Fontys University of Applied Science

Medical Imaging and Radiation Therapy (MIRT)

Comparison between the numerical output of the Maastro and Moffitt Radiomics software and the impact of image intensity resampling for textural features – a study in lung tumors
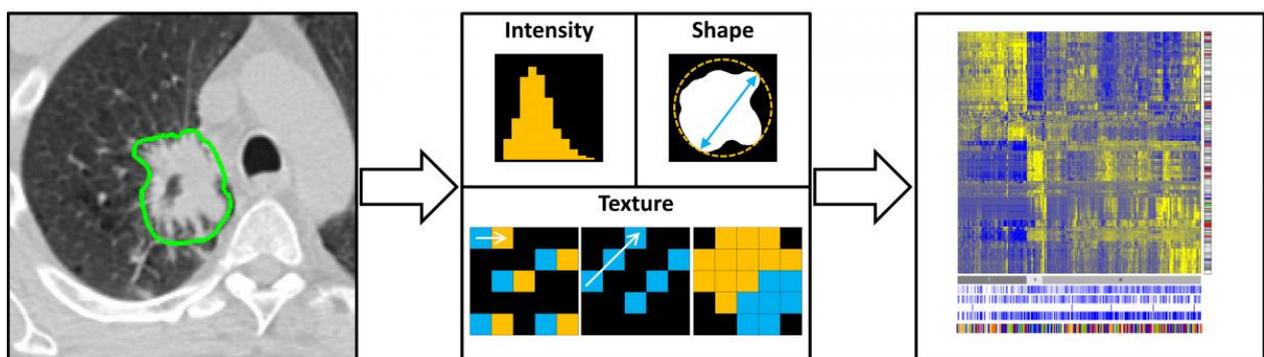
M.E. Pijls

Bachelor thesis, June 2016

Student number: 2207410

Institutes: Maastro Clinic & Moffitt Cancer Center and Research Institute

Supervisors: Ralph Leijenaar (Maastro) & Robert J. Gillies (Moffitt)

Mentors: Esther Bloemen – van Gurp (Maastro) & Eduardo Moros (Moffitt)

Preface

This thesis is written as completion of the study Medical Imaging and Radiation Therapy at Fontys University of Applied Science and was commissioned by Maastro Clinic (Maastricht, The Netherlands) and Moffitt Cancer Center and Research Institute (Tampa, Florida, USA). The aim of this study is to compare the developed Radiomic software tools between both institutes.

Writing this thesis in Florida, was a once in a lifetime experience! I am grateful that I was able to do this together with my friend and fellow student, Marlon van den Broek. Thanks for all the great memories we made here and supporting me in stressful periods.

I would love to thank my mum, Madelon Johannesma, for getting us in touch with Moffitt, Maastro and Moffitt for giving me the opportunity to do this graduation project at their institutes, Robert Gillies and his Radiomics department for the hospitality and help during this project, Eduardo Moros, Ralph Leijenaar and Esther Bloemen – van Gurp for the guidance, feedback and investment of time to answer my questions during this project. Last but not least, I want to thank Shonagh Russell for correcting my thesis on English language and grammar and for being an amazing friend!

Tampa, June 2016

Mirthe Pijls
Bachelor Student of Medical Imaging and Radiation Therapy – Graduation Class of 2016

Abstract

Introduction:

Radiomic tools make it possible to extract large numbers of quantitative features from tumors using medical images (e.g. CT/ MRI/ PET) that, in contrast to biopsies, are able to characterize the whole tumor in a non-invasive way. To ensure interoperability of Radiomics, it is important to compare different tools.

Objective:

To assess (1) if there is a difference between the implementation and numerical output of the Maastro and Moffitt Radiomic algorithms and (2) if a specific setting in the Maastro tool make the numerical values of texture features between two tools comparable.

Method:

A retrospective quantitative comparison study was performed on a cohort of NSCLC patients (N=42). The features of both tools were matched by name and algorithm. RT-STRUCTS and CT-scans of the patients were loaded in the tools to gain numerical values. To evaluate the similarity of the numerical values between the tools, Pearson's correlation coefficient and Lin's concordance correlation coefficient (CCC) were used. Subsequently, textural feature values of Moffitt were compared to those calculated using different intensity resampling methods with the Maastro tool.

Results:

34 features were matched. The strength of linear association (SA) for the Pearson correlation coefficient was strong for 20 features, moderate for 5 features, weak for 1 feature, very weak for 1 feature and there was no correlation for 7 features. The CCC of 9 features was almost perfect, the CCC of 1 feature was moderate and the CCC of 24 features was poor. There is no resampling method which made the output of textural features comparable and all CCC values were poor.

Conclusion:

This study has shown that comparing Radiomics results using different software implementations is not straightforward. To make comparisons in multicentre settings, and to ensure optimal decision support systems for lung cancer in the future, standardization of Radiomics is needed.

Samenvatting

Inleiding:

Radiomic tools kunnen grote aantallen kwantitatieve kenmerken (features) van tumoren uit medische beelden (bijvoorbeeld CT / MRI / PET) extraheren die, in tegenstelling tot biopten, de gehele tumor kunnen karaktiseren op een niet-invasieve manier. Om interoperabiliteit van Radiomics te waarborgen, is het belangrijk om verschillende instrumenten te vergelijken.

Doelstelling:

Onderzoeken (1) of er een verschil is tussen de implementatie en numerieke output van de Maastro en Moffitt Radiomic algoritmen en (2) het vinden van een specifieke instelling in de Maastro tool om numerieke waarden van textuur kermerken vergelijkbaar te maken tussen de tools.

Methode:

Een kwantitatieve retrospectieve vergelijkingsstudie werd uitgevoerd op een cohort van NSCLC-patiënten (N=42). De features van beide tools werden gematched met behulp van namen en algoritmes. RT-STRUCTS en CT-scans van de patiënten werden in beide tools geladen om numerieke waarden te verkrijgen. Om de gelijkenis van de numerieke waarden tussen de tools te evalueren, zijn de Pearson correlatiecoëfficiënt en de concordantie correlatiecoëfficiënt (CCC) gebruikt. Vervolgens werden de waarden van textuur features van Moffitt vergeleken met de verschillende berekende intensiteit resampling instellingen van de Maastro tool.

Resultaten:

Vierendertig features zijn gematched. De Pearson correlatiecoëfficiënt toonde een sterke lineaire overeenkomst aan voor 20 features, een matige overeenkomst voor 5 features, een zwakke overeenkomst voor 1 feature, een zeer zwakke overeenkomst voor 1 feature en er was geen correlatie voor 7 features. De CCC van 9 features was bijna perfect, de CCC van 1 feature was matig en de CCC van 24 features was slecht. Alle resampling instellingen gaven slechte CCC's en geen vergelijkbare output voor textuur features.

Conclusie:

Deze studie laat zien dat het niet eenvoudig is om Radiomics resultaten met verschillende softwareimplementaties te vergelijken. Om vergelijkingen in een multicentrische setting te maken en om te zorgen voor optimale beslissingsondersteunende systemen voor longkanker in de toekomst, is standaardisatie van Radiomics nodig.

Table of contents

Personnel:

Mirthe Pijls

Student number: 2207410

Institute: Fontys University of Applied Science

Function: Fourth year student of Medical Imaging and Radiation Therapy (MIRT)

Phone number: (+31) 0621182728

E-mail: mirthepijls@hotmail.com

Ralph Leijenaar

Function: Supervisor

Institute: Maastro Clinic

Phone number: +31 (088) 445 5666

E-mail: ralph.leijenaar@maastro.nl

Esther Bloemen – van Gurp

Function: Mentor

Institute: Maastro Clinic and Fontys University of Applied Science

Phone number: (+31) 0653230749

E-mail: e.bloemenvangurp@fontys.nl

Claudia Offerman

Function: Graduation coordinator

Institute: Maastro Clinic

Phone number: +31 (088) 445 5666

E-mail: claudia.offerman@maastro.nl

Robert J. Gillies

Function: Supervisor/ Mentor

Institute: H. Lee Moffitt Cancer Center and Research Institute

Phone number: +1 (813) 745 8355

E-mail: Robert.Gillies@Moffitt.org

Eduardo Moros

Function: Mentor

Institute: H. Lee Moffitt Cancer Center and Research Institute

Phone number: +1 (813) 745 1075

E-mail: Eduardo.Moros@moffitt.org

Introduction

Current cancer treatments are often inadequate, with 75% being ineffective in the clinic. This is thought to be due to the strong intra- and inter patient heterogeneity.[1,2] Therefore, over the past decade, investment into new diagnostic and treatment modalities has been carried out. This has led to an increased amount of patient specific information, including identification of novel disease biomarkers using genomics, proteomics and non-invasive imaging approaches, in order to classify subgroups of patients to more adequately refer patients to the right treatment.[3] Genomics and proteomic approaches are successfully used in the clinic currently to obtain the molecular characteristics of diseased tissues, but come with high associated risks and costs for the patient, because of biopsies. Therefore there is much room for improvement within clinical practice.[4]

