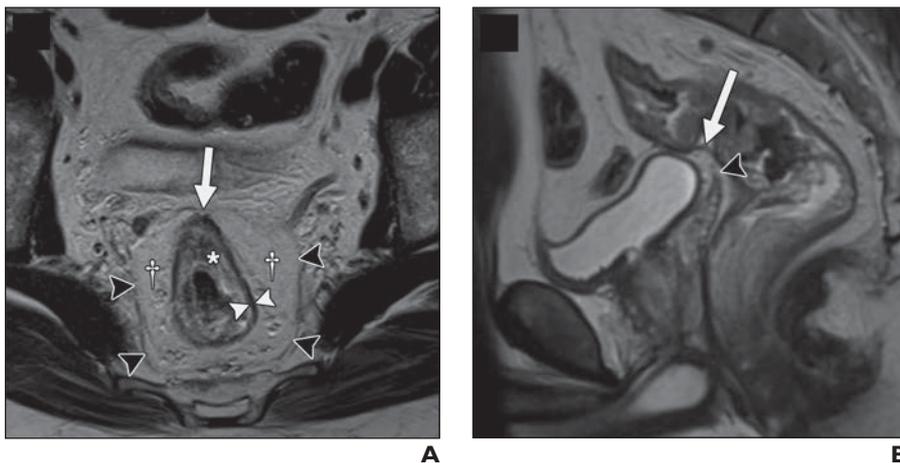


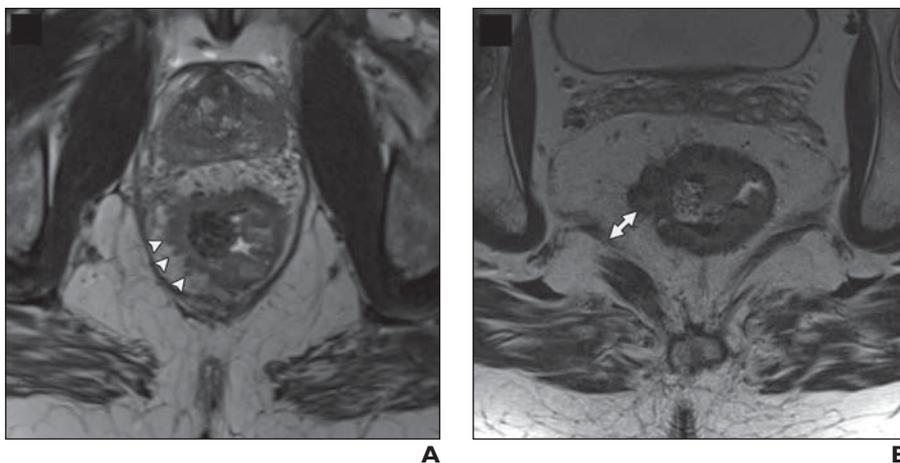
Figure 1 – Rectal anatomy on MRI. A, Axial T2-weighted image in 65-year-old man shows muscularis propria as hypointense band (white arrowheads) between mesorectal fat and submucosa (asterisk). Mesorectal fascia is depicted as thin hypointense line (black arrowheads) surrounding mesorectal fat (daggers). Note peritoneal attachment to anterior aspect of rectum (arrow). B, Sagittal T2-weighted image in 52-year-old man shows peritoneal attachment (arrow) above tip of seminal vesicles (arrowhead). (From Furey E, Jhaveri KS. MRI in rectal cancer. *Magn Reson Imaging Clin N Am* 2014; 22:165–190, v–vi).

T Stage

On T2-weighted imaging, the muscularis propria appears as a hypointense line between the hyperintense mesorectal fat and the inner submucosa and mucosa, which show intermediate to mild hyperintensity (Figure 1).

The signal intensity of a rectal tumor on T2-weighted images is typically intermediate between the

signal intensity of the muscularis propria and mucosa. Differentiation of T1 tumors from T2 tumors on MRI is usually not reliable without an endorectal coil, and tumors should be generally staged as “T1/T2.” A tumor is staged as T3 when it invades the mesorectal fat. Disruption of the muscularis propria because of the penetrating vessels should not be overstaged as T3. Spiculation of the mesorectal fat can be caused by either a benign desmoplastic reaction, seen as low signal intensity on T2-weighted images (T2 tumor), or tumor extension, seen as intermediate signal intensity on T2-weighted images (early T3 tumor), and may not be easily differentiated from one another on MRI. For T3 tumors, the shortest distance between the most penetrating parts of the tumor and the MRF should be measured (Figure 2).



*Figure 2: T3 rectal tumors on T2-weighted MR images. A, Low rectal tumor in 58-year-old man with tumoral spiculations (intermediate signal intensity) of mesorectal fat (arrowheads). B, Low rectal tumor in 63-year-old man with nodular extension to mesorectal fat. Double-headed arrow shows shortest distance from most penetrating part of tumor and mesorectal fascia. (From Furey E, Jhaveri KS. MRI in rectal cancer. *Magn Reson Imaging Clin N Am* 2014; 22:165–190, v–vi).*

The MRF is not circumferential at or above the peritoneal reflection and here it covers the posterior or posterolateral aspects of mesorectal fat of the rectum. The MRF is best visualized on T2-weighted images as a hypointense line surrounding the mesorectal fat especially at the proximal and posterior portions of the rectum where fat tissue is more abundant. A tumor-MRF distance of more than 1 mm is a reliable predictor for negative margins after total mesorectal excision. In the presence of satellite nodules, the shortest distance between the nodules and the MRF should also be reported. The extramural depth of invasion refers to extension of tumor beyond the muscularis propria and is

a prognostic factor. The American Joint Committee on Cancer suggested an optional stratification of T3 tumors based on the extramural depth of invasion: less than 5 mm, T3a; 5–10 mm, T3b; and more than 10 mm, T3c. An extramural depth of invasion of less than 5 mm confers a significantly higher survival rate, and these early T3 tumors may be adequately managed with surgery alone and have a prognosis comparable to that of tumors characterized as “T1/T2”.

If a tumor invades the visceral peritoneum, it is staged as T4a. Therefore, accurate depiction of the peritoneal reflection on MRI is crucial for proper staging of rectal tumors.

Tumors that invade other structures or adjacent organs are staged as T4b.

Particular attention should be paid for low rectal tumors. The surgical approach for low rectal tumors is more complex than for middle and high rectal tumors because of the thinner mesorectum and the presence of the surrounding sphincter complexes. The internal sphincter is a smooth-muscle ring formed by the inner (circular) muscle layer of the rectum. The external sphincter complex is a group of voluntary muscles in continuation of the levator muscles and consisting of the puborectalis and external sphincter muscles. Given the shape and anatomic location of the levator muscles, their accurate assessment in relation to the tumor is optimal through an evaluation of coronal and sagittal images.

Low rectal tumors typically undergo standard LAR, intersphincteric resection, or APR. The proximity of the inferior border of the tumors to the top border of the anal sphincters accounts for the selection of sphincter-preservation surgery and should be measured for low rectal tumors. Accordingly, a staging system geared toward staging low rectal cancers has been devised (Table 1). Consideration should be given to assessing the lateral extent of the tumor in the perineum so that a wide APR can be planned to ensure negative margins. APR is required for advanced T2 tumors, T3 tumors, and high rectal tumors that involve the levator muscles. In patients with adjacent organ invasion, pelvic exenteration may be indicated.

Stage	Description
T stage for middle tumors and high tumors ^a	
T1	Tumor invades submucosa
T2	Tumor invades muscularis propria
T3	Tumor invades through muscularis propria to pericolorectal tissues
a	Tumor < 5 mm into the perirectal fat or extramural
b	Tumor 5–10 mm into the perirectal fat or extramural
c	Tumor > 10 mm into the perirectal fat or extramural
T4	Organ invasion
a	Tumor penetrates to surface of visceral peritoneum
b	Tumor directly invades or is adherent to other organs or structures
T stage for low tumors ^b	
T1	Tumor confined to bowel wall but does not extend through full thickness; intact outer muscle coat
T2	Tumor replaces muscle coat but does not extend into intersphincteric plane
T3	Tumor invades intersphincteric plane or lies within 1 mm of levator muscle
T4	Tumor invades external anal sphincter and is within 1 mm and beyond levator muscle with or without invading adjacent organs
N stage	
Nx	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in 1–3 regional lymph nodes
N2	Metastasis in > 3 regional lymph nodes
M stage	
M0	No distant metastasis
M1	Distant metastasis
a	Metastasis confined to 1 organ or 1 site
b	Metastasis in more than 1 organ, 1 site, or peritoneum

Table 1: Rectal Cancer Staging

N Stage

The extent of nodal disease is important for both choosing and planning preoperative CRT and surgery. In the TNM system, disease involving only the regional nodes, including the mesorectal and internal iliac nodes, accounts for the N stage; involvement of other nodes is regarded as metastasis (Table 1). Mesorectal nodes are often the first and the most common group of nodes that are involved. Nodal metastases are usually within the proximal 5 cm of the tumor. Extramesorectal nodes are generally involved in locally advanced cancers. Inguinal nodal metastases, which are more typical of anal cancer than of rectal cancer, are uncommon even in low rectal cancers and imply poor prognosis.

Currently, size (i.e., short axis) and morphologic criteria are used with variable sensitivities (56–94%) and specificities (67–83%) for the differentiation of benign from malignant nodes. Although metastatic nodes are generally larger than benign nodes, malignant disease may be present even in very small nodes. The most commonly advocated size cutoffs for the diagnosis of malignancy are in the range of 5–8 mm; however, adding morphologic features, such as irregular contour and mixed signal intensity, to a size cutoff increases the diagnostic accuracy. Craniocaudal localization and clock-face localization of suspicious lymph nodes relative to the tumor are necessary; for mesorectal nodes, the distance to the MRF should also be measured.

DWI has shown low to moderate sensitivity (67–78%) and specificity (60–67%) for detecting involved lymph nodes.

FDG PET has high specificity (85–95%) for the identification of malignant nodes, but its role in nodal staging is limited because of low sensitivity (29–63%) and the inability to accurately localize the involved nodes. FDG PET is more suitable for the assessment of extramesorectal nodes because high uptake of tracer by the primary tumor may prevent the visualization of mesorectal nodes. Application of CT with PET improves the ability of the modality for anatomic localization. PET/CT has shown 63–70% accuracy for the detection of regional lymph node metastasis, and the combination of high-resolution MRI with PET/CT increases the accuracy up to 90%.

Extramural Vascular Invasion

EMVI refers to the extension of rectal tumor into the veins beyond the muscularis propria and can be detected on MRI with moderate sensitivity and high specificity by visualizing the vessels close to the tumor; EMVI is suggested when vessels close to the tumor are obviously irregular or expanded by tumoral signal intensity. EMVI has been accepted as an independent prognostic indicator in colorectal cancer that is associated with a higher incidence of metastasis, local recurrence, poorer response to preoperative CRT, and overall lower survival rate.

EMVI assessment is not included in the TNM staging system. However, EMVI status on initial MRI staging has been suggested by some studies as a prognostic factor for the stratification of patients for selecting the appropriate treatment, especially for indicating adjuvant therapy and its intensity.

Assessment of Response

Preoperative neoadjuvant CRT has improved the survival of patients possibly by increasing the CRM-negative resections. In addition, it may enable sphincter-preserving resection in patients with low rectal tumors through downstaging of locally advanced disease. The post-CRT restaging MRI examination is performed with the intent to, again, ensure negative margins, select patients with preexisting morbidities for local or less radical excision, and reassess patients for interval development of metastasis and extramesorectal lymphadenopathy. Tumor restaging involves correlating the posttreatment images with the pretreatment images with respect to all the elements assessed in the initial staging and necessitates image acquisition with almost the same protocol and on the same planes .

One of the most important parameters in restaging is reassessment of the MRF. MRI has shown approximately 76% sensitivity and 86% specificity for the assessment of the MRF in the irradiated pelvis. However, the accuracy of MRI for restaging is generally lower than the accuracy of MRI for initial staging mainly owing to overstaging of nodal disease, failure to differentiate tumoral infiltration or residual tumor from desmoplastic reaction or radiation fibrosis, and misinterpretation of radiation proctitis as local invasion. Evaluation of mucinous adenocarcinomas on posttreatment

MRI is also considerably challenging because these tumors remain hyperintense on T2-weighted imaging regardless of their response to treatment.

Although MRI restaging of rectal cancer with the conventional protocol is based on morphologic findings and changes in anatomic measurements, DWI and DCE-MRI may potentially provide functional information that can be correlated with changes at the cellular level in response to treatment. After CRT, the decrease in cellularity and the development of fibrosis or necrosis in responders result in an increase in diffusion and increase in the ADC value. A recent meta-analysis has shown that DWI is more sensitive than (62–94%) and is almost as specific as (74–91%) conventional MRI in restaging rectal tumors after CRT (van der Paardt et al. Radiology 2013).

The persistence of EMVI after CRT can be detected on MRI and, regardless of the final pathologic staging, may predict a higher risk of metastatic disease and an overall shorter disease-free survival. Therefore, EMVI status after CRT may possibly be used as an imaging biomarker for counseling patients for postoperative chemotherapy or more intensive surveillance.

Overall, the diagnostic performance of MRI for restaging rectal cancer after CRT is heterogeneous. Although adding the multiparametric sequences may partly improve MRI accuracy, issues with nodal staging still exist.

Changes in cellular viability and metabolic activity after CRT can be depicted on PET and can be interpreted as a response to treatment. PET has a high diagnostic performance for the interim assessment of response (sensitivity and specificity \approx 80%) but is less specific in post-CRT response assessment (\approx 60%) and is less sensitive in the evaluation of mucinous tumors (\approx 50%), which uptake less tracer because of their lower cellular density.

Locoregional Recurrence

The incidence of recurrent rectal cancer has started declining in the most recent 2 decades because of the advent of adjuvant CRT and improvements in surgery. Pelvic recurrence occurs in approximately 4–8% of patients who undergo surgery performed with a curative intent, and most cases of pelvic recurrence are seen within the first 3 years after treatment. The risk factors for local

recurrence include CRM positivity, no preoperative radiotherapy, EMVI, perforation of the tumor at surgery, and close proximity of the tumor to the anal verge.

Most intraluminal recurrent tumors are diagnosed by either rectal examination or direct visualization on rectosigmoidoscopy; however, detection of extraluminal recurrence and differentiating extraluminal recurrence from postoperative changes may be possible only on imaging. MRI is the most accurate imaging modality for the evaluation of these patients, but its cost has limited its use for routine follow-up. Annual pelvic CT is included in some guidelines for the surveillance of patients after surgical resection of rectal cancer. In other centers, pelvic imaging is performed only in the presence of suggestive clinical or laboratory findings (e.g., elevated serum carcinoembryonic antigen value).

Although T2-weighted imaging is considered the main sequence for the initial staging of rectal cancer, it is not specific enough for the assessment of recurrence. On T2-weighted images, recurring tumor appears hyperintense and is not distinguishable from inflammation or edema, which may persist for a few months after radiotherapy or surgery. On the other hand, fibrotic tissue may appear homogeneously hypointense on T2-weighted imaging while containing microscopic tumoral foci. Currently, the most accurate MRI sequence for the differentiation of treatment-related changes from recurrent rectal cancer is contrast-enhanced imaging. Enhancement in tumoral tissue has been shown to occur earlier and to be more intense and heterogeneous than enhancement in benign posttreatment fibrosis. Besides its high diagnostic performance for the identification of distant metastatic disease in recurrent rectal cancer (sensitivity, 91%; specificity, 83%), PET/CT has also shown high sensitivity (89–94%) and variable specificity (69–94%) for diagnosing the local recurrence on the basis of the shape, location, and intensity of tracer uptake [88–90]. Biopsy is indicated whenever imaging or clinical findings are equivocal and the diagnosis of recurrent disease cannot be confirmed.

Surgical resection with negative margins is established as the most efficient treatment for longer survival in recurrent rectal cancer. However, assessment of local extension to adjacent structures is challenging because pelvic fat planes are no longer present or intact after surgery or radiotherapy.

In this setting, local invasion can be suggested only when anatomic destruction or tumoral signal intensity is seen in the adjacent tissue. Detection of sacral invasion is crucial for achieving a clear posterior margin after resection. Whether radical exenterative surgery is a treatment option partly depends on the level of sacral involvement. In most institutions, invasion above the S2–S3 junction is a relative contraindication to resection, although high sacrectomy is also practiced in some centers.

Conclusion

MRI is the modality of choice for staging rectal cancer to assist surgeons in obtaining negative surgical margins. MRI facilitates the accurate assessment of MRF and the sphincter complex for surgical planning. Multiparametric MRI may also help in the prediction and estimation of response to treatment and in the detection of recurrent disease.

Texture Analysis

There has been growing interest in quantitative imaging biomarkers in recent years, particularly in the evaluation of tumors and cancer response to therapy. However, as our knowledge of the molecular signatures of different tumor types improves and therapies become increasingly varied and individualized, characterization of tumor and assessment of response to therapy become increasingly complex and in many cases need to be tailored to the specific tumor and therapy type. As a result, a profusion of imaging biomarkers have emerged using advanced imaging techniques, many of which are still being validated.

Tumors are heterogeneous on the gross and cellular levels, as well as the genetic and phenotypic levels, with spatial heterogeneity in cellular density, angiogenesis, and necrosis. This heterogeneity may affect prognosis and treatment, as more heterogeneous tumors may be associated with more biologically aggressive behavior and increased resistance to treatment.

Tumor heterogeneity can be difficult to capture and quantify with traditional imaging tools, subjective assessment of images, or random sampling biopsy, which evaluates only a small part of the tumor. Therefore, although it is not a new tool, there is renewed interest in computed tomographic (CT) texture analysis (CTTA), a potentially useful biomarker that allows assessment and quantification of tumor spatial heterogeneity. CTTA is just one part of the growing field of radiomics, which comprises high-throughput extraction, analysis, and interpretation of quantitative features from medical images.

CTTA has shown promise in a variety of tumor types, including colorectal cancer, head and neck cancer, esophageal cancer, lung cancer, and renal cell carcinoma (RCC), in predicting pathologic features, response to therapy, and prognosis. In addition, data are emerging to show that CTTA may have utility in a variety of nononcologic applications, including assessment and quantification of hepatic fibrosis, pulmonary fibrosis, interstitial pneumonia, and emphysema. However, there are a

variety of unknowns, challenges, and limitations related to CTTA that require further investigation before this potentially valuable tool is ready for mainstream application.

The purpose of this chapter is to review the basic concepts of texture analysis, describe oncologic and potential nononcologic applications of CTTA, evaluate histopathologic correlation with individual CTTA parameters, and discuss unknowns, limitations, and challenges around CTTA.

