



**Fontys Paramedische Hogeschool**  
Medical Imaging and Radiation Therapy

**Comparison between the numerical output of  
the Maastricht and Moffitt Radiomic software  
implementations for lung and head and neck  
tumors.**

Thesis, June 2016

Marlon van den Broek

Student number: 2199470

Principals: Maastricht Clinic Maastricht and Moffitt Cancer Center Tampa

Mentor: Esther Bloemen (Maastricht), Eduardo Moros (Moffitt)

Supervisors: Ralph Leijenaar (Maastricht), Robert Gillies (Moffitt)

## Preface

This thesis is written as completion of the study Medical imaging and radiation therapy at the Fontys University of Applied Sciences. The purpose of this study is to compare the radiomic tools of Maastricht clinic Maastricht, The Netherlands and of Moffitt cancer center Tampa, the United States.

Together with Mirthe Pijls I went to Tampa to write this thesis. I would love to thank her for the partnership and for being there for me.

I would like to thank Maastricht and Moffitt to give me the opportunity to do my thesis at their institutes, Madelon Johannesma for get me in contact with both institutes, Robert Gillies for the hospitality and the help of his radiomics department. Ralph Leijenaar and Esther Bloemen for the help and feedback, Shonagh Russel for the English and spelling control.

Marlon van den Broek

May 2016

# Abstract

## Introduction

Radiomics is a tool which aims to utilize the full potential of medical imaging by extracting a large amount of quantitative features by using mathematical algorithms, for example tumor intensity, shape and texture. Maastricht clinic and Moffitt cancer center developed their own radiomics software tool. In order to facilitate interoperability of Radiomics, their differences must be elucidated.

## Objective

The difference in numerical values will be explored for radiomics features of lung and head/neck tumors between the Maastricht and Moffitt software tool and we will investigate whether this difference is disease site dependent.

## Methods

The features of Maastricht and Moffitt are matched by name, algorithm or both. The CT scans and RT structures of the 42 lung and the 52 head/neck patients were loaded into the Maastricht and Moffitt tool. To quantify the degree of agreement, Lin's Concordance Correlation Coefficient (CCC) was calculated. The CCC's of the lung and head/neck features were compared to each other.

## Results

A total of 34 features matched. The lung cohort had 24 features with a poor CCC ( $<0,90$ ), 1 feature had a moderate CCC ( $0,90-0,95$ ) and 9 features had an almost perfect CCC ( $>0,99$ ). For the head/neck cohort 24 features had a poor CCC, 1 feature had a moderate CCC, 3 features had a substantial CCC ( $0,95-0,99$ ) and 6 features had an almost perfect CCC. The comparison between the CCC of lung and the CCC of head/neck showed that 19 features have an equal CCC and 15 features do not have an equal CCC.

## Conclusion

We found a substantial difference in feature values for both lung and head and neck cancer patients between the Maastricht and Moffitt radiomics software tool. Generally there is a poor agreement, in terms of CCC's, between both tools. We found that for 15 features the difference between both tools may depend on the disease site. To ensure the optimal use of radiomics tools for lung and head/neck cancer in the future, standardization is required.

# Samenvatting

## Inleiding

Radiomics heeft ten doel de volle potentie van medische beeldvorming te benutten door het extraheren van een grote hoeveelheid kenmerken (features). Mathematische algoritmes beschrijven bijvoorbeeld de tumor intensiteit, vorm en textuur. Maastric clinic en Moffitt cancer center hebben beide een eigen radiomics tool ontwikkeld. Om de data analyse per tool te verbeteren zullen de verschillen tussen beide tools moeten worden onderzocht.

## Doelstelling

Onderzoeken wat het verschil in de numerieke waarden van radiomic features voor long en hoofd/hals tumoren tussen de Maastric en Moffitt software tool is. Daarnaast wordt onderzocht of dit verschil afhankelijk is van de plaats van de tumor.

## Methode

De features van Maastric en Moffitt zijn gematcht op naam en algoritme. De CT scans en de RT structuren van de 42 long en de 52 hoofd/hals patiënten werden geladen in de Maastric en Moffitt tools. Om de mate van overeenkomst tussen de Maastric en Moffitt tool te kwantificeren is de concordantie correlatie coëfficiënt (CCC) van Lin berekend. De CCC's van de long en de hoofd/hals features worden met elkaar vergeleken.

## Resultaten

In totaal zijn 34 features gematcht. De long cohort had 24 features met een zwakke CCC ( $<0,90$ ), 1 feature met een matige CCC ( $0,90-0,95$ ) en 9 features met een bijna perfecte CCC ( $>0,99$ ). De hoofd/hals cohort had 24 features met een zwakke CCC, 1 feature met een matige CCC, 3 features met een wezenlijke CCC ( $0,95-0,99$ ) en 6 features met een bijna perfecte CCC. De vergelijking tussen de CCC's van long en hoofd/hals laat zien dat 19 features een gelijke CCC hebben, bij 15 features verschilde de CCC.

## Conclusie

Een verschil in feature waarden tussen de Maastric en Moffitt software tool voor zowel long als hoofd/hals kanker patiënten werd aangetoond. De CCC toonde een zwakke overeenkomst tussen beide tools. Voor 15 features is het verschil tussen beide tools mogelijk afhankelijk is van de tumor plaats. Om het optimale gebruik van de radiomics tools voor long en hoofd/hals kanker te verzekeren is er standaardisatie vereist.

# Contents

Preface.....	2
Abstract.....	3
Introduction.....	3
Objective .....	3
Methods .....	3
Results .....	3
Conclusion .....	3
Samenvatting.....	4
Inleiding .....	4
Doelstelling .....	4
Methode .....	4
Resultaten .....	4
Conclusie.....	4
Personnal details.....	6
Introduction.....	7
Method .....	9
Research model .....	9
Data collection.....	9
Measurements .....	9
Data analysis .....	10
Ethics .....	10
Results .....	11
Discussion .....	16
Conclusion.....	18
References .....	19
Appendices.....	21
Appendix I.....	21
Appendix II.....	22
Appendix III.....	36
Appendix IV .....	47
Appendix V .....	48
Appendix VI .....	49

## Personnal details

Marlon van den Broek

Student number: 2199470

Fontys University of applied science

Medical Imaging and Radiation Therapy

E-mail: [Marlonannacatharina@gmail.com](mailto:Marlonannacatharina@gmail.com)

Phone number: (+31) 649983262

## Principals: MAASTRO clinic Maastricht & Moffitt Cancer Center Tampa

### MAASTRO clinic and Fontys University of applied science

Esther Bloemen

Function: Mentor

E-mail: [Esther.bloemen@maastro.nl](mailto:Esther.bloemen@maastro.nl)

Phone number: (+31) 0653230749

### MAASTRO clinic

Ralph Leijenaar

Function: Supervisor

E-mail: [Ralph.leijenaar@maastro.nl](mailto:Ralph.leijenaar@maastro.nl)

Phone number: +31 (088) 445 5666

Claudia Offermann

Function: Graduation coordinator

E-mail: [Claudia.offermann@maastro.nl](mailto:Claudia.offermann@maastro.nl)

Phone number: +31 (088) 445 5666

*Disclaimer (Appendix I)*

### Moffitt Cancer Center

Robert Gillies

Function: Supervisor/mentor

E-mail: [Robert.Gillies@moffitt.org](mailto:Robert.Gillies@moffitt.org)

Phone number: +1 (813) 745 8355

Eduardo Moros

Function: Mentor

E-mail: [Eduardo.moros@moffitt.org](mailto:Eduardo.moros@moffitt.org)

Phone number: +1 (813) 745 7231

## Introduction

In 2012 there were 14.1 million new cancer cases and 8.2 million cancer related deaths worldwide. Lung cancer accounts for 1.5 million deaths per year and is the leading cause of cancer related death in men. It is the most frequently diagnosed cancer type with more than 1.8 million cases per year. Head and neck cancer, although less frequently diagnosed than lung cancer with only 550,00 cases reported worldwide per year, still accounted for approximately 300,000 deaths in 2012.<sup>1</sup>

There are many ways to diagnose cancer, one such way is through the use of medical imaging technology. Over the past decade these technologies have greatly increased in number and capabilities. Although medical imaging in the past was primarily used as a diagnostic tool, it is now used more centrally in oncology to provide screening, prediction, treatment guidance and treatment response evaluation. Compared to other sources of information, e.g. demographics, pathology, blood biomarkers, genomics and proteomics, medical imaging gives complementary and interchangeable information.<sup>2</sup>

Although genomic and proteomic technologies provide detailed information about the molecular characteristics of diseased tissues, they can have limited applicability in the clinic and are often invasive procedures associated with high risk and cost to the patient. Non-invasive tissue characterization using medical imaging is a great opportunity to minimize the use of such invasive procedures whilst still obtaining all the information required to treat the patient. Combining medical imaging with other sources, such as pathology, genomics, blood biomarkers and proteomics, could improve individualized treatment selection and monitoring.<sup>2,3,4</sup>

Radiomics is one of the latest innovations in medical imaging analysis. It is a tool that aims to utilize the full potential of medical imaging. In the image the macroscopic (observable with the eye) tumor is defined, which is called gross tumor volume (GTV), either by an experienced radiologist or radiation oncologist or with an automated segmentation method. From the previously defined tumor region, radiomics extracts a large amount of quantitative features by using mathematical algorithms. Radiomics can describe for example tumor intensity, shape and texture. These features can help depict the complex intra-tumoral heterogeneity of the entire tumor, in contrast to invasive biopsies.<sup>2,3,4,5</sup> Radiomics anticipates that these features will convey both prognostic and predictive information for cancer patients.<sup>6</sup>

Maastro clinic (Maastricht, The Netherlands) and Moffitt cancer center (Tampa, United States) collaborate in research of lung- and head-neck tumors. Both institutions share the vision that quantitative features extracted from medical images have the potential to provide improved prognostic and predictive values. Maastro and Moffitt each developed their own radiomics tool. Tumor size, shape and texture features information is obtained by the complex radiomic software.<sup>7</sup> To allow for meaningful comparison of multi-centric radiomic results, interoperability of radiomics has to be ensured. Therefore, differences between these tools must be elucidated.