In the last ten years the use of medical imaging technologies (e.g. Computed Tomography (CT)/ Magnetic Resonance Imaging (MRI)/ Positron Emission Tomography (PET)) has greatly expanded. Primarily these technologies are used as a diagnostic tool however they play an increasingly crucial role in individualized patient care. Compared to other variables such as demographics, blood biomarkers, pathology and genomics, it is expected that imaging gives valuable complementary information, which can for instance be used to predict outcome for a specific patient.[5] This new use of imaging is a major step towards personalized medicine and provides valuable diagnostic, prognostic and predictive information.[6]

Current imaging technology makes it possible to extract large numbers of quantitative features from tumors using medical images (e.g. CT/ MRI/ PET) that, in contrast to biopsies, are able to characterize the whole tumor in a non-invasive way.[5] This field of research is referred to as Radiomics. Radiomic features can for instance, provide information about tumor image intensity, shape and texture. In order to extract these variables from medical images, software tools are being developed which implement Radiomic algorithms. The combination of specific variables as mentioned above (demographics, blood biomarkers, pathology and genomics) in combination with Radiomics has the potential to improve a patient's prognostic information and make a step towards personalized medicine. However, combining these data is overly complex for clinicians. Therefore the use of decision support systems will be able to strive towards the right treatment for a specific patient in order to improve outcomes and Quality of Life (QoL) as well as efficiency of care for different cancer sites, e.g. lung cancer, the leading cause of cancer related deaths amongst males and accounts for 13% of all cancer diagnosis.[7] In 2012 there were 1.8 million new cases reported and it is projected that by 2030 the number of lung cancer deaths will have surpassed 10 million.[8]

Radiomics is a rapidly expanding field of research that could have large clinical impact, providing an unprecedented opportunity to improve decision-support for physicians at low cost. Knowledge on Radiomic features of patients with non small cell lung carcinoma (NSCLC) is already available from earlier studies.[5,8,9,10] Because Radiomics is a relative new study, there is still much to explore in this

research field. In particular it has become apparent that there is a need for standardization of Radiomics.[11]

Most of the research concerning Radiomics was undertaken in specific centers in the U.S., The Netherlands and Canada. Two of these centers, Moffitt Cancer Center (Tampa, FL, U.S.) and Maastro Clinic (Maastricht, The Netherlands), participate in the Quantitative Imaging Network (QIN) of the National Cancer Institute (NCI) of the U.S.[12]

Moffitt has a vast database of patient samples including tissues and gene expression profiling. Maastro is a known world leader in radiation oncology.[8] Gillies et al.[8] describe the vision of both institutions as follows: "Moffitt and MAASTRO have shared a vision that quantitative features extracted from standard of care images have the potential to provide higher prognostic and predictive value."

Both institutes developed their own Radiomic tools and it is important to compare Radiomic results from different institutes. The algorithms implemented in the tools provide numerical values. In the Maastro tool it is possible to change certain parameters of feature calculations. Before calculating textural features, the image intensities are typically resampled into a reduced number of discrete bins. To achieve this, two methods can be applied: (1) dividing the Houndsfield Units (HU) range into a fixed number of equally spaced bins, where the bin size (i.e. intensity resolution) depends on the range of HU values in each image; and (2) maintaining a constant bin size for each image. The choice of resampling methodology was shown to have a crucial effect on textural features and their interpretation in a study on PET images of NSCLC patients.[13] The number of bins can be set to, e.g., 8, 16 or 32 discrete values (this means that the HU values of every image will be "resampled" to the same range of values). The bin size (bin width) can be set to any value (in HU). The Maastro and Moffitt implementations both use a different resampling scheme. The default for the Maastro tool is a bin width of 25 HU, whereas the default for the Moffitt tool is to use 256 number of bins. In contrast to the Maastro tool, the resampling scheme cannot be altered in the Moffitt tool.

Since this is a relatively new research area in lung cancer, it is valuable to compare the tools developed by different institutes, e.g., between Maastro and Moffitt, in order to potentially standardize Radiomics and to facilitate multi-institutional studies to optimize decision support systems for lung cancer globally in the future.

Therefore, this project focused on the comparison of the implementation and numerical output of the Radiomic features algorithms of two research groups, Maastro and Moffitt. In conjunction with the differences observed between the two tools, parameters of texture features will be changed in the Maastro tool, to see if this affects numerical values obtained.

The aim of this project is to assess (1) if there is a difference between the implementation and numerical output of the Maastro and Moffitt Radiomic algorithms in an identical dataset and (2) if specific settings in the Maastro tool have to be applied to make numerical values of texture features between two tools comparable. This effort will lead to the following research question:

What is the difference in numerical values for Radiomic features obtained by the Maastro and Moffitt software tool in a cohort of NSCLC patients and can certain settings be applied to make the output in textural features comparable?

Method

Design
This research is a retrospective quantitative comparison study. In this study we investigated the difference between the numerical values for Radiomic features of NSCLC patients between Maastro and Moffitt and determined if changing the resampling scheme for textural features affects this difference.

Population
In this study data from a cohort of 52 NSCLC patients, treated between July 2015 and March 2016, have been included for analysis. Patients were randomly chosen from all patients treated with curative intent; all stages of NSCLC were included and there were no restrictions for age or gender. Patients with lung metastasis from a non-lung primary tumor were excluded from this study. For all included patients the treatment CT-scans and the corresponding structure sets (RTSTRUCTS) have been retrieved from the database of Maastro. In this process all data were anonymized by the IT department of Maastro, removing all privacy sensitive information.

Measuring instruments
All CT-scans were performed before radiation therapy at Maastro, according to standard clinical scanning protocols. In this population, patients were scanned on one of the following three scanners: a Sensation 10 CT scanner RT (Siemens, 2004), a Sensation Open CT scanner RT (Siemens, 2006), or a Biograph TruePoint 40 (PET)-CT RT (Siemens, 2006). A spiral CT was performed covering the complete thoracic region. Images were reconstructed with an in plane pixel spacing of 0.977 mm x 0.977 mm, with a slice thickness of 3 mm. The primary lung tumor was delineated manually for treatment planning purposes on Eclipse (Varian Medical System, Palo Alto, CA, USA). There was no contrast used.

Data collection
The CT-images and RTSTRUCTS were used for this comparison study. On the CT-images, the tumors can be defined with an ROI (Region Of Interest). These ROIs were defined by a physician. From the RTSTRUCTS, all Gross Tumor Volumes (GTVs) have been used in the Radiomic tools to extract the Radiomic features (Appendix I and II).

The Radiomic tool of Maastro is designed by Ralph Leijenaar.[13,14] The Maastro tool runs on MATLAB R2014a (8.3). The software is able to extract a large number of quantitative features, however, for this study we focused on the features of the following groups: first order statistics (N=15), shape and size based features (N=12) and gray-level co-occurrence matrix (GLCM) and gray-level run-length matrix (RLGL) textural features (N=44). Appendix I describes the definitions as follows: "First-order statistics describe the distribution of voxel intensities within the CT image through commonly used and basic metrics. In shape and size based features we included descriptors of the three-dimensional size and shape of the tumor region. Textural features describe patterns or the spatial distribution of voxel

intensities." Textural features have different groups, namely GLCM and RLGL. If these abbreviations are standing in front of a feature, for example GLCM, this means that they belong to this particular group. Appendix I also describes corresponding algorithms of the features, used in this tool.[14,15]

First, the data was loaded into the Maastro tool. Once loaded, the ROI of interest (in this case the GTV) was selected for each patient from the list of available ROIs (e.g. the lungs, the heart, etc.). The GTV was selected automatically whenever possible, but since there is no naming convention for ROIs (the GTV can be named, for instance, GTV-1, GTVp1, GTV-tumor, tumor, etc.), the software also allowed manual selection of the GTV for each patient. The features of interest, first order statistics, shape and size based features and the required textural features have been selected. The tool extracted these features from the selected GTVs of each patient and the results were exported into Excel spreadsheets, Microsoft Excel 2010. Most numerical values of the features are unitless.

