RADIOMICS TO IMPROVE OUTCOMES IN CERVICAL CANCER

A thesis submitted to The University of Manchester for the degree of

Master of Philosophy

in the Faculty of Biology, Medicine and Health

2022

Chelmis M. Thiong'o
School of Medical Sciences
Division of Cancer Sciences

Contents

Word Count: 11180

List of Tables

List of Figures

Abbreviations

ADC Apparent Diffusion Coefficient.

BFT Belief Function Theory.

CC Cervical Cancer.

CET1W Contrast Enhanced T1-weighted images.

CT Computed Tomography.

CTV Clinical Target Volume.

CTV-HR High Risk Clinical Target Volume.

CTV-IR Intermediate Risk Clinical Target Volume.

CTV-IR Intermediate Risk Clinical Target Volume.

CTV-LR Low Risk Clinical Target Volume.

DCE Dynamic Contrast-Enhanced.

DWI Difussion-weighted imaging.

EBRT External Beam Radiation Therapy.

FBN Fixed Bin Number.

FBS Fixed Bin Size.

FIGO Fédération Internationale de Gynécologie et d'Obstétrique.

GLCM Grey Level Co-occurence Matrix.

GLDM Grey Level Dependence Matrix.

GLRLM Grey Level Run Length Matrix.

GLSZM Grey Level Size Zone Matrix.

GTV Gross Tumour volume.

HIV Human Immunodeficiency Virus.

HPV Human Papilloma Viruses.

IARC International Agency for Research on Cancer.

IBSI Image Biomarker Standardisation Initiative.

ICC Intraclass Correlation Coefficient.

ITV Internal Target Volume.

LASSO Least Absolute Shrinkage and Selection Operator.

LNM Lymph Node Metastasis.

LoG Laplacian of Gaussian.

LVSI Lymph Vascular Space Invasion.

mpMR Multiparametric-Magnetic Resonance Imaging.

MRI Magnetic Resonance Imaging.

NGTDM Neighbouring Grey Tone Difference Matrix.

OARs Organs at Risk.

PET Positron Emission Tomography.

PTV Planning Target Volume.

ROIs Regions of Interest.

RQS Radiomics Quality Score.

RT Radiation Therapy.

SVM-RFE Support Vector Machine - Recursive Feature Elimination method.

T2W T2-weighted.

TRIPOD Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

VOI Volume of Interest.

WHO World Health Organisation.

Abstract

Radiomics studies identify imaging biomarkers that can be used to predict treatment outcomes. One limitation is model over-fitting due to the high number of radiomic features extracted from a single image. Careful feature selection is therefore paramount to ensure redundant information is discarded, useful information retained. The aim of this study is to evaluate the test-retest repeatability of radiomic features from the GTV and peritumoral regions of T2-weighted and Dixon water-only images in cervical cancer patients.

21 scans, each comprising of 2 T2-weighted and 2 Dixon sequence images, were included in this study. This resulted in a total of 42 test-retest pairs of images. The GTV was contoured by a single observer on each of the T2W images and transferred through rigid registration to the Dixon sequence images. For each image, radiomic features were extracted from 2 regions: the GTV and a peritumoral region extending 2 voxels (1.875mm for T2-weighted images and 1.923mm for Dixon images) from the GTV contour but restricted to remain inside the cervix/uterus so as not to sample surrounding tissues. PyRadiomics was used to extract first-order, shape, and texture radiomic features from the original T2W and Dixon images, and their Laplacian of Gaussian-filtered (σ = 3, 4, and 5) and Wavelet-filtered counterparts. Spearman's $|\rho|$ was used to exclude features highly correlated with volume ($|\rho| \geq 0.9$). Of the remaining features, those with minimal variation in a test-retest pair of images were selected using one-way random effects Intraclass Correlation Coefficient (ICC), to assess absolute agreement of feature values. The criteria ICC(1,1) ≥ 0.9 was used. Remaining features were grouped by feature class and the feature with the lowest Spearman's correlation to volume within the class was considered stable. A total of 1130 features were extracted from each image. Features with Spearman's $|\rho| \geq 0.9$ in relation to volume were discarded. For the T2W images, 86% and 87.1% of the 1130 features extracted from the GTV and peritumoral regions respectively remained. For the Dixon images, 86.1% and 87.8% of the 1130 features extracted from the GTV and peritumoral regions respectively remained. Features with ICC(1,1) ≥ 0.9 were considered to have excellent test-retest repeatability and were retained. For the T2W images, these were 24.9% and 7.6% of the 1130 radiomic features extracted from the GTV and peritumoral region, respectively. For the Dixon images, they were 4.6% and 4.0% of the 1130 radiomic features extracted from the GTV and peritumoral region, respectively.

In conclusion, stable features were identified for each feature class and can be used in larger data-sets to build reliable predictive models. To confirm generalisability of our findings, validation of our results on a similar data-set is required.

Declaration

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `https://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations/`) and in The University's policy on Presentation of Theses

Acknowledgements

I would like to thank my supervisors, Dr. Andrew Green, Dr. Alan McWilliam and Dr. Gareth Price, and advisor, Prof Peter Hoskin. Most especially, Dr. Green for his invaluable feedback and guidance throughout the program. I've learned a lot.
I would also like to thank:
The UK Newton Fund through the funding of the Development in Africa with Radio Astronomy (DARA) Big Data project for funding my MPhil programme.
Linzi Stirrup for the tremendous support, especially at the start of the programme when I faced problems with my BRP. Words aren't enough.
The entire RRR group for the amazing support and feedback. I enjoyed being a part of this group.
Angie Davey and Chelsea Sargeant for helping me make sense of radiomics and everything after the fact.
The RRR PhD students, for being an amazing support group in terms of starting and going through a research programme and helping me to adjust to life in Manchester.
Finally, my friend Lorna Njanja and my family for their support.

The Author

- BSc. Microprocessor Technology and Instrumentation, The University of Nairobi, 2017

Chapter 1

Introduction

1.1 Cervical Cancer

Cervical Cancer (CC) is the fourth most common cancer in women, with slightly more than half of diagnosed women dying in 2020 [1] worldwide. Figure 1.1.1 shows the incidence and mortality rates of the most common cancers in women as of 2020, according to the World Health Organisation (WHO) International Agency for Research on Cancer (IARC) 2020. The incidence and mortality of CC is estimated to grow by 21% and 27% respectively by the year 2030. This is especially in middle and low-income countries where incidence is higher and mortality is twice the rate in high income countries [2]. With intervention such as the elimination strategy employed by WHO, however, it is projected that the incidence and mortality could reduce significantly [2]–[4].

According to WHO, cervical cancer can be controlled through primary preven-

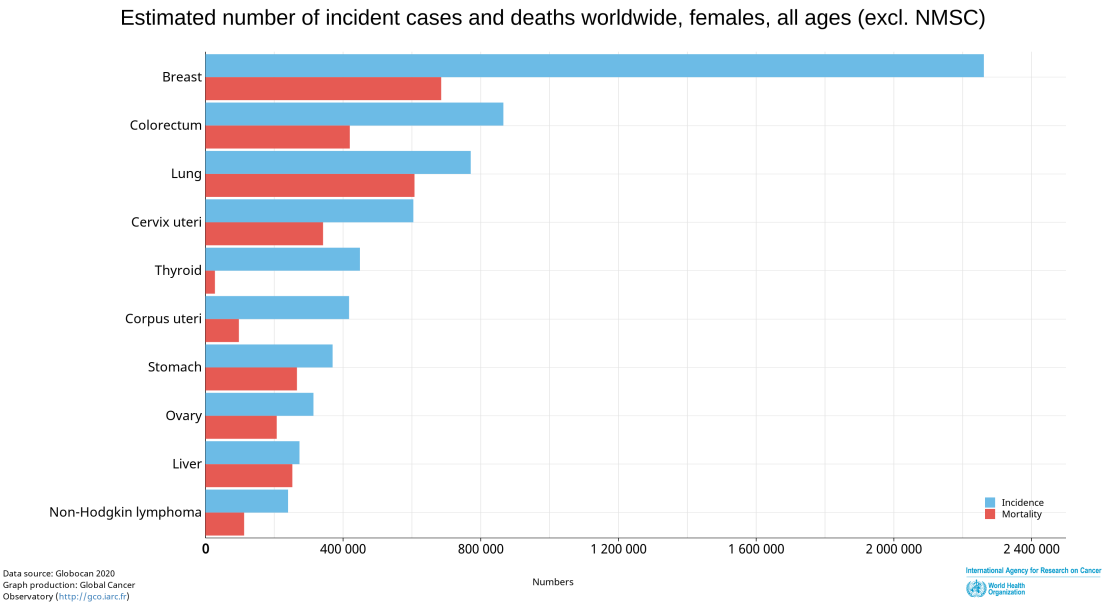


Figure 1.1.1: Estimated number of cancer incidence and mortality worldwide in females, of all ages in the year 2020 [5]

tion, secondary prevention, tertiary prevention and palliative care [2], [6]. Primary prevention entails vaccination against Human Papilloma Viruses (HPV). Secondary prevention entails screening and treating pre-cancerous lesions while tertiary prevention entails treatment of cervical cancer and palliative care. These prevention mechanisms are, however, not easily accessible to populations in low and middle income countries, resulting in the current projected rise in incidence and mortality rates.

A majority, (over 95%), of cervix cancer cases are attributed to persistent infection with the sexually transmitted HPV [1]. There are various types of HPV, with the most notorious being types 16 and 18 which have been attributed to 70% of cervical cancer cases [1]. While most HPV infections and pre-cancerous lesions clear up on their own, HPV infections can become chronic and pre-cancerous lesions can progress to invasive cancer. Women with a weakened immune system, such as those living with HIV are therefore at a higher risk [7]. HPV, however, can be vaccinated against and efforts by the World Health Organisation (WHO) are being increased to ensure vaccine uptake [8], especially in young adolescent girls [9].

The most common screening test is the Pap smear test, which detects precancerous cells or cancerous cells (indicative of cervical cancer). A positive result is followed by a histopathological assessment of a cervical biopsy and imaging [10] to check extent of disease. Cervical cancer staging is determined clinically based on tumour size and the degree of pelvic extension. Figure 1.1.2 show a summary of cervical cancer staging according to the Fédération Internationale de Gynécologie et d'Obstétrique (FIGO).

Early-stage cervical cancer can be managed surgically while the recommended treatment options for late-stage cervical cancer also includes chemotherapy, external beam radiotherapy, and brachytherapy [6]. Section 1.3 explains more about radiotherapy. The survival rate for cervical cancer patients depends on disease stage, among other factors [4], [11], [12]. For instance, Lymph Node Metastasis (LNM), where there's nodal involvement of disease, is a risk factor associated with survival of cervical cancer patients.

1.2 Imaging

Imaging plays an integral role in the detection, staging [13], treatment planning and treatment of tumours as well as assessment of treatment response. Lately, imaging has been seen to provide a non-invasive way to capture tumour heterogeneity [14], which, consequently gives huge potential for targeted treatment. This makes imaging potentially preferable to biopsy, pending further research into repeatability and correlation to histopathology [14], [15].

Stage I:

The carcinoma is strictly confined to the cervix uteri (extension to the corpus should be disregarded)

- **IA** Invasive carcinoma that can be diagnosed only by microscopy, with maximum depth of invasion <5 mm[a]
  - **IA1** Measured stromal invasion <3 mm in depth
  - **IA2** Measured stromal invasion ≥3 mm and <5 mm in depth
- **IB** Invasive carcinoma with measured deepest invasion ≥5 mm (greater than stage IA), lesion limited to the cervix uteri[b]
  - **IB1** Invasive carcinoma ≥5 mm depth of stromal invasion and <2 cm in greatest dimension
  - **IB2** Invasive carcinoma ≥2 cm and <4 cm in greatest dimension
  - **IB3** Invasive carcinoma ≥4 cm in greatest dimension

Stage II:

The carcinoma invades beyond the uterus, but has not extended onto the lower third of the vagina or to the pelvic wall

- **IIA** Involvement limited to the upper two-thirds of the vagina without parametrial involvement
  - **IIA1** Invasive carcinoma <4 cm in greatest dimension
  - **IIA2** Invasive carcinoma ≥4 cm in greatest dimension
- **IIB** With parametrial involvement but not up to the pelvic wall

Stage III:

The carcinoma involves the lower third of the vagina and/or extends to the pelvic wall and/or causes hydronephrosis or non-functioning kidney and/or involves pelvic and/or paraaortic lymph nodes[c]

- **IIIA** Carcinoma involves the lower third of the vagina, with no extension to the pelvic wall
- **IIIB** Extension to the pelvic wall and/or hydronephrosis or non-functioning kidney (unless known to be due to another cause)
- **IIIC** Involvement of pelvic and/or paraaortic lymph nodes, irrespective of tumor size and extent (with r and p notations)[c]
  - **IIIC1** Pelvic lymph node metastasis only
  - **IIIC2** Paraaortic lymph node metastasis

Stage IV:

The carcinoma has extended beyond the true pelvis or has involved (biopsy proven) the mucosa of the bladder or rectum. A bullous edema, as such, does not permit a case to be allotted to stage IV

- **IVA** Spread of the growth to adjacent organs
- **IVB** Spread to distant organs

[a]Imaging and pathology can be used, when available, to supplement clinical findings with respect to tumor size and extent, in all stages.

[b]The involvement of vascular/lymphatic spaces does not change the staging. The lateral extent of the lesion is no longer considered.

[c]Adding notation of r (imaging) and p (pathology) to indicate the findings that are used to allocate the case to stage IIIC. For example, if imaging indicates pelvic lymph node metastasis, the stage allocation would be stage IIIC1r and, if confirmed by pathological findings, it would be Stage IIIc1p. The type of imaging modality or pathology technique used should always be documented. When in doubt, the lower staging should be assigned.

Figure 1.1.2: Cervical Cancer (CC) diagnosis and staging according to FIGO (2018) [13]

The main imaging modalities used are Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). The modality used is dependent on the location of the tumour and the kind of information being gathered about the tumour. CT scans give high resolution structural information about the location and shape of tumours [10]. MRI scans, due to their excellent soft tissue contrast, give good visualisation of the primary tumour and the extent of soft tissue disease [16] as well as tumour functional and molecular characteristics [10]. PET imaging, on the other hand, is a type of nuclear medicine imaging where a radioactive tracer is used [10]. In oncology, F-fluorodeoxyglucose (FDG) radioactive tracer is commonly used to assess tumour metabolism, specifically glucose uptake in the tumour.

In this study, we will be using multiparametric-MRI scans. MRI is increasingly becoming a common imaging modality for cervical cancer, because it offers high accuracy and detailed evaluation of pelvic tissues and organs as well as good contrast

resolution ([17]–[21]) with no radiation [22]. It is extremely versatile and can provide multiparametric information from both morphological and functional signals [23]. Multiparametric-Magnetic Resonance (mpMR) Imaging involves MR imaging where multiple MRI sequences are combined[24], [25]. It gives a more accurate assessment of the tissue micro-environment than the conventional MRI [25]–[27]. In this study, we will be looking at 2D T2 weighted Turbo Spin Echo (TSE) images and 3D DIXON gradient echo water-only images. T2-weighted imaging (T2WI) offers good signal-to-noise ratio which gives good soft tissue contrast [28], [29]. In cervical cancer, T2W scans are good for showing the extent of disease [30]. Water-only Dixon imaging is where images are acquired with uniform suppression of the fat signal in a single acquisition [31], [32]. In cervical cancer imaging, fat suppression allows for better tissue mapping and consequently, better co-registration of images. This is because the bright fat signal may reduce the tumour's conspicuity leading to mis-registration [32].

1.3 Radiation Therapy (RT)

Treatments for cervical cancer include surgery, Radiation Therapy (RT)/radiotherapy, chemotherapy or a combination of the methods. Treatment method used depends on cervical cancer stage and other factors such as whether a patient wishes to maintain fertility [33]. For early-stage cervical cancer, surgery and/or radiotherapy is used for treatment while for advanced-stage cervical cancer, radiation and/or chemotherapy and sometimes surgery is used for treatment [34].

Radiotherapy uses ionising radiation to kill cancer cells and constitutes a major part of treatment for 50% of cancer patients [35], [36]. For cervical cancer, it can be delivered internally (brachytherapy) and/or externally (External Beam Radiation Therapy (EBRT)) and is used as part of the main treatment or to treat local recurrence and distant failure. Radiotherapy can be used for curative purposes (curative/radical radiotherapy), administered before local treatment (neoadjuvant radiotherapy) [37] and given after primary treatment to eradicate residual disease and reduce risk of recurrence (adjuvant radiotherapy) [38].

If radiotherapy is chosen as part of the treatment, a radiotherapy planning CT scan is acquired. This scan is used to design a personalised treatment plan for the patient. It covers the target regions where radiation will be delivered as well as nearby regions at risk of receiving some radiation, Organs at Risk (OARs). Expectations of motion - a challenge in radiotherapy delivery - such as of the cervix and uterus in cervical cancer are also defined in this scan [39]. In cervical cancer, OARs include the bladder, rectum and bowel. Target volumes include, the Gross Tumour volume (GTV) which is the visible disease region at diagnosis, the Clinical Target Volume

(CTV) which covers both the GTV as well as areas that may have microscopic disease, the Internal Target Volume (ITV) which is the CTV and internal margins to compensate for internal motions, and the Planning Target Volume (PTV) which covers the CTV and the uncertainties in the positioning and set-up for treatment delivery [39]–[42]. The CTV the High Risk CTV, CTV-HR, (the region suspected to have the highest risk of recurrence - typically, the entire cervix), Intermediate Risk CTV, CTV-IR, (the region including the CTV-HR and a margin surrounding the cervix borders) and the Low Risk CTV, CTV-LR, (the areas at risk for microscopic spread from the GTV) [39].
Radiotherapy dose directly relates to tumour control but also incurs risk to healthy tissues and organs, so RT delivery while reducing normal tissue damage and toxicity is paramount [36], [43]. This concept is represented by a "Therapeutic Ratio, which is the ratio of the tumour control probability (TCP) and normal tissue complication probability (NTCP) at a specified level of response (usually <5% or 0.05) for normal tissue" [44]. Adaptive radiotherapy has gained interest recently as it entails a radiation treatment plan that accounts for changes such as internal motion, tumour shrinkage, patient setup or machine delivery deviations and it estimates the actual delivered dose to a patient as the treatment progresses [45].
In this work, we will be looking at radiomic features from the GTV and a peritumoral region.

1.4 Radiomics

Radiomics is "the high-throughput extraction and analysis of large amounts of image features from radiographic images" [14], [18], [46], [47]. It is an emerging field of image analysis that has the potential to assist in clinical decision making and patient risk stratification [48]–[50]. Radiomics aims at identifying image biomarkers ("tests that can be measured objectively and evaluated as an indicator of a biologic process" [51]) that can be used to predict treatment outcomes. This is through exploring how various image characteristics or features, describe aspects of the target area, such as the tumour. The features, described in sub-section 1.4.1 below, non-invasively give information, that may be imperceivable by the human eye and specific to tumours which consequently gives potential for precise diagnosis, staging, treatment response assessment, evaluation of the tumours and survival prediction [19], [50], [52]–[59]. One benefit of radiomics is that it takes advantage of an existing cancer treatment protocol - imaging - which is used to diagnose, stage, plan treatment and monitor disease progression [60].
A typical radiomics workflow follows the following steps: Image acquisition, image segmentation, pre-processing, feature extraction and finally, data analysis. These

steps are described in detail in subsection 1.4.3.