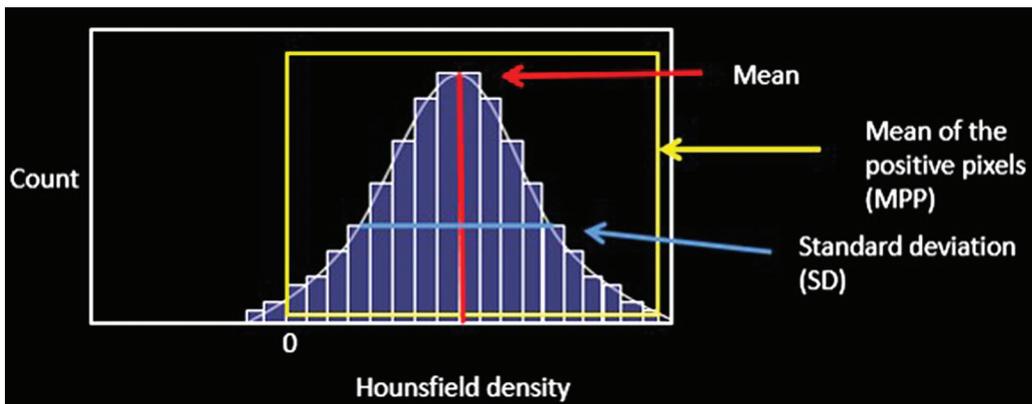
Basic Concepts of Texture Analysis

Texture analysis provides an objective, quantitative assessment of tumor heterogeneity by analyzing the distribution and relationship of pixel or voxel gray levels in the image. Different methods of texture analysis have been applied, including statistical-, model-, and transform-based methods. Statistical-based techniques have been most commonly applied, either through commercially available or inhouse software tools, to describe the relationship of gray-level values in the image.

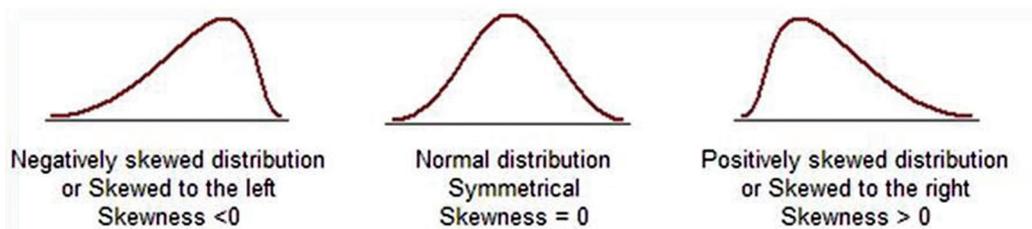
In a statistical-based model, first-order statistics evaluate the gray-level frequency distribution from the pixel intensity histogram in a given area of interest, including mean intensity, threshold (percentage of pixels within a specified range), entropy (irregularity), standard deviation, skewness (asymmetry), and kurtosis (peakedness/ flatness of pixel histogram) (Fig 1).

First-order histogram analysis does not account for the location of the pixels and lacks any reference to the spatial interrelationship between gray values.

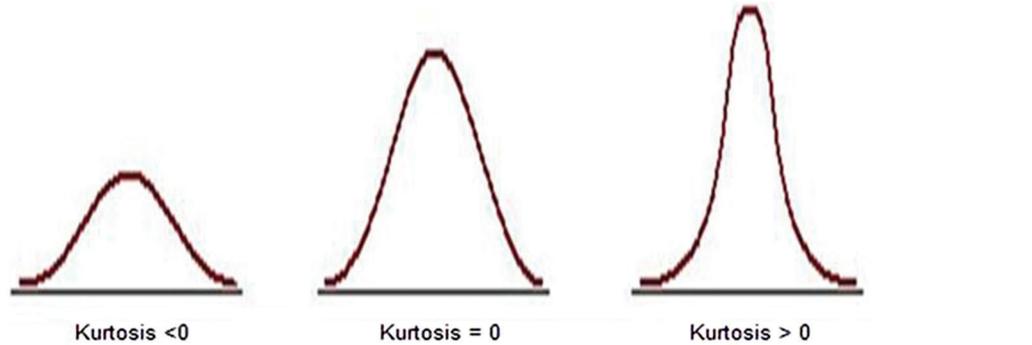
Second-order statistics can be based on a co-occurrence matrix and include things like second-order entropy, energy, homogeneity, dissimilarity, and correlation. Second-order statistics can also be derived using a run-length matrix, which analyzes texture in a specific direction. Higher-order statistics, such as contrast, coarseness, and busyness, can be calculated using neighborhood gray-tone difference matrices, which examine location and relationships between three or more pixels. Higher-order features have the advantage of evaluating voxels in their local context, taking the relationship with neighboring voxels into account.



a.



b.



c.

Figure 1: First-order statistical-based CT texture parameters. (a) Plot of the pixel histogram, where the x-axis represents gray-level values or attenuation and the y-axis represents the frequency of occurrence. First-order parameters include mean or mean gray-level intensity of the histogram (vertical red line), standard deviation of the histogram (horizontal blue line), and mean of the positive pixels (MPP) (average gray-level intensity above threshold of zero) (yellow box). (b) Sample histograms show skewness, which is negative when skewed values are less than zero (left) and positive when skewed values are greater than zero (right). (c) Sample histograms show kurtosis, which describes the peakedness or pointiness of the pixel histogram. A pointier or more peaked histogram is seen with positive and progressively higher kurtosis values (right). From Lubern et al, Radiographics 2017.

To perform CTTA, postprocessing software is needed, which can be either a commercially available tool or an in-house design, most of which are CT vendor neutral. This can be performed retrospectively on images obtained in the same phase of contrast enhancement with similar technique; in general, no prospective acquisition is needed. CTTA can be performed on either single-section (eg, largest cross-sectional area) or volumetric datasets to assess tumor heterogeneity.

An optional image filtration step can be performed. There are a wide variety of imaging filtration methods. A Laplacian or Gaussian bandpass filter is a commonly used advanced image filtration method that alters the image pixel intensity patterns and allows extraction of specific structures corresponding to the width of the filter. Lower filter values correspond to fine texture features, while higher filter values emphasize medium or coarse texture features. In addition, this filtration step is designed to remove noise and enhance edges, which may make measurements less susceptible to small differences in technique. Denoising or gray-level standardization steps have been used as well as a premeasurement step to help eliminate differences that are technical rather than biologic and to aid in reproducibility.

Model-based texture analysis uses advanced mathematical methods such as fractal analysis, while transform-based methods convert spatial information into frequency and/or scale (wavelet) information.

At CT, there has been some concern that heterogeneity related to technique (photon noise) may mask underlying true biologic heterogeneity, but studies have demonstrated that texture analysis at CT is feasible by reducing the effect of photon noise using image filtration. Other studies have shown that some but not all texture features applied to unfiltered or filtered images have high interobserver agreement.

Oncologic Applications

Potential oncologic applications of CTTA remain an active area of research. These applications seem to fall into three main categories: lesion characterization, pretreatment assessment (of both primary tumors and metastatic disease), and initial posttreatment assessment in reference to baseline tumor heterogeneity to predict therapeutic response.

Lesion Characterization

A major diagnostic challenge is to accurately differentiate benign from malignant lesions using noninvasive methods. For a variety of tumors, there are quantifiable texture differences between benign and malignant lesions (generally with greater heterogeneity in malignant lesions), possibly

allowing pathologic differentiation in certain clinical scenarios. CTTA has the potential to function as a “virtual biopsy” of indeterminate masses.

Numerous organs and lesions have been evaluated using texture analysis. CTTA has been used to differentiate renal cysts, oncocytomas, clear cell RCC, and papillary RCC with good diagnostic accuracy. Several groups have evaluated the utility of CTTA in differentiating lipid-poor angiomyolipomas (AMLs) from RCC.

Another important imaging task is differentiating benign from malignant portal vein thrombus. Texture features performed better than mean Hounsfield unit attenuation alone (AUC = 0.91) and radiologists’ subjective interpretation (AUC = 0.61). Another group looked at use of CTTA in differentiating hypervascular liver lesions including focal nodular hyperplasia, hepatic adenomas, and hepatocellular carcinoma. Using a random forest model, they were able to differentiate these lesion types. Predicted classification performance accuracy was 91.2% for adenoma, 94.4% for focal nodular hyperplasia, and 98.6% for hepatocellular carcinoma.

CTTA was used to differentiate pathologically proven intraductal papillary mucinous neoplasms (IPMNs) with high-grade dysplasia (n = 34) from those with low-grade dysplasia (n = 19). CTTA performed better in identifying lesions with high-grade dysplasia (higher risk for developing malignancy) than did use of imaging features based on the Fukuoka criteria.

The best texture feature had an AUC of 0.82, with sensitivity of 85% and specificity of 68% at the optimum threshold. The best logistic regression model had an AUC of 0.96, sensitivity of 97%, and specificity of 88%. Use of the Fukuoka criteria had a false-positive rate of 36%.

Multiple studies have evaluated the utility of CTTA in assessing colorectal polyps, which can have a variety of underlying histopathologic features, as well as a spectrum of biologic behavior. Traditionally, colonic polyps have been assessed using size and morphologic features (eg, flat versus nonflat). In a study of 384 polyps, CT texture features had an AUC of 0.80 for differentiating neoplastic from nonneoplastic polyps.

CTTA has also been applied to differentiating benign from malignant entities in the chest. It has been used to differentiate benign from malignant pulmonary nodules and mediastinal lymph nodes and to differentiate radiation fibrosis from recurrent lung cancer.

A summary of studies on lesion characterization with CTTA is presented in Table 1

Study and Reference	Organ	Lesion Type and Number	Imaging Technique	2D versus 3D Imaging	Texture Measure	Statistical Correction or Validation	Comments
Raman et al (20)	Kidney	Renal cyst (20), oncocytoma (20), ccRCC (20), pRCC (20)	CECT (renal mass protocol): arterial, venous, delayed	Multiple sections (maximum = 10)	GLH (TexRAD, Cambridge, England)	RFM, external validation in 19 cases, OOB error calculation	RFM classification: cyst 100%, oncocytoma 89%, ccRCC 91%, pRCC 100%
Yan et al (21)	Kidney	FP AML (18), ccRCC (18), pRCC (14)	Nonenhanced CT, CECT (CMP, NP); gray-level normalization	One to four sections	GLH, RLM, GLCM, gradient, autoregressive model, wavelet (MaZda)	No	Histogram and GLCM features: excellent classification of FP AML vs ccRCC, ccRCC vs pRCC, FP AML vs pRCC; better than subjective analysis
Hogdon et al (22)	Kidney	FP AML (16), RCC (84) (51 ccRCC, 20 chRCC, 13 pRCC pooled)	Nonenhanced CT, scanner vendor heterogeneity, image intensity normalization	Three sections	GLH, GLCM, RLM (MaZda, Technical University of Lodz, Poland)	10-fold cross-validation, Holm-Bonferroni procedure	Low homogeneity and high entropy in RCC; CTTA accuracy 83%–91%, better than subjective assessment; AUC 0.85–0.89; reproducibility of contour (inter-, intra-); reproducibility of measures (CT1 vs CT2)
Takahashi et al (23)	Kidney	AML (38), RCC (83)	Nonenhanced CT	Two sections (single-section small lesions)	GLH (Matlab, MathWorks, Natick, Mass)	No	Six or more pixels < -30 HU and skewness < -0.4 in 20 of 38 AMLs vs one of 83 RCCs; slightly better than subjective analysis (52% sensitivity, 99% specificity)
Leng et al (24)	Kidney	158 SRMs (<4 cm): ccRCC (98), pRCC (36), AML (24)	CECT ± denoising algorithm	Single section	GLH: SD, entropy, uniformity (Matlab)	No	ccRCC more heterogeneous (ccRCC vs pRCC AUC 0.91, 0.81, 0.78, 0.78 for subjective analysis/SD/entropy/uniformity); no difference between pRCC and AML; slight improvement in ccRCC vs pRCC AUC with weak denoising
Canellas et al (25)	Liver	117 PVTs (63 malignant, 54 bland)	CECT (portal venous)	Single section	GLH (TexRAD)	No	MPP (AUC 0.97), entropy (AUC 0.93), combined (AUC 0.99), attenuation (AUC 0.91) all better than subjective analysis (AUC 0.61)
Hanania et al (27)	Pancreas	53 IPMNs (34 HG, 19 LG)	CECT	Not reported	GLCM	Cross-validation	Best single-marker AUC 0.82 for LG vs HG, best logistic regression AUC 0.96, better than use of Fukuoka criteria
Hu et al (28)	Colon	384 CR polyps	Nonenhanced CT (CTC)	Volume	Intensity, gradient, curvature, GLCM	Training/testing dataset, OOB error calculation	RFM AUC 0.80 for differentiating neoplastic from nonneoplastic polyps
Song et al (29)	Colon	148 total: 35 nonneoplastic, 72 TAs, 36 TVAs, five adenomas	Nonenhanced CT (CTC)	Volume	Intensity, gradient, curvature, GLCM	SVM model, training and test set	AUC 0.74 for differentiating polyps on the basis of image intensity alone, improved to 0.85 when texture features added
Pooler et al (30)	Colon	63 CR masses ≥ 3 cm (32 adenocarcinomas, 31 adenomas)	Nonenhanced CT (CTC)	Volume	Intensity, gradient, curvature, GLCM	Previously established RFM method	CTTA AUC 0.936, better than human readers' AUC of 0.917

Note.—AML = angiomyolipoma, ccRCC = clear cell RCC, CECT = contrast-enhanced CT, chRCC = chromophobe RCC, CMP = corticomedullary phase, CR = colorectal, CTC = CT colonography, FP = fat-poor, GLCM = gray-level co-occurrence matrix, GLH = gray-level histogram, HG = high grade, IPMN = intraductal papillary mucinous neoplasm, LG = low grade, NP = nephrographic phase, OOB = out-of-bag, pRCC = papillary RCC, PVT = portal vein thrombus, RFM = random forest model, RLM = run-length matrix, SD = standard deviation, SRM = small renal mass, SVM = support vector machine, TA = tubular adenoma, TVA = tubulovillous adenoma, 3D = three-dimensional, 2D = two-dimensional.

Table 1: *Studies investigating lesions characterization with CTTA, from Lubern et al. Radiographics 2017.*

Primary Tumor Assessment

There is a growing body of literature evaluating the biology of tumors before treatment (Table 2). CTTA features are associated with histopathologic features and clinical outcomes in a variety of primary and metastatic tumors. In general, a change in tumor heterogeneity (either increased or decreased) may be associated with treatment response and improved prognosis/outcome. These data suggest that texture features may be useful in treatment planning and prognostication.

One study evaluated 157 patients with a large RCC (>7 cm) and found that texture features including entropy, standard deviation, and MPP were associated with histologic subtype (clear cell, papillary, chromophobe) and nuclear grade. Those same texture features were associated with time to recurrence and overall survival. In another study, kurtosis was associated with neovascularity (CD135/CD31) and Ki-67 in small renal masses (Scrima et al, 2016).

Zhang et al (Zhang et al, 2017) evaluated 105 patients with urothelial carcinomas (106 high grade, 18 low grade). They found that low-grade tumors were less heterogeneous, with significantly lower mean gray-level intensity, entropy, and MPP than high- grade tumors. MPP less than 24.1 on nonenhanced images was the optimal texture parameter for differentiating high-grade from low-grade tumors, with AUC of 0.78, sensitivity of 72%, and specificity of 85%.

Sandrasegaran et al (Sandrasegaran et al, 2017) used CTTA to assess 60 patients with nonmetastatic pancreatic ductal adenocarcinoma before treatment. They found that low kurtosis correlated with poor overall survival and that higher MPP (threshold of >29.4) was associated with better progression-free survival.

A study of 59 patients with pancreatic neuroendocrine tumors (PNETs) found that CTTA parameters including mean attenuation, MPP, skewness, kurtosis, and entropy allowed distinction between low-grade and high-grade PNETs (AUC ranging from 0.75 to 0.85 for different levels of filtration, all $P < .001$) (Canellas et al, 2016).

CTTA has been found to be useful in imaging assessment of hepatocellular carcinoma (HCC). One study evaluated 130 large HCCs (>5 cm) treated with liver resection ($n = 86$) or transarterial

chemoembolization (TACE) ($n = 44$) and found that texture features could be used to predict overall survival and may be useful in making treatment decisions (surgery vs embolization) (Li et al, 2016).

Studies have evaluated the value of CTTA in esophageal, gastric, and colonic cancers. Ganeshan et al (45) evaluated 21 patients with esophageal cancer using nonenhanced CT images with texture analysis and positron emission tomography (PET) standardized uptake values (SUVs) and found that tumor heterogeneity correlated with fluorodeoxyglucose (FDG) uptake, SUV_{max} , and SUV_{mean} . Heterogeneity was greater in patients with clinical stage III or IV disease (entropy, uniformity) and was an independent predictor of survival.

Multiple studies have evaluated use of CTTA features in identifying KRAS mutations in colorectal cancer. About 30%–40% of colon cancers exhibit a KRAS mutation, which is clinically significant because colorectal cancers bearing a KRAS mutation are resistant to drugs targeted against epidermal growth factor receptors (EGFRs). CTTA has shown some association with KRAS status, although limited data are available to date. MPP was used in combination with fluorodeoxyglucose uptake and CT perfusion to identify KRAS mutants and to categorize their phenotype (hypoxic vs proliferative).

Multiple studies have evaluated pretreatment assessment of lung cancer, particularly non-small cell lung cancer (NSCLC). Some of these studies included more detailed analyses of histopathologic correlates underlying certain texture features, and several studies have suggested that texture features may be related to the tumor microenvironment and the presence of features such as hypoxia and angiogenesis.

In a study of 45 patients with Hodgkin lymphoma ($n = 18$) and high-grade non-Hodgkin lymphoma ($n = 27$), CT texture features were associated with interim PET response and progression-free survival (Ganeshan et al, 2016).