A recent study by Aerts et al.<sup>8</sup> has been performed on radiomics describing feature extraction in both lung and head/neck cancer patients, this study present a radiomic analysis on 1019 lung and head/neck cancer patients. Another study of Parmar et al.<sup>3</sup> demonstrated that the radiomic features obtained from lung cancer and head/neck cancer patients, cluster in a different way.<sup>3</sup> Disease specific clustering of features are likely influenced by, e.g., tissue localization, cellular phenotype and genetic mutational status. With this study we aim to extend the knowledge, as valuable information could be gathered concerning interoperability of radiomics. We hypothesize that the difference in numerical values of the features between the Maastricht and Moffitt tools may as well be influenced by the disease site. The main research question is: **What is the difference in numerical values for radiomics features for lung and head/neck tumors between the Maastricht and Moffitt software tool and does this difference depend on the disease site?**



## Method

### Research model

We used a retrospective quantitative measurement study to explore the difference in numerical values for radiomics features in lung and head/neck tumors between the Maastricht and Moffitt software tool.

### Data collection

For this study Non-Small Cell Lung Carcinoma (NSCLC) patients (n=52) and head/neck cancer patients (n=52), treated between July 2015 and March 2016 were selected from the Maastricht clinic database. There was no exclusion for gender, age or tumor stage. All patients underwent a CT scan for treatment planning purposes at Maastricht clinic, according to standard clinical scanning protocols. From each of the 104 patients we obtained the treatment planning CT scan and the Radiation Therapy (RT) structure set. The CT scans were generated on a Sensation CT (Siemens), a Sensation open CT (Siemens), or a Biograph True Point 40 Positron emission tomography (PET) -CT (Siemens). The CT Images were reconstructed with an in plane pixel spacing of 0.977 mm x 0.977 mm, with a slice thickness of three mm. The primary gross tumor volume (GTV) was delineated manually for treatment planning purposes on Eclipse (Varian Medical System, Palo Alto, CA, USA).

### Measurements

This study used the standard settings of the Maastricht and Moffitt tools.

The features used for this study are; first order statistics, shape and size based features and gray-level co-occurrence (GLCM) and gray-level run-length (GLRL) textural features. First-order statistics describe the distribution of voxel intensities within the CT image through commonly used, basic metrics. Shape and size based features includes descriptors of the three-dimensional size and shape of the tumor region. Textural features describing patterns or the spatial distribution of voxel intensities, which were calculated from respectively gray level co-occurrence and gray level run-length texture matrices. Determining texture matrix representations requires the voxel intensity values within the voxel of interest to be discretized. See Appendix II and III for the feature definition and the algorithm per feature.<sup>8</sup>

The MAASTRO tool is designed by Ralph Leijenaar and runs on MATLAB R2014a (8.3).<sup>9</sup> The software is able to extract a vast amount of quantitative features. See Appendix II for the definitions and algorithms of the Maastricht features. For each patient the treatment planning CT scan and the RT structure set was loaded into the tool. We selected the GTV of the RT structures and the features we wanted to extract from the GTV; first order statistics, shape and size based and the specified textural features. The outputs of numerical values per feature per patient were exported in Microsoft Excel 2010. Most features are unitless.

Moffitt uses MATLAB R2016a to extract features out of the GTV.<sup>10</sup> See Appendix III for the definitions and algorithms of the Moffitt features. The RT structures of the patients were loaded into MATLAB. The GTV of the RT structures was chosen and a MATLAB script transformed the GTV from a Digital Imaging and Communications in Medicine (DICOM) to a Portable Network Graphics (PNG) format.

The CT scan and the PNG format GTV's were loaded into MATLAB, the implemented code with algorithms extracted the features from the GTV of the CT scan. The outputs of numerical values per feature per patient were exported in Microsoft Excel 2010 (Excel). Most features are unitless.

### Data analysis

The first step in comparing both tools was to identify which radiomic features from Maastricht could be compared with radiomic features of Moffitt. We matched the features by name, algorithm or both. The feature definitions lists in Appendix II (Maastricht) and Appendix III (Moffitt) were used to do the matching. Radiomic features that didn't match were excluded from the study.

The CT scans and RT structures of the lung and the head/neck patients were loaded into both the Maastricht and Moffitt tool to acquire the numerical values of each features of each patient. To quantify the degree of agreement between the Maastricht and Moffitt tool, Lin's Concordance Correlation Coefficient (CCC) was calculated for the numerical values of each feature for both the lung and head/neck patient cohort. The CCC evaluates the degree to which pairs of observations fall on the 45° line through the origin. With a CCC value of < 0,90 the strength of agreement is poor, by 0,90 – 0,95 it is Moderate, by 0,95 – 0,99 it is substantial and by >0,99 it is almost perfect.<sup>11,12,13,14,15,16</sup> We chose the CCC because it has emerged as the best measure of agreement for two methods of measuring in the same continuous variable according to Diederhofen et al.<sup>15</sup>

The CCC's of the lung and head/neck features were compared to each other with the Concor package for the R programming language for which a web interface is available (<http://comparingcorrelations.org/>). The lung and head/neck CCC's are from two independent groups and we used a two-tailed test. With  $p \leq 0,05$  it is statistically plausible that the lung and head/neck CCC are different.<sup>15,17</sup>

To model the relationship between the feature values of Maastricht and Moffitt, we performed a linear regression for the first order statistics features values.<sup>18</sup> See Appendix V for the formula.

### Ethics

Patient records are all anonymous to ensure patient privacy is upheld. Therefore no personal information can be obtained from these images.

## Results

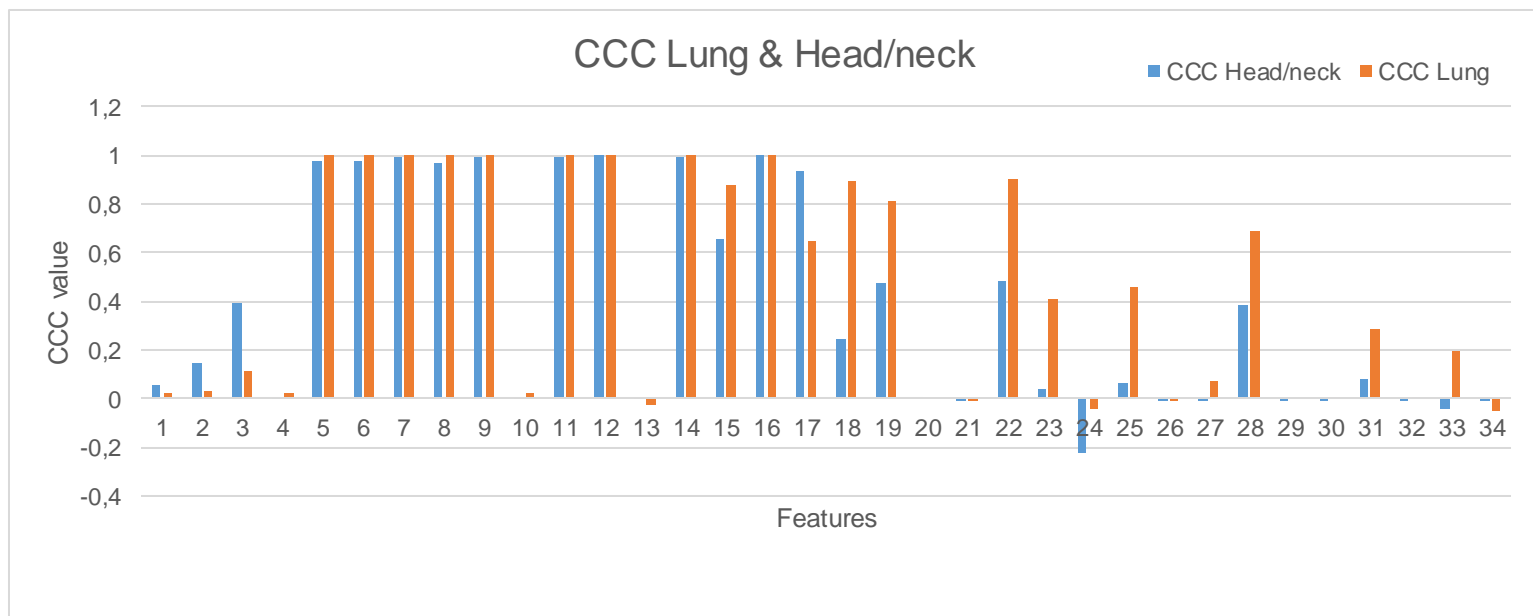
We matched 34 features, including 14 first order statistics, 3 shape and size based features and 17 GLCM and GLRL texture features. See **Table 1** for the names of the matched features. For both tools the numerical values per feature were calculated for 42 lung and 52 head/neck patients. We had to exclude 10 lung patients because the GTV-1 was not defined and therefore feature extraction was not possible.

**Table 1.** Lung and head/neck feature names for Maastricht and Moffitt with the calculated CCC and the Correlation comparison p-value for the two-tailed test of equal CCC values.