1.4.1 Radiomic Features

Radiomic features are the features extracted from radiographic images [14], [19]. They describe the distribution of intensity values of individual voxels, spatial relationships between various intensity levels, texture heterogeneity patterns, descriptors of tumour geometry, relations of tumour to surrounding tissues [14], among others. They also aid in prediction of outcomes such as treatment response since they have been seen to change with continued treatment [18], [52].
There are various platforms for radiomic features extractions, including PyRadiomics [61], CERR [62], TexRAD [63], LIFEx [64] and IBEX [65]. It is considered best practice to specify the software used and the version of the software and to also ensure the specific software version adheres to the Image Biomarker Standardisation Initiative (IBSI) guidelines as suggested in a study by Fornacon-Wood et al (2020) [66]. These guidelines have been set to standardise radiomic feature definition and calculation [67]. In this work, we will be using PyRadiomics [61]. Radiomic features - as defined in the PyRadiomics documentation [61] - are categorised as follows:
First-Order Statistics Features. These are also referred to as intensity features and they describe the distribution of grey-level signal intensities of individual voxels within a Regions of Interest (ROIs) [68]–[70]. There are 19 first order statistics features in PyRadiomics. These features have potential to predict Lymph Node Metastasis (LNM) which is associated with poor prognosis in cervical cancer patients [71], [72].
Shape Features. They describe the geometrical characteristics of the ROIs and include shape-based 2D and 3D features. They can also be referred to as morphological [73] or geometric features [74]. There are 16 3D and 10 2D shape features.
Texture Features. These are matrix features, based on the joint probability distribution of pairs of voxels, that examine the spatial relationship between grey-level signal intensities in an image. First, a matrix is defined: Grey Level Co-occurence Matrix (GLCM), Grey Level Run Length Matrix (GLRLM), Grey Level Size Zone Matrix (GLSZM), Neighbouring Grey Tone Difference Matrix (NGTDM) and Grey Level Dependence Matrix (GLDM), after which some metrics on the matrix are evaluated [70]. Texture features have been seen to predict the stage [75] and give the histological type of patients with cervical cancer [76].
There are 24 GLCM features and these describe the statistical information about how pixel pairs are distributed in the image by showing the distribution characteristics of brightness and locations between image pixels with similar brightness [70], [77]. Figure 1.4.1 shows an example of how to find the Grey-Level Co-Occurrence matrix of an image in the 0° direction.

There are 16 GLRLM features. Galloway et al (1975) described GLRLM features

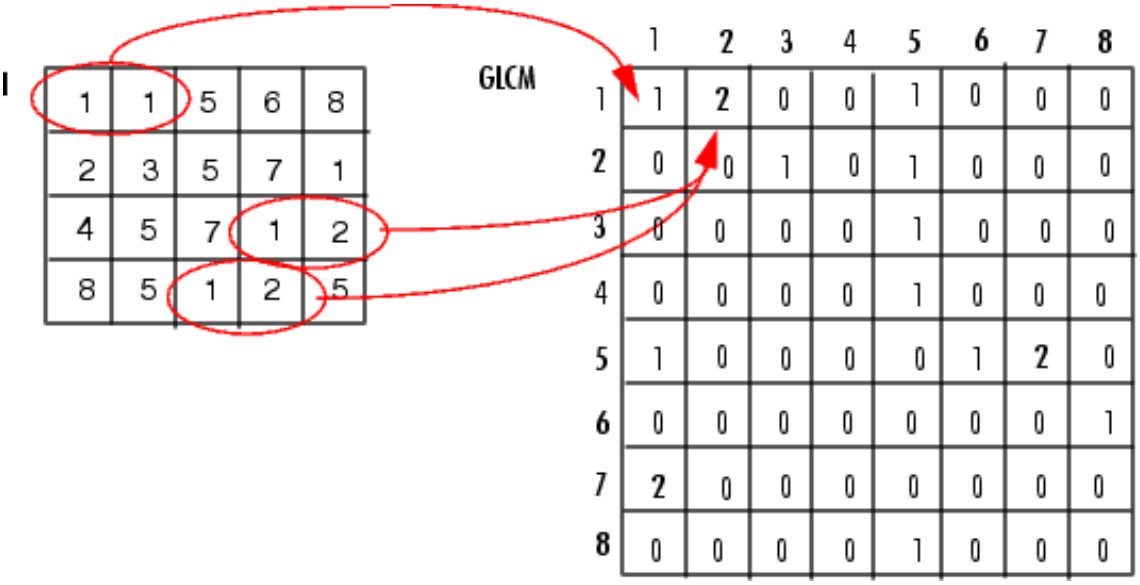


Figure 1.4.1: Grey-Level Co-Occurrence Matrix example. With I being the image and GLCM being its corresponding GLCM matrix, element (1,1) on GLCM contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. element (1,2) on GLCM contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2 [78].

as those showing the length of consecutive voxels having the same intensity in a pre-set direction in the image. This means texture in a specific direction, where fine texture has more short runs while coarse texture has more long runs with different intensity values [68], [69]. Figure 1.4.2 shows an example of how to find the Grey-Level Run-Length matrices of an image in 4 directions (0°, 45°, 90° and 135°.

There are 16 GLSZM features. Thibault et al (2014) described GLSZM features as those that quantify gray level zones in an image, which are defined as the number of connected voxels that share the same gray level intensity. The region of interest is homogeneous when the matrix is wide and flat, and it is heterogeneous when the matrix is narrow. GLSZM is rotation-independent, unlike GLCM and GLRLM. This means that, it results in 1 matrix for each image, unlike rotation-dependent features that result in a matrix for each rotation direction. Figure 1.4.3 shows an example, from Van Griethuysen et al (2017) [61], of how to find the GLSZM of an image.

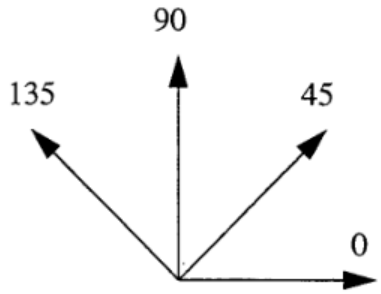
There are 5 NGTDM features and these quantify the difference between a grey value and the average grey value of its neighbours within a specified distance. The resultant matrix contains the sum of absolute differences for the grey levels. Figure 1.4.4 illustrates an example of how to find the neighbourhood grey tone difference matrix of an image, I.

The 14 GLDM features quantify gray level dependencies in an image, which are defined as the number of connected voxels within a certain distance that are depen-

Original image

1	1	1	2	2
3	4	2	2	3
4	4	4	4	4
5	5	3	3	3
1	1	3	4	5

Four directions



Run length

Gray level	0	1	2	3	4	5
	1	0	1	1	0	0
	2	0	2	0	0	0
	3	3	0	1	0	0
	4	2	0	0	0	1
	5	1	1	0	0	0

Run length

45	1	2	3	4	5
1	5	0	0	0	0
2	0	2	0	0	0
3	4	0	1	0	0
4	5	1	0	0	0
5	3	0	0	0	0

Gray level	90	1	2	3	4	5
	1	5	0	0	0	0
	2	2	1	0	0	0
	3	4	1	0	0	0
	4	5	1	0	0	0
	5	3	0	0	0	0

135	1	2	3	4	5
1	5	0	0	0	0
2	4	0	0	0	0
3	6	0	0	0	0
4	5	1	0	0	0
5	3	3	0	0	0

Figure 1.4.2: Grey-Level Run-Length Matrix example in 4 directions: 0°, 45°, 90° *and* 135° *[69]*

dent on the center voxel. IBSI names GLDM features as Neighbouring Grey Level Dependence Matrix features. This feature is rotationally independent [47]. Figure 1.4.5, is an example of how to find the grey level dependence matrix, P, of an image, I. Table 1.4.1 shows the MR sequence, the distribution of features, the site looked at, the number of patients, and the feature selection method used in a few MRI-based radiomic studies for cervical cancer. In summary, Wang et al (2019), Shi et al (2022) and Wu et al (2019) used MRI-based radiomic features and showed that radiomic features could predict lymph node metastasis in cervical cancer [68], [80], [81]. Fiset et al (2019) used MRI-based radiomic features in a test-retest analysis to find stable radiomic features [82]. They found over 200 features to be stable in images acquired with 14-47 minutes of each other. Hua et al (2020) found that considering radiomic

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1.4.3: I is a 5X5 image with 5 discrete grey levels and P is the resultant Grey-Level Size Zone matrix. In the grey level size zone matrix P(i,j), the (i,j)th element equals the number zones with grey level i and size j in the ROI. [61]

$$\mathbf{I} = \begin{bmatrix} 1 & 2 & 5 & 2 \\ 3 & 5 & 1 & 3 \\ 1 & 3 & 5 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix}$$

i	n_i	p_i	s_i
1	6	0.375	13.35
2	2	0.125	2.00
3	4	0.25	2.63
4	0	0.00	0.00
5	4	0.25	10.075

Figure 1.4.4: I is a 4X4 image with 5 discrete grey levels, but missing voxels of discrete value 4. the resultant matrix is such that i is the grey level; n_i is the number of voxels with grey level i; p_i is the grey level probability and equal to n_i/N_v (N_v is the total number of voxels with a neighbour) while s_i is the sum of absolute differences for grey level i [79]

features of cervical cancer peritumor tissues could contribute to improving Lymph Vascular Space Invasion prediction performance [83]. Fang et al (2020) showed an MRI-based radiomics score could be used as a prognostic biomarker for patients with early-stage (IB-IIA) cervical cancer and facilitate clinical decision [84]. Liu et al (2019) carried out a reproducibility study of apparent diffusion coefficient in cervical cancer [85]. Traverso et al (2020) found that applying normalisation prior

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 2 & 3 & 0 \\ 1 & 4 & 4 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix}$$

Figure 1.4.5: I is a 5X5 image with 5 discrete grey levels and P is the resultant Grey-Level Dependence matrix. In the grey level dependence matrix P(i,j), the (i,j)th element equals the number zones with grey level i and size j in the ROI. [79]

to features extraction increased the reproducibility of ADC-based radiomics features [86]. Takada et al (2020) used MRI radiomics to predict recurrence [87]. First order features and textural features are most used, apart from studies investigating a specific radiomic feature as in the study [88].

In this study, we will extract first-order, shape and texture features on the original images, images that have gone through Laplacian of Gaussian (LoG) transformation with sigma values (σ = 3, 4, and 5) and images that have undergone Wavelet filtering.

Reference	MR Sequence	Site	First-Order	Shape	Texture	Filtered	No. of patients	Selection method
[68]	T2W & DWI	CC	✓	✓	✓	✓	96	LASSO
[82]	T2W	CC	✓	✓	✓	✓	56	ICC
[83]	CET1W & T2W	CC	✓	✓	✓	X	111	SVM-RFE
[84]	CET1W & T2W	CC	✓	✓	✓	✓	248	Cox model
								LASSO
[85]	T2W & DWI (ADC)	CC	✓	X	✓	✓	160	LASSO
[86]	T2W & DWI (ADC)	CC	✓	✓	✓	X	81	ICC
[87]	T2W & DWI (ADC)	CC	✓	✓	✓	X	87	
[81]	T2W & DWI (ADC)	CC	✓	✓	✓	✓	189	BFT
[89]	CET1W & T2W	CC	✓	✓	✓	✓	169	Mann-Whitney U test
								LASSO

Continued on next page

Table 1.4.1 – continued from previous page

Reference	MR Sequence	Site	First-Order	Shape	Texture	Filtered	No. of patients	Selection method

Table 1.4.1: Some MRI-based radiomic features in the literature.
✓*- features in the study*
X - features not mentioned in the study
T2W - T2-weighted
DWI - Difussion-weighted imaging
CET1W - Contrast Enhanced T1-weighted images
ADC - Apparent Diffusion Coefficient
CC - Cervical Cancer
LASSO - Least Absolute Shrinkage and Selection Operator
DCE - Dynamic Contrast-Enhanced MR
r - Pearson's Correlation Coefficient
ICC - Intraclass Correlation Coefficient Filtered - radiomic features extracted from pre-processed or filtered images
No. of patients - number of patients used in the study
SVM-RFE - Support Vector Machine - Recursive Feature Elimination method
Cox model - univariate Cox proportional hazards model
BFT - Belief Function Theory

1.4.2 Applications of Radiomics

Radiomics has been seen to hypothesised to predict treatment response and outcomes [15], [90]. For instance, a study by Yip et al (2016) [18], found that first-order features changed according to treatment response, that NGTDM derived coarseness, busyness, and contrast could better differentiate between responders and non-responders to chemo-radiotherapy and that coarseness predicted patient overall survival. Some studies, Shukla- Dave et al (2012), King et al (2013), Peng et al (2013) have shown a relationship between radiomic feature changes and tumour changes in response to treatment. Coroller et al (2015) also showed that radiomic features from CT images could potentially predict potential for metastasis in lung cancer patients. Other studies, [54], [60], have shown that radiomics can be used to characterise tumour aggressiveness, hence tailor an intensive treatment regimen which consequently improves prognosis.

Gillies et al (2016) found that radiomics can be used for image guided biopsy collection due to its ability to show tumour heterogeneity. This way biopsies are collected from the most informative part of the tumour a priori [53], such as in [91].

Studies have shown that radiomics can allow for the differentiation between early stage and advanced stage tumours [18]. For instance, Mu et al (2015), used PET radiomic features to classify CC patients into early and advanced stage CC. Early patient stratification can aid in individualised treatment[18].

Radiomics can allow for the differentiation between malignant and benign tissue. For example, [92] showed that radiomics had potential to differentiate between cancerous and non-cancerous prostate tissue. Nie et al (2008), used MR radiomic features to differentiate between malignant and benign tumours in breasts.

Image biomarkers have been seen to have a relationship with tumour genetics. For instance, Nair et al (2012, 2014), showed that some first-order radiomic features strongly correlated with gene signatures and expressions relating to patient survival.

In summary, the table 1.4.3 shows that MR radiomics in CC has been used to predict disease-free survival, to predict lymph-vascular space invasion (LVSI), investigate repeatability and reproducibility of MR-based radiomic features, investigate radiomic feature sensitivity to inter-observer variability, segmentation methods, reconstruction algorithms and image processing, investigate treatment response and disease free survival.

1.4.3 Radiomics Workflow

The process of a radiomics study is as summarised in figure 1.4.6 [93]. Each step in the workflow affects the results of a radiomics study as shown in Figure 1.4.7 and

Table 1.4.5.

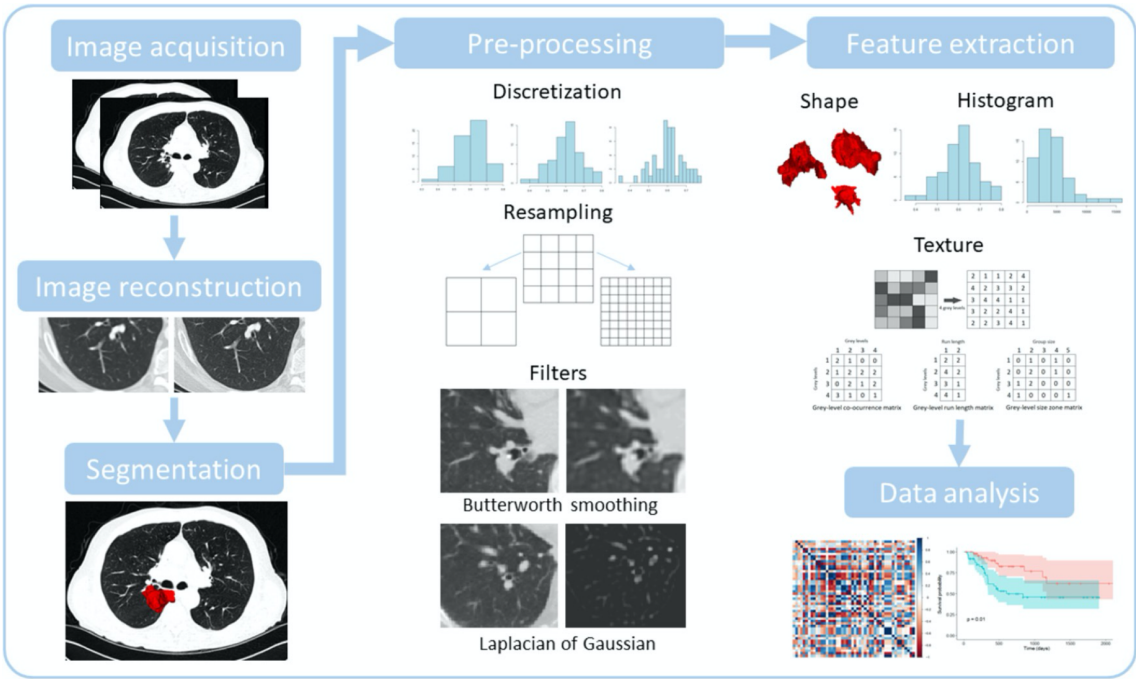


Figure 1.4.6: A radiomics workflow [93]

Image Acquisition

This is the first step in the workflow. It entails acquiring high quality and standardised images. Each imaging modality has its own characteristics that could affect a radiomics study [50] as shown in figure 1.4.7.

The acquisition mode, reconstruction parameters, smoothing, segmentation threshold, as described in Figure 1.4.7, could affect the outcomes of a radiomics study. For instance, in MR imaging, acquisition parameters greatly affect stability of radiomics [82]. This is because MR images do not have fixed tissue-specific numeric voxel intensity units, so MR imaging under constant conditions could give varying results [94]. The optimal way of solving this would be to standardise all image acquisition protocols [95] or by performing careful image processing before feature extraction is done [96]. For instance, performing normalisation before quantitative image analysis is commonly carried out [86], [94]. Another challenge with MR-based radiomics studies is that different MR scan protocols may have different voxel sizes. In such a case, performing resampling prior to quantitative image analysis is advised [97], [98]. Table 1.4.2, adapted from a publication looking into radiomics by Scapicchio et al (2021) [70], shows some pre-processing techniques that could be applied to images before quantitative image analysis.