Study and Reference	Tumor Type and Number	Imaging Technique	2D versus 3D Imaging	Texture Measure	Statistical Correction	Comments
Lubner et al (37)	157 RCCs (131 ccRCC, 13 pRCC, four chRCC)	Nonenhanced CT, portal venous CECT, heterogeneity in CT vendor and technique	Single section	GLH (TexRAD)	Bonferroni correction	Entropy AUC 0.94 for ccRCC vs other types; entropy, MPP, and SD associated with disease recurrence and death from disease
Schieda et al (39)	RCC (25 ccRCC, 20 sarcomatoid RCC)	Nonenhanced CT, CECT (renal protocol), heterogeneity in CT vendor; CTTA only nonenhanced CT, image intensity normalization	Three axial images	GLCM, RLM (MaZda)	Bonferroni correction	Combined texture features AUC 0.81 for sarcomatoid RCC vs ccRCC; increased run-length nonuniformity and increased gray-level nonuniformity in sarcomatoid RCC
Zhang et al (40)	TCC (106 HG, 18 LG)	Nonenhanced CT, CECT	Single section	GLH (TexRAD)	No	Mean, entropy, MPP, and SD lower in LG tumors; MPP AUC 0.78 for LG vs HG
Sandrasegaran et al (41)	Pancreatic cancer (60 patients)	CECT	Single section	GLH (TexRAD)	Holm correction	Low kurtosis correlated with low OS, high MPP associated with better PFS
Ganeshan et al (45)	Esophageal cancer (21 patients)	Nonenhanced CT	Unclear	GLH (entropy, uniformity)	No	Heterogeneity correlated with SUV_{max} , SUV_{mean} , increased with increasing tumor stage; uniformity independent predictor of survival
Ng et al (6)	CRC (55 patients, primary tumor)	CECT	Primary tumor volume	GLH	No	More homogeneous tumors (lower entropy, higher uniformity, lower SD) had poorer prognosis
Lubner et al (47)	CRC (77 patients, hepatic metastatic disease)	CECT	Single section	GLH (TexRAD)	Bonferroni correction	Entropy, MPP, and SD negatively associated with tumor grade; entropy negatively associated with survival
Ganeshan et al (53)	NSCLC (14 patients)	Nonenhanced CT, CECT	Three sections	GLH	Holm correction	SD and MPP associated with pimonidazole staining (hypoxia); uniformity negatively associated with Glut-1 (hypoxia); MPP negatively associated with CD34 (angiogenesis)
Ganeshan et al (4)	NSCLC (54 patients)	Nonenhanced CT	No data	GLH	No data	PET stage and tumor heterogeneity were independent predictors of survival
Ahn et al (59)	NSCLC (98 unresectable cases treated with CCRT)	CECT	Whole tumor volume	GLH (in-house program)	No	Higher entropy, higher skewness, and higher mean gray-level intensity associated with decreased 3-year OS
Hayano et al (60)	NSCLC (35 patients with metastatic disease treated with antiangiogenic therapy)	Nonenhanced CT for CTTA, CECT (CTP)	Single section	GLH (TexRAD)	No	High MPP and low entropy = favorable PFS and OS; low SUV_{max} = favorable OS; CTP not associated with survival; entropy independent predictor of OS in MV analysis
Zhang et al (8)	SCC of head and neck (72 patients)	CECT, some vendor heterogeneity of CT scanners	Single section	GLH (TexRAD)	No	High entropy and high skewness associated with poor OS; tumor size and nodal stage also associated with OS

Note.—ccRCC = clear cell RCC, CCRT = chemotherapy–radiation therapy, CECT = contrast-enhanced CT, chRCC = chromophobe RCC, CRC = colorectal cancer, CTP = CT perfusion, GLCM = gray-level co-occurrence matrix, GLH = gray-level histogram, HG = high grade, LG = low grade, MV = multivariate, NSCLC = non–small cell lung cancer, OS = overall survival, PFS = progression-free survival, pRCC = papillary RCC, RLM = run-length matrix, SCC = squamous cell carcinoma, SD = standard deviation, TCC = transitional cell carcinoma.

Table 2: Pretreatment assessment with CTTA, from Lubner et al. Radiographics 2017.

Response to Therapy

Multiple studies have investigated use of CTTA as an adjunct to conventional imaging findings, like size or CT attenuation, to determine the response of tumors to therapy. CTTA features are associated with histopathologic features and clinical outcomes in a variety of primary and metastatic tumors. In general, a change in tumor heterogeneity (either increased or decreased) may be associated with treatment response and improved prognosis/outcome.

Smith and coworkers studied 42 patients with melanoma being treated with antiangiogenic therapy who had stable disease according to the Response Evaluation Criteria in Solid Tumors (RECIST) (Smith AD et al. AJR 2015). Absolute change in MPP, change in tumor size, and baseline lactate dehydrogenase level were predictors of overall survival. A prognostic index incorporating these three factors was highly accurate for predicting overall survival at 18 months (AUC = 0.91).

Goh and coworkers (Goh V et al. Radiology 2011) assessed 39 patients with metastatic RCC being treated with tyrosine kinase inhibitors and found that entropy decreased and uniformity increased as tumors were treated, suggesting decreasing heterogeneity. Texture uniformity was an independent predictor of time to progression. Kaplan-Meier curves using a uniformity change threshold performed better in stratifying patients without disease progression than standard response assessments including RECIST, Choi criteria, and modified Choi criteria.

Use of texture to assess response to therapy has also shown some promise in lung, pancreas, and colorectal cancer.

Nononcologic Applications of CTTA

In addition to the oncologic applications, there are a number of emerging nononcologic applications of CTTA, including assessing lung disease such as fibrosis and emphysema, assessing hepatic fibrosis or risk for hepatic failure after resection, assessing abdominal aortic aneurysms for risk of rupture, and evaluating osseous trabecular texture to detect anorexia nervosa. The number of studies is too limited to draw conclusions about the value of CTTA in assessing diffuse disease of the lungs and liver.

Artificial Intelligence in Radiology

Artificial intelligence (AI) has recently made substantial strides in perception (the interpretation of sensory information), allowing machines to better represent and interpret complex data. This has led to major advances in applications ranging from web search and self-driving vehicles to natural language processing and computer vision – tasks that until a few years ago could be done only by humans. Deep learning is a subset of machine learning that is based on a neural network structure loosely inspired by the human brain. Such structures learn discriminative features from data automatically, giving them the ability to approximate very complex nonlinear relationships (BOX 1). While most earlier AI methods have led to applications with subhuman performance, recent deep learning algorithms are able to match and even surpass humans in task-specific applications (Fig. 1). This is owing to recent advances in AI research, the massive amounts of digital data now available to train algorithms and modern, powerful computational hardware. Deep learning methods have been able to defeat humans in the strategy board game of Go, an achievement that was previously thought to be decades away given the highly complex game space and massive number of potential moves. Following the trend towards a human-level general AI, researchers predict that AI will automate many tasks, including translating languages, writing best-selling books and performing surgery – all within the coming decades.

Within health care, AI is becoming a major constituent of many applications, including drug discovery, remote patient monitoring, medical diagnostics and imaging, risk management, wearables, virtual assistants and hospital management. Many domains with big data components such as the analysis of DNA and RNA sequencing data are also expected to benefit from the use of AI. Medical fields that rely on imaging data, including radiology, pathology, dermatology and ophthalmology, have already begun to benefit from the implementation of AI methods (Box 2).

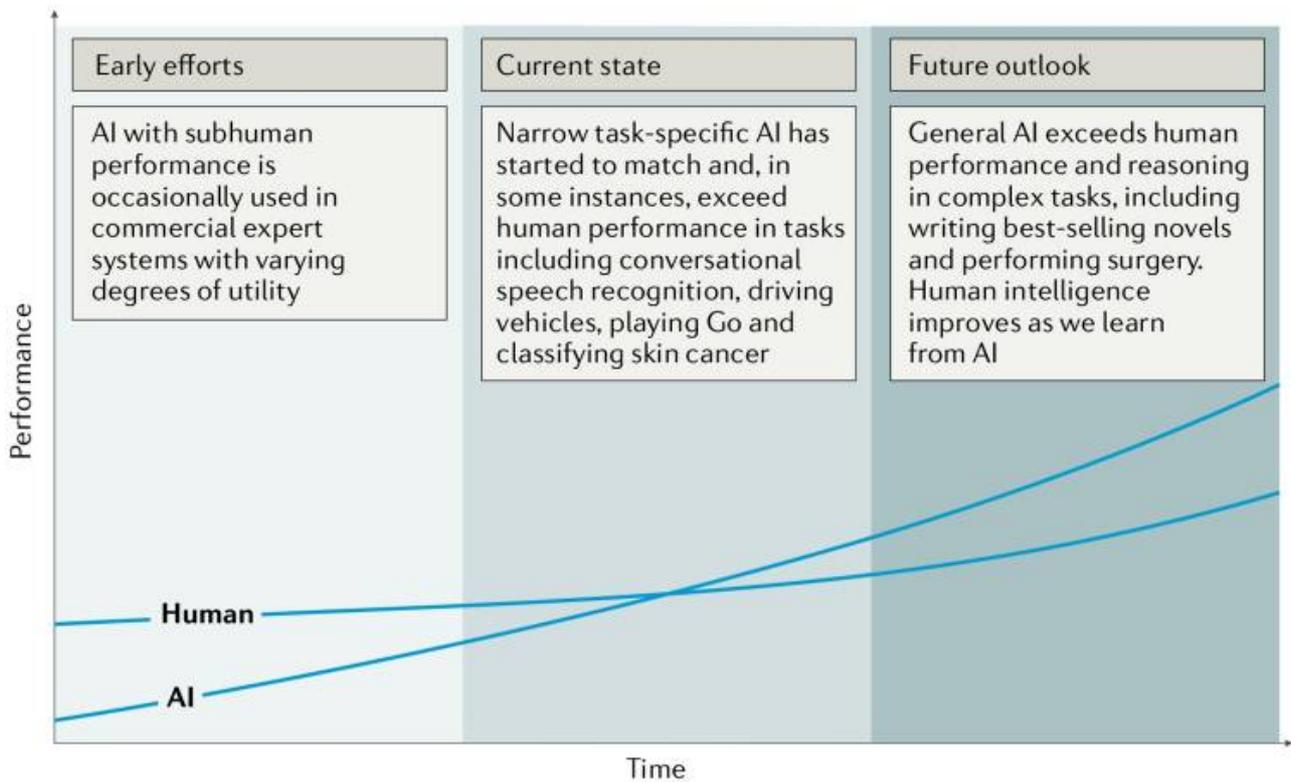


Figure 1: Artificial versus human intelligence. This plot outlines the performance levels of artificial intelligence (AI) and human intelligence starting from the early computer age and extrapolating into the future. Early AI came with a subhuman performance and varying degrees of success. Currently, we are witnessing narrow task-specific AI applications that are able to match and occasionally surpass human intelligence. It is expected that general AI will surpass human performance in specific applications within the coming years. Humans will potentially benefit from the human-AI interaction, bringing them to higher levels of intelligence.

Within radiology, trained physicians visually assess medical images and report findings to detect, characterize and monitor diseases. Such assessment is often based on education and experience and can be, at times, subjective. In contrast to such qualitative reasoning, AI excels at recognizing complex patterns in imaging data and can provide a quantitative assessment in an automated fashion. More accurate and reproducible radiology assessments can then be made when AI is integrated into the clinical workflow as a tool to assist physicians.

Box 1 | Artificial intelligence methods in medical imaging

Machine learning algorithms based on predefined engineered features

Traditional artificial intelligence (AI) methods rely largely on predefined engineered feature algorithms (Fig. 2a) with explicit parameters based on expert knowledge. Such features are designed to quantify specific radiographic characteristics, such as the 3D shape of a tumour or the intratumoural texture and distribution of pixel intensities (histogram). A subsequent selection step ensures that only the most relevant features are used. Statistical machine learning models are then fit to these data to identify potential imaging-based biomarkers. Examples of these models include support vector machines and random forests.

Deep learning algorithms

Recent advances in AI research have given rise to new, non-deterministic, deep learning algorithms that do not require explicit feature definition, representing a fundamentally different paradigm in machine learning. The underlying methods of deep learning have existed for decades. However, only in recent years have sufficient data and computational power become available. Without explicit feature predefinition or selection, these algorithms learn directly by navigating the data space, giving them superior problem-solving capabilities. While various deep learning architectures have been explored to address different tasks, convolutional neural networks (CNNs) are the most prevalent deep learning architecture typologies in medical imaging today. A typical CNN comprises a series of layers that successively map image inputs to desired end points while learning increasingly higher-level imaging features (Fig. 2b). Starting from an input image, 'hidden layers' within CNNs usually include a series of convolution and pooling operations extracting feature maps and performing feature aggregation, respectively. These hidden layers are then followed by fully connected layers providing high-level reasoning before an output layer produces predictions. CNNs are often trained end-to-end with labelled data for supervised learning. Other architectures, such as deep autoencoders and generative adversarial networks, are more suited for unsupervised learning tasks on unlabelled data. Transfer learning, or using pre-trained networks on other data sets, is often utilized when dealing with scarce data.

As imaging data are collected during routine clinical practice, large data sets are – in principle – readily available, thus offering an incredibly rich resource for scientific and medical discovery. Radiographic images, coupled with data on clinical outcomes, have led to the emergence and rapid expansion of radiomics as a field of medical research. Early radiomics studies were largely focused on mining images for a large set of predefined engineered features that describe radiographic aspects of shape, intensity and texture. More recently, radiomics studies have incorporated deep learning techniques to learn feature representations automatically from example images, hinting at the substantial clinical relevance of many of these radiographic features. Within oncology, multiple efforts have successfully explored radiomics tools for assisting clinical decision making related to

Box 2 |. Examples of clinical application areas of artificial intelligence in oncology

Radiology-based

Thoracic imaging.

Lung cancer is one of the most common and deadly tumors. Lung cancer screening can help identify pulmonary nodules, with early detection being lifesaving in many patients. Artificial intelligence (AI) can help in automatically identifying these nodules and categorizing them as benign or malignant.

Abdominal and pelvic imaging.

With the rapid growth in medical imaging, especially computed tomography (CT) and magnetic resonance imaging (MRI), more incidental findings, including liver lesions, are identified. AI may aid in characterizing these lesions as benign or malignant and prioritizing follow-up evaluation for patients with these lesions.

Colonoscopy.

Colonic polyps that are undetected or misclassified pose a potential risk of colorectal cancer. Although most polyps are initially benign, they can become malignant over time. Hence, early detection and consistent monitoring with robust AI-based tools are critical.

Mammography.

Screening mammography is technically challenging to expertly interpret. AI can assist in the interpretation, in part by identifying and characterizing microcalcifications (small deposits of calcium in the breast).

Brain imaging.

Brain tumors are characterized by abnormal growth of tissue and can be benign, malignant, primary or metastatic; AI could be used to make diagnostic predictions.

Radiation oncology.

Radiation treatment planning can be automated by segmenting tumors for radiation dose optimization. Furthermore, assessing response to treatment by monitoring over time is essential for evaluating the success of radiation therapy efforts. AI is able to perform these assessments, thereby improving accuracy and speed.

the diagnosis and risk stratification of different cancers. For example, studies in non-small-cell lung cancer (NSCLC) used radiomics to predict distant metastasis in lung adenocarcinoma and tumour histological subtypes as well as disease recurrence, somatic mutations, gene-expression profiles and overall survival. Such findings have motivated an exploration of the clinical utility of AI-generated biomarkers based on standard-of-care radiographic images – with the ultimate hope of better supporting radiologists in disease diagnosis, imaging quality optimization, data visualization, response assessment and report generation.

AI in Medical Imaging

The primary driver behind the emergence of AI in medical imaging has been the desire for greater efficacy and efficiency in clinical care. Radiological imaging data continues to grow at a disproportionate rate when compared with the number of available trained readers, and the decline in imaging reimbursements has forced health-care providers to compensate by increasing productivity. These factors have contributed to a dramatic increase in radiologists' workloads. Studies report that, in some cases, an average radiologist must interpret one image every 3–4 seconds in an 8-hour workday to meet workload demands. As radiology involves visual perception as well as decision making under uncertainty, errors are inevitable – especially under such constrained conditions.

A seamlessly integrated AI component within the imaging workflow would increase efficiency, reduce errors and achieve objectives with minimal manual input by providing trained radiologists with pre-screened images and identified features. Therefore, substantial efforts and policies are being put forward to facilitate technological advances related to AI in medical imaging. Almost all image-based radiology tasks are contingent upon the quantification and assessment of radiographic characteristics from images. These characteristics can be important for the clinical task at hand, that is, for the detection, characterization or monitoring of diseases. The application of logic and statistical pattern recognition to problems in medicine has been proposed since the early 1960s. As computers became more prevalent in the 1980s, the AI-powered automation of many clinical tasks has shifted radiology from a perceptual subjective craft to a quantitatively computable domain. The rate at which AI is evolving radiology is parallel to that in other application areas and is proportional to the rapid growth of data and computational power. There are two classes of AI methods that are in wide use today (BOX 1; Figure 2). The first uses handcrafted engineered features that are defined in terms of mathematical equations (such as tumor texture) and can thus be quantified using computer programs. These features are used as inputs to state-of-the-art machine learning models that are trained to classify patients in ways that can support clinical decision making. Although such features are perceived to be discriminative, they rely on expert definition

and hence do not necessarily represent the most optimal feature quantification approach for the discrimination task at hand. Moreover, predefined features are often unable to adapt to variations in imaging modalities, such as computed tomography (CT), positron emission tomography (PET) and magnetic resonance imaging (MRI), and their associated signal-to-noise characteristics.

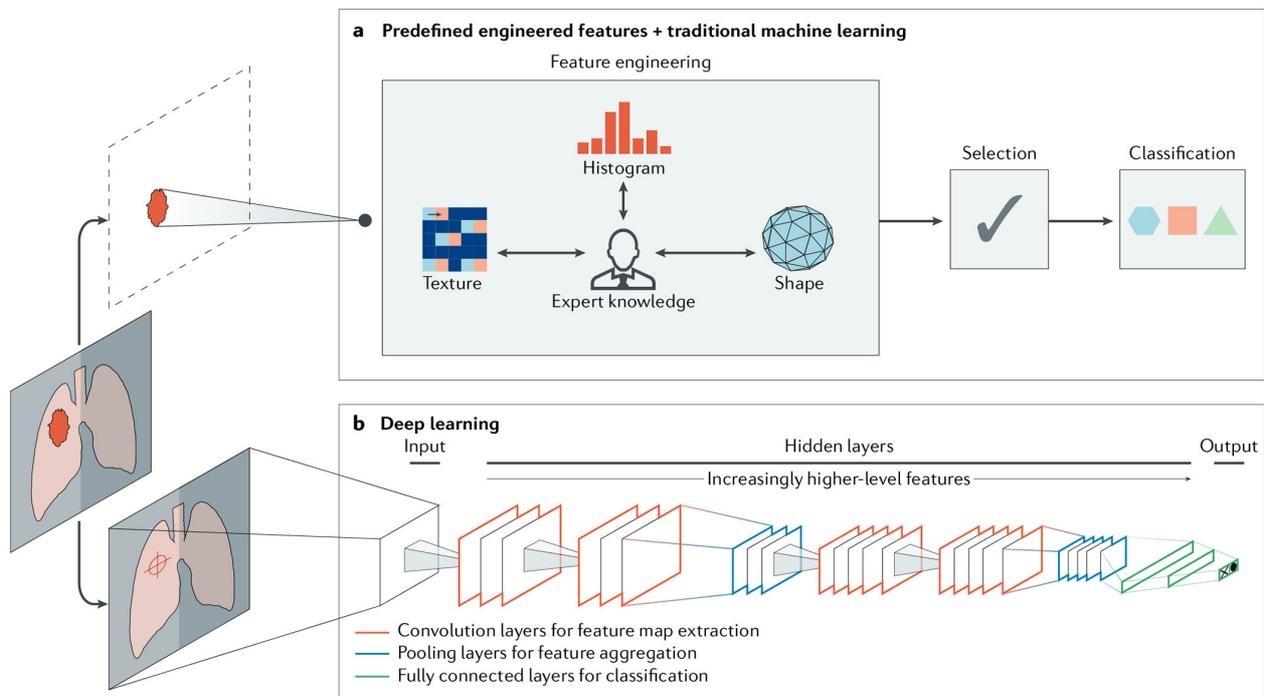


Figure 2: Artificial intelligence methods in medical imaging.