#	Feature category	Moffitt features	Maastricht features	CCC Lung	CCC Head/neck	Comparing CCC
1	First order statistics	Energy	Stats_energy	0,019	0,054	0,874
2	First order statistics	Min HU	Stats_min	0,029	0,149	0,573
3	First order statistics	Max HU	Stats_max	0,116	0,396	0,159
4	First order statistics	Mean HU	Stats_mean	0,022	0,001	0,922
5	First order statistics	Variance HU	Stats_var	0,999	0,976	0,000
6	First order statistics	Skewness HU	Stats_skewness	0,999	0,979	0,000
7	First order statistics	Kurtosis HU	Stats_kurtosis	0,999	0,994	0,000
8	First order statistics	Standard Deviation HU	Stats_std	0,999	0,968	0,000
9	First order statistics	Mean Absolute Deviation HU	Stats_md	0,999	0,990	0,000
10	First order statistics	Median HU	Stats_median	0,022	0,001	0,918
11	First order statistics	Range HU	Stats_range	0,999	0,996	0,001
12	First order statistics	Entropy	Stats_entropy	0,999	0,997	0,001
13	First order statistics	RMS HU	Stats_rms	-0,031	0,003	0,874
14	First order statistics	Uniformity HU	Stats_uniformity	0,999	0,994	0,000
15	Shape and size	Connected 3D components	Shape_volumeNumber	0,876	0,658	0,008
16	Shape and size	Volume	Shape_volume	0,999	0,999	1,000
17	Shape and size	Longest Diameter	Shape_maxDiameter3D	0,651	0,932	0,000
18	Texture	Short Run Emphasis (SRE)	RLGL_shortRunEmphasis	0,894	0,243	0,000
19	Texture	Long Run Emphasis (LRE)	RLGL_longRunEmphasis	0,810	0,472	0,004
20	Texture	Gray-Level Nonuniformity (GLN)	RLGL_grayLevelNonuniformity	6,86E-06	1,56E-05	1,00
21	Texture	Run-Length Nonuniformity (RLN)	RLGL_runLengthNonuniformity	-6,24E-06	-2,77E-05	1,00
22	Texture	Run Percentage (RP)	RLGL_runPercentage	0,906	0,486	0,00
23	Texture	Low Gray-Level Run Emphasis (LGRE)	RLGL_lowGrayLevelRunEmphasis	0,411	0,038	0,063
24	Texture	High Gray-Level Run Emphasis (HGRE)	RLGL_highGrayLevelRunEmphasis	-0,042	-0,223	0,389
25	Texture	Short Run Low Gray-Level Emphasis (SRLGE)	RLGL_shortRunLowGrayLevEmpha	0,460	0,061	0,042
26	Texture	Short Run High Gray-Level Emphasis (SRHGE)	RLGL_shortRunHighGrayLevEmpha	-0,000	-0,002	0,993

27	Texture	Long Run Low Gray-Level Emphasis (LRLGE)	RLGL_longRunLowGrayLevEmpha	0,075	-0,007	0,702
28	Texture	Long Run High Gray-Level Emphasis (LRHGE)	RLGL_longRunHighGrayLevEmpha	0,686	0,382	0,041
29	Texture	avgCooccurrence-Energy	GLCM_energy	0,006	-0,007	0,952
30	Texture	avgCooccurrence-CONTRAST	GLCM_contrast	0,005	-0,001	0,978
31	Texture	avgCooccurrence-HOMOGENEITY	GLCM_homogeneity1	0,286	0,082	0,323
32	Texture	avgCooccurrence-MAXPROB	GLCM_maxProb	0,010	-0,001	0,963
33	Texture	avgCooccurrence-ENTROPY	GLCM_entrop2	0,195	-0,046	0,256
34	Texture	avgCooccurrence-SUMMEAN	GLCM_sumAvg	-0,049	-0,004	0,834

CCC = Concordance Correlation Coefficient.



**Figure 1.** The Concordance correlation coefficient (CCC) between the lung features of Maastro and Moffitt and the CCC between head/neck features of Maastro and Moffitt. See **Table 1** for the numbers per feature.

Per feature the CCC was calculated. Most features had a low CCC, in particular textural features e.g. Long Run Emphasis of Moffitt and RLGL\_longRunEmphasis of Maastrro. For the lung cancer cohort, 24 features were found to have a poor CCC(<0.90), 1 feature had a moderate CCC (0.90-0.95) and 9 features had an almost perfect CCC (>0.99). For the head/neck cohort, 24 features were found to have a poor CCC (<0.90), 1 feature had a moderate CCC (0.90-0.95), 3 features had a substantial CCC (0.95-0.99) and 6 features had an almost perfect CCC (>0.99) (**Table 1 & Figure 1**).

First order statistics features (stats\_min, stats\_max, stats\_mean and stats\_median) from Maastrro show differences in the numerical values matched with Moffitt statistic features (min HU, max HU, mean HU and Median HU). There is a difference of + 1000 HU, in both the lung and head/neck features (**Table 2**). **Figure 2** shows the scatterplots of max HU and stats\_max for the lung and head/neck cohort. The matched features Energy with stats\_energery and RMS HU with stats\_rms have a low CCC and  $R^2$  compared to the other features of first order statistics (except for the above four matched features).

**Table 2.** Intercept, slope and  $R^2$  calculated for the first order statistic features of head/neck (A) and lung (B).

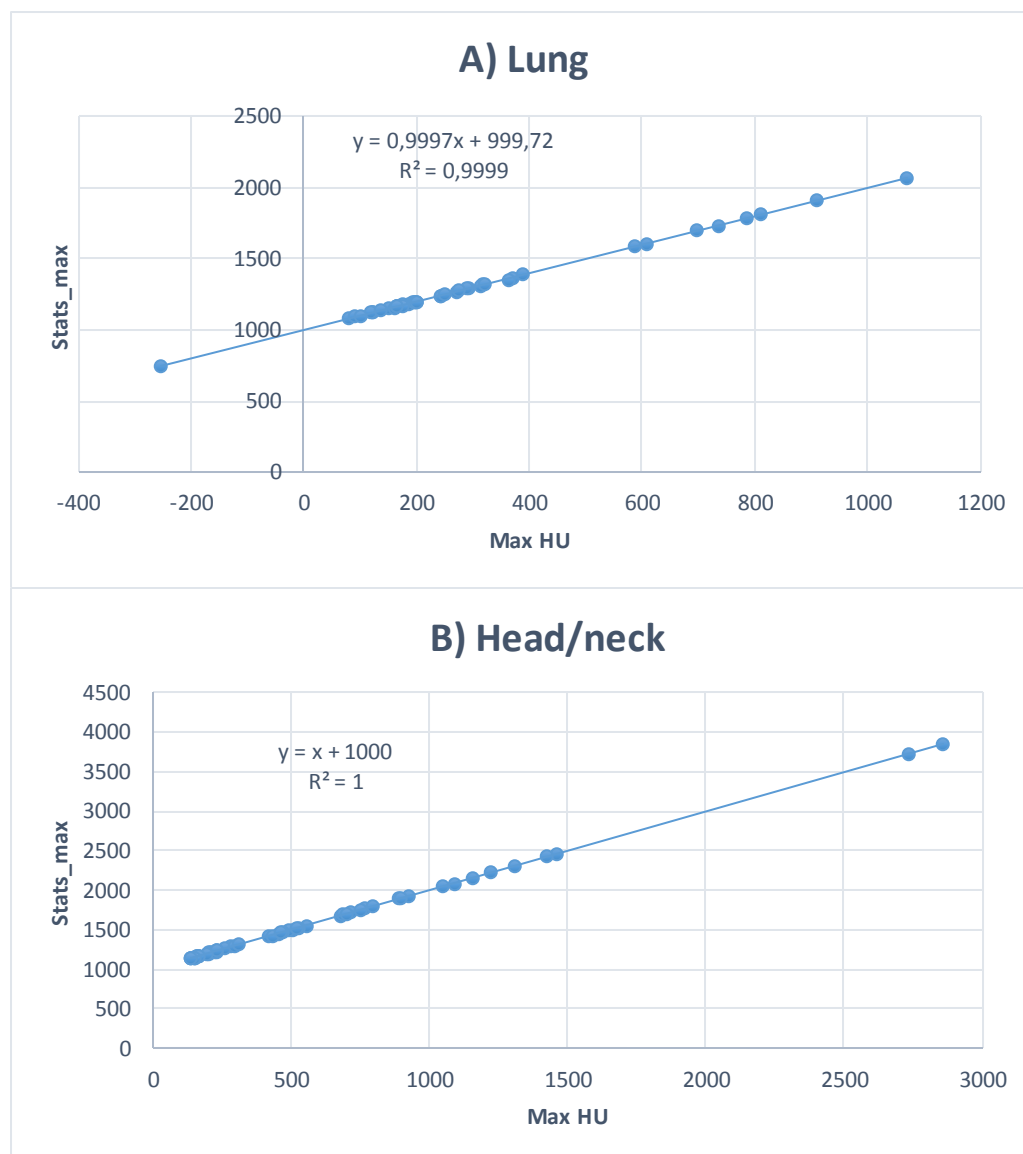
A)

Head/neck Moffitt features	Head/neck Maastrro features	Intercept	slope	$R^2$
Energy	Stats_energery	2E+09	27,42	0,93
Min HU	Stats_min	984,38	0,95	0,96
Max HU	Stats_max	1000,00	1,00	1,00
Mean HU	Stats_mean	1001,80	0,99	0,98
Variance HU	Stats_var	-312,72	0,96	0,95
Skewness HU	Stats_skewness	0,07	0,97	0,96
Kurtosis HU	Stats_kurtosis	0,22	0,99	0,99
Standard Deviation HU	Stats_std	1,45	0,94	0,94
Mean Absolute Deviation HU	Stats_md	-0,60	0,99	0,98
Median HU	Stats_median	999,80	1,00	1,00
Range HU	Stats_range	-2,13	0,99	0,99
Entropy	Stats_entropy	0,04	0,99	0,99
RMS HU	Stats_rms	1018,40	0,48	0,66
Uniformity HU	Stats_uniformity	0,00	1,01	0,99

B)

Lung Moffitt features	Lung Maastrro features	Intercept	Slope	$R^2$
Energy	Stats_energery	1,00E+10	14,61	0,21
Min HU	Stats_min	965,24	0,95	0,99
Max HU	Stats_max	999,72	0,99	0,99
Mean HU	Stats_mean	999,87	0,99	0,99
Variance HU	Stats_var	49,58	0,99	0,99

Skewness HU	Stats_skewness	0,00	0,99	0,99
Kurtosis HU	Stats_kurtosis	-0,02	1,00	0,99
Standard Deviation HU	Stats_std	0,08	0,99	0,99
Mean Absolute Deviation HU	Stats_md	-0,02	0,99	0,99
Median HU	Stats_median	999,95	0,99	0,99
Range HU	Stats_range	8,89	0,99	0,99
Entropy	Stats_entropy	0,00	1,00	0,99
RMS HU	Stats_rms	1120,20	-0,86	0,77
Uniformity HU	Stats_uniformity	0,00	0,99	0,99



**Figure 2.** Scatterplots for lung (A) and head/neck (B) for the features Max HU of Moffitt and Stats\_max of Maastro.

Moffitt calculates total 3D diameter and Maastro calculates 2D transversal, 2D coronal, 2D sagittal and full 3D diameter. Comparing the Moffitt feature Longest Diameter with the Full 3D feature of Maastro

(matched by name) showed that the CCC for lung is 0,651 (poor) and for head/neck 0,932 (moderate). For all the CCC values of all diameter features see Appendix VI.