Moffitt uses MATLAB R2016a to extract Radiomic features. Appendix II describes the definitions of the features with corresponding algorithms used in this tool.[10,16] Excel spreadsheets derived from the Maastro tool with the names of the selected GTVs were used. The RTSTRUCTS of the patients were loaded in MATLAB and the provided GTVs were selected. Moffitt extracted the GTVs of interest identified in the Maastro tool and converted them from Digital Imaging and Communications in Medicine (DICOM) format to Portable Network Graphics (PNG) images with a MATLAB script. The selected GTVs of both tools had to be the same to make a correct comparison. Converting an RTSTRUCT into a binary pixel mask is different in each tool, which may lead to differences in output. Then the CT-images and transformed GTVs were loaded in MATLAB so the features could be extracted with an implemented code. Most features were unitless and were imported into Excel spreadsheets, Microsoft Excel 2010.

Data analysis

To compare the Radiomic tools, first a list of included features was created. To make this list, the Radiomic features of Maastro and Moffitt were matched by name and by algorithm. Features that could not be matched were not included in this research. The included features and their respective feature groups (i.e. first order statistics, shape and size based features or textural features) are given in Table 1.

To obtain the numerical values from the tools, the CT-images and RTSTRUCTS of each patient were processed by both tools. Each feature therefore had two values for each patient; one for the Maastro tool and one for the Moffitt tool.

To evaluate the similarity of the Moffitt numerical values with the Maastro numerical values, the Pearson correlation coefficient and Lin's concordance correlation coefficient (CCC) were used.[10]

The Pearson correlation coefficient has been calculated for each matching feature between Maastro and Moffitt, in order to be implemented in the CCC calculation. This coefficient has been calculated with the standard formula for correlation (=CORREL) in Excel. Correlation coefficient size is defined follows; If the correlation range (CR) is 0.81 to 1.00, the strength of linear association (SA) is strong, if CR is 0.61 to 0.80 the SA is moderate, if CR is 0.41 to 0.60 the SA is weak, if CR is 0.21 to 0.40 the SA is very weak and if CR is 0.00 to 0.20 there is no SA.[17]

Let x and y be the feature values of the Maastro and Moffitt tool, respectively. 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 ρ is the Pearson correlation, σ the standard deviation, σ^2 the variance, and μ the mean.[18]

The CCC has four different strength-of-agreements. A CCC >0.99 is almost perfect, a CCC between 0.95 – 0.99 is substantial, a CCC between 0.90 – 0.95 is moderate and a CCC<0.90 is poor.[19] First all components (σ, σ^2 and μ) have been calculated separately from each other and then the formula of the CCC was made in Excel combining all individual components. Because the manual input of the formula in Excel is prone to errors, the CCC was verified with an online calculator of the CCC.[20]

The Pearson correlation coefficient and CCC values from -1 to +1 have been visualized in a column chart, with features on the x-axis and similarity on the y-axis (Figure 1).

The CCC measures reliability between two variables, based on covariation and correspondence. The Pearson correlation coefficient measures the linear covariation between two variables, independently of correspondence. The Pearson correlation coefficient thus evaluates if there is a linear continuity between the numerical values of Maastro and Moffitt, whereas the CCC also evaluates the degree to which numerical values of matching features fall on the 45° line through the origin.[10,19,21,22,23]

To visualize the linearity between features, a scatter plot of the linear fit was made in Excel (Figure 2). For each feature a linear fit is calculated with formula y = ax + b, where a is the slope and b the intercept between the Maastro and Moffitt feature. R^2 is the same as the Pearson correlation coefficient squared, so the linear correlation between both features expressed as a value between 0 and 1, where 0 is no correlation and 1 is a perfect correlation. A value of +1 is a perfect correlation and -1 is a perfect negative correlation.

For the second part of this study we determined if changing the intensity resampling scheme for texture in the Maastro tool would have an impact on the outcome[13] and if matching this setting between both tools would make their outputs comparable. We changed the bin width in the Maastro tool to 1, 5, 10, 25 and 50 HU and the number of bins to 8, 16, 32, 64, 128 and 256, of which the latter matches the default setting of the Moffitt tool. The CCC was calculated for each feature, for all different

intensity resampling settings (i.e., for each setting the feature values were compared to those determined by the Moffitt software).

Ethics

All patient related data in this study was anonymized. Therefore, ethical approval by the Medical Committee was not required and it was not necessary to seek consent for its use. Attached is the signed cooperation contract from Maastro (Appendix III). In agreement with signing this contract all data has been contained within Maastro and remain anonymous (Article 12).

Results

We matched 34 features from Maastro and Moffitt by name and algorithm; first order statistics (N=14), shape and size based features (N=3) and GLCM and RLGL textural features (N=17). See Table 1 for the list of matched features.

Out of the 52 selected patients loaded in the Maastro tool, 10 patients did not have a defined GTV. Without a GTV the features cannot be extracted from the image, therefore these patients were excluded. The patients with a defined GTV (n=42) were included in this study.

The Pearson correlation coefficient and CCC were calculated for all features using the default settings of the Maastro tool (bin width 25 HU) and the Moffitt tool (256 number of bins). The strength of association (SA) for the Pearson correlation coefficient was found to be strong for 20 features, moderate for 5 features, weak for 1 feature, very weak for 1 feature and there was no correlation for 7 features. The CCC of 9 features was almost perfect, the CCC of 1 feature is moderate and the CCC of 24 features is poor (Table 1). See Figure 1 for a column chart of these results.

The Pearson correlation coefficient shows that 20 features have a strong SA and of these 20 features, 9 have an almost perfect CCC. Four features have a moderate Pearson correlation coefficient, and just one of those features has a moderate CCC. All other features have poor agreement (Table 1).
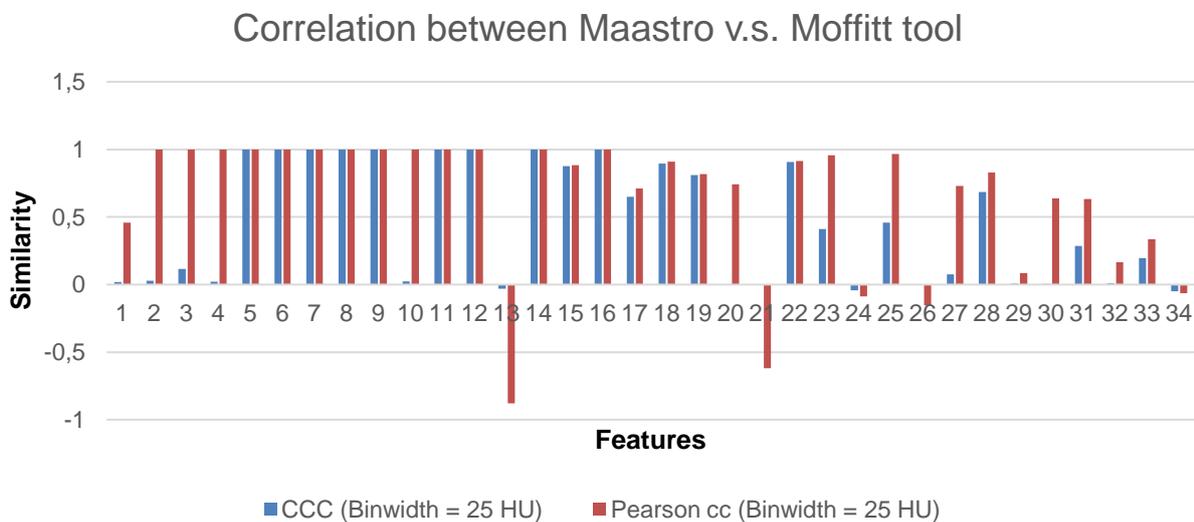

Figure 1: Column chart of the CCC and Pearson correlation coefficient (cc) between the Maastro and Moffitt tool, with the default setting of the Maastro and Moffitt tool. The feature names of these numbers are shown in Table 1.

Table 1: Matching features of Maastro and Moffitt with the calculated CCC and Pearson correlation coefficient (cc), with a default setting for both tools.