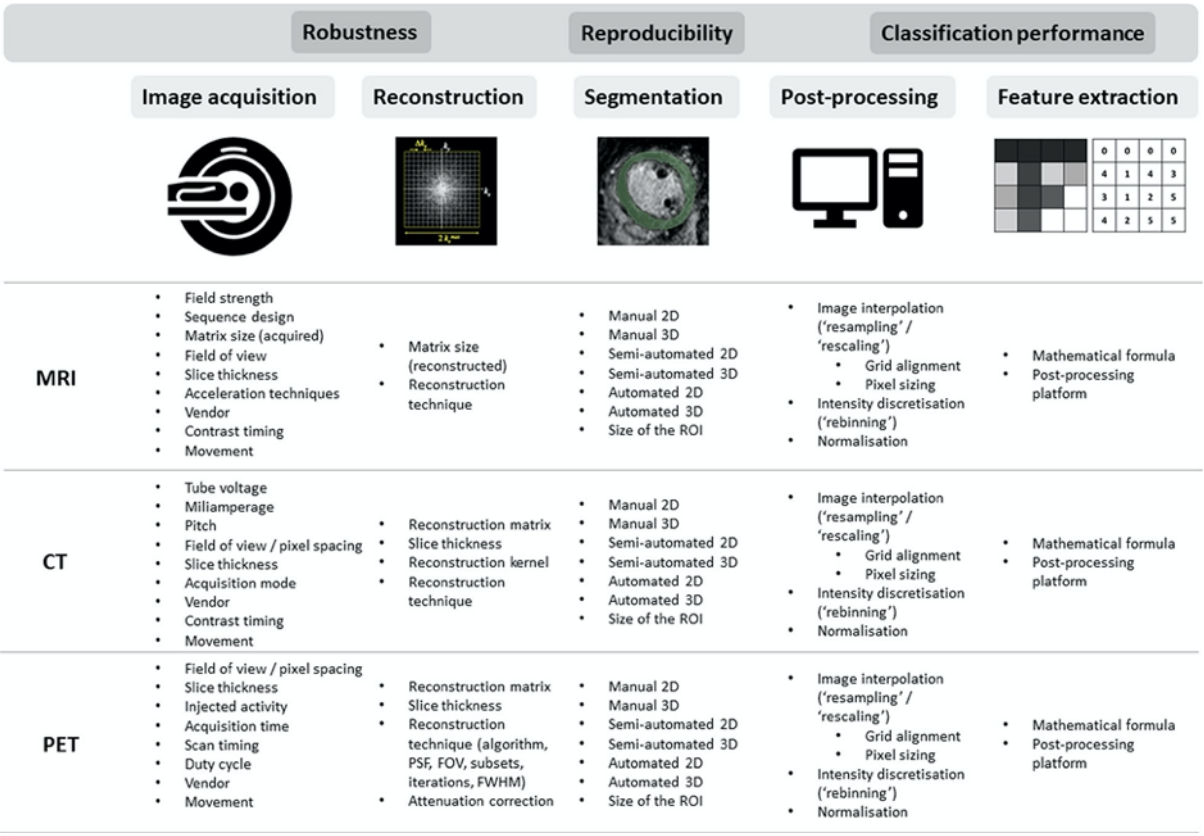


Figure 1.4.7: Factors influencing radiomics stability in each step of a radiomics workflow for MRI, CT and PET [50].

Pre-processing technique	**Effect on image**
Resampling	Changing the number of pixels in the image using interpolation (linear, polynomial, spline, etc.).
Normalisation or intensity standardisation	Changing the range of pixel intensity values, in order to remove bias, scaling factors and outliers from the image.
Quantisation of gray levels	Reduction of gray levels used to represent the image.
Motion correction	Reduction of motion confounds.
Filtering to remove noise and/or improve image characteristics	Laplacian: bringing out area of rapid intensity change and usually used for edge detection. Gaussian: smoothing the image and reducing noise. Wavelet filtering or transform methods: decomposing the original image and offering some advantages, such as variation of the spatial resolution (to represent textures at the most appropriate scale), enhancement of the texture appearance and a very wide range of choices for the wavelet function that can be adjusted for specific applications.

Table 1.4.2: Examples of pre-processing techniques that could be applied before quantitative image analysis.

Segmentation

Segmentation entails delineating the Regions of Interest (ROIs) in 2D scans or Volume of Interest (VOI) in 3D scans. These areas are such as those which contain the tumour, microscopic disease as well as the Organs at Risk (OARs). The areas delineated are determined by the region in which the tumour exists as well as surrounding organs.
Segmentation can be manual (carried out by an oncologist or a radiologist), automatic, or semi-automatic (incorporating both a human observer and an automated method), and so the choice of segmentation is based upon the method that gives the best accuracy. Manual delineation, which is the current gold standard, has the disadvantage of introducing inter-observer variability [99] and is also time consuming. An automated method of segmentation is less time consuming and offers consistent results, especially where many scans are involved [50] and has been found to increase reproducibility and robustness of the radiomic features [18]. However, one study found that automated segmentation produced false positives [100]. Overall, studies agree that a semi-automated method would be preferable as it combines human intervention, which reduces false positives, while maintaining the benefits of automatic segmentation [18], [99], [101].
Proper delineation of the tumour area improves radiotherapy delivery [102], [103] as it improves lesion targeting, which consequently leads to lower recurrence rates [103], [104]. Proper segmentation also minimises radiation toxicities to Organs at Risk [105], such as bowel toxicity during Cervical Cancer. In radiomic studies, segmentation provides the mask with which radiomics feature extraction can be applied.

Image Processing

Image processing is done for various reasons such as to enhance image quality. MR images, for instance, contain Gaussian noise [106], so they require denoising [47]. They may also have non-uniform intensities that need to be corrected [107]. Some image processing techniques that can be applied include: Image normalisation, resampling, image discretisation and applying filtering on the original images [50], [61]. Improving uniformity of image features improves repeatability of radiomics features across various patients [108].
The type of image processing applied depends on the data being used in the study and the kind of study being carried out. For instance, in this study where multiparametric MR images will be used, the image processing that will be employed includes, image normalisation, resampling, applying filtering such as Laplacian of Gaussian and Wavelet filtering, and image discretisation with a fixed bin count. Filtering or performing mathematical transforms on the original images prior to radiomic feature extraction, allows for the identification of patterns and details in

the image that may have been imperceivable [53], [94]. In PyRadiomics, filtering specifications are specified in the parameters file's settings section after which the original images undergo filtering or transformation, after which first-order, shape and second-order radiomic features are extracted from the resulting images. All image processing parameters to be applied on a data-set are defined in the PyRadiomics documentation [61].

Since various image processing techniques exist with each offering different outcomes, outlining the processing done in a study is crucial to ensure results can be verified and reproduced on a different data-set.

Table 1.4.3 shows various image processing techniques done in Cervical Cancer radiomics studies, the extraction platform used and the study's endpoint.

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint
MR	[74]	CC	380	Unspecified	unspecified	Comparison of radiomics models built through ML in a multicentric context
MR	[82]	CC	62	resampling	PyRadiomics	Investigating repeatability and reproducibility of MRI-based radiomic features in cervical cancer
MR	[83]	CC	111	Normalisation	PyRadiomics	Predicting LVSI in CC
MR	[84]	CC	248	z-score normalisation	PyRadiomics	Relating MRI-derived radiomic biomarker with disease free survival in early stage CC patients
MR	[85]	CC	160	in-house based on MatLab 2016a		Radiomics analysis of ADC in CC

Continued on next page

Table 1.4.3 – continued from previous page

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint
MR	[86]	CC	81	Normalisation	PyRadiomics	Investigating the sensitivity of radiomic features to inter-observer variability and image processing in ADC maps of CC patients
MR	[87]	CC	87	intensity rescaling	LIFEx	identifying the most appropriate VOI setting in prognostic prediction.
MR	[81]	CC	189	z-score normalisation	PyRadiomics	utilising radiomics analysis of MRI to improve diagnostic performance of LNM in cervical cancer patients.

Continued on next page

Table 1.4.3 – continued from previous page

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint
MR	[89]	CC	169	Unspecified	PyRadiomics	preoperative prediction of lymph node metastasis in early-stage cervical cancer.
MR	[109]	CC	105	Normalisation	PyRadiomics	MR-based radiomics nomogram of CC in predicting LVSI pre-operatively
MR	[72]	CC	233	Normalisation	PyRadiomics	mpMRI-based radiomics nomogram for the prediction of LNM in early-stage CC
MR	[110]	CC	479	Rectangular bounding boxes	unspecified	Deep Learning model to identify LNM on MRI in patients with CC
MR	[111]	CC	120	Normalistion	in-house software based on MatLab 2017a	Radiomics to predict treatment response in locally advanced CC

Continued on next page

Table 1.4.3 – continued from previous page

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint
MR	[112]	CC	66	unspecified	unspecified	Pre-treatment minimum ADC value as a prognostic imaging biomarker in CC patients treated with chemoradiation
MR	[113]	CC	160	unspecified	in-house software based on MATLAB 2016a	Potential use of (DCE) MRI parameters as radiomic features of CC
CT	[114]	CC	226	Unspecified	PyRadiomics	Predicting lymph node status in operable cervical cancer
PET	[115]	NSCLC, CC	100	Unspecified	Matlab vR2016a	Predicting distant failure after radiotherapy in NSCLC and CC

Continued on next page

Table 1.4.3 – continued from previous page

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint
PET	[116]	CC	80	Rescaling radiomic features, Correction for volume, discretisation	in-house software	Investigating multi-radiomic models for enhancing prediction power of CC treatment outcomes
PET-CT	[117]	CC	97	z-score normalisation	Unspecified	integrating primary tumor and peritumoral areas to predict E-cadherin expression and correlate with pelvic lymph node metastasis in early-stage cervical cancer.

Continued on next page

Table 1.4.3 – continued from previous page

Modality	Reference	Site	Patients	Image Processing done	Extraction platform	Study Endpoint

Table 1.4.3: Radiomics studies in cervical cancer with the imaging modalities used, number of patients in the studies, image processing techniques employed, extraction platform used and the study's endpoint.
Cervical Cancer (CC)
Head-Neck cancer (HN)
Magnetic Resonance Imaging (MRI)
Multiparametric-Magnetic Resonance Imaging (mpMR)
Lymph Node Metastasis (LNM)
Lymph Vascular Space Invasion (LVSI)
Apparent Diffusion Coefficient (ADC)
Dynamic Contrast-Enhanced (DCE)
Volume of Interest (VOI)
Oropharyngeal squamous cell carcinoma (OPSCC)
Small Cell Lung Cancer (SCLC)
Non-small Cell Lung Cancer (NSCLC)
Head-Neck Squamous Cell Carcinoma (HNSCC)
Computer Controlled Radiation Therapy (CCRT)
Machine Learning (ML)

There are various image processing techniques. Each of these techniques influences the outcomes of a radiomics study and can be dependent on the type of imaging modality used and the intended goal [47]. For this reason, proper reporting of all techniques applied in the study allows for validation of results and reproducibility of the results on a different data-set.

Some image processing techniques applied on MR images include normalisation prior to features extraction, which was seen to increase the reproducibility of Apparent diffusion coefficient (ADC)-based radiomics features [86]. Resampling is another technique that aims at maintaining a consistent isotropic voxel spacing across different measurements and devices which gives good reproducibility [82], [118], [119]. Grey-level discretisation is another image processing technique dependent on the imaging modality used. The main methods are using a Fixed Bin Number (FBN) or a Fixed Bin Size (FBS). Choice of grey-level discretisation method depends on the imaging modality and the intended purpose for image intensities discretisation. Since grey-level discretisation impacts image intensities within the ROIs, radiomic feature values and the robustness of features extracted are also impacted [120]. FBS allows for comparison to be made for discretised data with different calibrated ranges, such as PET and CT data, since the bins belonging to the overlapping range will represent the same data interval images [47]. It can also be referred to as the absolute discretisation method [121]. FBN normalises images and is especially beneficial in data with arbitrary intensity units such as MRI data [47], [50]. It can also be referred to as the relative discretisation method [121]. Table 1.4.4, adapted from IBSI documentation [47], is a table with recommendations of when to use FBN or FBS. Since we will be using MR images, whose imaging intensity units are arbitrary, Fixed Bin Number (FBN) will be used in this study.

Imaging Intensity Units	Re-Segmentation range	FBN	FBS
calibrated	$[a, b]$	✓	✓
	$[a, \infty]$	✓	✓
	none	✓	X
arbitrary	none	✓	X

Table 1.4.4: PET and CT images contain calibrated imaging intensity units, MR images contain arbitrary imaging intensity units[1]*. While FBN uses the actual range of intensities within the identified ROI* [2]*, FBS uses the lower bound of the resegmentation range as the minimum set value* [3] *[47], [122].*
Checkmarks (✓) - recommended combinations of resegmentation range and discretisation algorithm.
Crossmarks (X) - non-recommended combinations.

Feature Extraction

This is the stage at which radiomic features are extracted. The features are termed "quantitative imaging biomarkers" [60]. Some features can be derived from others, and since there are various definitions for feature calculations, the Image Biomarker Standardisation Initiative (IBSI) guidelines were set up to standardise feature calculations [14], [47], [123]. Of great importance to note is that the reliability and value of radiomic features is highly dependent on the choice of the radiomics package and the package version [66], so following standardisation guidelines such as those set by the IBSI [47] significantly strengthens any radiomics study. This is because different radiomics packages and different versions of the same radiomics package may define radiomic features differently as was shown by a study carried out by Fornacon-Wood et al (2020) [66].

At this stage of a radiomics workflow, ensuring the features extraction software used adheres to IBSI guidelines and is readily available to other researchers to allow for verification of results is considered best practice [66].

Extraction Platform

There are various platforms for radiomic features extractions, including PyRadiomics [79], CERR [62], TexRAD [63], LIFEx [64] and IBEX [65]. It is considered best practice to specify the software used and the version of the software and to also ensure the specific software version adheres to IBSI guidelines as suggested in a study by Fornacon-Wood et al (2020) [66]. This is because computation of radiomic features has been defined in various ways on various platforms. Features may have different names and/or computation across various feature extraction platforms. Open-source tool-kits are preferred as they allow for easy verification of results [66], as opposed to in-house and commercial software, especially those not made readily available. Table 1.4.3 shows the main tool-kits used in the various studies. In this project, PyRadiomics, which is open-source and adheres to IBSI guidelines, will be used to extract radiomic features.

Feature Selection and Analysis

Extracted radiomic features are not all informative and feature usefulness can depend on the study being carried out. These features are usually many which carries the risk of model over-fitting. The aim of this multi-step recursive process, therefore, is to exclude features that are non-reproducible, redundant (they depend on other extracted features [124]) and non-relevant to the study from the data-set and also to analyse retained features in relation to intended outcomes. Features retained are those that are considered most informative based on their independence from others, their robustness and prominence on the data [14].

The choice of statistical methods employed depends on the study being carried out,

but since there are various statistical methods for feature selection and analysis, proper reporting is also important to allow for reproducibility of results by other researchers. Validation of the methods on external data is important for verification of results [15], [125], [126].

In this study, we will be looking at feature selection. One way of reducing the number of features is to drop features that are confounded, especially to volume. This is because volume is a known prognostic indicator (large tumours generally mean more advanced cancer with poorer outcomes). Features that correlate with volume would, therefore, be redundant in a study [82]. Another way is to find correlation in the remaining features and dropping features that give similar information. Test-retest repeatability analysis has also been used to select stable features. This is because test-retest analysis relies on the assumption that images acquired within a short duration of each other should not have varying radiomic feature values [82], [108], [125]. In radiomics, *repeatability* refers to features that remain the same when imaged multiple times on the same subject [108], while *reproducibility* refers to features that remain the same in one subject at different visits or features that remain the same when imaged using different equipment, different software, different image acquisition settings, or different operators such as in different clinics, whether in the same subject or in different subjects [108], [127]. Repeatability is a term used when the radiomics features extracted are considered to be "stable" or "robust" [82], [121], [125], [128], [129]. Radiomics features-based models combine repeatable features with clinical parameters and information such as patient demographics to predict outcomes [90].

Other feature selection methods that could be used include using multiple observers in a radiomic study and looking for agreement of features between observers could be used to select features that are robust to inter-observer variability [130]. The retained features are then used in models [131] to, say, investigate associations to outcomes.

In this study, radiomic features retained will be those that are not highly correlated to volume and remain robust within a test-retest setting.

In conclusion, there are various factors that could affect a radiomic features' repeatability. These factors range from the image acquisition modalities and settings, image reconstruction algorithms, image processing carried out, and the platform used to extract radiomic features [108]. Table 1.4.5 shows a summary of these factors as well as ways to mitigate the factors. Each of these factors could potentially give different results. It is therefore paramount to employ proper reporting of every aspect of the study as insufficient reporting of radiomics studies greatly impedes study reproducibility and utilisation of results in clinical settings [132]. For this reason, studies have agreed that systematic documentation of methods and processes is paramount to allow for replication of results.

There are some guidelines that have been established to improve the quality of radiomics studies, such as the Image Biomarker Standardisation Initiative (IBSI) guidelines, the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines and the Radiomics Quality Score (RQS). All these guidelines are continously evolving and aim to improve the quality of radiomic studies. The IBSI guidelines govern the technical aspects such as the implementation of a radiomics study [47], [50]. The TRIPOD guidelines cover the standards for reporting of models. The RQS divides a study into 5 phases and seeks to find the quality of the study by assigning a score based on the methodology and quality of the study [122]. Figure 1.4.8 is a summary of an RQS workflow. In

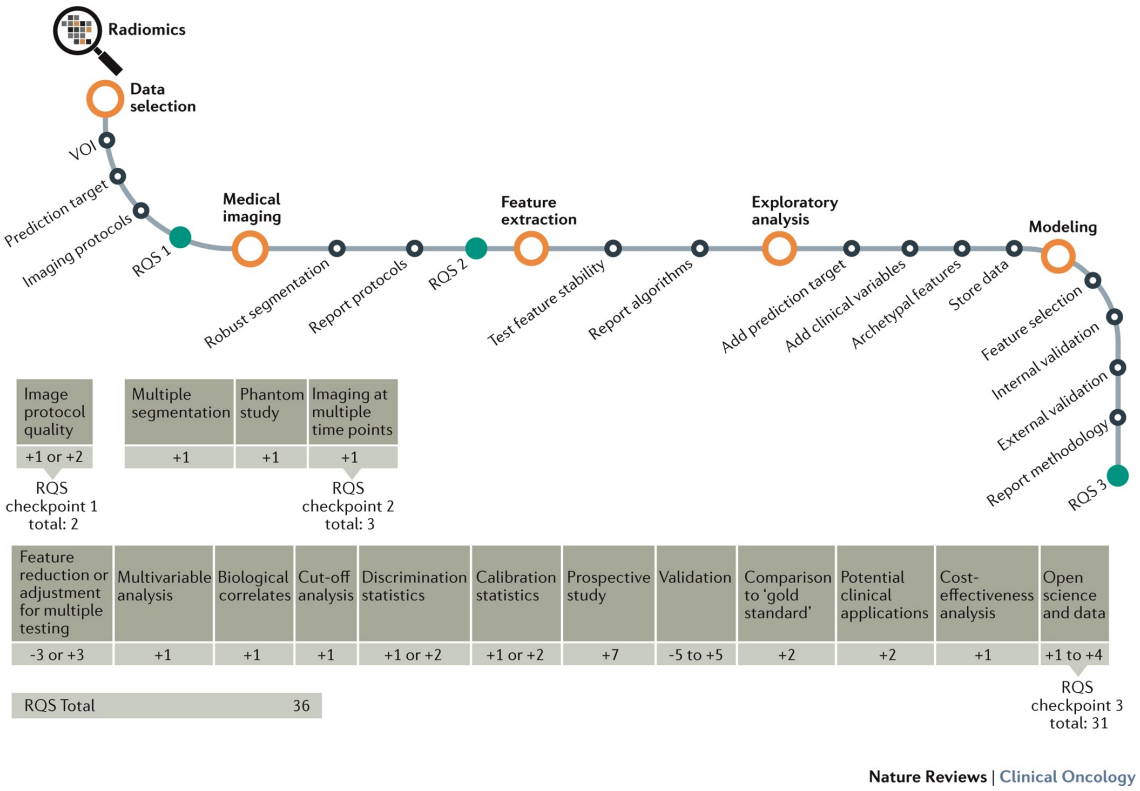


Figure 1.4.8: An RQS workflow [122]

summary, there are various factors that could affect a radiomics study [14], [18]. These factors manifest in all steps of the radiomics workflow [50] as outlined in Figure 1.4.7. Table 1.4.5, adapted from a study by Fornacon-Wood et al (2020) [93] summarises the potential problems and solutions.