This schematic outlines two artificial intelligence (AI) methods for a representative classification task, such as the diagnosis of a suspicious object as either benign or malignant. a | The first method relies on engineered features extracted from regions of interest on the basis of expert knowledge. Examples of these features in cancer characterization include tumor volume, shape, texture, intensity and location. The most robust features are selected and fed into machine learning classifiers. b | The second method uses deep learning and does not require region annotation — rather, localization is usually sufficient. It comprises several layers where feature extraction, selection and ultimate classification are performed simultaneously during training. As layers learn increasingly higher-level features (Box 1), earlier layers might learn abstract shapes such as lines and shadows, while other deeper layers might learn entire organs or objects. Both methods fall under radiomics, the data-centric, radiology-based research field.

The second method, deep learning, has gained considerable attention in recent years. Deep learning algorithms can automatically learn feature representations from data without the need for prior definition by human experts. This data-driven approach allows for more abstract feature definitions, making it more informative and generalizable. Deep learning can thus automatically quantify

phenotypic characteristics of human tissues, promising substantial improvements in diagnosis and clinical care. Deep learning has the added benefit of reducing the need for manual preprocessing steps. For example, to extract predefined features, accurate segmentation of diseased tissues by experts is often needed. Because deep learning is data driven (Box 1), with enough example data, it can automatically identify diseased tissues and hence avoid the need for expert-defined segmentations. Given its ability to learn complex data representations, deep learning is also often robust against undesired variation, such as the inter-reader variability, and can hence be applied to a large variety of clinical conditions and parameters. In many ways, deep learning can mirror what trained radiologists do, that is, identify image parameters but also weigh up the importance of these parameters on the basis of other factors to arrive at a clinical decision.

Given the growing number of applications of deep learning in medical imaging¹⁴, several efforts have compared deep learning methods with their predefined feature-based counterparts and have reported substantial performance improvements with deep learning. Studies have also shown that deep learning technologies are on par with radiologists' performance for both detection and segmentation tasks in ultrasonography and MRI, respectively. For the classification tasks of lymph node metastasis in PET-CT, deep learning had higher sensitivities but lower specificities than radiologists. As these methods are iteratively refined and tailored for specific applications, a better command of the sensitivity/specificity trade-off is expected. Deep learning can also enable faster development times, as it depends solely on curated data and the corresponding metadata rather than domain expertise. On the other hand, traditional predefined feature systems have shown plateauing performance over recent years and hence do not generally meet the stringent requirements for clinical utility. As a result, only a few have been translated into the clinic. It is expected that high-performance deep learning methods will surpass the threshold for clinical utility in the near future and can therefore be expeditiously translated into the clinic.

Impact on Oncology Imaging

In this section, we focus on three main clinical radiology tasks that specifically pertain to oncology: abnormality detection, followed by characterization and subsequent monitoring of change (FIG. 3). These tasks require a diversified set of skills: medical, in terms of disease diagnosis and care, as well as technical, for capturing and processing radiographic images. Both these skills hint at the ample opportunities where up-and-coming AI technologies can positively impact clinical outcomes by identifying phenotypic characteristics in images.

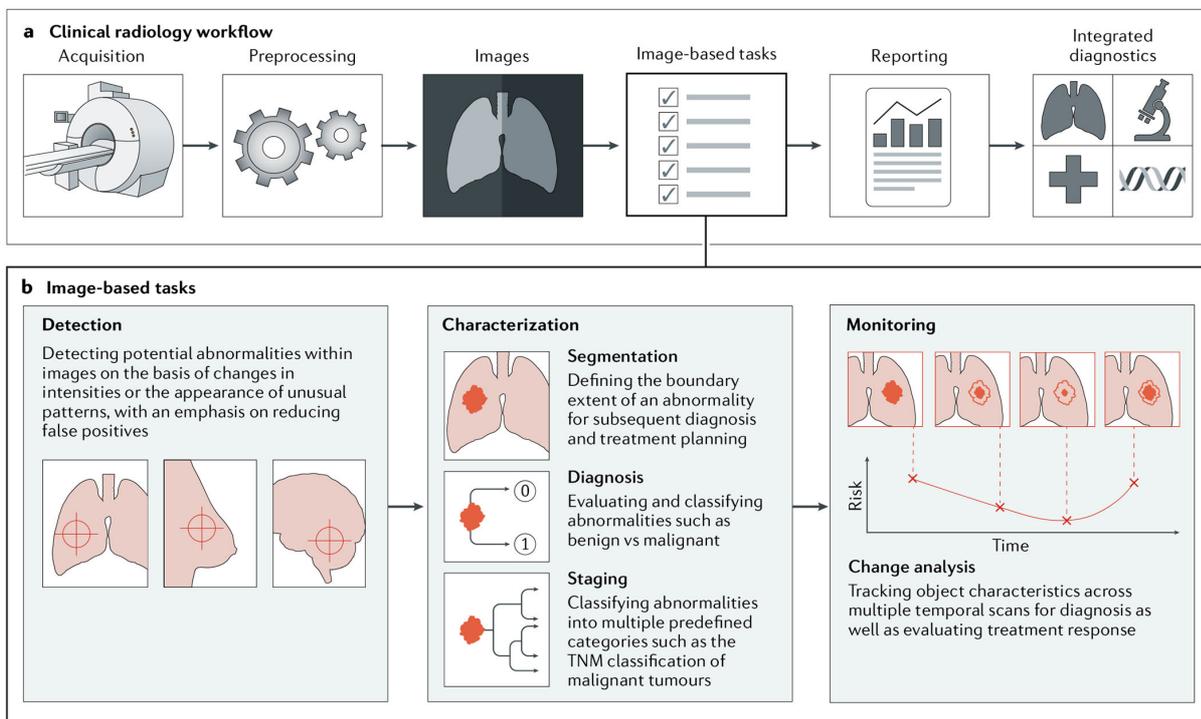


Figure 3: Artificial intelligence impact areas within oncology imaging. This schematic outlines the various tasks within radiology where artificial intelligence (AI) implementation is likely to have a large impact. a | The workflow comprises the following steps: preprocessing of images after acquisition, image-based clinical tasks (which usually involve the quantification of features either using engineered features with traditional machine learning or deep learning), reporting results through the generation of textual radiology reports and, finally, the integration of patient information from multiple data sources. b | AI is expected to impact image-based clinical tasks, including the detection of abnormalities; the characterization of objects in images using segmentation, diagnosis and staging; and the monitoring of objects for diagnosis and assessment of treatment response. TNM, tumour–node–metastasis

Detection.

Within the manual detection workflow, radiologists rely on manual perceptive skills to identify possible abnormalities, followed by cognitive skills to either confirm or reject the findings.

Radiologists visually scan through stacks of images while periodically adjusting viewing planes and window width and level settings. Relying on education, experience and an understanding of the healthy radiograph, radiologists are trained to identify abnormalities on the basis of changes in imaging intensities or the appearance of unusual patterns. These criteria, and many more, fall within a somewhat subjective decision matrix that enables reasoning in problems ranging from detecting lung nodules to breast lesions and colon polyps. As dependence on computers has increased, automated methods for the identification and processing of these predefined features – collectively known as computer-aided detection (CADe) – have long been proposed and occasionally utilized in the clinic. Radiologist-defined criteria are distilled into a pattern-recognition problem where computer vision algorithms highlight conspicuous objects within the image. However, these algorithms are often task-specific and do not generalize across diseases and imaging modalities. Additionally, the accuracy of traditional predefined feature-based CADe systems remains questionable, with ongoing efforts to reduce false positives. It is often the case that outputs have to be assessed by radiologists to decide whether a certain automated annotation merits further investigation, thereby making it labour intensive. In examining mammograms, some studies have reported that radiologists rarely altered their diagnostic decisions after viewing results from predefined, feature-based CADe systems and that their clinical integration had no statistical significance on the radiologists' performance. This is owing, in part, to the subhuman performance of these systems. Recent efforts have explored deep learning-based CADe to detect pulmonary nodules in CT and prostate cancer in multiparametric imaging, specifically multiparametric MRI. In detecting lesions in mammograms, early results show that utilizing convolutional neural networks (CNNs; deep learning algorithms; Box 1) in CADe outperforms traditional CADe systems at low sensitivity while performing comparably at high sensitivity and shows similar performance

compared with human readers. These findings hint at the utility of deep learning in developing robust, high-performance CAdE systems.

Characterization.

Characterization is an umbrella term referring to the segmentation, diagnosis and staging of a disease. These tasks are accomplished by quantifying radiological characteristics of an abnormality, such as the size, extent and internal texture. While handling routine tasks of examining medical images, humans are simply not capable of accounting for more than a handful of qualitative features. This is exacerbated by the inevitable variability across human readers, with some performing better than others. Automation through AI can, in principle, consider a large number of quantitative features together with their degrees of relevance while performing the task at hand in a reproducible manner every time. For instance, it is difficult for humans to accurately predict the status of malignancy in the lung owing to the similarity between benign and malignant nodules in CT scans. AI can automatically identify these features, and many others, while treating them as imaging biomarkers. Such biomarkers could hence be used to predict malignancy likelihood among other clinical end points including risk assessment, differential diagnosis, prognosis and response to treatment.

Within the initial segmentation step, while non-diseased organs can be segmented with relative ease, identifying the extent of diseased tissue is potentially orders of magnitude more challenging. Typical practices of tumor segmentation within clinical radiology today are often limited to high-level metrics such as the largest in-plane diameter. However, in other clinical cases, a higher specificity and precision are vital. For instance, in clinical radiation oncology, the extents of both tumor and non-tumor tissues have to be accurately segmented for radiation treatment planning. Attempts at automating segmentation have made their way into the clinic, with varying degrees of success. Segmentation finds its roots in earlier computer vision research carried out in the 1980s, with continued refinement over the past decades. Simpler segmentation algorithms used clustered imaging intensities to isolate different areas or utilized region growing, where regions are

expanded around user-defined seed points within objects until a certain homogeneity criterion is no longer met. A second generation of algorithms saw the incorporation of statistical learning and optimization methods to improve segmentation precision, such as the watershed algorithm, where images are transformed into topological maps with intensities representing heights. More advanced systems incorporate previous knowledge into the solution space, as in the use of a probabilistic atlas – often an attractive option when objects are ill-defined in terms of their pixel intensities. Such atlases have enabled more accurate automated segmentations, as they contain information regarding the expected locations of tumors across entire patient populations. Applications of probabilistic atlases include segmenting brain MRI for locating diffuse low-grade glioma, prostate MRI for volume estimation and head and neck CT for radiotherapy treatment planning, to name a few.

Recently proposed deep learning architectures for segmentation include fully convolutional networks, which are networks comprising convolutional layers only, that output segmentation probability maps across entire images. Other architectures, such as the U-net, have been specifically designed for medical images. Studies have reported that a single deep learning system is able to perform diverse segmentation tasks across multiple modalities and tissue types, including brain MRI, breast MRI and cardiac CT angiography (CTA), without task-specific training. Others describe deep learning methods for brain MRI segmentation that completely eliminate the need for image registration, a required preprocessing step in atlas-based methods.

Multiple radiographic characteristics are also employed in subsequent diagnosis tasks. These are critical to determine, for instance, whether a lung nodule is solid or whether it contains non-solid areas, also known as ground-glass opacity (GGO) nodules. GGO nodules are rather challenging to diagnose and often require special management protocols, mainly owing to the lack of associated characteristics of malignancy or invasiveness in radiographs. Generally, tumor radiographic characteristics may include information regarding size, maximum diameter, sphericity, internal texture and margin definition. The logic for diagnosis is based on these, often subjective, characteristics, enabling the stratification of objects into classes indicative of being benign or

malignant. Methods to automate diagnoses are collectively referred to as computer-aided diagnosis (CADx) systems. Similar to CADE, they often rely on predefined engineered discriminative features. Several systems are already in clinical use, as is the case with screening mammograms. They usually serve as a second opinion in complementing a radiologist's assessment, and their perceived successes have led to the development of similar systems for other imaging modalities, including ultrasonography and MRI. For instance, traditional CADx systems have been used on ultrasonography images to diagnose cervical cancer in lymph nodes, where they have been found to improve the performance of particularly inexperienced radiologists as well as reduce variability among them. Other application areas include prostate cancer in multiparametric MRI, where a malignancy probability map is first calculated for the entire prostate, followed by automated segmentation for candidate detection.

The accuracy of traditional predefined feature-based CADx systems is contingent upon several factors, including the accuracy of previous object segmentations. It is often the case that errors are magnified as they propagate through the various image-based tasks within the clinical oncology workflow. We also find that some traditional CADx methods fail to generalize across different objects. For instance, while the measurement of growth rates over time is considered a major factor in assessing risk, pulmonary nodule CADx systems designed around this criterion are often unable to accurately diagnose special nodules such as cavity and GGO nodules. Such nodules require further descriptors for accurate detection and diagnosis – descriptors that are not discriminative when applied to the more common solid nodules. This eventually leads to multiple solutions that are tailored for specific conditions with limited generalizability. Without explicit predefinition of these discriminative features, deep learning-based CADx is able to automatically learn from patient populations and form a general understanding of variations in anatomy – thus allowing it to capture a representation of common and uncommon cases alike.

Architectures such as CNNs are well suited for supervised diagnostic classification tasks (FIG. 2b). For both the breast lesion and lung nodule classification tasks, studies report a substantial performance gain of deep learning-based CADx methods – specifically those utilizing stacked

denoising autoencoders – over their traditional state-of-the-art counterparts. This is mainly owing to the automatic feature exploration mechanism and higher noise tolerance of deep learning. Such performance gain is assessed using multiple metrics, including the area under receiver operating characteristic curve (AUC), accuracy, sensitivity and specificity.

Monitoring.

Disease monitoring is essential for diagnosis as well as for evaluation of treatment response. The workflow involves an image registration preprocess where the diseased tissue is aligned across multiple scans, followed by an evaluation of simple metrics on them using predefined protocols – which is very similar to diagnosis tasks on single time-point images. A simple data comparison protocol follows and is used to quantify change. In oncology, for instance, these protocols define information regarding tumor size. Examples include the Response Evaluation Criteria in Solid Tumors (RECIST) and those created by the World Health Organization (WHO). Here, we find that the main goal behind such simplification is reducing the amount of effort and data a human reader must interact with while performing the task. However, this simplification is often based on incorrect assumptions regarding isotropic tumor growth. Whereas some change characteristics are directly identifiable by humans, such as moderately large variations in object size, shape and cavitation, others are not. These could include subtle variations in texture and heterogeneity within the object. Poor image registration, dealing with multiple objects and physiological changes over time all contribute to more challenging change analyses. Moreover, the inevitable interobserver variability remains a major weakness in the process. Computer-aided change analysis is considered a relatively younger field than CADe and CADx systems and has not yet achieved as much of a widespread adoption. Early efforts in automating change analysis workflows relied on the automated registration of multiple images followed by subtraction of one from another, after which changed pixels are highlighted and presented to the reader. Other more sophisticated methods perform a pixel-by-pixel classification – on the basis of predefined discriminative features – to identify changed regions and hence produce a more concise map of change. As the predefined

features used for registration differ from those used for the subsequent change analysis, a multistep procedure combining different feature sets is required. This could compromise the change analysis step, as it becomes highly sensitive to registration errors. With computer-aided change analysis based on deep learning, feature engineering is eliminated and a joint data representation can be learned. Deep learning architectures, such as recurrent neural networks, are very well suited for such temporal sequence data formats and are expected to find ample applications in monitoring tasks.