Feature number	Feature catagory	Maastro	Moffitt	CCC	Pearson cc
1	First order statistics	Stats_energy	Energy	0.02	0.46
2	First order statistics	Stats_min	Min HU	0.03	0.99
3	First order statistics	Stats_max	Max HU	0.12	0.99
4	First order statistics	Stats_mean	Mean HU	0.02	0.99
5	First order statistics	Stats_var	Variance HU	0.99	0.99
6	First order statistics	Stats_skewness	Skewness HU	0.99	0.99
7	First order statistics	Stats_kurtosis	Kurtosis HU	0.99	0.99
8	First order statistics	Stats_std	Standard Deviation HU	0.99	0.99
9	First order statistics	Stats_md	Mean Absolute Deviation HU	0.99	0.99
10	First order statistics	Stats_median	Median HU	0.02	0.99
11	First order statistics	Stats_range	Range HU	0.99	0.99
12	First order statistics	Stats_entropy	Entropy	0.99	0.99
13	First order statistics	Stats_rms	RMS HU	-0.03	-0.88
14	First order statistics	Stats_uniformity	Uniformity HU	0.99	0.99
15	Shape & size	Shape_volumeNumber	Connected 3D components	0.88	0.88
16	Shape & size	Shape_volume	Volume	0.99	0.99
17	Shape & size	Shape_maxDiameter3D	Longest Diameter	0.65	0.71
18	Textural features	RLGL_shortRunEmphasis	Short Run Emphasis (SRE)	0.89	0.91
19	Textural features	RLGL_longRunEmphasis	Long Run Emphasis (LRE)	0.81	0.82
20	Textural features	RLGL_grayLevelNonuniformity	Gray-Level Nonuniformity (GLN)	6.86E-06	0.74
21	Textural features	RLGL_runLengthNonuniformity	Run-Length Nonuniformity (RLN)	-6.24E-06	-0.62
22	Textural features	RLGL_runPercentage	Run Percentage (RP)	0.91	0.91
23	Textural features	RLGL_lowGrayLevelRunEmphasis	Low Gray-Level Run Emphasis (LGRE)	0.41	0.96
24	Textural features	RLGL_highGrayLevelRunEmphasis	High Gray-Level Run Emphasis (HGRE)	-0.04	-0.09
25	Textural features	RLGL_shortRunLowGrayLevEmpha	Short Run Low Gray-Level Emphasis (SRLGE)	0.46	0.97
26	Textural features	RLGL_shortRunHighGrayLevEmpha	Short Run High Gray-Level Emphasis (SRHGE)	-0.00	-0.15
27	Textural features	RLGL_longRunLowGrayLevEmpha	Long Run Low Gray-Level Emphasis (LRLGE)	0.07	0.73
28	Textural features	RLGL_longRunHighGrayLevEmpha	Long Run High Gray-Level Emphasis (LRHGE)	0.69	0.83
29	Textural features	GLCM_energy	avgCoocurrence-Energy	0.01	0.09
30	Textural features	GLCM_contrast	avgCoocurrence-CONTRAST	0.00	0.64
31	Textural features	GLCM_homogeneity1	avgCoocurrence-HOMOGENEITY	0.29	0.63
32	Textural features	GLCM_maxProb	avgCoocurrence-MAXPROB	0.01	0.16
33	Textural features	GLCM_entrop2	avgCoocurrence-ENTROPY	0.19	0.34
34	Textural features	GLCM_sumAvg	avgCoocurrence-SUMMEAN	-0.05	-0.06

Legend:

CCC	Pearson correlation coefficient
Almost perfect (> 0.99)	Strong (0.81 - 1.00)
Substantial (0.95 - 0.99)	Moderate (0.61 - 0.80)
Moderate (0.90 - 0.95)	Weak (0.41 - 0.60)
Poor (< 0.90)	Very weak (0.21 - 0.40)
	None (0.00 - 0.20)

The CCC values are generally poor (<0.90), indicating there is a substantial difference in feature values obtained with both tools. For the Pearson correlation coefficient all first order statistic features have a strong linear association, except the features Stats_energy and Stats_rms (Table 1).

For two features we observed a substantial negative Pearson correlation. The Pearson correlation coefficient is -0.88 for matched feature 13 and -0.62 for matched feature 21. This means that these features have an inverse linear relationship and act in the opposite way. The interpretation of a perfect negative correlation is that for an increase in X of 1, there is a decrease in Y of 1.

It appears that Maastro has four different ways to calculate the diameter, while Moffitt has just one feature to calculate the diameter. Moffitt calculates the diameter only full 3D and Maastro calculates the diameter in full 2D transversal, full 2D coronal and full 2D sagittal, in addition to the calculation of full 3D. We calculated the CCC for the four different diameters of Maastro and the diameter of Moffitt. The CCC values were all poor and the CCC of the match Shape_maxDiameter3D (full 3D) of Maastro and Longest Diameter of Moffitt was 0.65. See Appendix IV for all CCC values of the compared diameter features of Maastro and Moffitt.

A scaling difference was found in some features. First order statistics Maastro features Stats_min, Stats_max, Stats_median and Stats_mean have a scaling difference of +1000 HU compared to the Moffitt features, which is reflected in the intercept of a linear fit between the Maastro and Moffitt values (Appendix V). In Appendix V the intercept, slope and R^2 (which equals the Pearson correlation coefficient squared) for all first order statistic values are shown. A scatter plot of the maximum HU (Moffitt: "Max HU" and Maastro "Stats_max") visualizes the linear fit and the offset in HU (Figure 2). As illustrated in Appendix V, this also applies to Stats_min, Stats_median and Stats_mean.
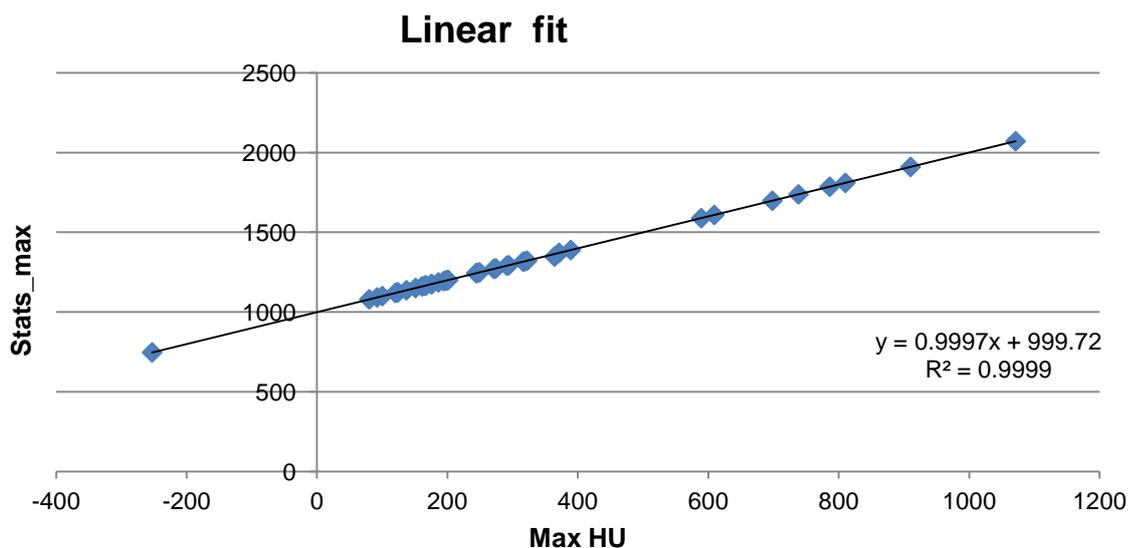
Figure 2: Scatterplot of linear fit of features Stats_max (Maastro) v.s. Max HU (Moffitt).

For the second part of this study, the resampling scheme was changed several times to assess whether changing this setting had an impact on the obtained results for the included textural features and to find out if matching the setting of the Moffitt tool (256 number of bins) and the Maastro tool results in comparable numerical values for both tools. We also identified the resampling schemes which produced CCC values closest to one. The results for each resampling scheme are provided in Appendix VI. In Table 2 the CCC values for the default settings of the Maastro tool (bin width 25 HU), the resampling scheme of the Moffitt tool (256 number of bins) and the resampling method that resembles the output of the Moffitt tool most closely, including the respective CCC values, are presented.

Indicated by low CCC values, there is no resampling method for which the values of all features match between both tools. When matching the resampling method (256 number of bins), the average for the CCC values of the RLGL features is 0.22 and 0.09 for GLCM features.