Validation of Results and Modelling

The final step of the radiomics workflow is the validation of results. This is where methods used in the study are employed on a different data-set to test for similar outcomes. Validation of results is then followed by model building. As with other

Problem area	Potential problems	Potential solutions
Image acquisition	Different scanners and acquisitions protocols affect feature reproducibility	Image phantoms on different scanners to provide baseline, establish credibility of scanners and protocols, catalogue reproducible features, model a correction algorithm, harmonise data
	Patient motion affects feature reproducibility	Set motion tolerances, reduce ROI boundaries, use single phase from 4D images, find robust features using 4DCT data
Image acquisition and reconstruction	Image resolution parameters (voxel size, slice thickness) affect feature values model performance	Control resolution parameters in prospective studies, resample to common resolution and voxel depth, apply smoothing image filters, apply deep learning methods
Image reconstruction	Image reconstruction algorithm and reconstruction parameters (kernel) affects features	Pre-processing image correction and harmonisation of acquisition techniques
Segmentation	Delineation variability affects features and is time consuming. Results from one disease are not necessarily transferable to another	Expert ROI definition, multiple observers, identification of stable features with respect to delineation, automated segmentation, image filtering
Pre-processing	Number of grey levels used to discretise histogram and texture features affects feature values, as does width	Texture features can be normalised to reduce dependency on the number of grey levels, number of grey levels used for discretisation should be recorded with feature formula. 128 grey levels may be optimal for texture features along with thresholding
Feature correlation	Strong correlations between tumour volume and radiomic features exist	Normalisation of features to volume, bit depth resampling, feature redesign, more robust statistics to check added value of radiomics signatures
Test re-test	Radiomic features may not be repeatable over multiple measurements, repeatable features are not generalisable to other disease sites	Test-retest data acquisition, use of multiple 4D phases, use of simulated retest by image perturbation
Modelling clinical outcome	Different modelling strategies affect model performance	Sample sizes above 50 give better predictive performance, as does normalising features. No consensus on best modelling strategies to use

Table 1.4.5: Table showing a summary of some factors affecting radiomics studies and potential solutions to the factors adapted from a study by Fornacon-Wood et al (2020) [93]

steps in the radiomics workflow, model building follows guidelines that have been devised on modelling and reporting in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) [133]. TRIPOD governs the proper reporting of any study that entails the development or validation of a prediction model for the purpose of diagnosis or prognosis [133].

1.5 Summary and Aims

The overall aim of this study is to identify multiparametric-MR-based radiomic features that are robust to test-retest in cervical cancer. These features can then be utilised in larger data-sets to build reliable predictive models. This will be achieved through the following sub-aims:

- Setting up a radiomics pipeline using PyRadiomics to extract radiomic features from T2-weighted and Dixon images for Cervical Cancer patients.
- Selecting stable radiomic features based on high Intraclass Correlation Coefficient value between features in a test-retest setting.
- Identifying a stable radiomic feature from each feature class.

Chapter 2

Test-Retest Repeatability of Multiparametric MR-based Radiomic Features in Cervical Cancer - GTV and peritumoral regions.

Authors

Chelmis M. Thiong'o[1], Alan McWilliam[1], Gareth Price[1], Andrew Green[1].

Affiliations

1. Division of Cancer Sciences, University of Manchester, Manchester, UK

This work is unpublished but will be adapted for publication.

Abstract

Purpose: The aim of this study is to evaluate the test-retest repeatability of radiomic features from the GTV and peritumoral regions of T2-weighted and Dixon water-only images in cervical cancer patients.

Methods: 21 scans, each comprising of 2 T2-weighted and 2 Dixon sequence images, were included in this study. This resulted in a total of 42 test-retest pairs of images. The GTV was contoured by a single observer on each of the T2W images and transferred through rigid registration to the Dixon sequence images. For each image, radiomic features were extracted from 2 regions: the GTV and a peritumoral region extending 2 voxels (1.875mm for T2-weighted images and 1.923mm for Dixon images) from the GTV contour but restricted to remain inside the cervix/uterus so as not to sample surrounding tissues.

PyRadiomics was used to extract first-order, shape, and texture radiomic features from the original T2W and Dixon images, and their Laplacian of Gaussian-filtered (σ = 3, 4, and 5) and Wavelet-filtered counterparts. Spearman's Rank Correlation Coefficient, $|\rho|$, was used to exclude features highly correlated with volume ($|\rho| \geq 0.9$). Of the remaining features, those with minimal variation in a test-retest pair of images were selected using one-way random effects Intraclass Correlation Coefficient (ICC), to assess absolute agreement of feature values. The criteria ICC(1,1) ≥ 0.9 was used. Remaining features were grouped by feature class and the feature with the lowest Spearman's correlation to volume within the class was considered stable.

Results: A total of 1130 features were extracted from each image. Features with Spearman's $|\rho| \geq 0.9$ in relation to volume were discarded. For the T2W images, 86% and 87.1% of the 1130 features extracted from the GTV and peritumoral regions respectively remained. For the Dixon images, 86.1% and 87.8% of the 1130 features extracted from the GTV and peritumoral regions respectively remained. Features with ICC(1,1) ≥ 0.9 were considered to have excellent test-retest repeatability and were retained. For the T2W images, these were 24.9% and 7.6% of the 1130 radiomic features extracted from the GTV and peritumoral region, respectively. For the Dixon images, they were 4.6% and 4.0% of the 1130 radiomic features extracted from the GTV and peritumoral region, respectively.

Conclusions: Stable features were identified for each feature class and can be used in larger data-sets to build reliable predictive models.To confirm generalisability of our findings, validation of our results on a similar data-set is required.

Keywords: Radiomics, MRI, Cervical Cancer, Test-retest, Repeatability, T2-Weighted, Dixon.

2.1 Introduction

Radiomics - "the high-throughput extraction and analysis of large amounts of image features from radiographic images" - promises to identify image characteristics that may predict treatment outcomes [14]. However, identification of robust radiomic features is first required before model building. The main challenge as identified by existing studies is the need for optimal analysis parameters and consistent feature reporting [82], [125]. For instance, pre-processing parameters have been shown to affect radiomic texture features [134] which presents a challenge especially since different image modalities require different quantisation methods, each producing varying results [47], [121]. The Image Biomarker Standardisation Initiative (IBSI) guidelines have been set to standardise radiomic feature definition and calculation [67], which significantly strengthens a radiomics study.

Studies investigating repeatability (comparison on the same subject under constant conditions) and reproducibility (comparison under varying conditions) of radiomic features have been used to improve model generalisability [82], [108], [125]. Test-retest analysis of radiomic features is one way of identifying feature repeatability [82], [125] and it compares scans acquired on the same patient within a short duration of time. Features that remain stable in the test-retest setting are considered robust and suitable for use in predictive models and generally result in better model generalisability. Studies have confirmed that features from scans acquired within minutes of each other are not all stable. Stable features are those that are considered most informative for the study based on their independence from other features, repeatability or prominence in the data. For instance, a study looking at test-retest repeatability by Fiset et al (2019) evaluated the stability of radiomic features from MR images of cervical cancer [82]. They found that 47.9% of the extracted features for their test-retest cohort were not stable.

While most radiomic studies look at CT data, there has been an increased appreciation of MRI-based radiomic studies due to the excellent soft-tissue contrast offered by MR imaging. For cervical cancer, MR imaging results in high accuracy and detailed evaluation of pelvic tissues and organs as well as good contrast resolution ([17]–[21]) with no radiation [22], and affords versatility such as the option for multiparametric information from both morphological and functional signals [23].

Most radiomic studies look at the Gross Tumour volume (GTV) region. However, evidence has shown microscopic disease near the GTV which is why the Clinical Target Volume (CTV) includes areas near the GTV that are suspected to have presence of microscopic disease [40]–[42]. Radiomic studies done on the peritumoral region in cervical cancer look at in-field recurrence [87], prediction of nodal metastasis and distant metastasis [115], by analysing the region near the GTV for microscopic disease [81], [89], [117].

In this work, we perform test-retest analysis to investigate repeatability of radiomic

features in cervical cancer from T2-weighted (T2W) and Dixon MR images for the GTV and peritumoral regions. This is the first cervical cancer study looking at test-retest repeatability in the peritumoral region as well as test-retest analysis using MR data that includes Dixon images. Overall, we aim to guide the selection of stable radiomic features for use in predictive models on larger data-sets.

2.2 Materials and Methods

2.2.1 Patient Demographics

Data from a cohort of 7 Cervical Cancer (CC) patients treated with concurrent chemoradiotherapy, between September 2017 and September 2018 was available. The data was collected and analysed with approval (Ethics: Greater Manchester South Research Ethics Committee: 17/NW/0300). 2 patients had FIGO stage IB2 CC and 5 patients had FIGO stage IIB CC.
A summary of patient demographics is outlined in Table 2.2.1.

2.2.2 Image Acquisition

All images were acquired on the same 1.5T MR system (MAGNETOM Aera; Siemens Healthcare, Erlangen, Germany). For this study the following sequences were used: 2D T2 weighted Turbo Spin Echo (TSE) (TR/TE = 5390/99 ms, in-plane resolution = 0.9375 x 0.9375 mm, slice thickness = 3mm) and a 3D DIXON gradient echo (TR/TE = 7.2/2.4 ms, in-plane resolution 0.9615 x 0.9615 mm, slice thickness = 3mm) water-only image.
Collectively, 21 scans were acquired, where each patient underwent 3 MR scans taken at least a week apart, with the first scan being acquired on the first day of radiotherapy treatment. At each MR scan, a T2-weighted image was taken at the start and at the end of the visit resulting in a test-retest pair. Dixon sequences were sequentially acquired over 10 minutes. The first and last Dixon images in the sequence were collected as a test-retest pair. Overall, 6 test-retest pairs were acquired for each patient, resulting in a total of 42 test-retest pairs of images. Scanner and imaging parameters are summarised in Table 2.2.1.
2 regions were identified for this study: the Gross Tumour volume (GTV) and the peritumoral region. Contouring of the GTV and Organs at Risk (OARs) was carried out using Raystation v6 (Raysearch AB, Stockholm, Sweden) on all the T2-weighted sequence slices. As in Fiset et al (2019), rigid registration was performed such that GTV contours on the T2-Weighted sequence images were propagated onto the Dixon sequence images with no modification. This was to minimise intra-observer contour-

Characteristics		Value
Number of patients		7
Age (*yrs*)		
	Mean ± SD	50.1±13.5
	Median[Min, Max]	48 [36, 75]
Height (*m*)		
	Mean ± SD	1.61±0.05
	Median[Min, Max]	1.6 [1.54, 1.68]
Weight (*kg*)		
	Mean ± SD	72.6±17.1
	Median[Min, Max]	64 [57, 106]
Comorbidity score		
	None	5
	Mild	1
	Moderate	1
Tumour characteristics		
	FIGO Stage (*n*)	
	IB2	2
	IIB	5
	Volume, mean ± SD (cm^3)	6.1±1.5

Image Parameters		Imaging sequence T2-Weighted	Dixon
Number of images		42	42
MR Scanner (*n*)			
	Siemens MAGNETOM Aera	42	42
Magnetic Field (*n*)			
	1.5 T	42	42
Sequence, median (*range*)			
	Slice Thickness (*mm*)	3	3
	TE (*ms*)	99	2.4
	TR (*ms*)	5390	7.2
	Scanning sequence	SE	GR
	Sequence variant	SK, SP, OSP	SP, OSP
Pixel Spacing (*mm*)		0.9375 x 0.9375	0.9615 x 0.9615

Table 2.2.1: A summary of the patient demographics for the cohort and Imaging parameters for the MR scanner used in this work.
SD - standard deviation
TE - Echo time
TR - Repetition time
SE - Spin echo
GR - Gradient echo
SK - Segmented k-space
SP - Spoiled
OSP - Oversampling phase
The same patients were used in the T2-weighted and Dixon MR sequences in the study.

ing variation between T2-weighted and Dixon MR images [82].

The GTV contour of each image was dilated by a value of 2 voxels (1.875mm for T2-weighted images and 1.923mm for Dixon images) to form the peritumoral region's contour using SciPy's multidimensional binary dilation module. This was done for

all T2-Weighted and Dixon images. This contour was restricted to remain inside the cervix/uterus (within the Intermediate Risk Clinical Target Volume (CTV-IR)) so as not to sample surrounding tissues [135]. The peritumoral region was then defined as the region between the GTV and peritumoral contours as shown in Figures 2.2.1 and 2.2.2. Wu et al (2021) explored various tumor sites to determine the predictive value of additional peritumoral regions based on deep learning and radiomics [135]. They found the size of the peritumor to be critical in giving additional predictive value in three tumor datasets (gastrointestinal stromal tumors, laryngeal carcinoma and nasopharyngeal carcinoma), and only the performance of 1.5 mm–4.5 mm peritumors was found to be stable.
Moving forward, 4 data-sets were identified:

- T2W images and their respective GTV contours
- T2W images and their respective peritumoral region contours
- Dixon images and their respective GTV contours
- Dixon images and their respective peritumoral region contours

Figure 2.2.1 shows T2W images while Figure 2.2.2 shows Dixon images acquired at the start and end of a single scan - forming 2 test-retest pairs - for 1 patient with FIGO Stage IIB cervical cancer; the GTV (region within the red contour) was delineated independently; and the peritumoral region (region between the red and green contours) was acquired through dilating the GTV by 2 voxels.

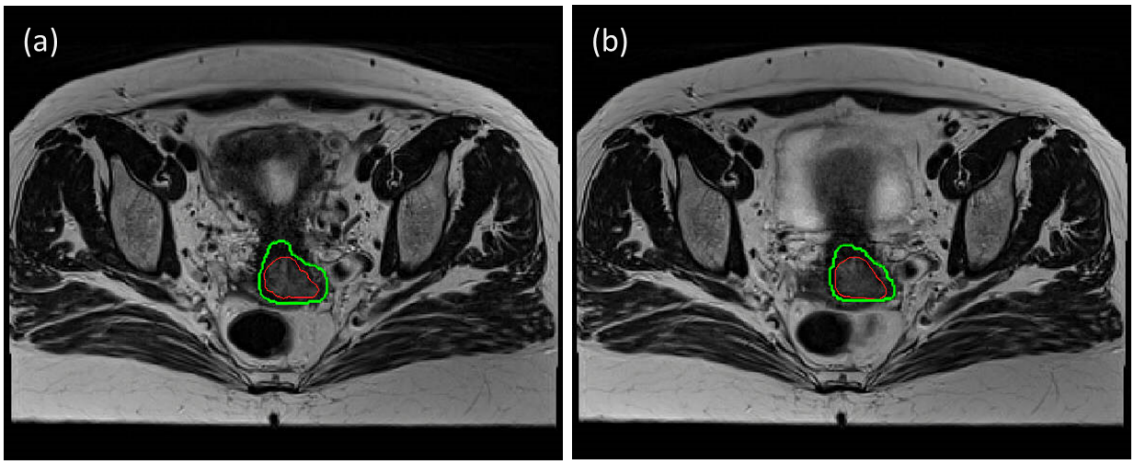


Figure 2.2.1: Axial T2-weighted images of a single patient with FIGO Stage IIB cervical cancer at the first visit. The GTV (in red), is contoured independently on each image. The peritumoral region (in green), is derived from the GTV by dilation of 2 voxels.
"a" T2-Weighted MR image at the start of the visit; "b" T2-Weighted MR image at the end of the visit.

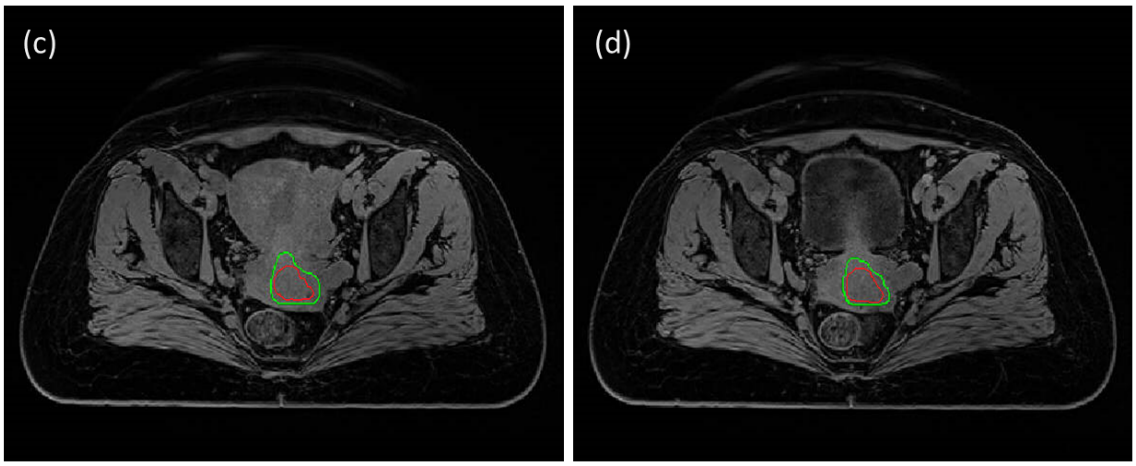


Figure 2.2.2: Axial Dixon images of a single patient with FIGO Stage IIB cervical cancer at the first visit. The GTV (in red), is propagated from the corresponding T2W image acquired within the same visit as in Figure 2.2.1 above. The peritumoral region (in green), is propagated from the corresponding T2W image's peritumoral region in Figure 2.2.1.
"c" Dixon MR image at the start of the visit; "d" Dixon MR image at the end of the visit.

2.2.3 Image Processing

All images and their respective contours underwent image processing as defined in the PyRadiomics package v.3.0.1 before feature extraction [50], [61] as follows:
All images and contour files were resampled to the same in-plane resolution of $0.9mm \times 0.9mm \times 3.0mm$ using B-Spline interpolation which has been seen to retain soft-tissue contrast and give good reproducibility [82].
Whole image intensity normalisation was also carried out since MR images have non-uniform intensities that need to be corrected to allow comparison between images [107]. This was by centering at the mean and dividing by standard deviation of the gray values in the image as per RyRadiomics standards. Improving uniformity of image intensity improves repeatability of radiomics features across various patients [108].
Since MR images have arbitrary intensity units, IBSI recommends performing the Fixed Bin Number (FBN) image discretisation method. To get the FBN value, first-order range feature values were divided by bin width (Fixed Bin Size (FBS)). IBSI recommends a bin width value of approximately 5. In this study, a bin number of 64 was used which corresponds to a bin size of 5 for an average first-order range value of 320.