The comparison between the CCC of lung and the CCC of head/neck showed that 19 features have an equal CCC ( $p \geq 0,05$ ) and 15 features do not have an equal CCC ( $p \leq 0,05$ ) (**Table 1**).

## Discussion

The purpose of this study was to evaluate differences in numerical values for radiomic features for lung and head/neck tumors between the Maastricht and Moffitt software tool and to see if the differences depend on the disease site.

In this study 34 features could be matched and were included. From the 34 features assessed in this study, most features show a poor CCC ( $<0.90$ ) (24 for lung and 24 for head/neck). The CCC's between lung and head/neck are overall different in strength. Lung has 9 almost perfect CCC's ( $>0.99$ ) and head/neck has 6 almost perfect CCC's ( $>0.99$ ). Generally, the numerical output of both tools is not comparable. This suggests that standardization is needed, which should be reflected in the software implementation.

The method described by Diedenhofen et al.<sup>15</sup> was used to compare the strength of CCC's between the two groups of patients (lung and head/neck). We found that from the 34 features, 19 features for lung and head/neck were not statistically significantly different ( $p \geq 0.05$ ). There is no evidence that for these 19 features the disease site has an impact on the difference between the output of the Maastricht and Moffitt radiomics implementations. The remaining 15 features did not have an equal CCC for the lung and head/neck cohort, which suggest that for these features the difference between Maastricht and Moffitt might be disease specific.

Image preprocessing is performed differently in both software implementations. Maastricht adds +1000 HU to each image to prevent negative values. The first order statistic features (Variance HU/Stats\_var, Skewness HU/Stats\_skewness, Kurtosis HU/Stats\_kurtosis, Standard Deviation HU/Stats\_std, Mean Absolute Deviation HU/Stats\_md, Range HU/Stats\_range, Entropy/Stats\_entropy and Uniformity HU/Stats\_uniformity) are not affected by this offset and almost equal between Maastricht and Moffitt ( $R^2 \sim 1$ ). Some feature values (stats\_min, stats\_max, stats\_mean and stats\_median) present low CCC values, but can easily be corrected for the offset in HU by subtracting 1000 HU, making them comparable to the feature values obtained by the Moffitt software. The CCC, unlike the  $R^2$  value, evaluates not only the linearity but also if the angle falls on the  $45^\circ$  line through the origin. This explains why for these features (stats\_min, stats\_max, stats\_mean, and stats\_median) there was a bad correlation between the Maastricht and Moffitt values but a high linear regression  $R^2$ . Features stats\_energy and stats\_rms cannot be corrected in the same way due to the quadratic term in the feature's mathematical definition. Additionally, we calculated the radiomic features without the +1000 HU offset for the Maastricht tool for one example case. Shape and textural features are not affected by the addition of 1000 HU, they are invariant to this offset in image intensity values.

This study is one of the first that explores the difference of radiomic tools for two institutes. Berthon et al.<sup>19</sup> did a study in texture analysis for Positron Emission Tomography in oesophageal cancer, this study shows the importance of comparing different methods and standardization in radiomics.<sup>19</sup>



We had to exclude 10 lung patients, since they did not have a segmented GTV, meaning that feature extraction was not possible. The features "Connected 3D components" of Moffitt and "Shape\_volumeNumber" of Maastricht had a poor CCC but the numerical values of these features are all similar except for one patient. This feature does not have a normal distribution. Therefore CCC is an unsuitable method to compare these features.

Both institutes developed their own radiomics implementations. As there is no defined standard, this inherently can lead to differences in numerical output, even for features that have the same name and mathematical definitions. This study shows that preprocessing and features calculations differ for several features, resulting in different numerical outputs. For standardization of radiomics features it is important to understand these differences. Parmar et al.<sup>3</sup> demonstrated a difference in clustering between lung and head/neck. With the comparison of the CCC this study found 15 features that do not have an equal CCC ( $p \leq 0.05$ ) for the lung and head/neck cohort between Maastricht and Moffitt. This suggests that there are differences between the different disease sites and this is important for further studies. Standardization is critical, because this can enable optimization of specific disease site analysis and treatment.

To optimize tools it is important that the results obtained can be directly compared across different institutes. This study showed that most features had a substantial difference in numerical output in both tools, partly due to preprocessing (+1000 HU) in first order statistics. For shape and texture features the implementations of the features and/or the settings in the tool are different in both institutes. Standardization is required to ensure interoperability of radiomics. We therefore recommend that further research identifies the details of the difference in implementations for features with the same name and definition. Here, standardization is needed for validation of multi-centric studies that use different radiomic software. We furthermore observed that not all institutes have the same features implemented in their software. For example Maastricht may have feature X, while Moffitt does not. On the other hand Moffitt may have feature Y but Maastricht does not have this feature. It is possible that feature X is better than feature Y as a biomarker for a specific situation, disease or diagnosis while Y is maybe better in another case. In this case multi-centric studies are only possible using either the same software, or by providing all necessary details on these institute specific features to implement them in other softwares, ensuring interoperability."

## Conclusion

We found a substantial difference in feature values for both lung and head and neck cancer patients between the Maastricht and Moffitt radiomics software tool. Some first order statistic features have a scaling difference of +1000 HU for the Maastricht features compared to the Moffitt features. This can be corrected afterwards to make them comparable between both tools. Generally there is a poor agreement in CCC's between both tools. We found that for 15 features the difference between both tools may depend on the disease site. To ensure the optimal use of radiomics tools for lung and head/neck cancer in the future, standardization is required.

## References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA a cancer J Clin* [Internet]. 2015;65(2):87–108. Available from: <http://onlinelibrary.wiley.com/doi/10.3322/caac.21262/abstract>
2. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6.
3. Parmar C, Leijenaar RTH, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci Rep* [Internet]. Nature Publishing Group; 2015;5:1–10. Available from: <http://dx.doi.org/10.1038/srep11044>
4. Panth KM, Leijenaar RTH, Carvalho S, Lieuwes NG, Yaromina A, Dubois L, et al. Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. *Radiother Oncol* [Internet]. Elsevier Ireland Ltd; 2015;116(3):462–6. Available from: <http://dx.doi.org/10.1016/j.radonc.2015.06.013>
5. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. *Front Oncol* [Internet]. 2015;5(December):272. Available from: <http://journal.frontiersin.org/article/10.3389/fonc.2015.00272/abstract>
6. Parmar C, Velazquez ER, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*. 2014;9(7):1–8.
7. Moffitt HL. Radiomics of NSCLC H . Lee Moffitt Cancer Center and Research Institute. :1–39.
8. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
9. Leijenaar RT. Oncoradiomics [Internet]. Available from: <http://www.oncoradiomics.com/>
10. MATLAB. 2016; Available from: <http://nl.mathworks.com/products/matlab/>
11. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl Oncol* [Internet]. 2014;7(1):72–87. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1936523314800107>
12. Lin L-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45. 1989;
13. Lin L-K. A note on the concordance correlation coefficient. *Biometrics* 54. 2000;324–5.
14. McBride G. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Rep. 2005;HAM2005-06:14.
15. Diedenhofen B, Musch J. Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS One*. 2015;10(4):1–12.
16. Concordance correlation coefficient [Internet]. [cited 2016 May 6]. Available from: [https://en.wikipedia.org/wiki/Concordance\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Concordance_correlation_coefficient)
17. Diedenhofen B. cocor comparing correlations. Available from: <http://www.comparingcorrelations.org/>
18. Linear regression [Internet]. Available from: [https://en.wikipedia.org/wiki/Linear\\_regression](https://en.wikipedia.org/wiki/Linear_regression)
19. Berthon B. Benchmarking texture analysis for PET in oesophageal cancer. *ESTRO 35 Abstract*

book [Internet]. 2016. p. 1020. Available from: <http://cld.bz/BdkOkGe#907/z>

## Appendices

### Appendix I

*Copyright / Disclaimer MAASTRO data*

*Deze informatie en/of beeldmateriaal is eigendom van MAASTRO CLINIC, alle rechten zijn voorbehouden aan MAASTRO CLINIC. Niets van deze informatie en/of beeldmateriaal mag worden verveelvoudigd en/of openbaar gemaakt via internet, door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van MAASTRO CLINIC. Aan eventuele onjuistheden of onvolledigheden kunnen geen rechten ontleend worden.*

*MAASTRO CLINIC kan evenmin aansprakelijk gesteld worden indien zich, door het gebruik van deze informatie en/of beeldmateriaal, bij de raadpleger schade zou voordoen van welke aard ook.*

*This information and/or images are the property of MAASTRO CLINIC. All rights are reserved for MAASTRO CLINIC. None of this information and/or images may be reproduced, stored in a retrieval system, transmitted or utilized in any form or by any means, electronic, mechanical, photocopying, microfilm, internet or otherwise, without permission in writing from MAASTRO CLINIC. MAASTRO CLINIC cannot be held responsible should a person making use of this information and/or images suffer loss or damage of whatever type or extent.*

## Supplement I: Radiomics Features

In this study we explore a feature-based approach to extract and quantify meaningful and reliable information from CT images. In this section we describe in detail the different groups of imaging traits assessed in our study, that were used to derive a prognostic model in non-small cell lung carcinoma. We evaluated a total number of 440 CT imaging features, which are divided in four groups as follows:

*Group 1. First order statistics*

*Group 2. Shape and size based features*

*Group 3. Textural features*

*Group 4. Wavelet features*

### **Group 1. First order statistics**

First-order statistics describe the distribution of voxel intensities within the CT image through commonly used and basic metrics. Let  $X$  denote the three dimensional image matrix with  $N$  voxels and  $P$  the first order histogram with  $N_l$  discrete intensity levels. The following first order statistics were extracted:

#### **1.1. Energy:**

$$energy = \sum_i^N X(i)^2$$

#### **1.2. Entropy:**

$$entropy = \sum_{i=1}^{N_l} P(i) \log_2 P(i)$$

#### **1.3. Kurtosis:**

$$kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (X(i) - \bar{X})^4}{\left( \sqrt{\frac{1}{N} \sum_{i=1}^N (X(i) - \bar{X})^2} \right)^2}$$

where  $\bar{X}$  is the mean of  $X$ .