References

- Adam IJ, Mohamdee MO, Martin IG, et al. Role of circumferential margin involvement in the local recurrence of rectal cancer. *Lancet* 1994; 344:707-711
- Aerts HJWL et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun* 5, 4006 (2014). [PMC free article] [PubMed] [Google Scholar]
- Aerts HJWL The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol.* 2, 1636-1642 (2016). [PubMed] [Google Scholar]
- Alberda WJ, Dassen HP, Dwarkasing RS, et al. Prediction of tumor stage and lymph node involvement with dynamic contrast-enhanced MRI after chemoradiotherapy for locally advanced rectal cancer. *Int J Colorectal Dis* 2013; 28:573-580
- Ambinder EP A history of the shift toward full computerization of medicine. *J. Oncol. Pract* 1, 54-56 (2005). [PMC free article] [PubMed] [Google Scholar]
- Ayer T, Ayvaci MU, Liu ZX, Alagoz O & Burnside ES Computer-aided diagnostic models in breast cancer screening. *Imag. Med* 2, 313-323 (2010). [PMC free article] [PubMed] [Google Scholar]
- Barbaro B, Vitale R, Valentini V, et al. Diffusion-weighted magnetic resonance imaging in monitoring rectal cancer response to neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2012; 83:594-599
- Beets-Tan RG, Lambregts DM, Maas M, et al. Magnetic resonance imaging for the clinical management of rectal cancer patients: recommendations from the 2012 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol* 2013; 23:2522-2531
- Bipat S, Glas AS, Slors FJ, Zwinderman AH, Bossuyt PM, Stoker J. Rectal cancer: local staging and assessment of lymph node involvement with endoluminal US, CT, and MR imaging – a meta-analysis. *Radiology* 2004; 232:773-783
- Boland GWL, Guimaraes AS & Mueller PR The radiologist's conundrum: benefits and costs of increasing CT capacity and utilization. *Eur. Radiol* 19, 9-12 (2009). [PubMed] [Google Scholar]
- Brown G, Richards CJ, Newcombe RG, et al. Rectal carcinoma: thin-section MR imaging for staging in 28 patients. *Radiology* 1999; 211:215-222 Taylor FG, Quirke P, Heald RJ, et al. One millimetre is the safe cut-off for magnetic resonance imaging prediction of surgical margin status in rectal cancer. *Br J Surg* 2011; 98:872-879
- Canellas R, Bhowmik S, Almeida RR, Burk KS, McDermott S, Sahani DV. Prediction of pancreatic neuroendocrine tumor grade on CT images using a software based texture analysis [abstr]. In: Radiological Society of North America Scientific Assembly and Annual Meeting Program. Oak Brook, Ill: Radiological Society of North America, 2016; 143.
- Castellino RA Computer aided detection (CAD): an overview. *Cancer Imag.* 5, 17-19 (2005). [PMC free article] [PubMed] [Google Scholar]
- Champaign JL & Cederbom GJ Advances in breast cancer detection with screening mammography. *Ochsner J* 2, 33-35 (2000). [PMC free article] [PubMed] [Google Scholar]
- Chen H, Zheng Y, Park J-H, Heng P-A & Zhou SK in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 487-495 (Athens, Greece, 2016). [Google Scholar]
- Cheng J-Z et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci. Rep* 6, 24454 (2016). [PMC free article] [PubMed] [Google Scholar]
- Cioffi U, Raveglia F, De Simone M & Baisi A Ground-glass opacities: a curable disease but a big challenge for surgeons. *J. Thorac. Cardiovasc. Surg* 154, 375-376 (2017). [PubMed] [Google Scholar]

- Cole EB et al. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *AJR Am. J. Roentgenol* 203, 909–916 (2014). [PMC free article] [PubMed] [Google Scholar]
- Cong GN, Qin MW, You H, et al. Diffusion weighted imaging combined with magnetic resonance conventional sequences for the diagnosis of rectal cancer [in Chinese]. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 2009; 31:200–205
- Coroller TP et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother. Oncol* 114, 345–350 (2015). [PMC free article] [PubMed] [Google Scholar]
- Cruz-Roa A et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci. Rep* 7, 46450 (2017). [PMC free article] [PubMed] [Google Scholar]
- de Brebisson A & Montana G in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 20–28 (Boston, MA, USA, 2015). [Google Scholar]
- DeVries AF, Piringer G, Kremser C, et al. Pre-treatment evaluation of microcirculation by dynamic contrast-enhanced magnetic resonance imaging predicts survival in primary rectal cancer patients. *Int J Radiat Oncol Biol Phys* 2014; 90:1161–1167
- Donmez FY, Tunaci M, Yekeler E, Balik E, Tunaci A, Acunas G. Effect of using endorectal coil in pre-operative staging of rectal carcinomas by pelvic MR imaging. *Eur J Radiol* 2008; 67:139–145
- Edey AJ & Hansell DM Incidentally detected small pulmonary nodules on CT. *Clin. Radiol* 64, 872–884 (2009). [PubMed] [Google Scholar]
- Edge SB, Byrd DR, Compton CC. AJCC cancer staging handbook: from the AJCC cancer staging. Editors, N. Auspicious machine learning. *Nat. Biomed. Engineer* 1, 0036 (2017). [Google Scholar]
- El-Baz A et al. Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *Int. J. Biomed. Imag* 2013, 942353 (2013). [PMC free article] [PubMed] [Google Scholar]
- Esteva A et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017). [PubMed] [Google Scholar]
- Fernandez-Esparrach G, Ayuso-Colella JR, Sendino O, et al. EUS and magnetic resonance imaging in the staging of rectal cancer: a prospective and comparative study. *Gastrointest Endosc* 2011; 74:347–354
- Fitzgerald R Error in radiology. *Clin. Radiol* 56, 938–946 (2001). [PubMed] [Google Scholar]
- Furey E, Jhaveri KS. Magnetic resonance imaging in rectal cancer. *Magn Reson Imaging Clin N Am* 2014; 22:165–190, v–vi
- Ganeshan B, Miles KA, Babikir S, et al. CT-based texture analysis potentially provides prognostic information complementary to interim FDG-PET for patients with Hodgkin’s and aggressive non-Hodgkin’s lymphomas. *Eur Radiol* 2017;27(3):1012–1020.
- Ghafoorian M et al. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *Sci. Rep* 7, 5110 (2017). [PMC free article] [PubMed] [Google Scholar]
- Ghose S et al. in 2012 19th IEEE International Conference on Image Processing 541–544 (Orlando, FL, USA, 2012). [Google Scholar]
- Giannini V et al. A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging. *Comput. Med. Imaging Graph* 46, 219–226 (2015). [PubMed] [Google Scholar]
- Gollub MJ, Cao K, Gultekin DH, et al. Prognostic aspects of DCE-MRI in recurrent rectal cancer. *Eur Radiol* 2013; 23:3336–3344
- Gollub MJ, Gultekin DH, Akin O, et al. Dynamic contrast enhanced-MRI for the detection of pathological complete response to neoadjuvant chemotherapy for locally advanced rectal cancer. *Eur Radiol* 2012; 22:821–831

- Grace K, Salvatier J, Dafoe A, Zhang B & Evans O When will AI exceed human performance? Evidence from AI experts. Preprint at arXiv, 1705.08807 (2017). [Google Scholar]
- Grau V, Mewes AUJ, Alcañiz M, Kikinis R & Warfield SK Improved watershed transform for medical image segmentation using prior information. *IEEE Trans. Med. Imag* 23, 447–458 (2004). [PubMed] [Google Scholar]
- Grossmann P et al. Defining the biological basis of radiomic phenotypes in lung cancer. *eLife* 6, e23421 (2017) [PMC free article] [PubMed] [Google Scholar]
- Gulshan V et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410 (2016). [PubMed] [Google Scholar]
- Han X et al. Atlas-based auto-segmentation of head and neck CT images. *Med. Image Comput. Comput. Assist. Interv* 11, 434–441 (2008). [PubMed] [Google Scholar]
- Haralick RM & Shapiro LG Image segmentation techniques. *Computer Vision Graph. Image Process* 29, 100–132 (1985). [Google Scholar]
- Haug PJ Uses of diagnostic expert systems in clinical care. *Proc. Annu. Symp. Comput. Appl. Med. Care*, 379–383 (1993). [PMC free article] [PubMed] [Google Scholar]
- Heijnen LA, Lambregts DM, Martens MH, et al. Performance of gadofosveset-enhanced MRI for staging rectal cancer nodes: can the initial promising results be reproduced? *Eur Radiol* 2014; 24:371–379
- Hotker AM, Garcia-Aguilar J, Gollub MJ. Multi-parametric MRI of rectal cancer in the assessment of response to therapy: a systematic review. *Dis Colon Rectum* 2014; 57:790–799
- Huang X, Shan J & Vaidya V in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) 379–383 (Melbourne, Australia, 2017). [Google Scholar]
- Hussain SM, Outwater EK, Siegelman ES. Mucinous versus nonmucinous rectal carcinomas: differentiation with MR imaging. *Radiology* 1999; 213:79–85
- Huynh E et al. Associations of radiomic data extracted from static and respiratory-gated CT scans with disease recurrence in lung cancer patients treated with SBRT. *PLoS ONE* 12, e0169172 (2017). [PMC free article] [PubMed] [Google Scholar]
- Intven M, Reerink O, Philippens ME. Diffusion-weighted MRI in locally advanced rectal cancer: pathological response prediction after neo-adjuvant radiochemotherapy. *Strahlenther Onkol* 2013; 189:117–122
- Intven M, Reerink O, Philippens ME. Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. *J Magn Reson Imaging* 2014 Aug 14 [Epub ahead of print]
- Jaffe CC Measures of response: RECIST, WHO, and new alternatives. *J. Clin. Oncol* 24, 3245–3251 (2006). [PubMed] [Google Scholar]
- Jung SH, Heo SH, Kim JW, et al. Predicting re-sponse to neoadjuvant chemoradiation therapy in locally advanced rectal cancer: diffusion-weight- ed 3 Tesla MR imaging. *J Magn Reson Imaging* 2012; 35:110–116
- Kang H, O’Connell JB, Leonardi MJ, Maggard MA, McGory ML, Ko CY. Rare tumors of the co- lon and rectum: a national review. *Int J Colorectal Dis* 2007; 22:183–189
- Kelly SB, Mills SJ, Bradburn DM, Ratcliffe AA, Borowski DW; Northern Region Colorectal Cancer Audit Group. Effect of the circumferential re- section margin on survival following rectal can- cer surgery. *Br J Surg* 2011; 98:573–581
- Khorasani R, Erickson BJ & Patriarche J New opportunities in computer-aided diagnosis: change detection and characterization. *J. Am. Coll. Radiol* 3, 468–469 (2006). [PubMed] [Google Scholar]
- Kim SH, Lee JM, Gupta SN, Han JK, Choi BI. Dynamic contrast-enhanced MRI to evaluate the therapeutic response to neoadjuvant chemoradiation therapy in locally advanced rectal cancer. *J Magn Reson Imaging* 2014; 40:730–737

- Kim SH, Lee JM, Hong SH, et al. Locally advanced rectal cancer: added value of diffusion-weighted MR imaging in the evaluation of tumor response to neoadjuvant chemotherapy and radiation therapy. *Radiology* 2009; 253:116–125
- Kolossváry M, Kellermayer M, Merkely B & Maurovich-Horvat P Cardiac computed tomography radiomics: a comprehensive review on radiomic techniques. *J. Thorac. Imag* 33, 26–34 (2018). [PubMed] [Google Scholar]
- Kooi T et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med. Image Anal* 35, 303–312 (2017). [PubMed] [Google Scholar]
- Kosinski L, Habr-Gama A, Ludwig K, Perez R. Shifting concepts in rectal cancer management: a review of contemporary primary rectal cancer treatment strategies. *CA Cancer J Clin* 2012; 62:173–202
- Kumar V et al. Radiomics: the process and the challenges. *Magn. Reson. Imag* 30, 1234–1248 (2012). [PMC free article] [PubMed] [Google Scholar]
- Lambin P et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446 (2012). [PMC free article] [PubMed] [Google Scholar]
- Lambregts DM, Vandecaveye V, Barbaro B, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. *Ann Surg Oncol* 2011; 18:2224–2231
- Ledley RS & Lusted LB Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* 130, 9–21 (1959). [PubMed] [Google Scholar]
- Lehman CD et al. Diagnostic accuracy of digital screening mammography with and without computer aided detection. *JAMA Intern. Med* 175, 1828–1837 (2015). [PMC free article] [PubMed] [Google Scholar]
- Li M, Fu S, Zhu Y, et al. Computed tomography texture analysis to facilitate therapeutic decision making in hepatocellular carcinoma. *Oncotarget* 2016;7(11):13248–13259.
- Lim JS, Kim D, Baek SE, et al. Perfusion MRI for the prediction of treatment response after preoperative chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol* 2012; 22:1693–1700
- Litjens G et al. A survey on deep learning in medical image analysis. *Med. Image Anal* 42, 60–88 (2017). [PubMed] [Google Scholar]
- Litjens G et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep* 6, 26286 (2016). [PMC free article] [PubMed] [Google Scholar]
- Lodwick GS, Keats TE & Dorst JP The coding of Roentgen images for computer analysis as applied to lung cancer. *Radiology* 81, 185–200 (1963). [PubMed] [Google Scholar]
- Long J, Shelhamer E & Darrell T in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 3431–3440 (Boston, MA, USA, 2015). [Google Scholar]
- Maas M, Lambregts DM, Lahaye MJ, et al. T-staging of rectal cancer: accuracy of 3.0 Tesla MRI compared with 1.5 Tesla. *Abdom Imaging* 2012; 37:475–481
- McDonald RJ et al. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad. Radiol* 22, 1191–1198 (2015). [PubMed] [Google Scholar]
- Mir N, Sohaib SA, Collins D, Koh DM. Fusion of high b-value diffusion-weighted and T2-weighted MR images improves identification of lymph nodes in the pelvis. *J Med Imaging Radiat Oncol* 2010; 54:358–364
- Mirsadraee S, Oswal D, Alizadeh Y, Caulo A & van Beek E Jr. The 7th lung cancer TNM classification and staging system: review of the changes and implications. *World J. Radiol* 4, 128–134 (2012). [PMC free article] [PubMed] [Google Scholar]
- Mnih V et al. Human-level control through deep reinforcement learning. *Nature* 518, 529–533 (2015). [PubMed] [Google Scholar]

- Moeskops P et al. in Medical Image Computing and Computer- Assisted Intervention – MICCAI 2016 478–486 (Athens, Greece, 2016). [Google Scholar]
- Moravčík M et al. DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science* 356, 508–513 (2017). [PubMed] [Google Scholar]
- Nagaraj S, Rao GN & Koteswararao K The role of pattern recognition in computer-aided diagnosis and computer-aided detection in medical imaging: a clinical validation. *Int. J. Comput. Appl* 8, 18–22 (2010). [Google Scholar]
- Nasu K, Kuroki Y, Minami M. Diffusion-weight- ed imaging findings of mucinous carcinoma arising in the ano-rectal region: comparison of appar- ent diffusion coefficient with that of tubular adenocarcinoma. *Jpn J Radiol* 2012; 30:120–127 Taylor FG, Swift RI, Blomqvist L, Brown G. A systematic approach to the interpretation of pre- operative staging MRI for rectal cancer. *AJR* 2008; 191:1827–1835
- O’Connor JPB et al. Imaging biomarker roadmap for cancer studies. *Nat. Rev. Clin. Oncol* 14, 169–186 (2017). [PMC free article] [PubMed] [Google Scholar]
- Oberholzer K, Menig M, Pohlmann A, et al. Rec- tal cancer: assessment of response to neoadjuvant chemoradiation by dynamic contrast-enhanced MRI. *J Magn Reson Imaging* 2013; 38:119–126
- Ozis SE, Soydal C, Akyol C, et al. The role of 18F- fluorodeoxyglucose positron emission tomogra- phy/computed tomography in the primary staging of rectal cancer. *World J Surg Oncol* 2014; 12:26
- Pan X, Sidky EY & Vannier M Why do commercial CT scanners still employ traditional, filtered backprojection for image reconstruction? *Inverse Probl* 25, 1230009 (2009). [PMC free article] [PubMed] [Google Scholar]
- Parisot S et al. A probabilistic atlas of diffuse WHO grade II glioma locations in the brain. *PLoS ONE* 11, e0144200 (2016). [PMC free article] [PubMed] [Google Scholar]
- Parmar C, Grossmann P, Bussink J, Lambin P & Aerts HJWL Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep* 5, 13087 (2015). [PMC free article] [PubMed] [Google Scholar]
- Patriarche JW & Erickson BJ Part 1. Automated change detection and characterization in serial MR studies of brain-tumor patients. *J. Digit. Imag* 20, 203–222 (2007). [PMC free article] [PubMed] [Google Scholar]
- Paul R et al. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography* 2, 388–395 (2016). [PMC free article] [PubMed] [Google Scholar]
- Pendleton SD et al. Perception, planning, control, and coordination for autonomous vehicles. *Machines* 5, 6 (2017). [Google Scholar]
- Pham DL, Xu C & Prince JL Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng* 2, 315–337 (2000). [PubMed] [Google Scholar]
- Pipatsrisawat T, Gacic A, Franchetti F, Puschel M & Moura JMF in Proceedings. (ICASSP ‘05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005 v/153–v/156 (Philadelphia, PA, USA, 2005). [Google Scholar]
- Rios Velazquez E et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res.* 77, 3922–3930 (2017). [PMC free article] [PubMed] [Google Scholar]
- Ronneberger O, Fischer P & Brox TU in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 234–241 (Munich, Germany, 2015). [Google Scholar]
- Rusk N Deep learning. *Nat. Methods* 13, 35–35 (2015). [Google Scholar]
- Sandrasegaran K, Lin C, Asare-Sawiri M, Lin Y. CT texture analysis of pancreatic cancer. Presented at the Annual Scientific Meeting and Educational Course of the Society of Abdominal Radiology, Waikoloa, Hawaii, March 13–18, 2016.
- Schieda N, Thornhill RE, Al-Subhi M, et al. Diagnosis of sarcomatoid renal cell carcinoma with CT: evaluation by qualitative imaging features and texture analysis. *AJR Am J Roentgenol* 2015;204(5):1013–1023.

- Scrima AN, Lubner MG, Abel EJ, Pickhardt PJ. CT textural analysis of small renal masses. Madison, Wis: University of Wisconsin School of Medicine and Public Health, 2016.
- Sharma N & Aggarwal LM Automated medical image segmentation techniques. *J. Med. Phys* 35, 3–14 (2010). [PMC free article] [PubMed] [Google Scholar]
- Shen D, Wu G & Suk H-I Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng* 19, 221–248 (2017). [PMC free article] [PubMed] [Google Scholar]
- Shiraishi J, Li Q, Appelbaum D & Doi K Computeraided diagnosis and artificial intelligence in clinical imaging. *Semin. Nucl. Med* 41, 449–462 (2011). [PubMed] [Google Scholar]
- Silver D et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489 (2016). [PubMed] [Google Scholar]
- Slater A, Halligan S, Taylor SA, Marshall M. Distance between the rectal wall and mesorectal fascia measured by MRI: effect of rectal distension and implications for preoperative prediction of a tumour-free circumferential resection margin. *Clin Radiol* 2006; 61:65–70
- Sohn K, Shang W & Lee H in *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (eds Ghahramani Z, Welling M, Cortes C, Lawrence ND & Weinberger KQ) 2141–2149 (Montreal, Canada, 2014). [Google Scholar]
- Suzuki C, Torkzad MR, Tanaka S, et al. The importance of rectal cancer MRI protocols on interpretation accuracy. *World J Surg Oncol* 2008; 6:89
- Szmulowicz UM, Wu JS. Squamous cell carcinoma of the anal canal: a review of the aetiology, presentation, staging, prognosis and methods available for treatment. *Sex Health* 2012; 9:593–609
- Thiesse P et al. Response rate accuracy in oncology trials: reasons for interobserver variability. *Groupe Français d’Immunothérapie of the Fédération Nationale des Centres de Lutte Contre le Cancer. J. Clin. Oncol* 15, 3507–3514 (1997). [PubMed] [Google Scholar]
- Tsehay YK et al. in *Proceedings of SPIE* <https://doi.org/10.1117/12.2254423> (2017).
- van Ginneken B, Schaefer-Prokop CM & Prokop M Computer-aided diagnosis: how to move from the laboratory to the clinic. *Radiology* 261, 719–732 (2011). [PubMed] [Google Scholar]
- Veeraraghavan H MO-A-207B-01: Radiomics: Segmentation & feature extraction techniques. *Med. Phys* 43, 3694–3694 (2016). [Google Scholar]
- Wang H et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res.* 7, 11 (2017). [PMC free article] [PubMed] [Google Scholar]
- Wu W et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front. Oncol* 6, 71 (2016). [PMC free article] [PubMed] [Google Scholar]
- Xiong W et al. Toward human parity in conversational speech recognition. *IEEE/ACM Trans. Audio Speech Language Process* 25, 2410–2423 (2017). [Google Scholar]
- Zhang GM, Sun H, Shi B, Jin ZY, Xue HD. Quantitative CT texture analysis for evaluating histologic grade of urothelial carcinoma. *Abdom Radiol (NY)* 2017;42(2):561–568.
- Zhang J, Wang Y, Yu B, Shi X & Zhang Y Application of computer-aided diagnosis to the sonographic evaluation of cervical lymph nodes. *Ultrason. Imag* 38, 159–171 (2016). [PubMed] [Google Scholar]
- Zhu B, Liu JZ, Cauley SF, Rosen BR & Rosen MS Image reconstruction by domain-transform manifold learning. *Nature* 555, 487–492 (2018). [PubMed] [Google Scholar]

Experimental Study

**Performance of texture analysis in predicting tumoural response
to neoadjuvant chemoradiotherapy in rectal cancer patients
studied with 3T MR**

Abstract

Purpose: To determine the performance of texture analysis (TA) in the prediction of tumoral response in Colon rectal Cancer (CRC) Patients.