2.2.4 Feature Extraction

PyRadiomics version 3.0.1, an open-source software that has been validated against the Image Biomarker Standardisation Initiative (IBSI) standards [61], was used to extract radiomic features. All the default radiomic features defined in the PyRadiomics documentation [50], [61] were extracted from the original image, its Laplacian of Gaussian (LoG)-filtered (with σ = 3, 4, and 5) and Wavelet-filtered (8 decompositions) counterparts. LoG filtering was performed to enhance image edges with the sigma values determining the coarseness of the texture features [61]. Sigma values were selected from literature to give a range of coarseness [82], [125]. A sigma of 3, relative to the other sigma values, emphasises on fine textures (change over a short distance), while 5, relative to the other sigma values emphasises coarse textures (gray level change over a large distance) [61]. Wavelet-filtering was done to denoise the images by applying all combinations of high- and low-pass filters on each image dimension [125].

Radiomic features extracted on the original images included 19 first order features, 16 3D shape features and texture features including: 24 Grey Level Co-occurence Matrix (GLCM) features, 16 Grey Level Run Length Matrix (GLRLM) features, 16 Grey Level Size Zone Matrix (GLSZM) features, 14 Grey Level Dependence Matrix (GLDM) features and 5 Neighbouring Grey Tone Difference Matrix (NGTDM) features. On each of the filtered images (LoG-filtered and Wavelet-filtered), 19 first order features, and texture features: 24 Grey Level Co-occurence Matrix (GLCM) features, 16 Grey Level Run Length Matrix (GLRLM) features, 16 Grey Level Size Zone Matrix (GLSZM) features, 14 Grey Level Dependence Matrix (GLDM) features and 5 Neighbouring Grey Tone Difference Matrix (NGTDM) features, were extracted.

Image processing and feature extraction as described above was carried out on all 4 datasets. The script used to extract radiomics features using Python version 3.6.9 is available in the Appendix, Section A.1. The parameters script describing the settings applied on the images prior as well as the radiomic features extracted and the image types is also available in the Appendix, Section A.1.

2.2.5 Feature Selection

Volume is a known prognostic indicator, so radiomic features that are highly correlated to volume do not add meaningful information to a radiomics study [82]. Volume confounded features were dropped as a first step in feature selection. Spearman's Rank Correlation Coefficient, ρ, was used to investigate the relationship between tumour volume and all the extracted radiomic features. Only features meeting the condition $|\rho| < 0.9$ were retained [82], [136].

Stability of the remaining features was evaluated using a one-way random effects Intraclass Correlation Coefficient, to assess absolute agreement of feature values (ICC(1,1)), with a confidence interval of 95%. ICC(1,1), recommended by Koo et al (2016) [137], [138], was used to select features with minimal variation in a test-retest pair of images. Features with ICC $\geq$ 0.9 were considered to have excellent repeatability and were retained.
Remaining features were grouped by feature class. Spearman's Rank Correlation Coefficient, ρ was then used to select a feature within a feature class that was least correlated to volume. Feature selection was carried out with Python version 3.6.9 and R package version 1.4.1717 (2021) on all 4 data-sets. The scripts used to select features are available in the Appendix, Section A.1.

2.3 Results

A total of 1130 features were extracted on each image. Figure 2.3.1 shows barplots of the percentage of features remaining at each feature selection step for each radiomic feature class. Figure 2.3.1(a) is a barplot of feature selection for T2W images on the GTV region while Figure 2.3.1(b) is a barplot of feature selection for Dixon images on the GTV region. Figure 2.3.1(c) is a barplot of feature selection for T2W images on the peritumoral region while Figure 2.3.1(d) is a barplot of feature selection for Dixon images on the peritumoral region.
Features not correlated with volume were retained. For the T2W images, 972 and 984 of the 1130 radiomic features extracted from the GTV and peritumoural regions respectively remained; For the Dixon images, 974 and 992 of the 1130 radiomic features extracted from the GTV and peritumoural regions respectively remained. Figure 2.3.1 shows in percentage, the number of features for each feature class that were correlated to volume (in orange) and those that reamined (in red and maroon). Tables showing the Spearman's rank correlation coefficient values for all 4 data-sets are available in the appendix A.1.
Radiomic features with minimal variation in a test-retest pair of images (ICC(1,1) $\geq$ 0.9) were retained. This was 281 and 86 features in the T2W datasets for the GTV and peritumoural region respectively (24.9% and 7.6% of the total number of radiomic features extracted from the GTV and peritumoral region respectively). In the Dixon data-sets, this was 52 and 45 features for the GTV and peritumoural region respectively (4.6% and 4.0% of the total number of radiomic features extracted from the GTV and peritumoral region, respectively). These features are indicated in maroon in Figure 2.3.1. Tables A.2.1 and A.2.2 show a breakdown of feature classes with excellent ICC values for both the GTV and peritumoral regions.
Table A.2.3 shows a summary of the number of features retained at each feature

selection step. Figures 2.3.2 and 2.3.3 shows Venn diagrams of remaining features. No feature within the GLCM and NGTDM feature classes had an excellent ICC for Dixon images. Shape features in all 4 data-sets exhibited the same behaviour in all feature selection steps. Finally, the percentage number of first-order and shape features that had excellent ICC was higher than that of texture features, except in the data-set looking at the GTV region of T2W images where the overall percentage of features with excellent ICC was higher than in other data-sets. Table A.2.4 in the Appendix shows the selected stable radiomic feature from each feature class in all 4 data-sets.

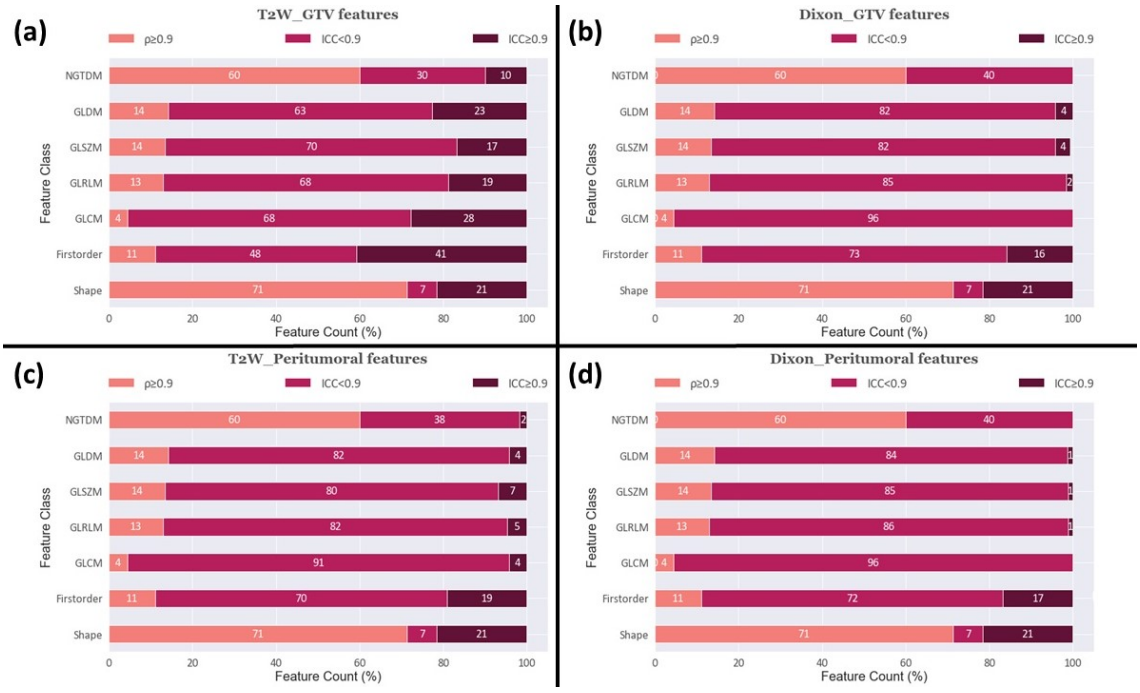


Figure 2.3.1: Radiomic features remaining after each feature selections step as a percentage of the total number of features extracted in each feature class, irregardless of image type, as follows: 14 shape features, 216 first-order features, 288 GLCM features, 192 GLRLM features, 192 GLSZM features, 168 GLDM features and 60 NGTDM features.
$|\rho| \geq 0.9$ - features dropped due to high correlation to volume.
$ICC < 0.9$ (in red) - features with poor Intraclass Correlation Coefficient were eliminated.
$ICC \geq 0.9$ (in dark red) - remaining features with excellent ICC were retained.

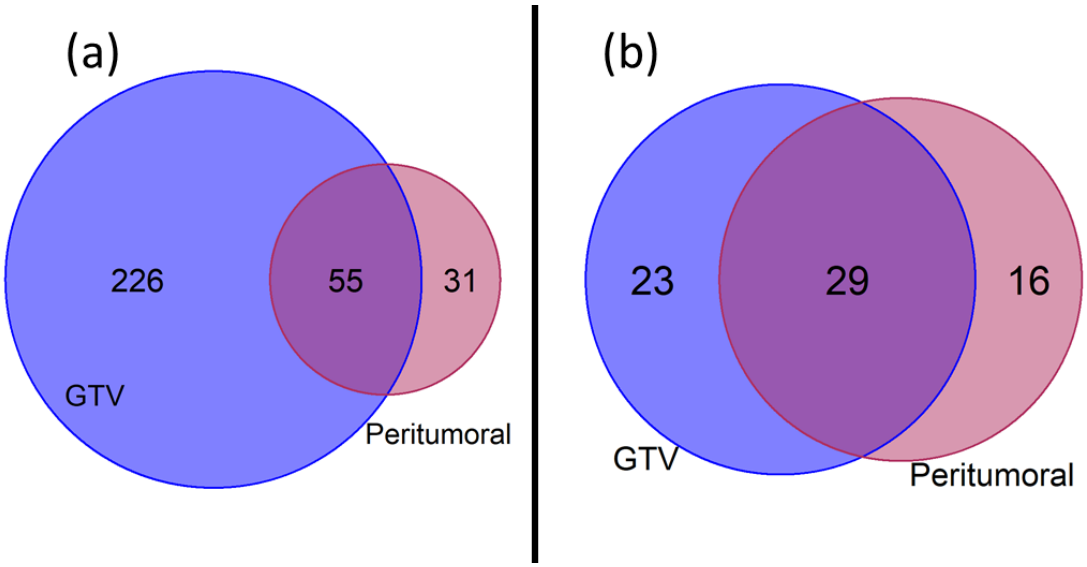


Figure 2.3.2: Venn diagrams showing relationships between radiomic features remaining for each MR sequence after all feature selection steps.
(a) Number of radiomic features remaining in the GTV and peritumoral regions from T2W images. (b) Number of radiomic features remaining in the GTV and peritumoral regions from Dixon images.

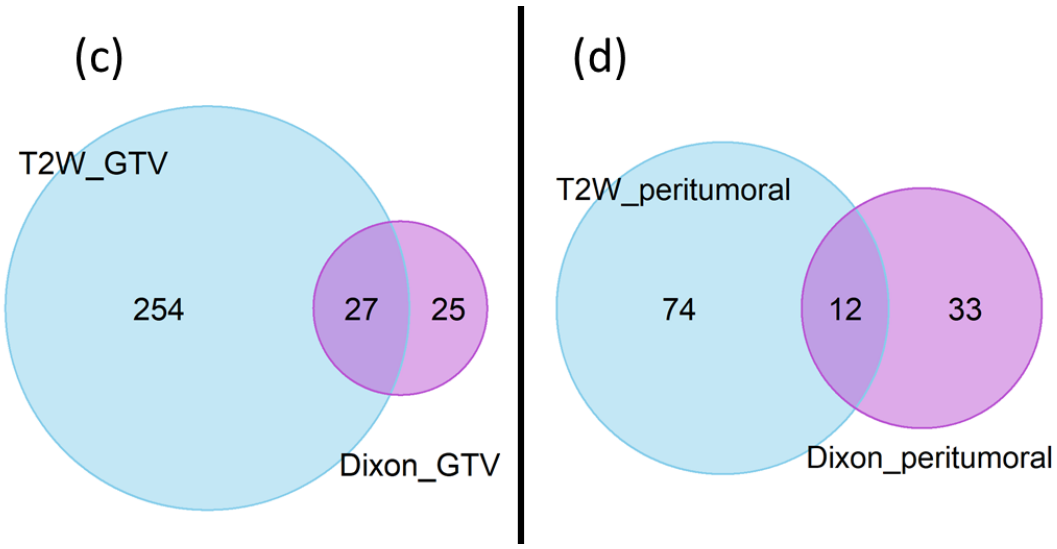


Figure 2.3.3: Venn diagrams showing relationships between radiomic features remaining in each location after all feature selection steps.
(c) Number of radiomic features remaining in the GTV region for both T2W and Dixon images. (d) Number of radiomic features remaining in the peritumoral region for both T2W and Dixon images.

2.4 Discussion

The aim of this thesis was to guide the selection of stable radiomic features in cervical cancer using test-retest repeatability analysis. These features would be used to build predictive models on a larger data-set. In this study we looked at multipara-

metric MR cervical cancer data (T2-Weighted and Dixon images) in two locations: the GTV and a peritumoral region. Our aim was met through setting up a radiomics pipeline that extracted radiomic features from the GTV and a peritumoral region for both T2-Weighted and Dixon water-only images. Stable radiomic features were then selected based on a 3-step selection process. The first step entailed selecting features not highly correlated with volume. Of these features, those with minimal variation in a test-retest pair of images were retained in the second step using the Intraclass Correlation Coefficient (ICC) and the criteria ICC(1,1)$\geq$0.9. The features that satisfied this criteria were then grouped based on their respective feature classes where the feature with the least correlation to volume in each feature class was selected in the third step as being stable and could be used to build reliable predictive models.

2.4.1 Novelty and Comparison to literature

Radiomic studies investigating test-retest repeatability of multiparametric MR data have mostly been carried out on prostate cancer [125], [139], [140] or using a phantom [141]. Fiset et al (2019) however carried a test-retest study in MR cervical data, but only one MR sequence was studied - axial T2-weighted turbo spin-echo (TSE) sequence [82]. So far no study has been carried out in test-retest analysis for cervical cancer using multiparametric-MR data even though most clinical settings combine MR sequences to reach a diagnosis [142]. Our study looked at 2 MR sequences, T2-Weighted and Dixon water-only, and highlighted the difference in radiomic feature repeatability for the different MR sequences in cervical cancer. We also showed radiomic feature stability also varied for different regions by looking at the GTV region and the peritumoral region.

In our study, we observed more T2-weighted image features as being stable compared to Dixon image features. This supports findings by Lecler et al (2019) [142] whose work showed that various MR sequences gave different radiomic feature stability results. In addition, features from the GTV region exhibited more stability than those from the peritumoral region, though the difference in this number for the 2 Dixon images' data-sets was small as shown in Figure A.2.3 in the Appendix section A.2.

Studies looking at test-retest radiomics have found shape and first-order features to be repeatable [82], [108], [125], [143]. Studies by Fiset et al (2019) and Schwier et al (2019), used the same assessment of feature repeatability that we used in our study. In line with their results, we found shape features showing consistent repeatability in all 4 data-sets. The number of shape features at each feature selection step was similar on all data-sets. This was expected since the contours were delineated by one observer, there was minimal or lack of geometric changes in the regions of interest

in our test-retest setting, and the individual T2W image contours were propagated onto the Dixon images acquired within a visit. For this reason, we hypothesise that shape features do not add valuable information to a test-retest repeatability study. This consistence in all shape features, however could be used to test intra-observer uncertainty. It could also be used to provide confidence that the clinician performed consistent segmentation between images in the test-retest setting.

Various clustering methods could be employed for selecting radiomic features that do not contain redundant information. We clustered our features based on their feature classes. Notably, for the 2 data-sets looking at Dixon images, we did not find any GLCM and NGTDM feature that satisfied our stability criteria (see Figure 2.3.1 and Table A.2.4). This could have been because we only allowed for features with excellent repeatability, ICC(1,1) ≥ 0.9.

While a high number of texture features had a good ICC, there were generally few texture features with excellent ICC in all 4 data-sets. This finding supported a systematic review by Traverso et al (2018) [108] that found texture features to be least reproducible, compared to first-order and shape features in CT images. Work by Fiset et al (2019) [82] also had the same finding for cervical cancer T2-weighted MRI data.

Texture features are mostly considered in studies looking at the peritumoral region. For instance Chong et al (2021) looked at the prediction of tumour budding in cervical cancer patients and selected filtered texture features to use in a prediction model [83], [144]. However, the definition of the peritumoural region in our data-sets resulted in an annuli which could be considered to have dimensions too small for meaningful calculation of texture features. For this reason, these features are likely to be unstable and could be excluded from the resulting analysis. However, first-order features from the peritumoral region could also be used for lymph-vascular space invasion prediction in cervical cancer as shown in a study by Li et al (2019) [109]

2.4.2 Limitations

This study is hampered by the relatively small cohort size of 7 patients, however, the density of the timeseries resulting in 42 test-retest pairs of images was sufficient to give substantial results and support the intention of this study - to guide the selection of radiomic features for similar data-sets. In addition, though we ensured that the selected stable features were the ones least correlated to volume, a known prognostic indicator, we did not evaluate the clinical relevance of these stable features.

In the image processing step, sub-section 2.2.3, we chose to do upsampling of all our images (Table 2.2.1) to the same in-plane resolution of $0.9mm \times 0.9mm \times 3.0mm$.

Though, according to IBSI [47] there is no consensus as to whether upsampling or downsampling is preferable for resampling, upsampling may have introduced artificial information to our data [47].
The third limitation is that we used a high Intraclass Correlation Coefficient, ICC(1,1) threshold of 0.9. Most studies looking at test-retest analysis use a lower threshold of either 0.75 or 0.8 ICC, corresponding to "good and excellent repeatability" respectively, thus allowing more features into the "excellent ICC" category [19], [82], [90], [125]. The rationale behind our high threshold was that the images in our test-retest cohort were acquired within minutes of each other. We therefore assumed that there would be very minimal variation between radiomic features from images acquired within minutes of each other, given the controlled environment lacking variations in image acquisition parameters. A limitation in this assumption is that having a high threshold could lead to a loss of features that may hold relevant information [142]. Despite the high ICC(1,1) threshold set, we still found a high number of features falling under the "excellent" category, Table A.2.4. Based on the number of scans used in this study, this high number of features makes model building more complex due to multiple testing [70].
Another limitation is that, while test-retest analysis offers valuable information in terms of determining radiomic features' stability under minimally varying conditions, it can be argued that a test-retest setting where comparison is drawn between images are acquired within minutes of each other, does not represent a clinical setting where images are typically acquired days apart [143].
Finally, there was no assessment of any anatomical motions or changes that may take place within the time between the 2 images in the test-retest setting, nor the impact these potential changes could have on the radiomic feature values.

2.4.3 Research impact and future work

Overall, despite the short duration of time between the images in our test-retest data-sets, less than 25% of features extracted from each of the data-sets had an ICC(1,1) ≥ 0.9 as shown in Table A.2.3. This highlights the need for careful feature selection in radiomics studies [82], [125]. We therefore support the recommendation that a test-retest analysis should be carried out in radiomic studies to minimise variabilities in acquisition parameters and improve reliability of radiomic features used in models [82], [108], [143].
The identified stable features can be used to build models for larger data-sets based on similar imaging protocols including MR sequences used in this study. For the GTV region, these could be models looking at treatment response. For the peritumoral region, these could be models predicting nodal or distant metastasis based on microscopic disease in this region.

Based on the data-set used in this study - 3 scans for each patient, investigation into changes in radiomic features across different time-points could be carried out to potentially evaluate treatment response.