Table 2: Different resampling methods with CCC values of the default settings of both tools and CCC values which are closest to one.

| Feature number | Maastro | Moffitt | CCC (BW = 25) | CCC (NOB = 256) | CCC closest to 1 | Resampling method | |
						NOB	BW
18	RLGL_shortRunEmphasis	Short Run Emphasis (SRE)	0.89	0.16	0.92	64	-
19	RLGL_longRunEmphasis	Long Run Emphasis (LRE)	0.81	0.08	0.82	64	-
20	RLGL_grayLevelNonuniformity	Gray-Level Nonuniformity (GLN)	0	0	0	-	1
21	RLGL_runLengthNonuniformity	Run-Length Nonuniformity (RLN)	-6.24E-06	-4.01E-06	-3.38E-06	-	1
22	RLGL_runPercentage	Run Percentage (RP)	0.91	0.13	0.89	64	-
23	RLGL_lowGrayLevelRunEmphasis	Low Gray-Level Run Emphasis (LGRE)	0.41	0.66	0.75	-	1
24	RLGL_highGrayLevelRunEmphasis	High Gray-Level Run Emphasis (HGRE)	-0.04	0	0.60	64	-
25	RLGL_shortRunLowGrayLevEmpha	Short Run Low Gray-Level Emphasis (SRLGE)	0.46	0.76	0.84	-	1
26	RLGL_shortRunHighGrayLevEmpha	Short Run High Gray-Level Emphasis (SRHGE)	0	0.01	0.01	256	-
27	RLGL_longRunLowGrayLevEmpha	Long Run Low Gray-Level Emphasis (LRLGE)	0.07	0.58	0.75	-	1
28	RLGL_longRunHighGrayLevEmpha	Long Run High Gray-Level Emphasis (LRHGE)	0.69	0.01	0.85	64	-
29	GLCM_energy	avgCoocurrence-Energy	0.01	0	0.37	8	-
30	GLCM_contrast	avgCoocurrence-CONTRAST	0	0.44	0.44	256	-
31	GLCM_homogeneity1	avgCoocurrence-HOMOGENEITY	0.29	0.06	0.70	16	-
32	GLCM_maxProb	avgCoocurrence-MAXPROB	0.01	0	0.35	8	-
33	GLCM_entrop2	avgCoocurrence-ENTROPY	0.19	0.02	0.49	32	-
34	GLCM_sumAvg	avgCoocurrence-SUMMEAN	-0.05	0.01	0.19	64	-

NOB = number of bins and BW = bin width.

Discussion

The purpose of this study was to investigate if there is a difference between the numerical values for Radiomic features obtained by the Maastro and Moffitt software tool and whether it is possible to apply specific settings for the output to be comparable – in this case regarding the intensity resampling for textural features.

We found substantial differences in numerical output between the Maastro and Moffitt tool. Differences in first order statistic features are due to the pre-processing. Shape and size features and textural features have differences likely due to research groups implementing algorithms of features differently. Alternatively, both tools have different settings for these feature groups which can also result in differences. There is no gold standard, but it is clear that standardization is necessary.

As mentioned earlier, the feature names of each tool are different and algorithm matching was necessary to identify similar features. Regardless of the difference in feature definition lists, we were able to match 34 features by name and algorithm. Name alone was not sufficient to identify matching features, as the algorithms may be different even if the name is the same.

Preprocessing of the images is different in both tools, but is not the only source of discrepancy. To prevent negative values, Maastro adds +1000 HU to the intensity values. This affects first order statistic features Stats_min, Stats_max, Stats_mean, Stats_median, Stats_energy and Stats_rms, which is reflected in the low CCC values for these features. The Pearson correlation coefficient (and R^2), however, show that Stats_min, Stats_max, Stats_mean and Stats_median have strong linear relationships (R>0.99). The feature values for these features can be corrected by subtracting 1000 HU after feature extraction by the Maastro tool, which will then give the same numerical values for both tools. Hereby these features can be directly compared to each other. Features Stats_energy and Stats_rms cannot be straightforwardly corrected in a similar fashion due to the quadratic term in the feature's mathematical definition. We calculated all features without the preprocessing of +1000 HU for one patient, to see if it affects other feature groups. We found that the addition of 1000 HU does not affect shape and size features and textural features, because the numerical values of these features remained the same.

The Longest Diameter feature in the Moffitt tool was defined as the diameter in full 3D. Maastro calculates the diameter in four different ways; full 3D, full 2D transversal, full 2D coronal and full 2D sagittal. The CCC comparing the full 3D feature from Maastro with the Moffitt Longest Diameter feature was poor (<0.90). Comparing Longest Diameter with the alternative Maastro features did not

improve the CCC value. Therefore we have to conclude that there is an underlying difference in calculation.

The numerical values of the features Shape_volumeNumber (Maastro) and Connected 3D components (Moffitt) are all similar except the numerical values of one patient. Although the agreement is good, this is not reflected in the CCC value. Hence, the CCC appears to be not a useful method to compare these features, due to the distribution of this feature.

There is no resampling possible in the Moffitt software, unlike in the Maastro tool. Moffitt uses 256 number of bins to generate textural features. Multiple resampling methods were calculated in the Maastro tool, to find out which setting of the Maastro tool matched with the default setting of the Moffitt tool. We changed the number of bins in the Maastro tool multiple times in 8, 16, 32, 64, 128 and 256. We also changed the bin width in the Maastro tool to 1, 5, 10, 25 and 50 HU. Above a certain bin width it will not be meaningful to calculate the features, however a maximum of 50 HU was chosen as a reasonable upper limit for this study. Only two features had reasonable agreement in numerical values for both tools using 256 number of bins, namely RLGL_shortRunHighGrayLevEmpha and GCLM_contrast. All resampling methods gave different results for textural features in this study, thus the change of intensity is an important factor. Leijenaar et al.[13] carried out a similar study which examined if SUV discretization in FDG-PET images affects textural features. This study showed that different discretization methods affect textural features.[13]

CCC values with the default setting of Maastro (bin width = 25 HU) were found to be poor (<0.90). However, even when matching the resampling scheme using 256 number of bins, the overall agreement in numerical values of the textural did not significantly approve, with an average for the CCC values of 0.22 for gray-level run-length features (RLGL) and 0.09 for gray-level co-occurrence features (GLCM).

This study is one of the first studies comparing Radiomic software tools of two institutes. A recent conference contribution, Berthon et al.[24], about texture analysis for Positiron Emission Tomography in oesophageal cancer shows that comparison between Radiomic methods is important to help standardize Radiomics.[24]

The usability of the Moffitt tool was less than the Maastro tool. The Moffitt software consists of several different programs and there was no resampling possible. This has an influence on the outcome of this study because we cannot compare multiple resampling settings.

In this study we chose to focus on a select group of features. There are more feature groups available in both tools. Although tools worldwide have similar features, each institution may apply them in a different way resulting in disparate results from the same sample analysis. We observed that Maastro and Moffitt investigated different features.[5,10,13,16] It is possible that other features correspond better. Therefore we recommend to do further research in matching features.

To make a Radiomic tool that can be used in the clinic, more research has to be performed. Firstly, standardization of terminology of similar features used in different tools is required. Therefore, further research in the details of the differences in implementation for features that have the same name and definition should be done. Next, it is also required to explore all different kind of features used in both tools, in order to define a standard set of features that should be minimal integrated in each tool. Finally, for future comparison and validation, it is necessary that the software used in both tools is compatible. We therefore recommend further research to investigate the differences between the tools and to ensure interoperability.

This comparative study has clearly pointed to the need of standardization of Radiomic algorithms, methods and tools before integration in decision support systems for lung cancer, and most important, validated in a multi-centric setting.

.

Conclusion:

A substantial difference between the Maastro and Moffitt tool in terms of feature values was observed. Some first order statistic features have a scaling difference of +1000 HU compared to the Moffitt features, but can be corrected afterwards to make them comparable between both tools. However, in general there is a poor agreement between both tools.

When applying the same 256 number of bins intensity resampling for texture calculation in the Maastro tool as is used by the Moffitt tool, the agreement between both tools was still poor. Even though changing the intensity resampling scheme does affect numerical values for textural features, we have shown that there are more underlying differences that have to be investigated.

This study has shown that comparing Radiomic results using different software implementations is not straightforward. To make comparisons in multicentre settings, and to ensure optimal decision support systems for lung cancer in the future, standardization of Radiomics is needed.

Literature

1. Spear BB, Heath-chiozzi M, Huff J, Heath-chiozzi M. Clinical application of pharmacogenetics. 2001;7(5):201–4.

2. Administration D. Paving the Way for Personalized Medicine Paving the Way for Personalized Medicine : 2013;

3. Fallis A. Multifactoral decision support systems in radiation oncology: Clinical predictors and Radiomics. Journal of Chemical Information and Modeling. 2013. 1689-1699 p.

4. 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

5. 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.

6. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. Magn Reson Imaging [Internet]. Elsevier Inc.; 2012;30(9):1234–48. Available from: http://dx.doi.org/10.1016/j.mri.2012.06.010

7. Mitra, S.; Shankar BU. Integrating Radio Imaging With Gene Expressions Toward a Personalized Management of Cancer. Human-Machine Syst [Internet]. vol.44:664,677. Available from: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6828742&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6828742

8. Gillies, Robert J.; Gatenby RA. Radiomics of NSCLC H . Lee Moffitt Cancer Center and Research Institute. Radiomics of NSCLC. 2011;1–39.

9. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Sci Rep. 2013;3:3529.

10. 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

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2015;278(2):151169.