Material and methods: We prospectively enrolled 40 consecutive CRC patients, who underwent pre-treatment and post-treatment 3T MRI. A region of interest was drawn manually around the tumour on Unenhanced axial obliqueT2-weighted images and analysed using TA (TexRAD), evaluating first order statistical texture parameters (Skewness, Kurtosis). After CRT, all patients underwent complete surgical resection and the surgical specimen served as the gold standard. Moreover, receiver operating characteristic (ROC) curve analysis was performed to assess the discriminatory power of each quantitative parameter to predict complete response. Artificial Intelligence software (Weka) was used to evaluate the impact of machine learning algorithms on texture parameters in rectal cancer aimed to combine all texture parameters in order to improve diagnostic accuracy.

Results: Entropy, Kurtosis and MPP showed significant differences before and after CRT in CR's; PR/NR Entropy and Skewness showed significant differences before and after CRT (all $p < 0.05$). Absolute changes among different texture parameters in CR and PR/NR patients before and after CRT showed significant differences in Entropy, Kurtosis and MPP (0.31 ± 0.35 , in CR, -0.02 ± 1.28 in PR/NR, ($p=0.04$); 1.87 ± 2.19 , in CR, -0.06 ± 3.78 in PR/NR($p=0.0005$); (107.91 ± 274.40 , in CR, -28.33 ± 202.91 in PR/NR, ($p=0.004$), respectively). Kurtosis and Entropy seem to be the texture parameters able to predict CR. ROC curves showed the optimal cutoff value for pretreatment kurtosis, the best parameter, ≤ 3.29 , resulting a sensitivity and specificity for pCR prediction of 81.5% and 61.5%, respectively. A decisional map was defined combining all texture parameters.

Conclusion: TA from T2w images can potentially have an important role as imaging biomarkers of tumoral response to neoadjuvant CRT in rectal cancer. Machine learning software can be apply in this setting to combined efficiently all texture parameters in order to improve diagnostic accuracy.

Introduction

Neoadjuvant ChemoRadiotherapy (CRT) and radical surgery with Total Mesorectal Excision (TME) is the standard of treatment in locally advanced rectal cancer (LARC), since it can improve local tumour control. However, modern neoadjuvant CRT can achieve complete pathological response (pCR) in around 18-25% of the cases ^{1,2}, meaning that surgery in this particular sub-group of patients, might be considered an overtreatment.

New evidences from the literature show that a “watch-and-wait” approach might be feasible instead of surgery^{3,4}. Main advantages are in organ sparing (preserving continence and bowel, sexual and urinary functions) and reduced post-operative morbidity. However, the main disadvantage is the correct selection of patients with pathological complete response (pCR). A wrong selection, in fact, might affect patient survival. Unfortunately, conventional MR Imaging (MRI), which is the best modality for pre-operative staging of patients with LARC has poor results when assessing response to therapy.

Recently, Texture Analysis (TA) has been introduced as a novel imaging biomarker into research and clinical practice. TA is able to assess quantitatively heterogeneity of tissues, considered an important feature of malignancy associated with biological tumoural behaviors. Recent evidences suggest the important role of TA parameters as imaging biomarkers. TA can be easily integrated in conventional MR and CT imaging to evaluate response during and after cancer treatment and has also an important role as a prognostic factors⁵⁻⁸.

Another emerging field in medicine and in radiology is Machine Learning (ML). ML is a subfield of artificial intelligence (AI) with a primary focus on developing predictive algorithms through unbiased identification of patterns within large datasets and without being explicitly programmed for a particular task. One very interesting potential application of these models is to expand the potential of extracting more knowledge from radiological imaging datasets, especially if large and complex like data that come from texture analysis.

With this in mind, the primary aim of the present study, was to evaluate whether MRI texture parameters can be used as imaging biomarkers in order to discriminate before chemo and radiation

therapy Patients Complete Responders, using both conventional analysis and Machine learning Software.

Materials & Methods

Study Population

Our institutional ethics committee approved this study and written informed consent was obtained from each patient.

We prospectively enrolled 44 consecutive patients affected by locally advanced rectal cancer (LARC) in particular histologically confirmed as adenocarcinoma subtype and tumour stages II (cT3-4, N0, M0) and III (cT1-4, N+, M0) according to the International Union Against Cancer classification⁹.

Were excluded patients with the following criteria: a) contraindication to MR examination (e.g., pacemaker, intra-ocular metal foreign bodies, etc.); b) MRI protocol not fully acquired or incomplete histopathological data; c) contraindication to the use of neoadjuvant radio-chemotherapy or surgical treatment or suspension of neoadjuvant treatment before undergoing surgery; d) coexistence of other known tumours, cardiovascular disease, history of neurological or psychiatric disorders, drug or alcohol abuse or previous pelvic radiation treatment; e) legal incapacity or restricted legal capacity.

Protocol Timeline

The study, part of a larger study founded by AIRC (Associazione Italiana per la Ricerca sul Cancro), comprises three consecutive phases in which 3 MRI examinations were performed. First MRI Study includes initial evaluation of local tumour status after histological confirmation. Second MRI Study aimed to evaluate an early response to CRT, performed after 21 days from the treatment.

Third MRI Study, performed at the end of the neoadjuvant therapy. Patients underwent Total mesorectal excision after 6-8 weeks from the end of RCT and histopathological assessment of the gross specimen (Figure 1).

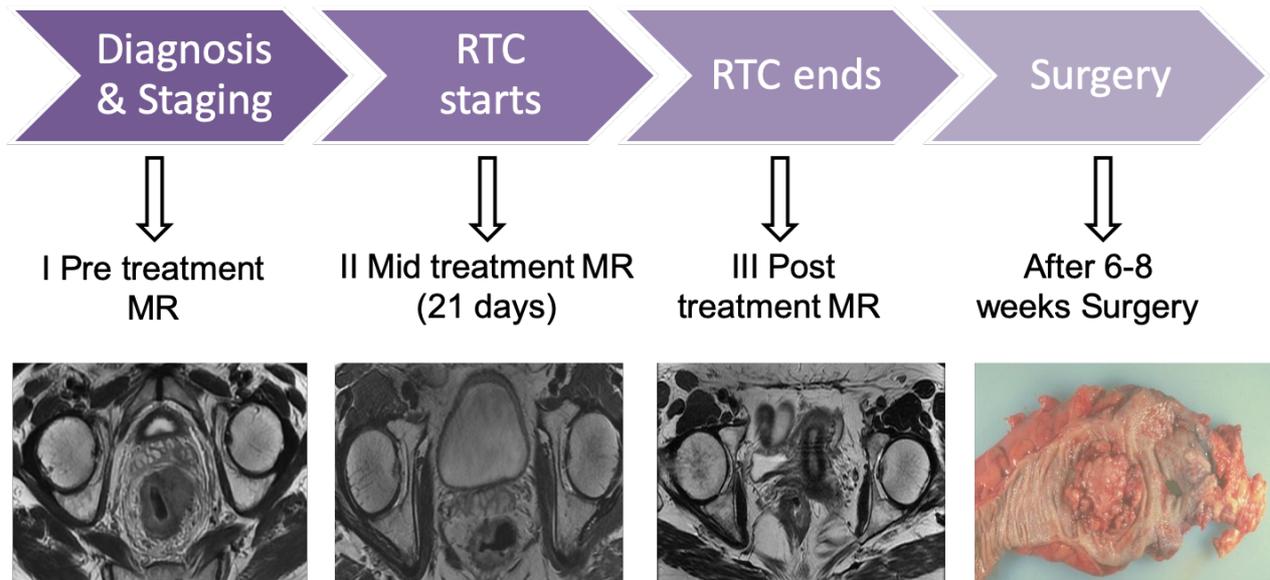


Figure 1: Timeline. RTC= Radiotherapy and Chemotherapy. MR= Magnetic resonance.

MR Examination

Magnetic resonance examinations were performed using a 3-T scanner (Discovery MR750; General Electrics, Milwaukee, WI) following a standard imaging protocol for rectal cancer evaluation was performed as described in other study¹⁰; for the specific purpose of the study has been analysed high-resolution T2w fast recovery fast-spin echo (2D) sequence (repetition time, 2086–4172 milliseconds; echo time, 11.4–122.3 milliseconds; Nex, 2; slice thickness, 4 mm; matrix, 512 x 512) acquired on dedicated axial oblique planes orthogonal to the long axis of the rectum.

Neoadjuvant Radiochemo-Therapy

For radiation therapy a total dose of 45 Gy was delivered to the whole pelvis with a 3D conformational multiple field technique (fractionated in 25 daily administrations of 1.8 Gy per day for 5 weeks); additional dose of 5.4 to 9 Gy was delivered to the tumour volume (fractionated in 3–5 per day administration of 1.8 Gy each), with 6 to 15 MV energy photons.

Chemotherapy protocol consisted in oxaliplatin and 5-fluoruracile i.v. administration after the placement of central venous access. Oxaliplatin was administered the first day of each RT weekly

session (2 hour infusion, 50 mg/m²) while 5-fluorouracil infusion continued for a total of 5 days (200 mg/m²/d).

Surgical Technique and Histopathological Assessment

Colorectal surgeons (at least 10 years of experience) performed a standardized total mesorectal excision in all patients¹¹.

Expert gastro-intestinal pathologists performed the histopathology assessment by evaluating the basic histopathology of the primary biopsied tumour (type and grade of the lesion) before neoadjuvant treatment. To allow a comparison as probable as possible with MRI analysis, the intestinal resected segment with the tumour was sectioned orthogonally to the long axis, obtaining macro-section thicknesses from 2 to 3 mm. Thanks to this technique was possible to preserve the anteroposterior and left-right orientation of the specimen and facilitates the localization of suspect pathological region seen on MRI. All specimens were assessed for T and N stage (according to 7th edition of the American Joint Commission on Cancer) and for surgical margins.

Texture Analysis

Heterogeneity of rectal tumors was assessed with TexRAD, a proprietary software algorithm (TexRAD Ltd). This was undertaken by a single operator (DB, with 5 years of experience in texture analysis).

An expert abdominal radiologist with 8 years of experience drawn a region of interest (ROI) around the largest tumour area appreciable on Axial oblique T2 image, avoiding cystic or necrotic regions, for pre and post treatment MRI examinations; the radiologist was blinded for histological results (Figure 2). The obtained ROIs were analysed for Texture Analysis using the image histogram (first order) statistical method that refers to the frequency of the intensity of pixels.

The in-plane filtration step used a Laplacian of Gaussian spatial band-pass filter to produce a series of derived images highlighting features at different anatomic spatial scales ranging from fine to coarse texture. The scale was selected by altering the spatial scale filter (SSF) value between 0 and 2

to extract MR intensity features of different sizes varying between 0 and 2 mm. Heterogeneity within this ROI was quantified with and without image filtration using the following histogram parameters:

kurtosis, skewness, entropy, and mean value of positive pixels (MPP) ^{6,12,13}.

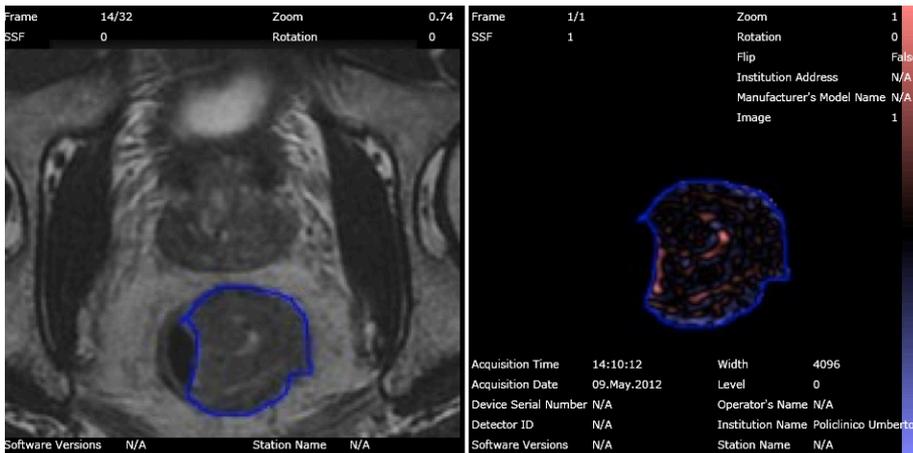


Figure 2: T2w images with the rectal tumor selected and analyzed with texture analysis software before chemoradiation treatment.

As described by Miles KA and colleagues in 2013 ¹⁴ entropy quantifies the irregularity of gray-level distribution. Kurtosis expresses peakedness and tailedness of the histogram, it is inversely related to the number of features highlighted and increases by intensity variations in highlighted features. Skewness evaluated pixel distribution asymmetry and MPP express the average brightness of positive pixel values within the image. Moreover, the absolute changes in the texture parameters after CRT was also calculated as the difference among pre-treatment and post-treatment texture values.

Statistical Analysis AND Outcomes

Descriptive statistics are provided as mean \pm SD for continuous variables. Nonparametric Mann-Whitney U test was used to compare texture parameters (kurtosis, skewness, entropy, MPP) and the response rate among pCR, PR, and NR groups before and after neoadjuvant therapy. Moreover, each parameter was also compared between the different patient subgroups.

We performed further analysis assessing absolute changes among different texture parameters in CR and PR/NR patients before and after neoadjuvant therapy.

To evaluate Performance of Texture parameters, a receiver operating characteristic (ROC) curve analysis was performed to assess the discriminatory power of texture parameters to predict pCR by calculating the areas under the ROC curve (AUCs) and the corresponding P values. ROC curves was computed from mean values of all filters combined and from all filters independently. Pre radiochemo-therapy datasets had been evaluated. Optimal cutoff values were calculated as the cutoff thresholds maximizing the Youden index J, where $J = \text{sensitivity} + \text{specificity} - 1$. Sensitivity and specificity were calculated for the thus determined optimal cutoff values. Statistical analysis was carried out using MedCalc version 12.7.2 (MedCalc Software, Ostend, Belgium). P values of <0.05 were considered statistically significant.

Artificial Intelligence Analysis

A dedicated online software was used (WEKA ®). Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. Moreover, it allows construction of classifier tree in order to correlate different parameters to define the best sequences for the proposed tasks, in our case, diagnostic accuracy and the ability to detect “complete responders”. We evaluate the impact of machine learning algorithms on texture parameters in rectal cancer combining all texture parameters.

Results

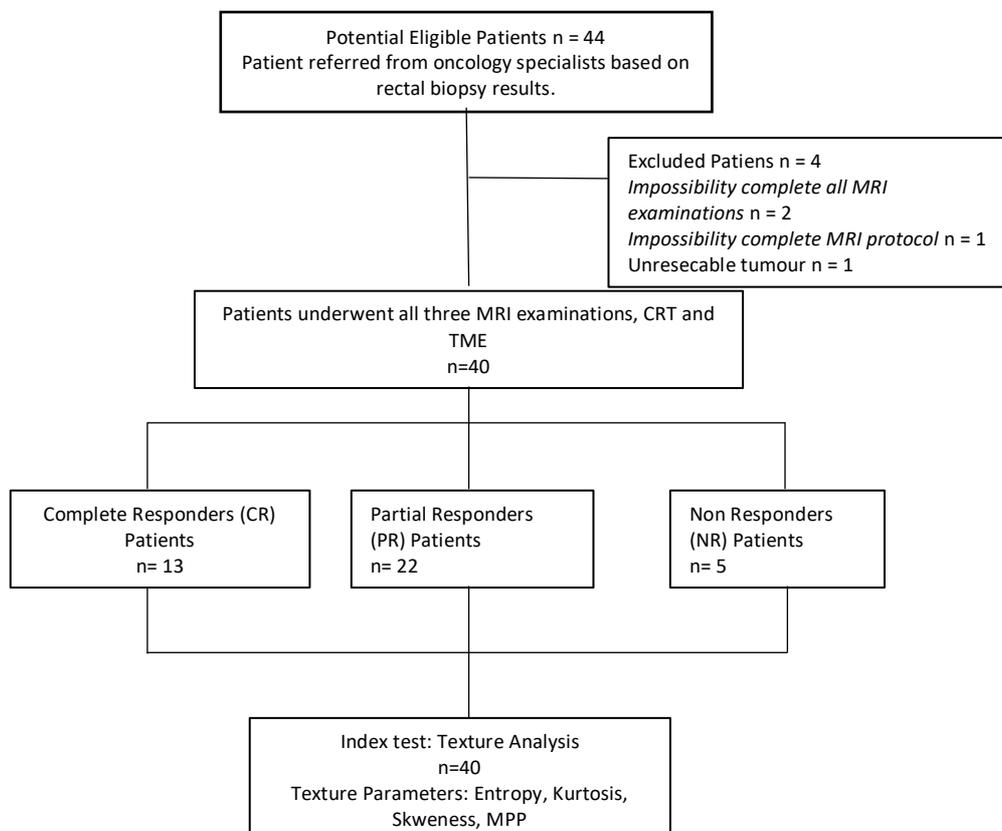
Patients population

The subjects' accrual flowchart is described in Figure 1. Four subjects were excluded from the study due to: (a) incomplete MRI examinations/MRI protocol (n = 3) (b) unresectable lesion thus, the impossibility to have histopathologic assessment (n=1).

The final study cohort consisted of 40 patients (mean age, 64 ± 9 years; age range, 39-82 years), including 24 men (mean age, 63 ± 9 years; age range, 39-82 years) and 16 women (mean age, 65 ± 10 years; age range, 47-81years).