2.5 Conclusion

In conclusion, multiparametric MRI-based radiomic features of cervical cancer were investigated for repeatability in a test-retest setting. A radiomics pipeline using PyRadiomics to extract radiomic features from T2-weighted and Dixon images for Cervical Cancer patients was set-up, stable radiomic features based on high Intraclass Correlation Coefficient value between features in a test-retest setting were selected and the stable radiomic feature from each feature class were identified.
It was determined that most radiomic features extracted from images acquired within minutes of each other varied greatly and were study-specific. We therefore suggest consideration for test-retest repeatability analysis for radiomics studies, which allows for the selection of features that remain stable under constant conditions, to improve model generalisability, rather than using pre-selected features from prior studies. Further work is required to validate our findings before the identified stable features can be used to build reliable models on a larger data-set.

2.6 Conflicts of interest

None.

2.7 Acknowledgements

Chelmis M. Thiong'o is supported by the UK Newton Fund through the funding of the Development in Africa with Radio Astronomy (DARA) Big Data project.

2.8 Supplementary Data

Supplementary data to this article can be found in the Appendix, Section A.

Chapter 3

Conclusion

Radiomics has potential to enhance clinical decision making through the identification of imaging biomarkers that can be used to predict treatment outcomes. One limitation is model over-fitting due to the high number of radiomic features extracted from a single image, which affects model generalisability. Test-retest analysis improves generalisability of models built using radiomic features since it allows for elimination of features that change at constant conditions. This way, only stable features are retained. Stable features are those that are considered most informative for the study based on their independence from other features, repeatability and prominence in the data.

The aim of this study was to investigate multiparametric MRI-based radiomic features of cervical cancer for repeatability in a test-retest setting. A radiomics pipeline using PyRadiomics to extract radiomic features from T2-weighted and Dixon images for Cervical Cancer patients was set-up, stable radiomic features based on high Intraclass Correlation Coefficient value between features in a test-retest setting were selected and the stable radiomic feature from each feature class were identified. It was determined that most radiomic features extracted from images acquired within minutes of each other were minimally stable and varied based on MR sequence used and the site the features were extracted from. We therefore advice region and MR sequence-specific test-retest analyses for selection of radiomic features, rather than using pre-selected features from prior studies, for reliable models and to improve model generalisability. Further work is required to validate our findings before the identified stable features can be used to build reliable models on a larger data-set.

Appendix A

Test-Retest Repeatability of Multiparametric MR-based Radiomic Features in Cervical Cancer.

This section contains supplementary material for the journal manuscript.

A.1 Online Repository

The radiomics pipeline used in this study is available on Github [145].
Tables showing Spearman's Correlation Coefficient values for extracted radiomic features in relation to the original_shape_MeshVolume for all of the 4 data-sets are also available on GitHub [145].

A.2 Intraclass Correlation Coefficient, ICC(1,1)

Table A.2.1 shows the number of features that were seen to be in the excellent ICC category (ICC ≥ 0.9), good ICC category ($0.9 >$ ICC ≥ 0.75) and poor ICC category (ICC < 0.75) in the Gross Tumour volume (GTV) region.
Table A.2.2 shows the number of features that were seen to be in the excellent ICC category (ICC ≥ 0.9), good ICC category ($0.9 >$ ICC ≥ 0.75) and poor ICC category (ICC < 0.75) in the peritumoral region.
Tables showing ICC values and p_values for all 4 data-sets are available on GitHub [145].

		Original		LoG		Wavelet	
Feature Class		T2W	Dixon	T2W	Dixon	T2W	Dixon
		n	*n*	*n*	*n*	*n*	*n*
3D Shape	Total	4	4	0	0	0	0
	ICC $\geq$ 0.9	1	1	0	0	0	0
	0.9 > ICC $\geq$ 0.75	1	1	0	0	0	0
	ICC < 0.75	2	2	0	0	0	0
First-order	Total	16	16	48	48	128	128
	ICC $\geq$ 0.9	11	4	43	0	34	30
	0.9 > ICC $\geq$ 0.75	5	0	5	19	38	22
	ICC < 0.75	0	12	0	29	56	76
GLCM	Total	24	24	72	72	179	180
	ICC $\geq$ 0.9	16	0	44	0	20	0
	0.9 > ICC $\geq$ 0.75	6	2	28	44	67	51
	ICC < 0.75	2	22	0	28	92	129
GLRLM	Total	14	14	41	42	112	112
	ICC $\geq$ 0.9	6	0	20	3	10	0
	0.9 > ICC $\geq$ 0.75	2	0	12	22	31	21
	ICC < 0.75	6	14	9	17	71	91
GLSZM	Total	14	14	40	39	112	113
	ICC $\geq$ 0.9	3	0	16	4	13	3
	0.9 > ICC $\geq$ 0.75	3	1	15	10	41	22
	ICC < 0.75	8	13	9	25	58	88
GLDM	Total	12	12	36	36	96	96
	ICC $\geq$ 0.9	6	0	23	7	9	0
	0.9 > ICC $\geq$ 0.75	2	0	6	17	35	28
	ICC < 0.75	4	12	7	12	52	68
NGTDM	Total	2	2	6	6	16	16
	ICC $\geq$ 0.9	1	0	4	0	1	0
	0.9 > ICC $\geq$ 0.75	1	0	2	3	7	7
	ICC < 0.75	0	2	0	3	8	9

Table A.2.1: GTV ICC values for Radiomics features remaining after dropping volume-confounded features.
T2W - T2-weighted
n - number of features

Feature Class		Original		LoG		Wavelet	
		T2W	Dixon	T2W	Dixon	T2W	Dixon
		n	n	n	n	n	n
3D Shape	Total	4	4	0	0	0	0
	ICC $\geq$ 0.9	3	3	0	0	0	0
	0.9 > ICC $\geq$	1	1	0	0	0	0
	ICC < 0.75	0	0	0	0	0	0
First-order	Total	16	16	48	48	128	128
	ICC $\geq$ 0.9	13	4	17	0	11	32
	0.9 > ICC $\geq$	3	3	23	5	31	24
	ICC < 0.75	0	9	8	43	86	72
GLCM	Total	24	24	72	72	184	184
	ICC $\geq$ 0.9	3	0	3	0	6	0
	0.9 > ICC $\geq$	10	0	31	8	25	36
	ICC < 0.75	11	24	38	64	153	148
GLRLM	Total	14	15	42	42	112	113
	ICC $\geq$ 0.9	0	1	1	0	8	1
	0.9 > ICC $\geq$	14	0	20	3	12	12
	ICC < 0.75	0	14	21	39	92	100
GLSZM	Total	14	14	45	44	112	144
	ICC $\geq$ 0.9	1	0	6	1	6	1
	0.9 > ICC $\geq$	10	2	9	2	11	11
	ICC < 0.75	3	12	30	41	95	102
GLDM	Total	12	13	36	36	96	97
	ICC $\geq$ 0.9	1	1	2	0	4	1
	0.9 > ICC $\geq$	10	3	15	4	12	9
	ICC < 0.75	1	9	19	32	80	87
NGTDM	Total	2	2	7	9	16	17
	ICC $\geq$ 0.9	0	0	1	0	0	0
	0.9 > ICC $\geq$	1	0	2	6	2	5
	ICC < 0.75	1	2	4	3	14	12

Table A.2.2: Peritumoral region ICC values for Radiomics features remaining after dropping volume-confounded features.
T2W - T2-weighted
n - number of features

	T2-weighted images		Dixon images	
1130	GTV (n)	Peritumoral region (n)	GTV (n)	Peritumoral region (n)
Volume correlation, Spearman's, $\|\rho\| < 0.9$	972	984	974	992
ICC $\geq$ 0.9	281	86	52	45

Table A.2.3: Number of radiomic features remaining after each feature selections step.
n - number of features
GTV - Gross Tumour Volume
ICC - Intraclass Correlation Coefficient

	T2-Weighted				Dixon			
	GTV		Peritumoral		GTV		Peritumoral	
Feature Class	Features (n)	Stable Feature	Features (n)	Stable Feature	Features (n)	Stable Feature	Features (n)	Stable Feature
3D Shape	1	MeshVolume	3	Elongation	1	MeshVolume	3	Elongation
First-order	88	90 Percentile (O)	41	Minimum (LLL)	34	10 Percentile (HLL)	36	90 Percentile (O)
GLCM	80	Autocorrelation (LoG3)	12	Maximum Probability (LLH)	0	nan	0	nan
GLRLM	36	Long Run High Gray Level Emphasis (LoG3)	9	Run Percentage (LLL)	3	Short Run Emphasis (LoG5)	2	Gray Level NonUniformity (LLL)
GLSZM	32	Zone Entropy (O)	13	Zone Variance (LoG3)	7	Zone Percentage (LoG3)	2	Size Zone NonUniformity (LoG5)
GLDM	38	Large Dependence Low Gray Level Emphasis (LLL)	7	Dependence NonUniformity Normalized (LLL)	7	Dependence Entropy (LoG4)	2	Gray Level NonUniformity (LLL)
NGTDM	6	Complexity (LoG5)	1	Busyness (LoG5)	0	nan	0	nan

Table A.2.4: Selected stable feature from each feature class.
n - number of stable features in each feature class.
O - original image
LoG3 - image with Laplacian of Gaussian (LoG) filtering with σ value 3
LoG4 - image with Laplacian of Gaussian (LoG) filtering with σ value 4
LoG5 - image with Laplacian of Gaussian (LoG) filtering with σ value 5
LLL - image with wavelet LLL filtering
HLL - image with wavelet HLL filtering

Appendix B

Alternative Peritumoral Region Analysis

B.1 Introduction

This chapter entails a test-retest study to investigate repeatability of radiomic features in cervical cancer from T2-weighted (T2W) and Dixon water-only images for the GTV and the peritumoral region. It is adapted from an abstract that will be presented as a digital poster at The European Society for Radiotherapy and Oncology Congress (ESTRO) 2022 Congress [146].

B.2 Materials and Methods

6 patients with FIGO stage IB2 and IIB cervical cancer had 3 MR scans each, with at least a week between scans. The scans were acquired on a single MR machine. At each scan, a T2-weighted image was taken at the start and end of the exam and Dixon sequences were sequentially acquired over 10 minutes. For this study, T2W images at the start and end of the scan and the first and last Dixon sequences in the scan were included, resulting in 4 images per scan. This selection resulted in a total of 36 test-retest pairs including both T2W and Dixon sequences.
For each image, radiomic features were extracted from 2 regions: the GTV and a peritumoral region. The GTV was contoured by a single observer on all the start and end T2W images, and transferred through rigid registration to the first and last Dixon sequence images. The peritumoral region was acquired through dilating the first scan's GTV by 2 voxels (∼1.8mm) and using this contour on the second and third scans. The regions are shown in Figures B.2.1 and B.2.2.

PyRadiomics was used to extract first-order, shape, and texture radiomic features from the original T2W and Dixon images, and their Laplacian of Gaussian-filtered (with σ = 3, 4, and 5) and Wavelet-filtered counterparts. This resulted in 1127

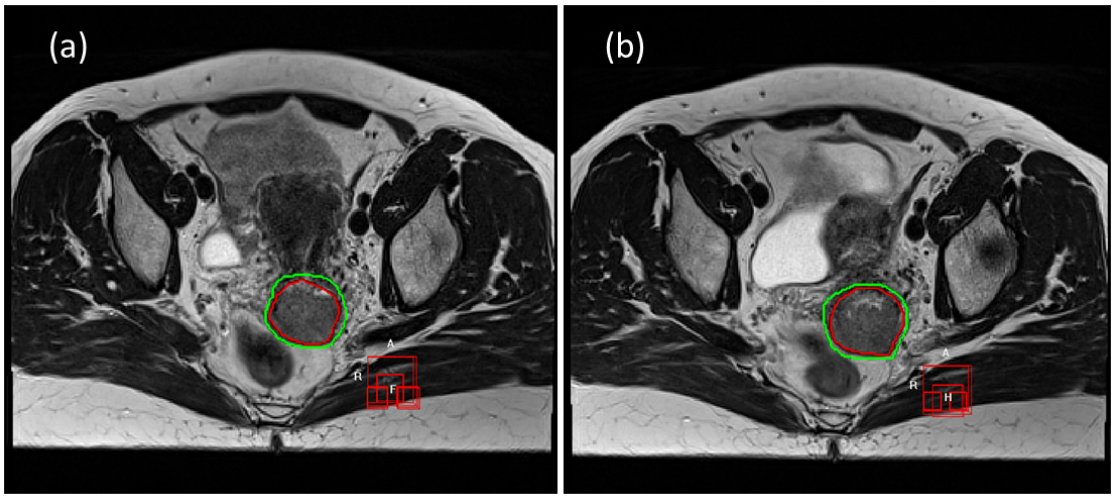


Figure B.2.1: Images from a single MR scan showing the GTV (red contour) and a peritumoral region extending 1.8mm from the GTV (region between the green and red contours). Figures "a" and "b" are T2-weighted images at the start and end of the scan, respectively

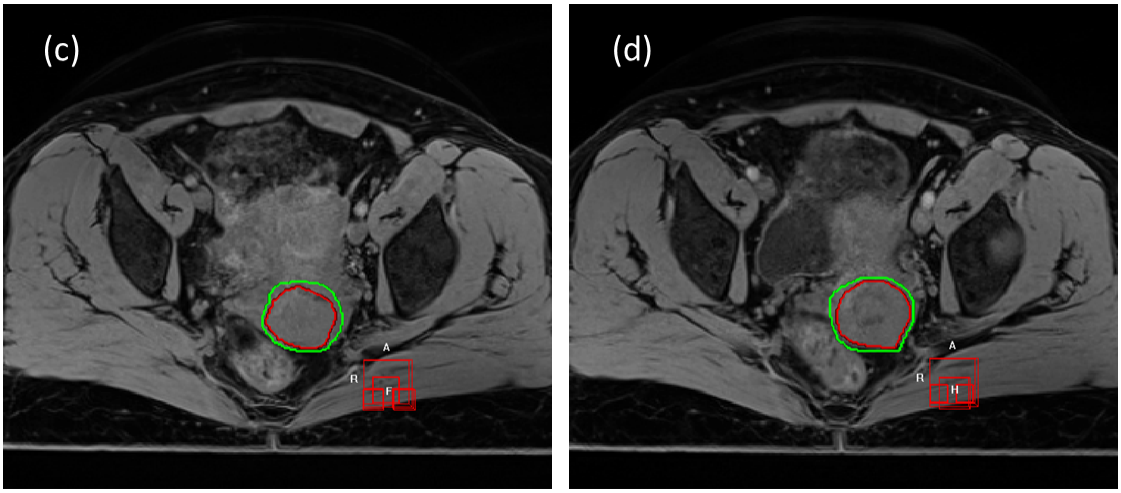


Figure B.2.2: Images from a single MR scan showing the GTV (red contour) and a peritumoral region extending 1.8mm from the GTV (region between the green and red contours). Figures "c" and "d" are Dixon water-only images at the start and end of the scan, respectively

features per image.

Spearman's Rank Correlation Coefficient, ρ, was used to exclude features highly correlated with volume ($|\rho| \geq 0.9$). Of the remaining features, those with minimal variation between the start and end of a single scan were selected as being stable using Intraclass Correlation Coefficient, ICC (1,1) ≥ 0.9. Spearman's, ρ, was then used to identify features that were not correlated with each other, with features having $|\rho| < 0.65$ being considered acceptable.

B.3 Results

In the GTV region, only original shape SurfaceVolumeRatio and MeshVolume features were repeatable in T2W and Dixon images. Original first-order InterquartileRange feature was also repeatable in T2W images.
In the peritumoral region, only original shape Elongation was repeatable in both T2W and Dixon images.
Table B.3.1a shows the number of radiomic features remaining at each step of the feature selection process. Less than 1% of radiomic features were stable and contained non-redundant information for potential predicitve modelling. Figure B.3.1 shows Venn diagrams of identified stable features in the GTV and peritumoral regions for both T2W and Dixon sequences.

	T2-weighted images		Dixon images	
1127	GTV (n)	Peritumoral region (n)	GTV (n)	Peritumoral region (n)
Volume correlation, Spearman's, $\|\rho\| < 0.9$	956	1003	976	1018
ICC ≥ 0.9	219	54	96	70
Correlation between features, Spearman's $\|\rho\| < 0.65$	3	1	2	1

Table B.3.1: Table showing number of radiomic features remaining after a feature selection method. 1127 is the total number of features extracted in each region for each MR sequence.
n - number of features
GTV - Gross Tumour Volume
ICC - Intraclass Correlation Coefficient

B.4 Conclusion

We show that features stable for one MR sequence are not necessarily stable in other sequences, and that the image region from which they are extracted can impact feature repeatability. We therefore advice region and modality specific test-retest analyses for selection of radiomic features for reliable models.

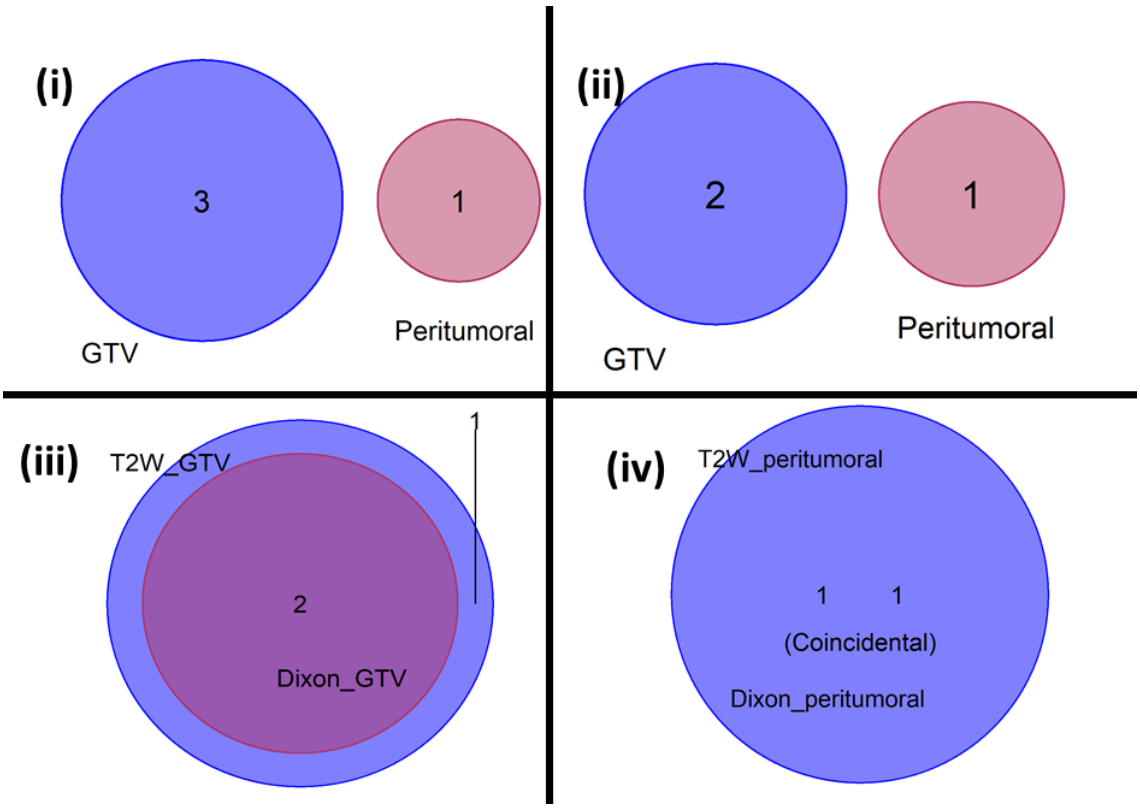


Figure B.3.1: Venn diagrams comparing features that were stable in T2W images (i) and Dixon-MR images (ii) for the GTV and peritumoral regions; and features that were stable in the GTV (iii) and in the peritumoral region (iv) for T2W and Dixon images.