#### **1.4. Maximum:**

The maximum intensity value of  $X$ .

**1.5. Mean:**

$$mean = \frac{1}{N} \sum_i^N X(i)$$

**1.6. Mean absolute deviation:**

The mean of the absolute deviations of all voxel intensities around the mean intensity value.

**1.7. Median:**

The median intensity value of  $X$ .

**1.8. Minimum:**

The minimum intensity value of  $X$ .

**1.9. Range:**

The range of intensity values of  $X$ .

**1.10. Root mean square (RMS):**

$$RMS = \sqrt{\frac{\sum_i^N X(i)^2}{N}}$$

**1.11. Skewness:**

$$skewness = \frac{\frac{1}{N} \sum_{i=1}^N (X(i) - \bar{X})^3}{\left( \sqrt{\frac{1}{N} \sum_{i=1}^N (X(i) - \bar{X})^2} \right)^3}$$

where  $\bar{X}$  is the mean of  $X$ .

**1.12. Standard deviation:**

$$standard\ deviation = \left( \frac{1}{N-1} \sum_{i=1}^N (X(i) - \bar{X})^2 \right)^{1/2}$$

where  $\bar{X}$  is the mean of  $X$ .

**1.13. Uniformity:**

$$uniformity = \sum_{i=1}^{N_l} P(i)^2$$

**1.14. Variance:**

$$variance = \frac{1}{N-1} \sum_{i=1}^N (X(i) - \bar{X})^2$$

where  $\bar{X}$  is the mean of  $X$ .

The standard deviation, variance and mean absolute deviation are measures of the histogram dispersion, that is, a measure of how much the gray levels differ from the mean. The variance, skewness and kurtosis are the most frequently used central moments. The skewness measures the degree of histogram asymmetry around the mean, and kurtosis is a measure of the histogram sharpness. As measures of histogram randomness we computed the uniformity and entropy of the image histogram.

## **Group 2. Shape and size based features**

In this group of features we included descriptors of the three-dimensional size and shape of the tumor region. Let in the following definitions  $V$  denote the volume and  $A$  the surface area of the volume of interest. We determined the following shape and size based features:

### **2.1. Compactness 1:**

$$compactness\ 1 = \frac{V}{\sqrt{\pi}A^{\frac{2}{3}}}$$

### **2.2. Compactness 2:**

$$compactness\ 2 = 36\pi \frac{V^2}{A^3}$$

### **2.3. Maximum 3D diameter:**

The maximum three-dimensional tumor diameter is measured as the largest pairwise Euclidean distance, between voxels on the surface of the tumor volume.

### **2.4. Spherical disproportion:**

$$spherical\ disproportion = \frac{A}{4\pi R^2}$$

Where  $R$  is the radius of a sphere with the same volume as the tumor.

### **2.5. Sphericity:**

$$sphericity = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$$

### **2.6. Surface area:**

The surface area is calculated by triangulation (i.e. dividing the surface into connected triangles) and is defined as:



$$A = \sum_{i=1}^N \frac{1}{2} |a_i b_i \times a_i c_i|$$

Where  $N$  is the total number of triangles covering the surface and  $a$ ,  $b$  and  $c$  are edge vectors of the triangles.

**2.7. Surface to volume ratio:**

$$\text{surface to volume ratio} = \frac{A}{V}$$

**2.8. Volume:**

The volume ( $V$ ) of the tumor is determined by counting the number of pixels in the tumor region and multiplying this value by the voxel size.

The maximum 3D diameter, surface area and volume provide information on the size of the lesion.

Measures of compactness, spherical disproportion, sphericity and the surface to volume ratio describe how spherical, rounded, or elongated the shape of the tumor is.

### Group 3. Textural features

The features shown above that resulted from group 1 (first-order statistics) provide information related to the gray-level distribution of the image; however they do not provide any information regarding the relative position of the various gray levels over the image. In this group we therefore included textural features describing patterns or the spatial distribution of voxel intensities, which were calculated from respectively gray level co-occurrence (GLCM), gray level run-length (GLRLM) and gray level size-zone (GLSZM) texture matrices. Determining texture matrix representations requires the voxel intensity values within the VOI to be discretized. Voxel intensities were therefore resampled into equally spaced bins using a bin-width of 25 Hounsfield Units. This discretization step not only reduces image noise, but also normalizes intensities across all patients, allowing for a direct comparison of all calculated textural features between patients. Texture matrices were determined considering 26-connected voxels (i.e. voxels were considered to be neighbors in all 13 directions in three dimensions).

#### Gray-Level Co-Occurrence Matrix based features

A GLCM is defined as  $P(i, j; \delta, \alpha)$ , a matrix with size  $N_g \times N_g$  describing the second-order joint probability function of an image, where the  $(i, j)$ th element represents the number of times the combination of intensity levels  $i$  and  $j$  occur in two pixels in the image, that are separated by a distance of  $\delta$  pixels in direction  $\alpha$ , and  $N_g$  is the number of discrete gray level intensities. As a two dimensional example, let the following matrix represent a 5x5 image, having 5 discrete gray levels:

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

For distance  $\delta = 1$  (considering pixels with a distance of 1 pixel from each other) in direction  $\alpha = 0$ , where 0 degrees is the horizontal direction, the following GLCM is obtained:

$$P(1,0) = \begin{bmatrix} 3 & 3 & 2 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \end{bmatrix}$$

In this study, distance  $\delta$  was set to 1 and direction  $\alpha$  to each of the 13 directions in three dimensions, yielding a total of 13 gray level co-occurrence matrices for each 3D image. From these gray-level co-occurrence matrices, several textural features are derived. Each 3D gray level co-occurrence based feature was then calculated as the mean of the feature calculations for each of the 13 directions.

Let:

$P(i, j)$  be the co-occurrence matrix for an arbitrary  $\delta$  and  $\alpha$ ,

$N_g$  be the number of discrete intensity levels in the image,

$\mu$  be the mean of  $P(i, j)$ ,

$p_x(i) = \sum_{j=1}^{N_g} P(i, j)$  be the marginal row probabilities,

$p_y(i) = \sum_{i=1}^{N_g} P(i, j)$  be the marginal column probabilities,

$\mu_x$  be the mean of  $p_x$ ,

$\mu_y$  be the mean of  $p_y$ ,

$\sigma_x$  be the standard deviation of  $p_x$ ,

$\sigma_y$  be the standard deviation of  $p_y$ ,

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j), i + j = k, k = 2, 3, \dots, 2N_g,$

$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j), |i - j| = k, k = 0, 1, \dots, N_g - 1,$

$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2[p_x(i)]$  be the entropy of  $p_x$ ,

$HY = -\sum_{i=1}^{N_g} p_y(i) \log_2[p_y(i)]$  be the entropy of  $p_y$ ,

$H = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log_2[P(i, j)]$  be the entropy of  $P(i, j)$ ,

$HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) \log(p_x(i)p_y(j)),$

$HXY2 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log(p_x(i)p_y(j)).$

### 3.1. Autocorrelation:

$$autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP(i, j)$$

### 3.2. Cluster Prominence:

$$cluster \text{ prominence} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x(i) - \mu_y(j)]^4 P(i, j)$$

### 3.3. Cluster Shade:

$$cluster \text{ shade} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x(i) - \mu_y(j)]^3 P(i, j)$$

### 3.4. Cluster Tendency:

$$cluster \text{ tendency} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i + j - \mu_x(i) - \mu_y(j)]^2 P(i, j)$$

### 3.5. Contrast:

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j|^2 P(i, j)$$

**3.6. Correlation:**

$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP(i,j) - \mu_i(i)\mu_j(j)}{\sigma_x(i)\sigma_y(j)}$$

**3.7. Difference entropy:**

$$difference\ entropy = \sum_{i=0}^{N_g-1} P_{x-y}(i) \log_2[P_{x-y}(i)]$$

**3.8. Dissimilarity:**

$$dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j|P(i,j)$$

**3.9. Energy:**

$$energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [P(i,j)]^2$$

**3.10. Entropy (H):**

$$entropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \log_2[P(i,j)]$$

**3.11. Homogeneity 1:**

$$homogeneity\ 1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1 + |i-j|}$$

**3.12. Homogeneity 2:**

$$homogeneity\ 2 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1 + |i-j|^2}$$

**3.13. Informational measure of correlation 1 (IMC1):**

$$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

**3.14. Informational measure of correlation 2 (IMC2):**

$$IMC2 = \sqrt{1 - e^{-2(HXY2 - HXY)}}$$

**3.15. Inverse Difference Moment Normalized (IDMN):**

$$IDMN = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1 + \left(\frac{|i-j|^2}{N^2}\right)}$$

**3.16. Inverse Difference Normalized (IDN):**

$$IDN = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1 + \left(\frac{|i-j|}{N}\right)}$$

**3.17. Inverse variance:**

$$inverse\ variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{|i-j|^2}, i \neq j$$

**3.18. Maximum Probability:**

$$maximum\ probability = \max\{P(i,j)\}$$

**3.19. Sum average:**

$$sum\ average = \sum_{i=2}^{2N_g} [iP_{x+y}(i)]$$

**3.20. Sum entropy:**

$$sum\ entropy = - \sum_{i=2}^{2N_g} P_{x+y}(i) \log_2 [P_{x+y}(i)]$$

**3.21. Sum variance:**

$$sum\ variance = \sum_{i=2}^{2N_g} (i - SE)^2 P_{x+y}(i)$$

**3.22. Variance:**

$$variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P(i,j)$$

*Gray-Level Run-Length matrix based features*

Run length metrics quantify gray level runs in an image. A gray level run is defined as the length in number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix  $p(i,j|\theta)$ , the  $(i,j)$ th element describes the number of times  $j$  a gray level  $i$  appears consecutively in the direction specified by  $\theta$ , and  $N_g$  is the number of discrete gray level intensities. As a two dimensional example, consider the following 5x5 image, with 5 discrete gray levels:

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

The GLRL matrix for  $\theta = 0$ , where 0 degrees is the horizontal direction, then becomes:

$$p(0) = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In this study, a GLRL matrix was computed for every of the 13 directions in three dimensions, from which the below textural features were derived. Each 3D GLRL feature was then calculated as the mean of the feature values for each of the 13 directions.