For all patients the response to therapy was confirmed histopathologically after TME. Thirteen patients showed pathological complete response(pCR), Twenty-two patients partial response(pPR) and Five patients were non responders (pNRs).

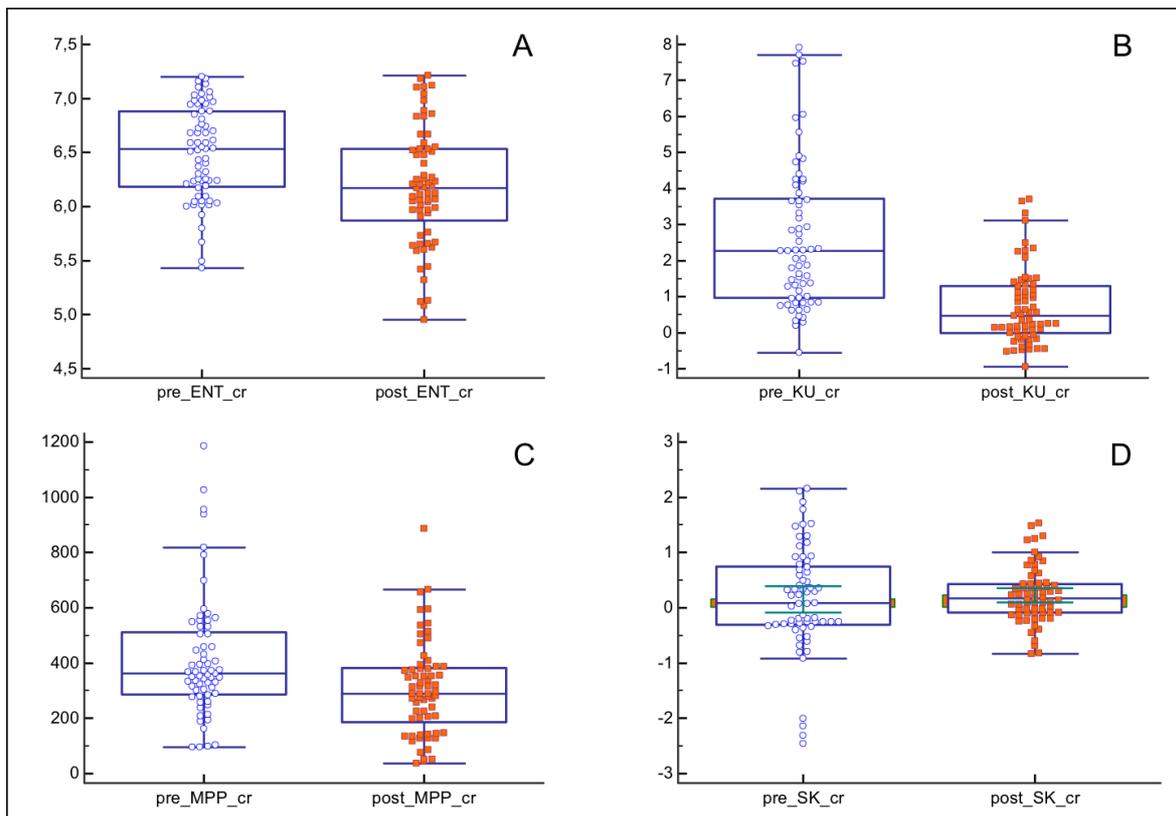
Figure 1: Patient's population flow chart.



Texture results

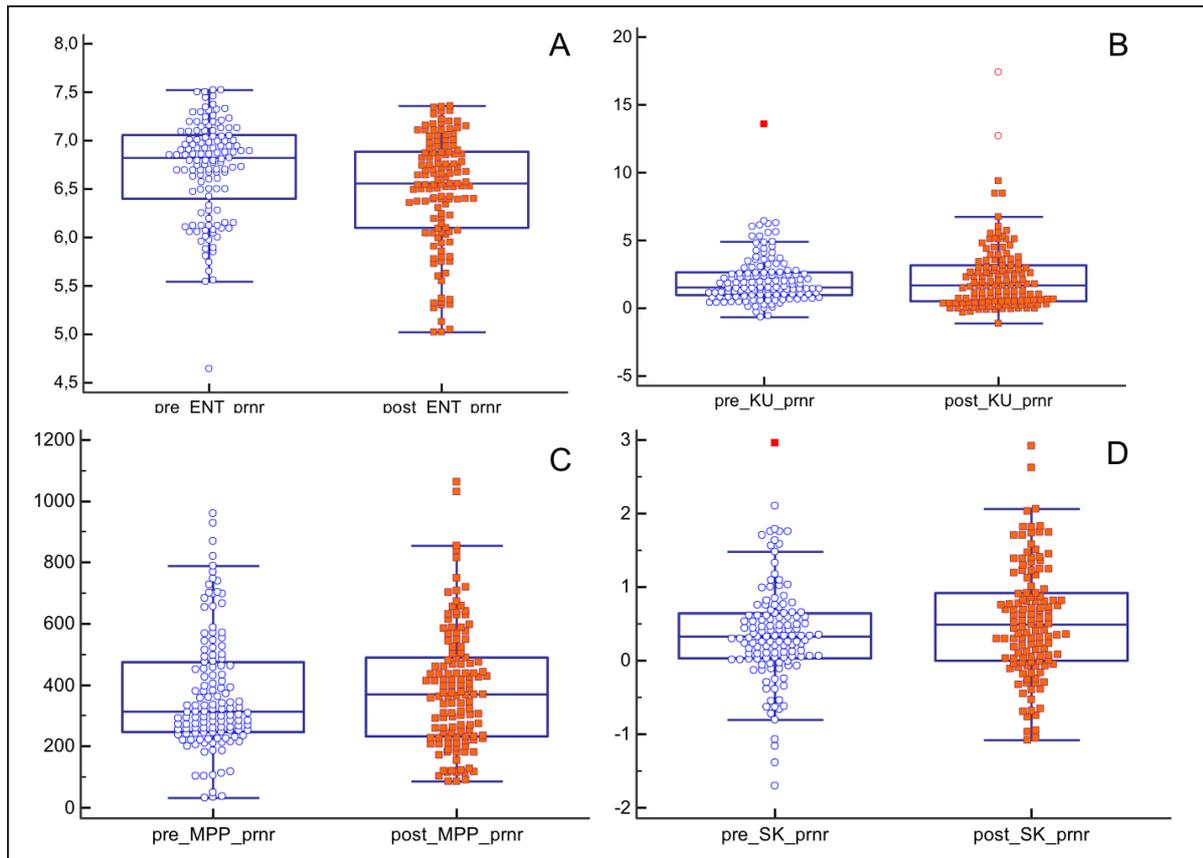
In CR patients' population, the mean values of Entropy, Kurtosis and MPP showed statistical significant differences before and after CRT (Entropy [pre CRT 6.49 ± 0.43 , post CRT 6.17 ± 0.54 ; $p\text{Value} < 0.0001$], Kurtosis [pre CRT 2.60 ± 2.01 , post CRT 0.72 ± 1.05 ; $p\text{Value} < 0.0001$]; MPP [pre CRT 414.24 ± 0219.26 , post CRT 306.33 ± 168.97 ; $p\text{Value} = 0.0023$]. All values significantly decreased after CRT (Figure 3). Skewness showed similar results before and after CRT [pre CRT 0.15 ± 0.96 , post CRT 0.22 ± 0.51 ; $p\text{Value} = 0.53$).

Figure 3: Differences in Texture Parameters before and after CRT in CR population. All differences are statistically significant except for Skewness.



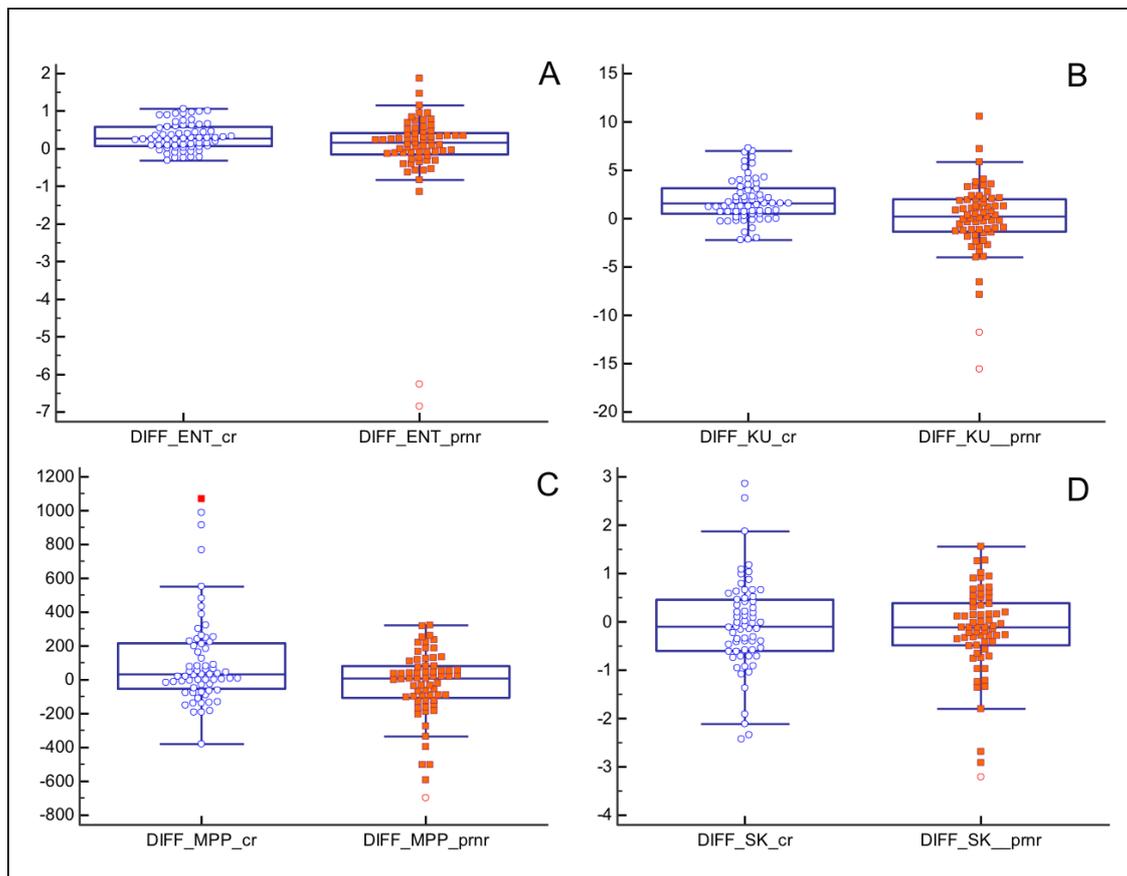
In PR/NR patients' population, only Entropy and Skewness showed statistical significant differences before and after CRT (Entropy [pre CRT 6.70 ± 0.50 , post CRT 6.47 ± 0.57 ; $p\text{Value} < 0.0001$], Skewness [pre CRT 0.35 ± 0.67 , post CRT 0.53 ± 0.75 ; $p\text{Value} = 0.029$] (Figure 4).

Figure 4: Differences in Texture Parameters before and after CRT in PR+NR population. Values of Entropy and Skewness are statistically different before and after CRT.



Absolute changes in Entropy, Kurtosis and MPP before and after CRT, comparing CR patients' population and PR/NR patients' population, showed statistically significant results. Texture parameters seems to decrease after CRT more in CR patients compared to PR/NR patient (Entropy [absolute reduction 0.31 ± 0.35 , in CR group, -0.02 ± 1.28 in PR/NR group; p Value= 0.04], Kurtosis [absolute reduction 1.87 ± 2.19 , in CR group, -0.06 ± 3.78 in PR/NR group; p Value= 0.0005]; MPP [absolute reduction 107.91 ± 274.40 , in CR group, -28.33 ± 202.91 in PR/NR group; p Value= 0.004] (Figure 5).

Figure 5: Absolute change in Texture Parameters before and after CRT in CR and PR/NR populations. All differences are statistically significant but Skeweness.



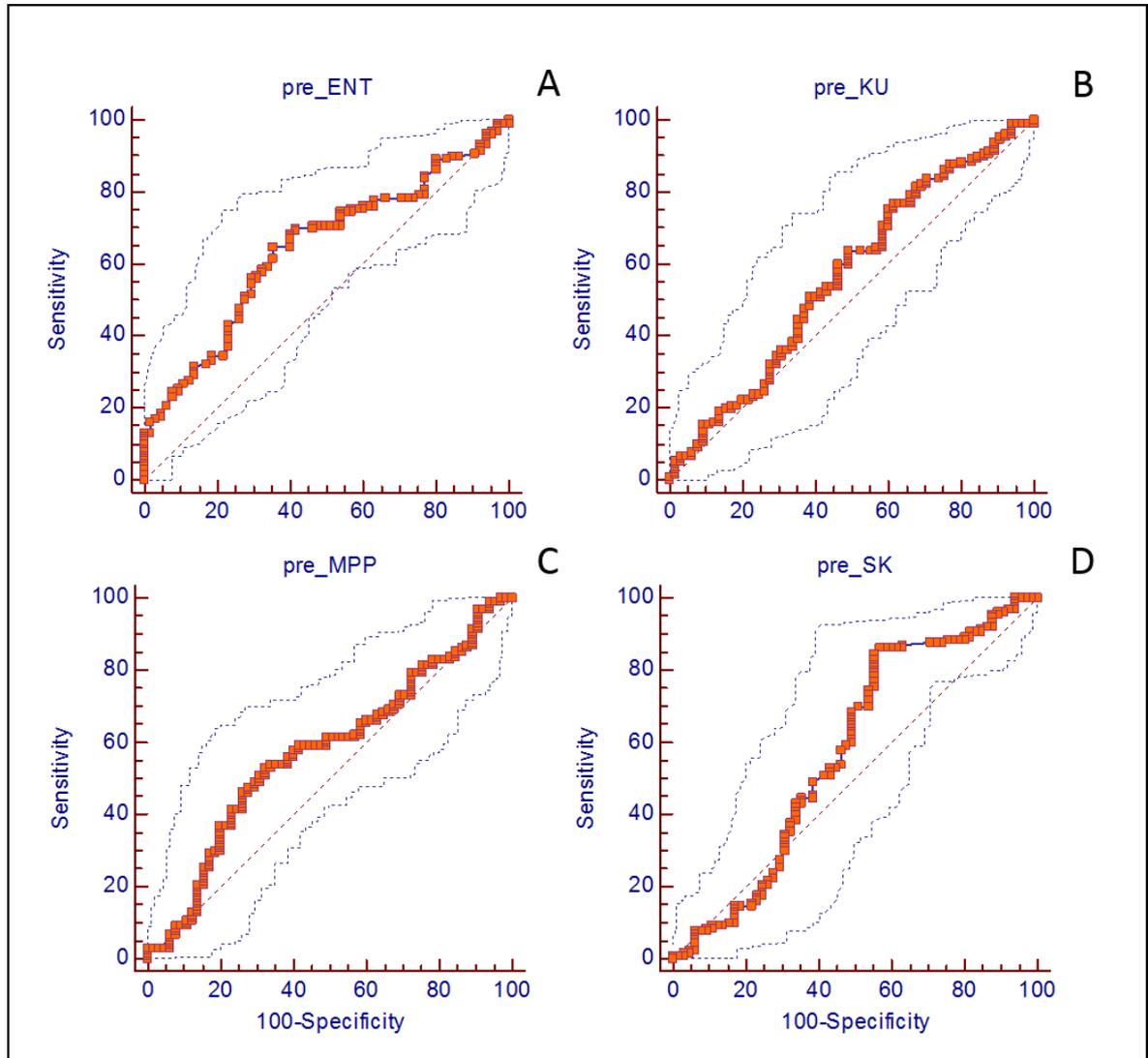
Kurtosis and Entropy seem to be the texture parameters able to predict CR. Pretreatment fine texture (SF1) quantified as kurtosis was significantly lower in patients with pCR in comparison with those with PR + NR ($P = 0.04$; Table 1). No significant difference in pretreatment skewness, and MPP was found for every texture filter applied (Table 1).

Filter	Pre CRT Texture Parameters								P Value
	pCR				PR/NR				
	Entropy	Kurtosis	Skewness	MPP	Entropy	Kurtosis	Skewness	MPP	
SF0	5.93 ±0.27	2.51 ±2.14	0.82 ±0.77	637.79 ±220.09	6.13 ±0.43	3.19 ±2.77	0.92 ±0.76	587.20 ±217.78	>0.05
SF1	6.56 ±0.30	3.52 ±2.04	-0.05±1.00	297.89 ±114.88	6.81 ±0.38	3.48 ±4.93	0.25±0.66	254.75 ±110.62	K: 0.04 E: 0.04
SF1.5	6.63 ±0.35	2.65 ±1.79	0.008 ±0.97	360.79 ±174.55	6.85 ±0.38	1.99 ±1.33	0.26 ±0.59	314.39 ±133.37	>0.05
SF1.8	6.65 ±0.37	2.28 ±2.04	0.008 ±0.95	383.31 ±205.67	6.87 ±0.42	1.60 ±1.48	0.24 ±0.57	333.76 ±149.64	>0.05
SF2	6.67 ±0.38	2.04 ±2.03	-0.017 ±0.92	391.43 ±220.21	6.90 ±0.43	1.33 ±1.49	0.24 ±0.57	345.12 ±158.18	>0.05

Pretreatment AUC for entropy to discriminate between CR and PR+NR was significantly higher than other parameters ($AUC_{entropy}=0.64$; $p=0.0004$; IC; 95% (0.57-0.71); $AUC_{kurtosis}=0.56$; $p=0.13$; IC 95% (0.49-0.63); $AUC_{MPP}=0.57$; $p=0.088$; IC95% (0.50-0.64); $AUC_{skewness}=0.57$; $p=0.11$; IC95% (0.50-0.64) (Figure 6). The optimal pretreatment entropy cut off was ≥ 6.68 , kurtosis cut-off was ≤ 2.78 , MPP cut off was ≤ 322.79 and skewness cut off was >-0.18 , and using this value, the sensitivity and specificity for CR for entropy, kurtosis, MP and skewness were 76.92% and 38.46%, 64.62% and 64.62%, 53.08% and 67.69%, 86.15% and 43.08% respectively.