Bibliography

[1] "World Health Organization (WHO)," WHO. (2022), [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cervical-cancer (visited on 03/11/2022).

[2] Geneva: World Health Organisation, *Global strategy to accelerate the elimination of cervical cancer as a public health problem*. 2020, ISBN: 9789240014107.

[3] M. Brisson, J. J. Kim, K. Canfell, *et al.*, "Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries," *The Lancet*, vol. 395, no. 10224, 2020, ISSN: 1474547X. DOI: 10.1016/S0140-6736(20)30068-4.

[4] K. Canfell, J. J. Kim, M. Brisson, *et al.*, "Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries," *The Lancet*, vol. 395, no. 10224, 2020, ISSN: 1474547X. DOI: 10.1016/S0140-6736(20)30157-4.

[5] "Global Cancer Observatory (GLOBOCAN) 2020." (2022), [Online]. Available: https://gco.iarc.fr/today/online-analysis-multi-bars?v=2020&mode=cancer&mode_population=countries&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=0&include_nmsc_other=1&type_multiple=%257B%2522inc%2522%253Atrue%252C%2522mort%2522%253Atrue%252C%2522prev%2522%253Afalse%257D&orientation=horizontal&type_sort=0&type_nb_items=%257B%2522top%2522%253Atrue%252C%2522bottom%2522%253Afalse%257D (visited on 03/18/2022).

[6] M. Vu, J. Yu, O. A. Awolude, and L. Chuang, *Cervical cancer worldwide*, 2018. DOI: 10.1016/j.currproblcancer.2018.06.003.

[7] S. Lin, K. Gao, S. Gu, *et al.*, "Worldwide trends in cervical cancer incidence and mortality, with predictions for the next 15 years," *American Cancer Society Journals*, Aug. 2021. DOI: 10.1002/cncr.33795.

[8] L. Bruni, A. Saura-Lázaro, A. Montoliu, *et al.*, "HPV vaccination introduction worldwide and WHO and UNICEF estimates of national HPV immunization coverage 2010–2019," *Preventive Medicine*, vol. 144, 2021, ISSN: 10960260. DOI: 10.1016/j.ypmed.2020.106399.

[9] "Human papillomavirus vaccines (HPV)," World Health Organization (WHO). (2022), [Online]. Available: https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/human-papillomavirus-vaccines-(HPV) (visited on 03/11/2022).

[10] S. Ranjbar and J. Ross Mitchell, "An introduction to radiomics: An evolving cornerstone of precision medicine," in *Biomedical Texture Analysis: Fundamentals, Tools and Challenges*, 2017. DOI: 10.1016/B978-0-12-812133-7.00008-9.

[11] "Survival," Cancer Research UK. (Jun. 18, 2020), [Online]. Available: https://www.cancerresearchuk.org/about-cancer/cervical-cancer/survival (visited on 03/11/2022).

[12] "Survival Rates for Cervical Cancer," American Cancer Society. (Mar. 1, 2022), [Online]. Available: https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/survival.html (visited on 03/11/2022).

[13] N. Bhatla, D. Aoki, D. N. Sharma, and R. Sankaranarayanan, "Cancer of the cervix uteri," *International Journal of Gynecology and Obstetrics*, 2018, ISSN: 18793479. DOI: 10.1002/ijgo.12611.

[14] P. Lambin, E. Rios-Velazquez, R. Leijenaar, *et al.*, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012, ISSN: 09598049. DOI: 10.1016/j.ejca.2011.11.036.

[15] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, 2014, ISSN: 20411723. DOI: 10.1038/ncomms5006.

[16] J. Kusmirek, J. Robbins, H. Allen, L. Barroilhet, B. Anderson, and E. A. Sadowski, *PET/CT and MRI in the imaging assessment of cervical cancer*, 2015. DOI: 10.1007/s00261-015-0363-6.

[17] A. Mahajan, R. Engineer, S. Chopra, *et al.*, "Role of 3T multiparametric-MRI with BOLD hypoxia imaging for diagnosis and post therapy response evaluation of postoperative recurrent cervical cancers," *European Journal of Radiology Open*, 2016, ISSN: 23520477. DOI: 10.1016/j.ejro.2015.11.003.

[18] S. S. Yip and H. J. Aerts, *Applications and limitations of radiomics*, 2016. DOI: 10.1088/0031-9155/61/13/R150.

[19] V. Kumar, Y. Gu, S. Basu, *et al.*, "Radiomics: The process and the challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, 2012, ISSN: 0730725X. DOI: 10.1016/j.mri.2012.06.010.

[20] P. Balcacer, A. Shergill, and B. Litkouhi, *MRI of cervical cancer with a surgical perspective: staging, prognostic implications and pitfalls*, 2019. DOI: 10.1007/s00261-019-01984-7.

[21] M. M. Otero-García, A. Mesa-Álvarez, O. Nikolic, *et al.*, *Role of MRI in staging and follow-up of endometrial and cervical cancer: pitfalls and mimickers*, 2019. DOI: 10.1186/s13244-019-0696-8.

[22] G. M. Rauch, H. Kaur, H. Choi, *et al.*, "Optimization of MR imaging for pretreatment evaluation of patients with endometrial and cervical cancer," *Radiographics*, vol. 34, no. 4, 2014, ISSN: 15271323. DOI: 10.1148/rg.344140001.

[23] S. Monti, M. Aiello, M. Incoronato, *et al.*, "DCE-MRI pharmacokinetic-based phenotyping of invasive ductal carcinoma: A radiomic study for prediction of histological outcomes," *Contrast Media and Molecular Imaging*, 2018, ISSN: 15554317. DOI: 10.1155/2018/5076269.

[24] R. M. Dunne, A. C. O'Neill, and C. M. Tempany, "Chapter 9 - Imaging Tools in Clinical Research: Focus on Imaging Technologies," in *Clinical and Translational Science (Second Edition)*, D. Robertson and G. H. Williams, Eds., Second Edi, Academic Press, 2017, pp. 157–179, ISBN: 978-0-12-802101-9. DOI: https://doi.org/10.1016/B978-0-12-802101-9.00009-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128021019000090.

[25] R. L. Wahl, "16 - Imaging," in *Abeloff's Clinical Oncology (Sixth Edition)*, J. E. Niederhuber, J. O. Armitage, M. B. Kastan, J. H. Doroshow, and J. E. Tepper, Eds., Sixth Edit, Philadelphia: Elsevier, 2020, pp. 254–283, ISBN: 978-0-323-47674-4. DOI: https://doi.org/10.1016/B978-0-323-47674-4.00016-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323476744000165.

[26] A. Stabile, F. Giganti, A. B. Rosenkrantz, *et al.*, *Multiparametric MRI for prostate cancer diagnosis: current status and future directions*, 2020. DOI: 10.1038/s41585-019-0212-4.

[27] V. Sawlani, M. D. Patel, N. Davies, *et al.*, "Multiparametric MRI: practical approach and pictorial review of a useful tool in the evaluation of brain tumours and tumour-like lesions," *Insights into Imaging*, vol. 11, no. 1, 2020, ISSN: 18694101. DOI: 10.1186/s13244-020-00888-1.

[28] G. B. Chavhan, *Appropriate selection of MRI sequences for common scenarios in clinical practice*, 2016. DOI: 10.1007/s00247-016-3556-4.

[29] Feger, J., "Prostate MRI protocol," *Radiopaedia.org*, 2019. DOI: https://doi.org/10.53347/rID-72497.

[30] K. Tsuda, T. Murakami, H. Kurachi, *et al.*, "Mr imaging of cervical carcinoma: Comparison among T2-weighted, dynamic, and postcontrast T1-weighted images with histopathological correlation," *Abdominal Imaging*, vol. 22, no. 1, 1997, ISSN: 09428925. DOI: 10.1007/s002619900151.

[31] C. F. Lins, C. E. G. Salmon, and M. H. Nogueira-Barbosa, "Applications of the dixon technique in the evaluation of the musculoskeletal system," *Radiologia Brasileira*, vol. 54, no. 1, 2021, ISSN: 01003984. DOI: 10.1590/0100-3984.2019.0086.

[32] J. Ma, *Dixon techniques for water and fat imaging*, 2008. DOI: 10.1002/jmri.21492.

[33] B. W. Wormald, S. J. Doran, T. E. Ind, J. D'Arcy, J. Petts, and N. M. deSouza, "Radiomic features of cervical cancer on T2-and diffusion-weighted MRI: Prognostic value in low-volume tumors suitable for trachelectomy," *Gynecologic Oncology*, vol. 156, no. 1, 2020, ISSN: 10956859. DOI: 10.1016/j.ygyno.2019.10.010.

[34] "Treatment Decisions — Cervical cancer," Cancer Research UK. (Mar. 18, 2020), [Online]. Available: https://www.cancerresearchuk.org/about-cancer/cervical-cancer/treatment/treatment-decisions (visited on 03/11/2022).

[35] P. Van Houtte, "RADIOTHERAPY," *Encyclopedia of Respiratory Medicine, Four-Volume Set*, pp. 610–614, Jan. 2006. DOI: 10.1016/B0-12-370879-6/00341-0.

[36] C. M. West and G. C. Barnett, "Genetics and genomics of radiotherapy toxicity: Towards prediction," *Genome Medicine*, vol. 3, no. 8, 2011, ISSN: 1756994X. DOI: 10.1186/gm268.

[37] E. L. Trimble, R. S. Ungerleider, J. A. Abrams, *et al.*, "Neoadjuvant therapy in cancer treatment," *Cancer*, vol. 72, no. 11 S, 1993, ISSN: 10970142. DOI: 10.1002/1097-0142(19931201)72:11+<3515::AID-CNCR2820721619>3.0.CO;2-A.

[38] S. Tangjitgamol, K. Katanyoo, M. Laopaiboon, P. Lumbiganon, S. Manusirivithaya, and B. Supawattanabodee, *Adjuvant chemotherapy after concurrent chemoradiation for locally advanced cervical cancer*, 2014. DOI: 10.1002/14651858.CD010401.pub2.

[39] Richard Pötter, Kari Tanderup, Jacob Lindegaard, Christian Kirisits, Ina Juergenliemk-Schulz, Astrid de Leeuw, Israël Fortin, Kathrin Kirchheiner, Dietmar Georg, Remi Nout, Yvette Seppenwoolde, Wolfgang Dörr, Thomas Liederer, Li Tee Tan, *Image guided intensity modulated External beam radiochemotherapy and MRI based adaptive BRAchytherapy in locally advanced CErvical cancer, EMBRACE-II protocol, v.1.0.*

[40] C. W. Williamson, H. C. Liu, J. Mayadev, and L. K. Mell, "Advances in External Beam Radiation Therapy and Brachytherapy for Cervical Cancer," *Clinical Oncology*, vol. 33, no. 9, 2021, ISSN: 14332981. DOI: 10.1016/j.clon.2021.06.012.

[41] K. E. Rosenzweig, C. P. Chen, S. S. Yom, and L. M. Krug, "35 - Tumors of the Lung, Pleura, and Mediastinum," in *Leibel and Phillips Textbook of Radiation Oncology (Third Edition)*, R. T. Hoppe, T. L. Phillips, and M. Roach, Eds., Third Edit, Philadelphia: W.B. Saunders, 2010, pp. 737–771, ISBN: 978-1-4160-5897-7. DOI: https://doi.org/10.1016/B978-1-4160-5897-7.00036-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781416058977000366.

[42] B. Thomadsen and R. Miller, "9.13 - Brachytherapy Physics," in *Comprehensive Biomedical Physics*, A. Brahme, Ed., Oxford: Elsevier, 2014, pp. 315–381, ISBN: 978-0-444-53633-4. DOI: https://doi.org/10.1016/B978-0-444-53632-7.00926-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780444536327009266.

[43] D. Vordermark, "Radiotherapy of Cervical Cancer," *Oncology Research and Treatment*, vol. 39, no. 9, pp. 516–520, 2016. DOI: 10.1159/000448902.

[44] T. V. Ajithkumar, "3 - Principles of radiotherapy," in *Specialist Training in Oncology*, T. V. Ajithkumar and H. M. Hatcher, Eds., Mosby, 2011, pp. 15–29, ISBN: 978-0-7234-3458-0. DOI: https://doi.org/10.1016/B978-0-7234-3458-0.00008-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780723434580000087.

[45] S. Dieterich, E. Ford, D. Pavord, and J. Zeng, "Treatment Planning and Quality Metrics," *Practical Radiation Oncology Physics*, pp. 189–206, 2016. DOI: 10.1016/B978-0-323-26209-5.00014-6.

[46] Y. Jiang, Q. Yuan, W. Lv, *et al.*, "Radiomic signature of 18F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits," *Theranostics*, vol. 8, no. 21, 2018, ISSN: 18387640. DOI: 10.7150/thno.28018.

[47] A. Zwanenburg, S. Leger, M. Vallières, and S. Löc, *Image biomarker standardisation initiative*, arXiv preprint arXiv:1612.07003, 2019.

[48] S. P. Haider, T. Zeevi, P. Baumeister, *et al.*, "Potential added value of PET/CT radiomics for survival prognostication beyond AJCC 8th edition staging in oropharyngeal squamous cell carcinoma," *Cancers*, vol. 12, no. 7, pp. 1–16, Jul. 2020, ISSN: 20726694. DOI: 10.3390/cancers12071778.

[49] Y. Luo, D. Mei, J. Gong, M. Zuo, and X. Guo, "Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymphovascular Space Invasion in Endometrial Carcinoma," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 4, pp. 1257–1262, Oct. 2020, ISSN: 15222586. DOI: 10.1002/jmri.27142.

[50] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, *Radiomics in medical imaging—"how-to" guide and critical reflection*, 2020. DOI: 10.1186/s13244-020-00887-2.

[51] V. G. De Gruttola, P. Clax, D. L. DeMets, *et al.*, "Considerations in the Evaluation of Surrogate Endpoints in Clinical Trials," *Controlled Clinical Trials*, vol. 22, no. 5, 2001, ISSN: 01972456. DOI: 10.1016/s0197-2456(01)00153-2.

[52] J. A. U. Perucho, M. Wang, K. Y. Tse, *et al.*, "Association between MRI histogram features and treatment response in locally advanced cervical cancer treated by chemoradiotherapy," *European Radiology*, vol. 31, no. 3, 2021, ISSN: 14321084. DOI: 10.1007/s00330-020-07217-6.

[53] A. HL, "The potential of radiomic-based phenotyping in precision medicine: A review," *JAMA Oncology*, vol. 2, no. 12, 2016, ISSN: 2374-2437.

[54] F. Davnall, C. S. Yip, G. Ljungqvist, *et al.*, *Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice?* 2012. DOI: 10.1007/s13244-012-0196-6.

[55] B. Zhang, J. Tian, D. Dong, *et al.*, "Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma," *Clinical Cancer Research*, vol. 23, no. 15, 2017, ISSN: 15573265. DOI: 10.1158/1078-0432.CCR-16-2910.

[56] H. Park, Y. Lim, E. S. Ko, *et al.*, "Radiomics signature on magnetic resonance imaging: Association with disease-free survival in patients with invasive breast cancer," *Clinical Cancer Research*, vol. 24, no. 19, 2018, ISSN: 15573265. DOI: 10.1158/1078-0432.CCR-17-3783.

[57] C. Sun, X. Tian, Z. Liu, *et al.*, "Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: A multicentre study," *EBioMedicine*, vol. 46, pp. 160–169, Aug. 2019, ISSN: 23523964. DOI: 10.1016/j.ebiom.2019.07.049.

[58] Z. Liu, S. Wang, D. Dong, *et al.*, *The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges*, 2019. DOI: 10.7150/thno.30309.

[59] Z. Liu, Z. Li, J. Qu, *et al.*, "Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: A multicenter study," *Clinical Cancer Research*, vol. 25, no. 12, 2019, ISSN: 15573265. DOI: 10.1158/1078-0432.CCR-18-3190.

[60] R. Cattell, S. Chen, and C. Huang, "Robustness of radiomic features in magnetic resonance imaging: review and a phantom study," *Visual Computing for Industry, Biomedicine, and Art*, 2019. DOI: 10.1186/s42492-019-0025-6.

[61] J. J. Van Griethuysen, A. Fedorov, C. Parmar, *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, 2017, ISSN: 15387445. DOI: 10.1158/0008-5472.CAN-17-0339.

[62] "A Computational Environment for Radiological Research Features." (2020), [Online]. Available: https://cerr.github.io/CERR (visited on 03/11/2022).

[63] "TexRAD," FEEDBACK Medical. (2022), [Online]. Available: https://fbkmed.com/texrad-landing-2 (visited on 03/11/2022).

[64] "LIFEx," LIFEx. (2022), [Online]. Available: https://lifexsoft.org (visited on 03/11/2022).

[65] "IBEX," NMMI Tools. (2019), [Online]. Available: https://nmmitools.org/2019/01/01/ibex_mda (visited on 03/11/2022).

[66] I. Fornacon-Wood, H. Mistry, C. J. Ackermann, *et al.*, "Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform," *European Radiology*, vol. 30, no. 11, pp. 6241–6250, Nov. 2020, ISSN: 14321084. DOI: 10.1007/s00330-020-06957-9.

[67] A. Zwanenburg, M. Vallières, M. A. Abdalah, *et al.*, "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, 2020, ISSN: 15271315. DOI: 10.1148/radiol.2020191145.

[68] T. Wang, T. Gao, J. Yang, *et al.*, "Preoperative prediction of pelvic lymph nodes metastasis in early-stage cervical cancer using radiomics nomogram developed based on T2-weighted MRI and diffusion-weighted imaging," *European Journal of Radiology*, vol. 114, pp. 128–135, May 2019, ISSN: 18727727. DOI: 10.1016/j.ejrad.2019.01.003.

[69] E. Scalco and G. Rizzo, *Texture analysis of medical images for radiotherapy applications*, 2017. DOI: 10.1259/bjr.20160642.

[70] C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, *A deep look into radiomics*, 2021. DOI: 10.1007/s11547-021-01389-x.

[71] W. C. Shen, S. W. Chen, J. A. Liang, T. C. Hsieh, K. Y. Yen, and C. H. Kao, "[18]Fluorodeoxyglucose Positron Emission Tomography for the Textural Features of Cervical Cancer Associated with Lymph Node Metastasis and Histological Type," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 10, 2017, ISSN: 16197089. DOI: 10.1007/s00259-017-3697-1.

[72] M. Xiao, F. Ma, Y. Li, *et al.*, "Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymph Node Metastasis in Early-Stage Cervical Cancer," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 3, pp. 885–896, Sep. 2020, ISSN: 15222586. DOI: 10.1002/jmri.27101.