Let:

$p(i, j|\theta)$  be the  $(i, j)$ th entry in the given run-length matrix  $p$  for a direction  $\theta$ ,

$N_g$  the number of discrete intensity values in the image,

$N_r$  the number of different run lengths,

$N_p$  the number of voxels in the image.

### 3.23. Short Run Emphasis (SRE)

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i, j|\theta)}{j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)}$$

### 3.24. Long Run Emphasis (LRE)

$$LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 p(i, j|\theta)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)}$$

### 3.25. Gray Level Non-Uniformity (GLN)

$$GLN = \frac{\sum_{i=1}^{N_g} \left[ \sum_{j=1}^{N_r} p(i, j|\theta) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)}$$

### 3.26. Run Length Non-Uniformity (RLN)

$$RLN = \frac{\sum_{j=1}^{N_r} \left[ \sum_{i=1}^{N_g} p(i, j|\theta) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i, j|\theta)}$$

### 3.27. Run Percentage (RP)

$$RP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i, j|\theta)}{N_p}$$

### 3.28. Low Gray Level Run Emphasis (LGLRE)

$$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta)}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

**3.29. High Gray Level Run Emphasis (HGLRE)**

$$HGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 p(i,j|\theta)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

**3.30. Short Run Low Gray Level Emphasis (SRLGLE)**

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta)}{i^2 j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

**3.31. Short Run High Gray Level Emphasis (SRHGLE)**

$$SRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta) i^2}{j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

**3.32. Long Run Low Gray Level Emphasis (LRLGLE)**

$$LRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta) j^2}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

**3.33. Long Run High Gray Level Emphasis (LRHGLE)**

$$LRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta) i^2 j^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

*Gray-Level size-zone matrix based features*

A gray level size-zone matrix describes the amount of homogeneous connected areas within the tumor volume of a certain size and intensity. In a gray level size-zone matrix  $p(i,j)$ , the  $(i,j)$ th element describes the number of times a homogeneous connected region with of size  $j$ , with intensity  $i$  appears. As a two dimensional example, consider the following 5x5 image, with 5 discrete gray levels:

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

The gray level size-zone matrix  $p(i,j)$  then becomes:

$$p = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let:

$p(i, j)$  be the  $(i, j)$ th entry in the given size-zone matrix  $p$ ,

$N_g$  the number of discrete intensity values in the image,

$N_z$  the size of the largest, homogeneous region in the volume of interest,

$N_a$  the number homogeneous areas in the image.

### 3.34. Small area Emphasis (SAE)

$$SAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[ \frac{p(i, j)}{j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i, j)}$$

### 3.35. Large area Emphasis (LAE)

$$LAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} j^2 p(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i, j)}$$

### 3.36. Intensity variability (IV)

$$IV = \frac{\sum_{i=1}^{N_g} \left[ \sum_{j=1}^{N_z} p(i, j) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i, j)}$$

### 3.37. Size-zone variability (SZV)

$$SZV = \frac{\sum_{j=1}^{N_z} \left[ \sum_{i=1}^{N_g} p(i, j) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i, j)}$$

### 3.38. Zone Percentage (ZP)

$$ZP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \frac{p(i, j)}{N_a}$$

### 3.39. Low intensity Emphasis (LIE)



$$LIE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[ \frac{p(i,j)}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

**3.40. High intensity Emphasis (HIE)**

$$HIE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} i^2 p(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

**3.41. Low intensity small area Emphasis (LISAE)**

$$LISAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[ \frac{p(i,j)}{i^2 j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

**3.42. High intensity small area Emphasis (HISAE)**

$$HISAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[ \frac{p(i,j) i^2}{j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

**3.43. Low intensity large area Emphasis (LILAE)**

$$LILAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[ \frac{p(i,j) j^2}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

**3.44. High intensity large area Emphasis (HILAE)**

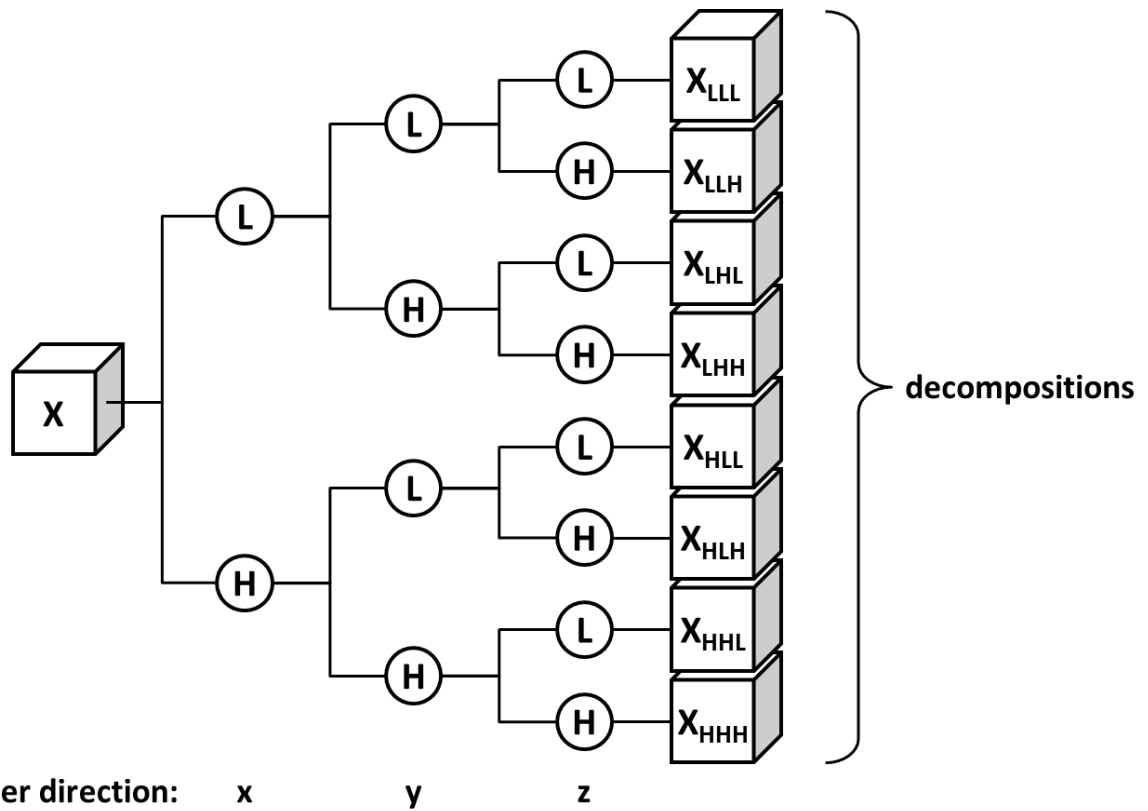
$$HILAE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j) i^2 j^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j)}$$

#### **Group 4. Wavelet features: first order statistics and texture of wavelet decompositions**

Wavelet transform effectively decouples textural information by decomposing the original image, in a similar manner as Fourier analysis, in low –and high-frequencies. In this study a discrete, one-level and undecimated three dimensional wavelet transform was applied to each CT image, which decomposes the original image  $X$  into 8 decompositions. Consider  $L$  and  $H$  to be a low-pass (i.e. a scaling) and, respectively, a high-pass (i.e. a wavelet) function, and the wavelet decompositions of  $X$  to be labeled as  $X_{LLL}, X_{LLH}, X_{LHL}, X_{LHH}, X_{HLL}, X_{HLH}, X_{HHL}$  and  $X_{HHH}$ . For example,  $X_{LLH}$  is then interpreted as the high-pass sub band, resulting from directional filtering of  $X$  with a low-pass filter along x-direction, a low pas filter along y-direction and a high-pass filter along z-direction and is constructed as:

$$X_{LLH}(i, j, k) = \sum_{p=1}^{N_L} \sum_{q=1}^{N_L} \sum_{r=1}^{N_H} L(p)L(q)H(r)X(i+p, j+q, k+r)$$

Where  $N_L$  is the length of filter  $L$  and  $N_H$  is the length of filter  $H$ . The other decompositions are constructed in a similar manner, applying their respective ordering of low or high-pass filtering in x, y and z-direction. Wavelet decomposition of the image  $X$  is schematically depicted in **Figure 1**. Since the applied wavelet decomposition is undecimated, the size of each decomposition is equal to the original image and each decomposition is shift invariant. Because of these properties, the original tumor delineation of the gross tumor volume (GTV) can be applied directly to the decompositions after wavelet transform. In this study “Coiflet 1” wavelet was applied on the original CT images. For each decomposition we computed the first order statistics as described in Group 1 and the textural features as described in Group 3.