12. Gillies R. Cancer Imaging Program. Natl Cancer Inst [Internet]. Available from: http://imaging.cancer.gov/programsandresources/specializedinitiatives/qin/moffitt

13. Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep [Internet]. Nature Publishing Group; 2015;5(August):11075. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4525145&tool=pmcentrez&rendertype=abstract

14. Leijenaar RTH. OncoRadiomics Research software. Available from: http://www.oncoradiomics.com/

15. 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.

16. Balagurunathan AY, Gu Y, Wang H, Kumar V, Grove O. Paper Supplemental : 2014;1–33.

17. Burns AC and, Bush RF. Chapter 18 : Determining and Interpreting Associations Among Variables. Pearson Prentice Hall; 2003. 150-179 p.

18. NIWA. Concordance correlation coefficient [Internet]. Available from: https://en.wikipedia.org/wiki/Concordance_correlation_coefficient

19. McBride G. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Rep. 2005;HAM2005-06:14.

20. NIWA. Statistical calculators, Lin's Concordance [Internet]. 2016. Available from: https://www.niwa.co.nz/node/104318/concordance

21. Neuendorf. The Pearson Correlation Coefficient (r) vs. Lin's Concordance Coefficient (rc). :10–2. Available from: http://academic.csuohio.edu/kneuendorf/c63310/CorrsCompared10.pdf

22. Lin L-K. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45. 1989;

23. Lin L-K. A note on the concordance correlation coefficient. Biometrics 54. 2000;324–5.

24. 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: Definition of imaging features Maastro Clinic

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:

$$A = \sum_{i=1}^{N}\frac{1}{2}|a_ib_i \times a_ic_i|$$

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**:

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

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 δ pixels in direction α, 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:

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

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:

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

In this study, distance δ was set to 1 and direction α 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 δ and α,

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

μ 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,

μ_x be the mean of p_x,

μ_y be the mean of p_y,

σ_x be the standard deviation of p_x,

σ_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:

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

3.3.   Cluster Shade:

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

3.4.   Cluster Tendency:

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

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:

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

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:

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

3.18.   Maximum Probability:

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

3.19.   Sum average:

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

3.20.   Sum entropy:

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

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 θ, 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:

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

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

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

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 θ,

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:

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

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

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

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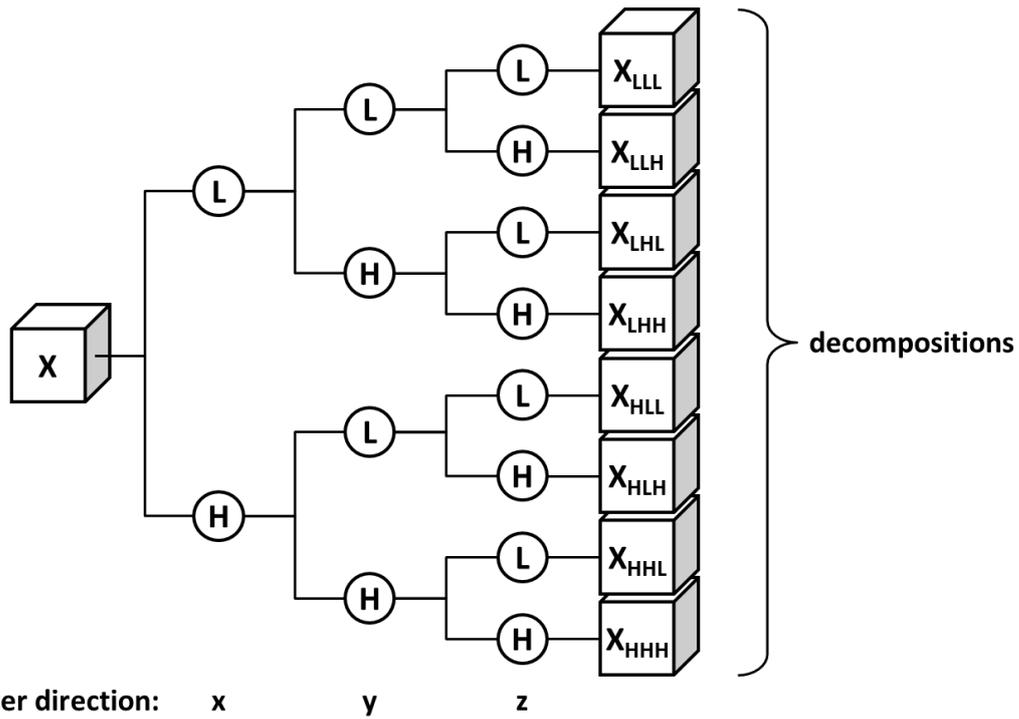
Supplement I Figure 1: Schematic of the undecimated three dimensional wavelet transform applied to each CT image. The original image X is decomposed into 8 decompositions, by directional low-pass (i.e. a scaling) and high-pass (i.e. a wavelet) filtering: X_{LLL}, X_{LLH}, X_{LHL}, X_{LHH}, X_{HLL}, X_{HLH},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

Appendix II: Definition of imaging features Moffitt Cancer Center

Image Feature List – Version 1.0

No.	Feature	Description	Calculation	Range
3D features				
1	Border Length	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. 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. For torus and image objects with holes the border length sums the inner and outer border	$$b_v = \left(\sum_{n=1}^{\#(slices)} b_v(\text{Slice}) \right) * u_{slices} + b_v(Z)$$ Parameters b_v: border length of image object v b_v(slice): border length of image object slice b_v(z): border length of image object in z-direction u_{slices}: spatial distance between slices in the coordinate system unit	[0, ∞]
2.	Length	The length of an image object is the largest 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. The length of an image object can be ≤ the largest of dimensions of the smallest rectangular 3D space enclosing the image object.		[0, ∞]
3.	Thickness	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 The thickness of an image object can be ≤ than the smallest of dimensions of the smallest rectangular 3D space enclosing the image object		[0, ∞]

4.	Width	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. 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.		[0, ∞]
5.	Length/Thickness	The length-to-thickness ratio of an image object.	Length/Thickness	[0, ∞]
6.	Length/Width	The length-to-width ratio of an image object.	Length/Width	[0, ∞]
7.	Number of Pixels	Number of pixels forming an image object.		[0, scene size]
8.	Volume	The number of voxels forming an image object rescaled by using unit information for x and y coordinates and distance information between slices.	$V_v = \#P_v * u^2 * u_{slices}$ V_v: volume of image object v $\#P_v$: total number of voxels contained in P_v u: size of a slice pixel in the coordinate system unit u_{slices}: spatial distance between slices in the coordinate system unit	[0, scene size]
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.	$$1 - \frac{\sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}}}$$ λmin: minimal eigenvalue λmax: maximal eigenvalue	[0, 1]
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:	$b_v/2(l_v + W_v)$ b_v: image object border length l_v: length of an image object v w_v : width of an image object v	[1, ∞] 1 = ideal.
11.	Compactness	A figure for the compactness of a 3D image object is calculated by a scaled product of its three eigenvalues	2 λ1*2 λ2*2 λ3/Vv	[0, ∞] 1 = ideal.

		2∗λ1, 2∗λ2, 2∗λ3 divided by the number of its pixel/voxel. We include a factor of 2 with each eigenvalue, since λi∗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.	λ1: eigenvalue 1 of a 3D image object v λ2: eigenvalue 2 of a 3D image object v λ3: eigenvalue 3 of a 3D image object v Vv: volume of image object v	
12.	Density	The Density feature describes the spatial distribution of the pixels of an image object. 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. It is calculated by the edge of the volume fitted cube divided by the fitted sphere radius.	$$\frac{\sqrt[3]{V_v}}{\sqrt{Var(X) + Var(Y) + Var(Z)}}$$ Vv: volume of image object v $\sqrt[3]{V}$: edge of the volume fitted cube $\sqrt{Var(X) + Var(Y) + Var(Z)}$: radius of the fitted sphere	[0, depended on shape of image object]
13.	Elliptic Fit	It describes how well an image object fits into an ellipsoid of similar size and proportions. While 0 indicates no fit, 1 indicates for a complete fitting image object. 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 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.	$\varphi = 2.\#\{x,y,z\} \in Pv : \varepsilon v(X,Y,Z) \leq 1 / \#Pv-1$ εv(x,y,z): elliptic distance at a pixel (x,y,z) Pv: set of pixels of an image object v #Pv: total number of pixels contained in Pv	[0, 1]; 1 = complete fitting, whereas 0 = only 50% or less voxels fit inside the ellipsoid
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 α between the resulting line of best fit and the z-axis is returned as		[0, 90]