[73] M. Fan, G. Wu, H. Cheng, J. Zhang, G. Shao, and L. Li, "Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients," *European Journal of Radiology*, vol. 94, 2017, ISSN: 18727727. DOI: 10.1016/j.ejrad.2017.06.019.

[74] T. Upadhaya, M. Vallieres, A. Chatterjee, *et al.*, "Comparison of Radiomics Models Built Through Machine Learning in a Multicentric Context With Independent Testing: Identical Data, Similar Algorithms, Different Methodologies," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, 2018, ISSN: 2469-7311. DOI: 10.1109/trpms.2018.2878934.

[75] Y. Ai, H. Zhu, C. Xie, and X. Jin, *Radiomics in cervical cancer: Current applications and future potential*, Aug. 2020. DOI: 10.1016/j.critrevonc.2020.102985.

[76] T. Tsujikawa, T. Rahman, M. Yamamoto, *et al.*, "18F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer," *Annals of Nuclear Medicine*, vol. 31, no. 9, pp. 678–685, Nov. 2017, ISSN: 18646433. DOI: 10.1007/s12149-017-1199-7.

[77] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, 1973, ISSN: 21682909. DOI: 10.1109/TSMC.1973.4309314.

[78] "Graycomatrix," The MathWorks, Inc. (2022), [Online]. Available: https://uk.mathworks.com/help/images/ref/graycomatrix.html (visited on 03/11/2022).

[79] "Pyradiomics." (2016), [Online]. Available: https://pyradiomics.readthedocs.io/en/latest/features.html (visited on 12/12/2020).

[80] J. Shi, L. Cui, H. Wang, *et al.*, "MRI-based intratumoral and peritumoral radiomics on prediction of lymph-vascular space invasion in cervical cancer: A multi-center study," *Biomedical Signal Processing and Control*, vol. 72, 2022, ISSN: 17468108. DOI: 10.1016/j.bspc.2021.103373.

[81] Q. Wu, S. Wang, X. Chen, *et al.*, "Radiomics analysis of magnetic resonance imaging improves diagnostic performance of lymph node metastasis in patients with cervical cancer," *Radiotherapy and Oncology*, vol. 138, 2019, ISSN: 18790887. DOI: 10.1016/j.radonc.2019.04.035.

[82] S. Fiset, M. L. Welch, J. Weiss, *et al.*, "Repeatability and reproducibility of MRI-based radiomic features in cervical cancer," *Radiotherapy and Oncology*, vol. 135, 2019, ISSN: 18790887. DOI: 10.1016/j.radonc.2019.03.001.

[83] W. Hua, T. Xiao, X. Jiang, *et al.*, "Lymph-vascular space invasion prediction in cervical cancer: Exploring radiomics and deep learning multilevel features of tumor and peritumor tissue on multiparametric MRI," *Biomedical Signal Processing and Control*, vol. 58, Apr. 2020, ISSN: 17468108. DOI: 10.1016/j.bspc.2020.101869.

[84] J. Fang, B. Zhang, S. Wang, *et al.*, "Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer," *Theranostics*, vol. 10, no. 5, 2020, ISSN: 18387640. DOI: 10.7150/thno.37429.

[85] Y. Liu, Y. Zhang, R. Cheng, *et al.*, "Radiomics analysis of apparent diffusion coefficient in cervical cancer: A preliminary study on histological grade evaluation," *Journal of Magnetic Resonance Imaging*, 2019, ISSN: 15222586. DOI: 10.1002/jmri.26192.

[86] A. Traverso, M. Kazmierski, M. L. Welch, *et al.*, "Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients," *Radiotherapy and Oncology*, vol. 143, 2020, ISSN: 18790887. DOI: 10.1016/j.radonc.2019.08.008.

[87] A. Takada, H. Yokota, M. Watanabe Nemoto, T. Horikoshi, J. Matsushima, and T. Uno, "A multi-scanner study of MRI radiomics in uterine cervical cancer: prediction of in-field tumor control after definitive radiotherapy based on a machine learning method including peritumoral regions," *Japanese Journal of Radiology*, vol. 38, no. 3, pp. 265–273, Mar. 2020, ISSN: 1867108X. DOI: 10.1007/s11604-019-00917-0.

[88] J. Wu, C. Lian, S. Ruan, *et al.*, "Treatment Outcome Prediction for Cancer Patients Based on Radiomics and Belief Function Theory," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2018, ISSN: 2469-7311. DOI: 10.1109/trpms.2018.2872406.

[89] J. Shi, Y. Dong, W. Jiang, *et al.*, "MRI-based peritumoral radiomics analysis for preoperative prediction of lymph node metastasis in early-stage cervical cancer: A multi-center study," *Magnetic Resonance Imaging*, 2021, ISSN: 0730-725X. DOI: https://doi.org/10.1016/j.mri.2021.12.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X21002563.

[90] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, Feb. 2016, ISSN: 15271315. DOI: 10.1148/radiol.2015151169.

[91] B. Klaeser, J. Wiskirchen, J. Wartenberg, *et al.*, "PET/CT-guided biopsies of metabolically active bone lesions: Applications and clinical impact," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 11, 2010, ISSN: 16197070. DOI: 10.1007/s00259-010-1524-z.

[92] A. Wibmer, H. Hricak, T. Gondo, *et al.*, "Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores," *European Radiology*, vol. 25, no. 10, 2015, ISSN: 14321084. DOI: 10.1007/s00330-015-3701-8.

[93] I. Fornacon-Wood, C. Faivre-Finn, J. P. O'Connor, and G. J. Price, "Radiomics as a personalized medicine tool in lung cancer: Separating the hope from the hype," *Lung Cancer*, vol. 146, no. March, pp. 197–208, 2020, ISSN: 18728332. DOI: 10.1016/j.lungcan.2020.05.028. [Online]. Available: https://doi.org/10.1016/j.lungcan.2020.05.028.

[94] S. Rizzo, F. Botta, D. Origgi, *et al.*, "Radiomics: the facts and the challenges of image analysis," *European radiology experimental*, vol. 2, no. 1, 2018, ISSN: 25099280. DOI: 10.1186/s41747-018-0068-z.

[95] M. Bologna, V. Corino, and L. Mainardi, "Technical Note: Virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain," *Medical Physics*, 2019, ISSN: 00942405. DOI: 10.1002/mp.13834.

[96] A. Traverso, M. Kazmierski, Z. Shi, *et al.*, "Stability of radiomic features of apparent diffusion coefficient (ADC) maps for locally advanced rectal cancer in response to image pre-processing," *Physica Medica*, 2019, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2019.04.009.

[97] S. H. Park, H. Lim, B. K. Bae, *et al.*, "Robustness of magnetic resonance radiomic features to pixel size resampling and interpolation in patients with cervical cancer," *Cancer Imaging*, vol. 21, no. 1, 2021, ISSN: 14707330. DOI: 10.1186/s40644-021-00388-5.

[98] M. Lacroix, F. Frouin, A. S. Dirand, *et al.*, "Correction for Magnetic Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential of MR Radiomic Features in Lung Cancer," *Frontiers in Oncology*, vol. 10, 2020, ISSN: 2234943X. DOI: 10.3389/fonc.2020.00043.

[99] C. Parmar, E. R. Velazquez, R. Leijenaar, *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLoS ONE*, 2014, ISSN: 19326203. DOI: 10.1371/journal.pone.0102107.

[100] M. D. Greer, N. Lay, J. H. Shih, *et al.*, "Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study," *European Radiology*, vol. 28, no. 10, 2018, ISSN: 14321084. DOI: 10.1007/s00330-018-5374-6.

[101] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, *Characterization of PET/CT images using texture analysis: the past, the present... any future?* 2017. DOI: 10.1007/s00259-016-3427-0.

[102] M. D. Bardis, R. Houshyar, P. D. Chang, *et al.*, "Applications of artificial intelligence to prostate multiparametric mri (Mpmri): Current and emerging trends," *Cancers*, vol. 12, no. 5, 2020, ISSN: 20726694. DOI: 10.3390/cancers12051204.

[103] I. M. Lips, U. A. van der Heide, K. Haustermans, *et al.*, "Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): Study protocol for a randomized controlled trial," *Trials*, vol. 12, 2011, ISSN: 17456215. DOI: 10.1186/1745-6215-12-255.

[104] N. Cellini, A. G. Morganti, G. C. Mattiucci, *et al.*, "Analysis of intraprostatic failures in patients treated with hormonal therapy and radiotherapy: Implications for conformal therapy planning," *International Journal of Radiation Oncology Biology Physics*, vol. 53, no. 3, 2002, ISSN: 03603016. DOI: 10.1016/S0360-3016(02)02795-5.

[105] Z. Liu, X. Liu, B. Xiao, *et al.*, "Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network," *Physica Medica*, vol. 69, 2020, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2019.12.008.

[106] A. Cárdenas-Blanco, C. Tejos, P. Irarrazaval, and I. Cameron, *Noise in magnitude magnetic resonance images*, 2008. DOI: 10.1002/cmr.a.20124.

[107] M. A. Balafar, A. R. Ramli, M. I. Saripan, and S. Mashohor, *Review of brain MRI image segmentation methods*, 2010. DOI: 10.1007/s10462-010-9155-0.

[108] A. Traverso, L. Wee, A. Dekker, and R. Gillies, "Repeatability and Reproducibility of Radiomic Features: A Systematic Review," *International Journal of Radiation Oncology Biology Physics*, vol. 102, no. 4, 2018, ISSN: 1879355X. DOI: 10.1016/j.ijrobp.2018.05.053.

[109] Z. Li, H. Li, S. Wang, *et al.*, "MR-Based Radiomics Nomogram of Cervical Cancer in Prediction of the Lymph-Vascular Space Invasion preoperatively," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 5, pp. 1420–1426, May 2019, ISSN: 15222586. DOI: 10.1002/jmri.26531.

[110] Q. Wu, S. Wang, S. Zhang, *et al.*, "Development of a Deep Learning Model to Identify Lymph Node Metastasis on Magnetic Resonance Imaging in Patients With Cervical Cancer," *JAMA network open*, vol. 3, no. 7, 2020, ISSN: 25743805. DOI: 10.1001/jamanetworkopen.2020.11625.

[111] M. Fang, Y. Kan, D. Dong, *et al.*, "Multi-Habitat Based Radiomics for the Prediction of Treatment Response to Concurrent Chemotherapy and Radiation Therapy in Locally Advanced Cervical Cancer," *Frontiers in Oncology*, vol. 10, May 2020, ISSN: 2234943X. DOI: 10.3389/fonc.2020.00563.

[112] D. G. Marconi, J. H. T. G. Fregnani, R. R. Rossini, *et al.*, "Pre-treatment MRI minimum apparent diffusion coefficient value is a potential prognostic imaging biomarker in cervical cancer patients treated with definitive chemoradiation," *BMC Cancer*, 2016, ISSN: 14712407. DOI: 10.1186/s12885-016-2619-0.

[113] Y. Lu, W. Peng, J. Song, *et al.*, "On the potential use of dynamic contrast-enhanced (DCE) MRI parameters as radiomic features of cervical cancer," *Medical Physics*, vol. 46, no. 11, pp. 5098–5109, Nov. 2019, ISSN: 00942405. DOI: 10.1002/mp.13821.

[114] T. Dong, C. Yang, B. Cui, *et al.*, "Development and Validation of a Deep Learning Radiomics Model Predicting Lymph Node Status in Operable Cervical Cancer," *Frontiers in Oncology*, vol. 10, 2020, ISSN: 2234943X. DOI: 10.3389/fonc.2020.00464.

[115] H. Hao, Z. Zhou, S. Li, *et al.*, "Shell feature: A new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer," *Physics in Medicine and Biology*, 2018, ISSN: 13616560. DOI: 10.1088/1361-6560/aabb5e.

[116] B. A. Altazi, D. C. Fernandez, G. G. Zhang, *et al.*, "Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes," *Physica Medica*, vol. 46, 2018, ISSN: 1724191X. DOI: 10.1016/j.ejmp.2017.10.009.

[117] X. r. Li, J. j. Jin, Y. Yu, X. h. Wang, Y. Guo, and H. z. Sun, "PET-CT radiomics by integrating primary tumor and peritumoral areas predicts E-cadherin expression and correlates with pelvic lymph node metastasis in early-stage cervical cancer," *European Radiology*, vol. 31, no. 8, 2021, ISSN: 14321084. DOI: 10.1007/s00330-021-07690-7.

[118] R. T. Larue, J. E. van Timmeren, E. E. de Jong, *et al.*, "Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study," *Acta Oncologica*, vol. 56, no. 11, 2017, ISSN: 1651226X. DOI: 10.1080/0284186X.2017.1351624.

[119] T. M. Lehmann, C. Gönner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, 1999, ISSN: 02780062. DOI: 10.1109/42.816070.

[120] B. A. Altazi, G. G. Zhang, D. C. Fernandez, *et al.*, "Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms," *Journal of Applied Clinical Medical Physics*, vol. 18, no. 6, pp. 32–48, Nov. 2017, ISSN: 15269914. DOI: 10.1002/acm2.12170.

[121] L. Duron, D. Balvay, S. V. Perre, *et al.*, "Gray-level discretization impacts reproducible MRI radiomics texture features," *PLoS ONE*, 2019, ISSN: 19326203. DOI: 10.1371/journal.pone.0213459.

[122] F. Mottaghy, T. Deist, W. VanElmpt, *et al.*, "Radiomics: The bridge between medical imaging and personalized medicine," *Nature Reviews Clinical Oncology*, vol. 14, no. 12, 2017, ISSN: 17594774. DOI: 10.1038/nrclinonc.2017.141.

[123] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, "Image biomarker standardisation initiative," 2016. DOI: 10.1148/radiol.2020191145. [Online]. Available: http://arxiv.org/abs/1612.07003%0Ahttp://dx.doi.org/10.1148/radiol.2020191145.

[124] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, and I. Buvat, "Validation of a method to compensate multicenter effects affecting CT radiomics," *Radiology*, vol. 291, no. 1, pp. 53–59, Apr. 2019, ISSN: 15271315. DOI: 10.1148/radiol.2019182023.

[125] M. Schwier, J. van Griethuysen, M. G. Vangel, *et al.*, "Repeatability of Multiparametric Prostate MRI Radiomics Features," *Scientific Reports*, 2019, ISSN: 20452322. DOI: 10.1038/s41598-019-45766-z.

[126] M. Scrivener, E. E. de Jong, J. E. van Timmeren, T. Pieters, B. Ghaye, and X. Geets, *Radiomics applied to lung cancer: A review*, 2016. DOI: 10.21037/tcr.2016.06.18.

[127] R. Berenguer, M. Del Rosario Pastor-Juan, J. Canales-Vázquez, *et al.*, "Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters," *Radiology*, vol. 288, no. 2, pp. 407–415, Aug. 2018, ISSN: 15271315. DOI: 10.1148/radiol.2018172361.

[128] A. Saha, M. R. Harowicz, and M. A. Mazurowski, "Breast cancer MRI radiomics: An overview of algorithmic features and impact of inter-reader variability in annotating tumors," *Medical Physics*, 2018, ISSN: 00942405. DOI: 10.1002/mp.12925.

[129] R. T. Leijenaar, S. Carvalho, E. R. Velazquez, *et al.*, "Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability," *Acta Oncologica*, vol. 52, no. 7, 2013, ISSN: 0284186X. DOI: 10.3109/0284186X.2013.812798.

[130] M. Pavic, M. Bogowicz, X. Würms, *et al.*, "Influence of inter-observer delineation variability on radiomics stability in different tumor sites," *Acta Oncologica*, vol. 57, no. 8, 2018, ISSN: 1651226X. DOI: 10.1080/0284186X.2018.1445283.

[131] "Random Forests," DeepAI. (2019), [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/random-forest (visited on 03/11/2022).

[132] J. E. Park, D. Kim, H. S. Kim, *et al.*, "Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement," *European Radiology*, vol. 30, no. 1, 2020, ISSN: 14321084. DOI: 10.1007/s00330-019-06360-z.

[133] K. G. Moons, D. G. Altman, J. B. Reitsma, *et al.*, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration," *Annals of Internal Medicine*, vol. 162, no. 1, 2015, ISSN: 15393704. DOI: 10.7326/M14-0698.

[134] B. Zhao, Y. Tan, W. Y. Tsai, *et al.*, "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Scientific Reports*, vol. 6, 2016, ISSN: 20452322. DOI: 10.1038/srep23428.

[135] X. Wu, D. Dong, L. Zhang, *et al.*, "Exploring the predictive value of additional peritumoral regions based on deep learning and radiomics: A multicenter study," *Medical Physics*, vol. 48, no. 5, 2021, ISSN: 00942405. DOI: 10.1002/mp.14767.

[136] M. L. Welch, C. McIntosh, B. Haibe-Kains, *et al.*, "Vulnerabilities of radiomic signature development: The need for safeguards," *Radiotherapy and Oncology*, vol. 130, pp. 2–9, Jan. 2019, ISSN: 18790887. DOI: 10.1016/j.radonc.2018.10.027.

[137] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, 2016, ISSN: 15563707. DOI: 10.1016/j.jcm.2016.02.012.

[138] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, 1979, ISSN: 00332909. DOI: 10.1037/0033-2909.86.2.420.

[139] H. Lu, N. A. Parra, J. Qi, *et al.*, "Repeatability of Quantitative Imaging Features in Prostate Magnetic Resonance Imaging," *Frontiers in Oncology*, vol. 10, 2020, ISSN: 2234943X. DOI: 10.3389/fonc.2020.00551.

[140] C. Xue, J. Yuan, D. M. Poon, *et al.*, "Reliability of MRI radiomics features in MR-guided radiotherapy for prostate cancer: Repeatability, reproducibility, and within-subject agreement," *Medical Physics*, vol. 48, no. 11, 2021, ISSN: 00942405. DOI: 10.1002/mp.15232.

[141] J. Lee, A. Steinmann, Y. Ding, *et al.*, "Radiomics feature robustness as measured using an MRI phantom," *Scientific Reports*, vol. 11, no. 1, 2021, ISSN: 20452322. DOI: 10.1038/s41598-021-83593-3.

[142] A. Lecler, L. Duron, D. Balvay, *et al.*, *Combining Multiple Magnetic Resonance Imaging Sequences Provides Independent Reproducible Radiomics Features*, 2019. DOI: 10.1038/s41598-018-37984-8.

[143] J. E. van Timmeren, R. T. Leijenaar, W. van Elmpt, *et al.*, "Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific?" *Tomography*, vol. 2, no. 4, 2016, ISSN: 2379-139X. DOI: 10.18383/j.tom.2016.00208.

[144] G. O. Chong, S. H. Park, N. J. Y. Park, *et al.*, "Predicting tumor budding status in cervical cancer using mri radiomics: Linking imaging biomarkers to histologic characteristics," *Cancers*, vol. 13, no. 20, 2021, ISSN: 20726694. DOI: 10.3390/cancers13205140.

[145] "CC_radiomics_project," GitHub Inc. (2022), [Online]. Available: https://github.com/rrr-uom-projects/CC_radiomics_project (visited on 03/11/2022).

[146] "The European Society for Radiotherapy and Oncology," ESTRO. (), [Online]. Available: https://www.estro.org/Congresses/ESTRO-2022 (visited on 03/11/2022).