**Supplement I Figure 1:** Schematic of the undecimated three dimensional wavelet transform applied to each CT image. The original image  $X$  is decomposed into 8 decompositions, by directional low-pass (i.e. a scaling) and high-pass (i.e. a wavelet) filtering:  $X_{LLL}$ ,  $X_{LLH}$ ,  $X_{LHL}$ ,  $X_{LHH}$ ,  $X_{HLL}$ ,  $X_{HLH}$ ,  $X_{HHL}$  and  $X_{HHH}$ .

## Image Feature List – Version 1.0

No	Feature	Description	Calculation	Range
<b>3D features</b>				
1	Border Length	<p>The border length of an image object is defined as the sum of edges of the image object that are shared with other image objects or are situated on the edge of the entire scene.</p> <p>The border length of a 3D image object is the sum of border lengths of all image object slices multiplied by the spatial distance between the slices.</p> <p>For torus and image objects with holes the border length sums the inner and outer border</p>	$b_v = \left( \sum_{n=1}^{\#(\text{slices})} b_v(\text{Slice}) \right) * u_{\text{slices}} + b_v(Z)$ <p>Parameters</p> <p><math>b_v</math>: border length of image object <math>v</math></p> <p><math>b_v(\text{slice})</math>: border length of image object slice</p> <p><math>b_v(z)</math>: border length of image object in z-direction</p> <p><math>u_{\text{slices}}</math>: spatial distance between slices in the coordinate system unit</p>	[0, ∞]
2.	Length	The length of an image object is the largest of three eigenvalues of a rectangular 3D space that is defined by		[0, ∞]

		the same volume as the image object and the same proportions of eigenvalues as the image object.		
		The length of an image object can be $\leq$ the largest of dimensions of the smallest rectangular 3D space enclosing the image object.		
3.	Thickness	<p>The thickness of an image object is the smallest of three eigenvalues of a rectangular 3D space that is defined by the same volume as the image object and the same proportions of eigenvalues as the image object</p> <p>The thickness of an image object can be <math>\leq</math> than the smallest of dimensions of the smallest rectangular 3D space enclosing the image object</p>		$[0, \infty]$
4.	Width	<p>The width of an image object is the middle of three eigenvalues of a rectangular 3D space that is defined by the same volume as the image object and the same proportions of eigenvalues as the image object.</p> <p>The width of an image object can be smaller or equal than the middle of dimensions of the smallest rectangular 3D space enclosing the image object.</p>		$[0, \infty]$
5.	Length/Thickness	The length-to-thickness ratio of an image object.	Length/Thickness	$[0, \infty]$
6.	Length/Width	The length-to-width ratio of an image object.	Length/Width	$[0, \infty]$
7.	Number of Pixels	Number of pixels forming an image object.		$[0, \text{scene size}]$
8.	Volume	The number of voxels forming an image object rescaled by using unit information for x	$V_v = \#P_v \cdot u^2 \cdot u_{\text{slices}}$	$[0, \text{scene size}]$

and y coordinates and distance information between slices.

Vv: volume of image object v

#Pv: total number of voxels contained in Pv

u: size of a slice pixel in the coordinate system unit

Uslices: spatial distance between slices in the coordinate system unit

## 9. Asymmetry

The more longish an image object, the more asymmetric it is. The feature value increases

with the asymmetry.

The asymmetry is calculated from the ratio between the smallest and largest

eigenvalues of the image object.

$$1 - \frac{\sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}}}$$

[0, 1]

$\lambda_{min}$ : minimal eigenvalue

$\lambda_{max}$ : maximal eigenvalue

## 10. Border Index

The more rough or

jagged an image object is, the higher its border index.

Similar to Shape index feature, but border index uses a rectangular approximation

instead of a square. The smallest rectangle enclosing the image object is created. The

border index is then calculated as the ratio of the Border length feature of the image object to the border length of this smallest enclosing rectangle.

Expression:

$$bv/2(lv+Wv)$$

[1,  $\infty$ ]

1 = ideal.

bv: image object border length

lv: length of an image object v

ww : width of an image object v

11.	Compactness	<p>A figure for the compactness of a 3D image object is calculated by a scaled product of its three eigenvalues <math>2\lambda_1, 2\lambda_2, 2\lambda_3</math> divided by the number of its pixel/voxel. We include a factor of 2 with each eigenvalue, since <math>\lambda_i</math>*eigenvectors represent otherwise half axes of an ellipsoid defined by its covariance matrix. The chosen approach thus provides an estimate of a cuboid occupied by the object.</p>	$2 \lambda_1^2 \lambda_2^2 \lambda_3 / V_v$ <p><math>\lambda_1</math>: eigenvalue 1 of a 3D image object v  <math>\lambda_2</math>: eigenvalue 2 of a 3D image object v  <math>\lambda_3</math>: eigenvalue 3 of a 3D image object v  <math>V_v</math>: volume of image object v</p>	<p>[0, <math>\infty</math>]  1 = ideal.</p>
12.	Density	<p>The Density feature describes the spatial distribution of the pixels of an image object.</p> <p>The ideal compact shape on a pixel raster is the cube. The more the shape of an image object is like a cube, the higher its density. The more the shape of an image object is like a filament, the lower its density.</p> <p>It is calculated by the edge of the volume fitted cube divided by the fitted sphere radius.</p>	$\frac{\sqrt[3]{V_v}}{\sqrt{\text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)}}$ <p><math>V_v</math>: volume of image object v  <math>\sqrt[3]{V}</math>: edge of the volume fitted cube  <math>\sqrt{\text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)}</math>: radius of the fitted sphere</p>	<p>[0, depended on shape of image object]</p>
13.	Elliptic Fit	<p>It describes how well an image object fits into an ellipsoid of similar size and proportions.</p> <p>While 0 indicates no fit, 1 indicates for a complete fitting image object.</p> <p>The calculation is based on an ellipsoid with the same volume as the considered image object. The proportions of the ellipsoid are equal to the proportions of the length to</p>	$\varphi = 2 \cdot \#\{x, y, z\} \in P_v : \varepsilon_v(X, Y, Z) \leq 1\} / \#P_v - 1$ <p><math>\varepsilon_v(x, y, z)</math>: elliptic distance at a pixel (x,y,z)  <math>P_v</math>: set of pixels of an image object v  <math>\#P_v</math>: total number of pixels contained in <math>P_v</math></p>	<p>[0, 1]; 1 = complete fitting, whereas 0 = only 50% or less voxels fit inside the ellipsoid</p>

width to thickness of the image object. The volume of the image object outside the

ellipsoid is compared with the volume inside the ellipsoid that is not filled out with the

image object.

#### 14. Main direction

Main direction feature of a three-dimensional image object is computed as follows:

1. For each image object slice (a 2D pieces of the image object in a slice) the centers of

gravities are calculated.

2. The coordinates of all centers of gravities are used to calculate a line of best fit

according to the Weighted Least Square method.

3. The angle  $\alpha$  between the resulting line of best fit and the z-axis is returned as

feature value

[0, 90]

#### 15. Radius of Largest Enclosed Ellipse

It describes how much the shape of an image object is similar to an ellipsoid. The calculation is based on an ellipsoid with the same volume as the object and based on the covariance matrix. This ellipsoid is scaled down until it is totally enclosed by the

image object. The ratio of the radius of this largest enclosed ellipsoid to the radius of the

original ellipsoid is returned as feature value.

$\varepsilon_v(x_0, y_0, z_0)$

with  $(x_0, y_0, z_0) = \min \varepsilon_v(x, y, z), (x, y, z) \in P_v$

$\varepsilon_v(x, y, z)$ : elliptic distance at a pixel  $(x, y, z)$

Expression:

[0,  $\infty$ ]

#### 16. Radius of Smallest Enclosing Ellipse

The calculation is based on an ellipsoid with the same volume as the image object and

$\varepsilon_v(x_0, y_0, z_0)$

with  $(x_0, y_0, z_0) = \max \varepsilon_v(x, y, z), (x, y, z) \in \sigma P_v$

[0,  $\infty$ ]



	<p>based on the covariance matrix. This ellipsoid is enlarged until it encloses the image</p> <p>object in total. The ratio of the radius of this smallest enclosing ellipsoid to the radius of</p> <p>the original ellipsoid is returned as feature value.</p>	<p><math>\varepsilon_v(x,y,z)</math>: elliptic distance at a pixel <math>(x,y,z)</math></p>	
17. Rectangular Fit	<p>It describes how well an image object fits into a cuboid of similar size and proportions.</p> <p>While 0 indicates no fit, 1 indicates for a complete fitting image object. The calculation is based on a cuboid with the same volume as the considered image object. The proportions of the cuboid are equal to the proportions of the length to width to thickness of the image object. The volume of the image object outside the rectangle is compared with the volume inside the cuboid that is not filled out with the image object.</p>	$\frac{\#\{(X,Y,Z) \in P_v : \rho_v(X,Y,Z) \leq 1\}}{\#P_v}$ <p><math>\rho_v(x,y,z)</math>: rectangular distance at a pixel <math>(x,y)</math></p> <p><math>\#P_v</math>: total number of pixels contained in <math>P_v</math></p>	<p>[0, 1]; 1 = complete fitting, whereas 0 = 0% fits inside the rectangular approximation</p>
18. Roundness	<p>Describes how much the shape of an image object is similar to an ellipsoid. The more the shape of an image object is similar to an ellipsoid, the lower its roundness. It is calculated by the difference of the enclosing ellipsoid and the enclosed ellipsoid. The radius of the largest enclosed ellipsoid is subtracted from the radius of the smallest enclosing ellipsoid.</p>	$\varepsilon_v^{max} - \varepsilon_v^{min}$ <p><math>\varepsilon_v</math></p> <p>max: radius of smallest enclosing ellipsoid</p> <p><math>\varepsilon_v</math></p> <p>min: radius of largest enclosed ellipsoid</p>	<p>[0, <math>\infty</math>]; 0 = ideal.</p>
19. Shape Index	<p>The smoother the surface of an image object is, the lower its shape index. It is calculated from the Border length feature of the image object divided by four times the square root of its area.</p>	<p><math>B_v/V_v</math></p> <p><math>b_v</math>: image object border length</p> <p><math>V_v</math>: volume of image object <math>v</math></p>	<p>[1, <math>\infty</math>]; 1 = ideal.</p>
20. Histogram features	<p>The intensity histogram <math>h(a)</math> is the number of pixels occurred for brightness level "a" plotted against their</p>		

brightness level. The probability distribution of the brightness  $P(a)$  can be calculated as well. Six features: mean, standard deviation, skewness, kurtosis, energy, entropy were incorporated.