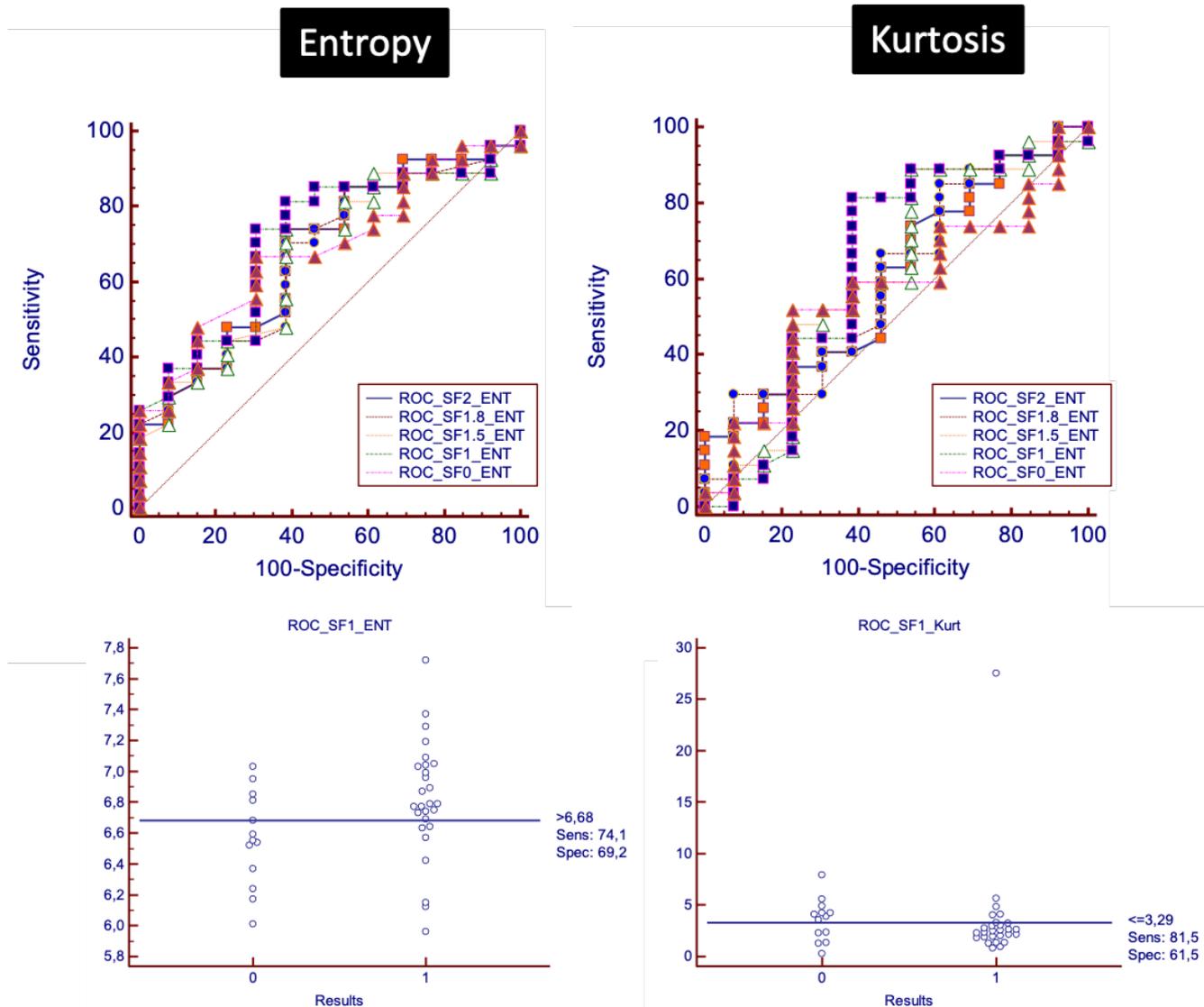
Figure 6: ROC Analysis from mean values combining all filters. ROC curves are shown analyzing the

discriminatory power of baseline texture parameters to distinguish between pCR and PR + NR.



Among all filters used, SF1 has the best diagnostic accuracy to discriminate between CR and PR+NR. The optimal pretreatment entropy cut off was ≥ 6.68 , kurtosis cut-off was ≤ 3.29 , MPP cut off was ≤ 261.39 , and using this value, the sensitivity and specificity for CR for entropy, kurtosis, MP and skewness were 74.1% and 69.4%, 81.5% and 61.5%, 63% and 69.2%,

The best filter for skewness was SF2, improving Sensitivity and Specificity up to 63% and 69.2% respectively (cut off > -0.07) (Figure 7).



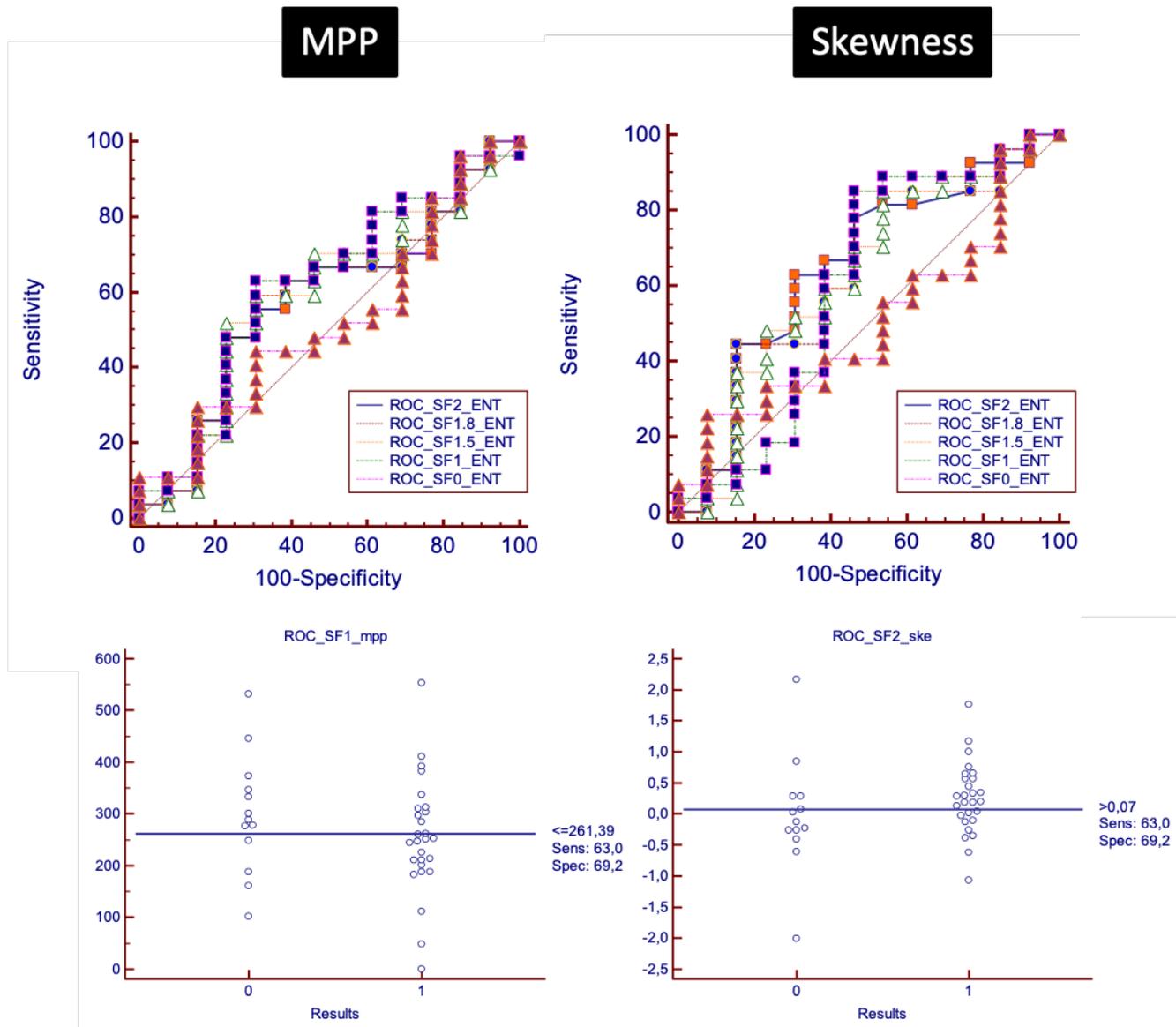


Figure 7: ROC Analysis. ROC curves for each filter are shown analyzing the discriminatory power of baseline texture parameters to distinguish between pCR and PR + NR.

Using Artificial Intelligence software, we obtain a decisional tree that correlate all filters and all parameters in order to obtain the best diagnostic accuracy (Figure 8) and to detect, during the first MRI, all patients that are going to respond completely to CRT.

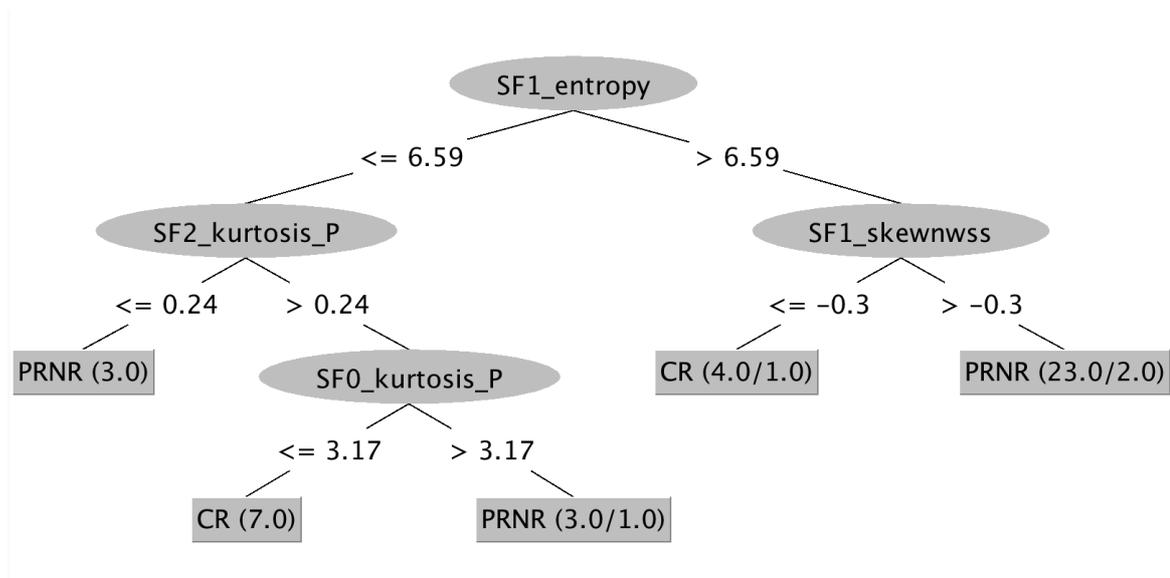


Figure 8: Decisional tree form Weka. All parameters and all Filters had been combined to obtained a map to detect, during the first MRI, the patients that are going to respond completely to CRT.

Discussion

According to our results, TA parameters can be considered accurate imaging biomarkers able to discriminate patients responders to non-responders after neoadjuvant chemotherapy and it can also be used as prognostic factor.

The mean values of Entropy, Kurtosis and MPP parameters show significant differences before and after CRT in CR's population with a decreasing of values after RCT; otherwise in PR/NR patients Entropy and Skewness showed statistical significant differences before and after RCT. Absolute changes among different texture parameters in CR and PR/NR patients before and after neoadjuvant therapy showed statistical significant differences in Entropy, Kurtosis and MPP. In fact, these Texture parameters after neoadjuvant therapy decrease more in CR patients compared to PR/NR patients. Our results are in line with previous results from Jalil and colleagues⁶ that reported a significant correlation between post-treatment kurtosis and entropy and disease free survival and recurrence free survival. Heterogeneity of a tissue is most likely related to oxidative stress and genomic instability that are typical in viable cancerous tissue; thus, its reduction as shown in CR patients, let us to infer the presence of a less active tissue such as fibrotic tissue according to histological complete response¹⁵.

Kurtosis and entropy seem to be the texture parameter able to predict CR (Pvalue = 0.004) According to De Cecco and colleagues^{10,16}, Kurtosis showed significant result in CR patients with a lower value compared to PR/NR. Ng and coworkers showed similar results at fine texture parameters, reporting that Kaplan-Meier survival plots for entropy, uniformity, kurtosis, skewness, and standard deviation of the pixel distribution histogram were significantly different for tumours above and below each respective threshold receiver operating characteristic (ROC) curve optimal cutoff value, with poorer prognosis for ROC optimal values less than 7.89 for entropy and less than 2.48 for kurtosis¹⁵. All these evidences underling the important role of Kurtosis in discriminating responders to non responders prior CHT therapy.

Performance of Texture analysis achieved with ROC curves showed Entropy as the best AUC as parameter to discriminate between CR and PR+NR with a value of 0.64, an entropy cut off ≥ 6.68

(sensitivity and specificity of 76.92% and 38.46% respectively). This result is in contrast with results of De Cecco and colleagues that assessed Kurtosis as the parameter with the best AUC ^{10,16}. Concordance has been found with other study performed on breast ¹⁷ and prostate cancer ¹⁸ where T2 derived TA showed the best AUC in prediction of CR and tumor aggressiveness respectively. This results let infer that Entropy is an expression of pixel irregularity and correlates with higher heterogeneity and aggressiveness of tumor. Moreover, Entropy is a parameter with good repeatability as shown by Gourtsoyianni¹⁹ and colleagues thus this result has an increased value because it could be adopted in the future as a concrete parameter in clinical practise for patient management.

Among all filters used on images (SF0, SF1, SF1.5, SF1.8, SF2), the best performance of Texture analysis parameters with ROC curves discriminating between CR and PR+NR was obtained with SF1 for kurtosis, MPP and entropy, increasing AUCs up to 0.71 for entropy (delta 10.9%), 0.63 for kurtosis (delta 12.5%), and 0.61 for MPP (delta 7%). The best filter for skewness was SF2, improving AUC up to 0.65 (delta 14%).

Texture analysis allows to obtain optional images with filters. A Laplacian or Gaussian bandpass filter is a commonly used advanced image filtration method that alters the image pixel intensity patterns and allows extraction of specific structures corresponding to the width of the filter. Lower filter values correspond to fine texture features, while higher filter values emphasize medium or coarse texture features²⁰. In addition, this filtration step is designed to remove noise and enhance edges, which may make measurements less susceptible to small differences in technique. Denoising or gray-level standardization steps have been used as well as a premeasurement step to help eliminate differences that are technical rather than biologic and to aid in reproducibility. In our study, introducing a filter SF1 had improved diagnostic accuracy for most of the parameters, reducing noise and enhancing images feature related to the biological tissues.

Unfortunately, introducing filters means that the datasets enlarge considerably, resulting in difficulties to extract summary data and to compute results without errors. Just in this case, software using artificial intelligence algorithms may help. In our study machine learning software gave us a

map that combined the most relevant parameters and filters; that map should be intended as a decisional tree, able to detect in all cases Patients that are going to respond completely to CRT.

Despite the encouraging results, our study has some limitations. First, the sample size of analysed patients was limited. Thus, observed correlations among texture parameters, response to therapy and histologic results should be confirmed in larger studies. Second, we did not correlate T2w texture analysis with other imaging biomarker; thus, we could not compare its predictive power with other imaging biomarkers. Moreover, we did not perform inter or intra-reader agreement for the manual segmentation of the ROI; should be interesting in the future assess how results could change or not with different operator or with repetitive sampling. Second, the absence of standardized protocol for texture analysis such as selection of the filtration; more studies are needed to identify which protocol is the more accurate. Third, no longer follow-up has been performed at the moment of the analysis, with the lack of results regarding the predictive value of tumour heterogeneity on patient survival. In the future, a longitudinal study would be helpful in the assessment of the prognostic value of texture analysis in rectal cancer. Finally, the decisional tree that comes from AI software has a perfect diagnostic accuracy only if applied in a population identical to our study population. It is necessary to validate the map thought a larger and different sample to test the real clinical value. However, we applied the decisional tree to a subgroup of patients extracted from our population, obtaining promising results in terms of diagnostic accuracy.

In conclusion, our preliminary results infer that texture parameters obtain from T2w images can potentially have an important role in the patient clinical management as imaging biomarkers of tumoral response to neoadjuvant therapy, in particular stratifying patients with CR from those with PR or NR during the baseline MRI.

References

1. Pham TT, Liney G, Wong K, et al. Study protocol: multi-parametric magnetic resonance imaging for therapeutic response prediction in rectal cancer. *BMC Cancer*. 2017;17(1):465.
2. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol*. 2010;11(9):835-844.
3. Renehan AG, Malcomson L, Emsley R, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol*. 2016;17(2):174-183.
4. Habr-Gama A, Sabbaga J, Gama-Rodrigues J, et al. Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis Colon Rectum*. 2013;56(10):1109-1117.
5. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology*. 2017;284(2):552-561.
6. 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 Dis*. 2017;19(4):349-362.
7. Li M, Fu S, Zhu Y, et al. Computed tomography texture analysis to facilitate therapeutic decision making in hepatocellular carcinoma. *Oncotarget*. 2016;7(11):13248-13259.
8. Liu L, Liu Y, Xu L, et al. Application of texture analysis based on apparent diffusion coefficient maps in discriminating different stages of rectal cancer. *J Magn Reson Imaging*. 2017;45(6):1798-1808.
9. Garajova I, Di Girolamo S, de Rosa F, et al. Neoadjuvant treatment in rectal cancer: actual status. *Chemother Res Pract*. 2011;2011:839742.
10. 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. *Abdom Radiol (NY)*. 2016;41(9):1728-1735.
11. Heald RJ, Ryall RD. Recurrence and survival after total mesorectal excision for rectal cancer. *Lancet*. 1986;1(8496):1479-1482.
12. Sieren JC, Smith AR, Thiesse J, et al. Exploration of the volumetric composition of human lung cancer nodules in correlated histopathology and computed tomography. *Lung Cancer*. 2011;74(1):61-68.
13. Liu L, Liu Y, Xu L, et al. Application of texture analysis based on apparent diffusion coefficient maps in discriminating different stages of rectal cancer. *J Magn Reson Imaging*. 2016.
14. Miles KA, Ganeshan B, Hayball MP. CT texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging*. 2013;13(3):400-406.
15. 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(1):177-184.
16. 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. *Invest Radiol*. 2015;50(4):239-245.

17. Henderson S, Purdie C, Michie C, et al. Interim heterogeneity changes measured using entropy texture features on T2-weighted MRI at 3.0 T are associated with pathological response to neoadjuvant chemotherapy in primary breast cancer. *Eur Radiol.* 2017.
18. Nketiah G, Elschot M, Kim E, et al. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. *Eur Radiol.* 2017;27(7):3050-3059.
19. Gourtsoyianni S, Doumou G, Prezzi D, et al. Primary Rectal Cancer: Repeatability of Global and Local-Regional MR Imaging Texture Features. *Radiology.* 2017:161375.
20. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging.* 2012;3(6):573-589.