		feature value		
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.	εv(xo,yo,zo) with (xo,yo,zo) = min εv(x,y,z), (x,y,z)∉Pv εv(x,y,z): elliptic distance at a pixel (x,y,z) Expression:	[0, ∞]
16.	Radius of Smallest Enclosing Ellipse	The calculation is based on an ellipsoid with the same volume as the image object and based on the covariance matrix. This ellipsoid is enlarged until it encloses the image object in total. The ratio of the radius of this smallest enclosing ellipsoid to the radius of the original ellipsoid is returned as feature value.	εv(xo,yo,zo) with (xo,yo,zo) = max εv(x,y,z), (x,y,z)∈σPv εv(x,y,z): elliptic distance at a pixel (x,y,z)	[0, ∞]
17.	Rectangular Fit	It describes how well an image object fits into a cuboid of similar size and proportions. 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.	$$\frac{\#\{(X,Y,Z) \in Pv : \rho v(X,Y,Z) \le 1\}}{\#Pv}$$ ρv(x,y,z): rectangular distance at a pixel (x,y) #Pv: total number of pixels contained in Pv	[0, 1]; 1 = complete fitting, whereas 0 = 0% fits inside the rectangular approximation
18.	Roundness	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.	$$\varepsilon_v^{max} - \varepsilon_v^{min}$$ εv max: radius of smallest enclosing ellipsoid εv min: radius of largest enclosed ellipsoid	[0, ∞]; 0 = ideal.
19.	Shape Index	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.	Bv/Vv bv: image object border length Vv: volume of image object v	[1, ∞]; 1 = ideal.

20.	Histogram features	The intensity histogram h(a) is the number of pixels occurred for brightness level "a" plotted against their 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. mean=$\sum_{i=1}^{range} i * prob(i)$ 2. sd=$\sqrt{\sum_{i=1}^{range}(i - mean)^2 * prob(i)}$ 3. skewness=$\frac{\sum_{i=1}^{range}(i-mean)^3*prob(i)}{(\sum_{i=1}^{range}(i-mean)^2*prob(i))^{1.5}}$ 4. kurtosis=$\frac{\sum_{i=1}^{range}(i-mean)^4*prob(i)}{(\sum_{i=1}^{range}(i-mean)^2*prob(i))^2}$ 5. energy=$\sum_{i=1}^{range} prob(i) * prob(i)$ 6. entropy=$-\sum_{i=1}^{range} prob(i) * Log(prob(i))$ Where intensity range is [0,range] (normalized) prob(i)=$\frac{hist(i)}{\sum hist(i)}$, hist(i) is the frequency of intensity i appears.	
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).	$p(i,j)$ 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. $GLN=\frac{1}{n_r}\sum_{i=1}^{M} p_g(i)^2$	

		11: Long Run High Gray-Level Emphasis (LGHGE).	4. $RLN=\frac{1}{n_r}\sum_{j=1}^{N}p_r(j)^2$ 5. $RP=\frac{n_r}{n_p}$ 6. $LGRE=\frac{1}{n_r}\sum_{i=1}^{M}\frac{p_g(i)}{i^2}$ 7. $HGRE=\frac{1}{n_r}\sum_{i=1}^{M}p_g(i)*i^2$ 8. $SRLGE=\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2*j^2}$ 9. $SRHGE=\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)*i^2}{j^2}$ 10. $LRLGE=\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)*j^2}{i^2}$ $LRHGE=\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}p(i,j)*i^2*j^2$	
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.	$p(i,j)$ 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$(p(i,j))$	
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	For each filtered images (125), the energy was calculated as following: $$Energy = \frac{1}{R}\sum_{i=N+1}^{I-N}\sum_{j=N+1}^{J-N}\sum_{k=N+1}^{k-N}h^2(i,j,k)\,,$$	

		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.	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.	$$Energy = \frac{1}{MxNxL}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L} I^2(i,j,k)$$ $$Entropy = \frac{-1}{MxNxL}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{L}(\frac{I^2(i,j,k)}{norm^2})\log(\frac{I^2(i,j,k)}{norm^2})$$ $I(i,j,k)$ 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(block) = 16 features 2 level: 2* 15(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.	Spherecity	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	Measure of distance from center of gravity to border of		

	gravity to border of tumor	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 III: Cooperation contract Maastro Clinic

SAMENWERKINGSOVEREENKOMST mv

Ondergetekenden, MAASTRO CLINIC, te Maastricht, Dr. Tanslaan 12, rechtsgeldig vertegenwoordigd door mevrouw drs M.J.G. Jacobs, Raad van Bestuur, verder te noemen MAASTRO en

Naam en voornamen:	Pijls, Mirthe Evelien
Adres en woonplaats:	Haghenstraat 63, 6461 VT te Kerkrade
Geboren op:	14 maart 1994 te Heerlen
Student aan:	Fontys Paramedische Hogeschool te: Eindhoven verder te noemen student, en
Onderwijsinstelling:	Fontys Paramedische Hogeschool te: Eindhoven verder te noemen, onderwijsinstelling.

verklaren hierbij een samenwerkingsovereenkomst te zijn aangegaan onder de volgende voorwaarden:

Artikel 1. Doel van de samenwerkingsovereenkomst
MBRT Afstudeerstage.

Artikel 2. De begeleiding door MAASTRO
MAASTRO wijst Claudia Offermann aan als contactpersoon, die verantwoordelijk is voor de opdracht en de begeleiding van de student/werkende.
Daarnaast fungeert de heer R. Leijenaar vanuit PTTeragnositic als extern adviseur.

Artikel 3. Tijdsduur en plaats van de opdracht
De student start met ingang van 15 februari 2016 tot en met 1 juli 2016 (voor gemiddeld 36 uur p/wk). Binnen de instelling wordt een 36-urige werkweek gehanteerd.

Artikel 4. Onderwijsactiviteiten
De opdracht kan worden onderbroken voor door de onderwijsinstelling georganiseerde terugkomactiviteiten.
Voor andere onderwijsactiviteiten (b.v. het afleggen van tentamens) kan de opdracht na voorafgaand overleg tussen onderwijsinstelling en MAASTRO worden onderbroken.

Artikel 5. De begeleiding door de onderwijsinstelling
De onderwijsinstelling wijst als aan mevrouw Esther Bloemen die zich bezighoudt met de begeleiding van de student bij de integratie van school- en praktijkervaring en de ontwikkeling van de beroepspersoonlijkheid van de student.

Artikel 6. Medische verklaring
De student dient, indien er sprake is van direct patiëntencontact, vóór aanvang van de opdracht een "verklaring van geen TBC" te overleggen. Deze verklaring mag niet ouder zijn dan 6 maanden (peildatum: 6 maanden voor aanvang van de werkzaamheden binnen MAASTRO). Indien deze verklaring niet overlegd wordt,

wordt deze overeenkomst ontbonden. De kosten van de verklaring worden op declaratiebasis gedragen door MAASTRO.

Artikel 7. Vergoedingen

De vergoeding bedraagt € 335,- bruto per maand op fulltime basis. De student ontvangt een reiskostenvergoeding woon-werkverkeer conform art. 12.1.1 van de cao, behoudens in die situatie dat de student over een ov-jaarkaart beschikt waarmee de reiskosten reeds zijn gedekt en onder de voorwaarde dat de stage tenminste 144 uur per studiejaar bedraagt. Ziekte- en verlofuren worden niet vergoed. De gewerkte uren zullen maandelijks op een werkstaat ingevuld worden, waarna de maand daaropvolgend de stagevergoeding wordt uitgekeerd.

Artikel 8. ID-card

De student ontvangt een ID badge van Maastro clinic voor de periode dat hij/zij hier werkzaam/aanwezig is en verleent daarbij Maastro clinic eenmalige machtiging om het bedrag van € 50,-- af te schrijven van bovengenoemd rekeningnummer indien de ID badge aan het einde van de samenwerking en/of aanwezigheid niet, binnen uiterlijk 1 week, geretourneerd wordt.

Artikel 9. Verzekeringen

MAASTRO sluit voor de student een verzekering af tegen het risico van aansprakelijkheid. De student is tijdens de opdracht verzekerd voor ZW en ZFW.

Student/werkende client zelf zorg te dragen voor aanmelding ziektekostenverzekering bij een door hem zelf gekozen ziektekostenverzekeraar voor bovengenoemde samenwerkingsperiode.

Artikel 10. De beoordeling van de opdracht

De beoordeling van de opdracht geschiedt conform de in het werkplan (van de onderwijsinstelling) van deze opdracht opgenomen beoordelingsregeling.