1.  $\text{mean} = \sum_{i=1}^{\text{range}} i * \text{prob}(i)$
2.  $\text{sd} = \sqrt{\sum_{i=1}^{\text{range}} (i - \text{mean})^2 * \text{prob}(i)}$
3.  $\text{skewness} = \frac{\sum_{i=1}^{\text{range}} (i - \text{mean})^3 * \text{prob}(i)}{(\sum_{i=1}^{\text{range}} (i - \text{mean})^2 * \text{prob}(i))^{1.5}}$
4.  $\text{kurtosis} = \frac{\sum_{i=1}^{\text{range}} (i - \text{mean})^4 * \text{prob}(i)}{(\sum_{i=1}^{\text{range}} (i - \text{mean})^2 * \text{prob}(i))^2}$
5.  $\text{energy} = \sum_{i=1}^{\text{range}} \text{prob}(i) * \text{prob}(i)$
6.  $\text{entropy} = - \sum_{i=1}^{\text{range}} \text{prob}(i) * \text{Log}(\text{prob}(i))$

Where intensity range is  $[0, \text{range}]$  (normalized)

$\text{prob}(i) = \frac{\text{hist}(i)}{\sum \text{hist}(i)}$ ,  $\text{hist}(i)$  is the frequency of intensity  $i$  appears.

$p(i, j)$  is the element of run-length matrix, let  $M$  be the number of gray levels,  $N$  be the maximum run length.  $n_r$  is the total number of runs,  $n_p$  is the number of pixels in the image. Define 3 new matrices first.

- (a)  $p_p(i, j) = p(i, j) * j$
- (b)  $p_g(i) = \sum_{j=1}^N p(i, j)$
- (c)  $p_r(j) = \sum_{i=1}^M p(i, j)$

1.  $SRE = \frac{1}{n_r} \sum_{j=1}^N \frac{p_r(j)}{j^2}$
2.  $LRE = \frac{1}{n_r} \sum_{j=1}^N p_r(j) * j^2$
3.  $GLN = \frac{1}{n_r} \sum_{i=1}^M p_g(i)^2$

## 21. Run length matrix features

Run-length texture features examine runs of similar gray values in an image. Runs may be labeled according to their length, gray value, and direction (either horizontal or vertical). Long runs of the same gray value correspond to coarser textures, whereas shorter runs correspond to finer textures. Texture content was quantified by computing 11 features derived from the run-length distribution matrix. They are

- 1: Short Run Emphasis (SRE).
- 2: Long Run Emphasis (LRE).
- 3: Gray-Level Nonuniformity (GLN).
- 4: Run Length Nonuniformity (RLN).
- 5: Run Percentage (RP).

- 6: Low Gray-Level Run Emphasis (LGRE).
- 7: High Gray-Level Run Emphasis (HGRE).
- 8: Short Run Low Gray-Level Emphasis (SRLGE).
- 9: Short Run High Gray-Level Emphasis (SRHGE).
- 10: Long Run Low Gray-Level Emphasis (LRLGE).
- 11: Long Run High Gray-Level Emphasis (LGHGE).

$$\begin{aligned}
 4. \quad RLN &= \frac{1}{n_r} \sum_{j=1}^N p_r(j)^2 \\
 5. \quad RP &= \frac{n_r}{n_p} \\
 6. \quad LGRE &= \frac{1}{n_r} \sum_{i=1}^M \frac{p_g(i)}{i^2} \\
 7. \quad HGRE &= \frac{1}{n_r} \sum_{i=1}^M p_g(i) * i^2 \\
 8. \quad SRLGE &= \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j)}{i^2 * j^2} \\
 9. \quad SRHGE &= \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) * i^2}{j^2} \\
 10. \quad LRLGE &= \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i,j) * j^2}{i^2}
 \end{aligned}$$

$$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i,j) * i^2 * j^2$$

$p(i,j)$  is the element of the co-occurrence matrix.

- 1. Contrast =  $\sum_{i,j} |i - j|^2 * p(i,j)$
- 2. Energy =  $\sum_{i,j} p(i,j) * p(i,j)$
- 3. Homogeneity =  $\sum_{i,j} \frac{p(i,j)}{1 + |i - j|}$
- 4. Entropy =  $-\sum_{i,j} p(i,j) * \log(p(i,j))$
- 5. Sum Mean =  $0.5 * \sum_{i,j} (i + j) * p(i,j)$

Max probability =  $\max(p(i,j))$

For each filtered images (125), the energy was calculated as following:

- 22. Co-occurrence matrix features  
The Co-occurrence matrix is a matrix that contains the frequency of one gray level intensity appearing in a specified spatial linear relationship with another gray level intensity within a certain range. Computation of features requires first constructing the co-occurrence matrix, then different measurements can be calculated based on the matrix. The measurements include: contrast, energy, homogeneity, entropy, mean and max probability.

- 23. Laws features  
Laws features are constructed from a set of five one-dimensional filters, each designed to reflect to a different type of structure in the image. These one-dimensional filters are defined as E5 (edges), S5 (spots), R5 (ripples), W5 (waves), and L5 (low pass, or average gray value). By using these 1-D convolution filters, 3D filters are

generated by convolving 3 types of 1D filter, such as L5L5L5, L5L5E5, L5L5S5, L5L5R5, L5L5W5, etc. The total number of 3-D filters is 125. After the convolution with the 3D filters for the image, the energy of the texture feature can be computed.

$$Energy = \frac{1}{R} \sum_{i=N+1}^{I-N} \sum_{j=N+1}^{J-N} \sum_{k=N+1}^{K-N} h^2(i, j, k),$$

Where R is a normalizing factor, I and J, K are image dimensions,  $h(i, j, k)$  is derived from the convolution filters and original image.

24. Wavelet decomposition The discrete wavelet transform can iteratively decompose an image (3D) into four components. Each iteration splits the image both horizontally and vertically into low-frequency (low pass) and high-frequency (high pass) components. Thus, four components are generated: a high-pass/high-pass component consisting of mostly diagonal structure, a high-pass/low-pass component consisting mostly of vertical structures, a low-pass/high-pass component consisting mostly of horizontal structure, and a low-pass/low-pass component that represents a blurred version of the original image. Subsequent iterations then repeat the decomposition on the low-pass/low-pass component from the previous iteration. These subsequent iterations highlight broader diagonal, vertical, and horizontal textures. And for each component, we calculated the energy feature.

$$Energy = \frac{1}{M \times N \times L} \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^L I^2(i, j, k)$$

$$Entropy = \frac{-1}{M \times N \times L} \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^L \left( \frac{I^2(i, j, k)}{norm^2} \right) \log \left( \frac{I^2(i, j, k)}{norm^2} \right)$$

$I(i, j, k)$  shows the subblock elements and M, N, and L are the dimensions of each subblock and

$$norm^2 = \sum_i \sum_j \sum_k I^2(i, j, k)$$

The number of features really depends on the number of decomposition level selected.

1 level:  $2^* 8(\text{block}) = 16$  features

2 level:  $2^* 15(\text{block}) = 30$  features

25. Fractional Anisotropy Fractional anisotropy of long vs. short axis

$$\frac{\sqrt{(L - W)^2 + (W - T)^2 + (T - L)^2}}{\sqrt{L^2 + W^2 + T^2}} * \sqrt{\frac{1}{2}}$$

L: Length

W: Width

T: Thickness

26. Spheredcity It describes how spherical is the image object.

$$\frac{\pi^{\frac{1}{3}} (6V_p)^{\frac{2}{3}}}{A_p}$$

Ap = surface area

Vp = volume

- 27 Number of Macrospiculations Number countable spiculations of tumor
31. Distance of center of gravity to border of tumor Measure of distance from center of gravity to border of the tumor. Reported as average, std dev, minimum and maximum.
32. Attachment of tumor to other anatomical structures Describes the attachment of the tumor to the other anatomical structures. Reported as relative border to lung, relative border to attached structure, ratio of free to attached surface area.
33. Intensity value of tumor in HU Brightness values of tumors measured in Hounsfield units (HU). Reported as mean(HU) and Std Dev (HU)



## Appendix IV

The feature values of the Maastro and Moffitt tool are  $x$  and  $y$ . The formula of the CCC then

is: 
$$CCC = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x + \mu_y)^2}$$

Where  $\rho$  is the Pearson correlation,  $\sigma$  the standard deviation,  $\sigma^2$  the variance, and  $\mu$  the mean.

## Appendix V

The linear regression formula is  $Y = a + bX$ . In this formula 'a' is the intercept, which is the value of Y when  $X=0$ . 'b' is the slope of the regression line and gives the average increase of Y when variable X increases by one unit.



## Appendix VI

The new calculated CCC for Longest Diameter of Moffitt with Shape\_maxDiameter2D1, Shape\_maxDiameter2D2, Shape\_maxDiameter2D3 and Shape\_maxDiameter3D of Maastro.

Feature Maastro	Feature Moffitt	CCC Lung	CCC head/neck
Shape_maxDiameter2D1	Longest Diameter	0,676453879	0,787669494
Shape_maxDiameter2D2	Longest Diameter	0,688697225	0,787670403
Shape_maxDiameter2D3	Longest Diameter	0,574404612	0,517999839
Shape_maxDiameter3D	Longest Diameter	0,651190078	0,931705241