Artikel 11. Gedragscode integriteit

Inzake interne en externe integriteit is medewerker gehouden aan de 'gedragscode integriteit' van MAASTRO clinic en diens onderliggende regelingen en bepalingen, waaronder het informatiebeveiligingsbeleid.

Het openbaar maken van onderzoeksgegevens en onderzoeksresultaten uit het afstudeerproject, zoals vermeld in de scriptie, artikel e.d. is alleen mogelijk na toestemming van de begeleider van MAASTRO clinic.

Artikel 12. Overige bepalingen

- De student is verplicht tot geheimhouding van hetgeen hem ter kennis komt, voor zover die verplichting uit de aard der taak volgt of hem uitdrukkelijk is opgelegd. Deze verplichting geldt ook na beëindiging van de opdracht.
- De student houdt zich aan de in MAASTRO geldende voorschriften omtrent geheimhouding en de voor de desbetreffende beroepsbeoefenaren geldende beroepsethiekregels. Ter verzekering van de geheimhouding en met in-

achtneming van de beroepsethiek, heeft MAASTRO het recht zich op de hoogte te stellen van alle schriftelijke stukken of op andere wijze opgeslagen informatie die door de student in het kader van de opdracht wordt opgesteld en niet uitsluitend voor eigen gebruik dient.

- De student houdt zich aan de in MAASTRO geldende voorschriften op het gebied van orde, veiligheid en gezondheid en volgt de aanwijzingen gerechtigd personeel van MAASTRO.
- Ziekteverzuim dient voor 09.00 uur te worden gemeld bij de werkbegeleider en de onderwijsinstelling.

Art 13. Slotbepalingen

- Wijziging of aanvulling van deze stageovereenkomst is slechts mogelijk voor zover door de wijziging of aanvulling niet in strijd wordt gehandeld met de gesloten samenwerkingsovereenkomst en betrokken partijen hiermee instemmen.
- De overeenkomst eindigt:
    a. na afloop van de onder artikel 3 bepaalde tijd;
    b. na schriftelijke aanzegging door een der partijen indien:
    1. de wederpartij niet of op onvoldoende wijze de verplichtingen na komt die voor hem uit deze overeenkomst voortvloeien:
    2. zich zodanige omstandigheden voordoen, dat van de andere partij redelijkerwijs niet verlangd kan worden, dat hij de overeenkomst voortzet.
- Deze overeenkomst is geen arbeidsovereenkomst in de zin van artikel 7:610 van het Burgerlijk Wetboek en is ook niet als zodanig bedoeld.

Overeengekomen en in tweevoud opgemaakt en ondertekend te Maastricht, 22 februari 2016.

Personeelszaken Maastro	Student/werkende	Contact Maastro (voor gezien)
Mw. J. Boumans	Mw. M. Pijls	Mw. C. Offermann

Appendix IV: CCC values of the compared diameter features

Maastro features	Moffitt feature	CCC
Shape_maxDiameter2D1	Longest Diameter	0.68
Shape_maxDiameter2D2	Longest Diameter	0.69
Shape_maxDiameter2D3	Longest Diameter	0.57
Shape_maxDiameter3D	Longest Diameter	0.65

Appendix V: Calculated intercept, slope and R^2 for all first order statistic features

Moffitt features	Maastro features	Intercept	Slope	R^2
Energy	Stats_energy	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	0.99	0.99
Kurtosis HU	Stats_kurtosis	-0.02	1	0.99
Standard Deviation HU	Stats_std	0.08	0.99	0.99
Mean Absolute Deviation HU	Stats_md	-0.02	1	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	1	0.99
RMS HU	Stats_rms	1120.20	-0.86	0.77
Uniformity HU	Stats_uniformity	0	0.99	0.99

Appendix VI: CCC values for each resampling scheme in textural features

Feature catagory	Maastro	Moffitt	CCC (BV = 1)	CCC (BV = 5)	CCC (BV = 10)	CCC (BV = 25)	CCC (BV = 50)	CCC (NOB = 8)	CCC (NOB = 16)	CCC (NOB = 32)	CCC (NOB = 64)	CCC (NOB = 128)	CCC (NOB = 256)
Textural features:	RLGL_shortRunEmphasis	Short Run Emphasis (SRE)	0,017121198	0,118923754	0,351698324	0,893933795	0,514932262	0,200602606	0,372072294	0,680443204	0,92337401	0,439336776	0,157148669
Textural features:	RLGL_longRunEmphasis	Long Run Emphasis (LRE)	0,008104943	0,054648546	0,161686446	0,810803489	0,236320978	0,020843807	0,055144066	0,487234728	0,824569587	0,230305422	0,077138959
Textural features:	RLGL_grayLevelNonuniformity	Gray-Level Nonuniformity (GLN)	9,52456E-05	2,08912E-05	1,1767E-05	6,85926E-06	6,71744E-06	4,48958E-06	4,99045E-06	5,89407E-06	6,67844E-06	9,17752E-06	1,54958E-05
Textural features:	RLGL_runLengthNonuniformity	Run-Length Nonuniformity (RLN)	-3,37555E-06	-3,79779E-06	-4,3576E-06	-6,24181E-06	-9,95787E-06	-1,71188E-05	-1,34153E-05	-9,10665E-06	-6,36627E-06	-4,80204E-06	-4,01316E-06
Textural features:	RLGL_runPercentage	Run Percentage (RP)	0,015008165	0,099505082	0,285183021	0,906088551	0,505659753	0,210022385	0,3517921	0,726262771	0,890249671	0,363040215	0,132171673
Textural features:	RLGL_lowGrayLevelRunEmphasis	Low Gray-Level Run Emphasis (LGRE)	0,748928463	0,684645136	0,614451457	0,410930769	0,171716698	0,029219543	0,104989035	0,286468202	0,456323446	0,579281305	0,658160618
Textural features:	RLGL_highGrayLevelRunEmphasis	High Gray-Level Run Emphasis (HGRE)	-1,57506E-05	-0,0004629	-0,002690474	-0,042324063	-0,003638374	0,001626712	0,007187481	0,050805547	0,60165825	0,026987501	0,004916145
Textural features:	RLGL_shortRunLowGrayLevEmpha	Short Run Low Gray-Level Emphasis (SRLGE)	0,835574297	0,790883204	0,720744735	0,458605651	0,177288084	0,037342256	0,118598467	0,322925596	0,550854656	0,69326168	0,755328094
Textural features:	RLGL_shortRunHighGrayLevEmpha	Short Run High Gray-Level Emphasis (SRHGE)	-0,036175189	-0,002075306	-0,000835118	-0,000233365	-7,91063E-05	0,000187893	0,000370601	0,000815325	0,001712623	0,003546209	0,007881659
Textural features:	RLGL_longRunLowGrayLevEmpha	Long Run Low Gray-Level Emphasis (LRLGE)	0,750351005	0,467699205	0,24702269	0,074756916	0,023583028	0,00421898	0,01038051	0,049359604	0,14084043	0,328448873	0,584859183
Textural features:	RLGL_longRunHighGrayLevEmpha	Long Run High Gray-Level Emphasis (LRHGE)	0,000710066	0,023447715	0,165497152	0,685587441	0,53518344	0,235316093	0,394991941	0,42587267	0,847295338	0,113670877	0,013075582
Textural features:	GLCM_energy	avgCoocurrence-Energy	-0,001316533	-0,0010957	-0,000354236	0,005723905	0,040173255	0,366020633	0,116877081	0,046617225	0,010949927	0,001640839	-0,000588235
Textural features:	GLCM_contrast	avgCoocurrence-CONTRAST	0,062307045	0,179099693	0,033430612	0,004865783	0,001192022	0,000295315	0,001139091	0,004600826	0,018956893	0,084855979	0,441530604
Textural features:	GLCM_homogeneity1	avgCoocurrence-HOMOGENEITY	0,009563847	0,041716707	0,089066311	0,285558289	0,606348381	0,516949744	0,704092671	0,565627329	0,270331538	0,122780762	0,059948641
Textural features:	GLCM_maxProb	avgCoocurrence-MAXPROB	-0,000358038	-0,000392675	0,000239245	0,009649263	0,077833959	0,354327008	0,122637186	0,052851927	0,013873938	0,002768821	0,000302682
Textural features:	GLCM_entrop2	avgCoocurrence-ENTROPY	-0,029912711	-0,003627371	0,032959632	0,194828787	0,357197655	0,220757798	0,387211279	0,487244404	0,220043526	0,074709669	0,01694994
Textural features:	GLCM_sumAvg	avgCoocurrence-SUMMEAN	-0,000345716	-0,002317276	-0,00682399	-0,048626887	-0,014739857	0,012430651	0,036720666	0,173935279	0,185147822	0,038669261	0,012928887