

# **Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics**

by

**Tamarisk du Plessis**

Submitted in fulfilment of the requirements for the degree  
Doctor of Philosophy in Nuclear Medical Sciences

in the

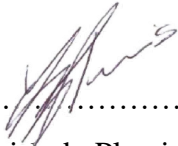
Department of Nuclear Medicine  
Faculty of Health Sciences  
**UNIVERSITY OF PRETORIA**

Supervisor: Prof MM Sathekge  
Co-Supervisor: Dr WID Rae  
Co-Supervisor: Prof NA Martinson

NOVEMBER 2023

## DECLARATION

I, Tamarisk du Plessis, hereby declare that this dissertation, *Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics*, submitted in accordance with the requirements for the Doctor of Philosophy degree at the University of Pretoria, is my own original work and has not previously been submitted to any other institution of higher learning. All sources cited or quoted in this research paper are indicated and acknowledged with a comprehensive list of references.



.....  
Tamarisk du Plessis

25 November 2023

## DEDICATION

I dedicate this study to my husband, Greg, and two children, Hayley and Gustav, for their endless support and for the time they allowed me to take to complete this study.

## ACKNOWLEDGEMENTS

I am deeply grateful to the following people who contributed on various levels to the success of this study. This study would not have been possible without their support, assistance, guidance and input.

- All the glory to God, my heavenly Father, for giving me everything I needed to complete this study.
- Co-supervisor, Dr William Rae, for introducing me to the field of radiomics, for his advice, motivation, his belief in my ability and his patient teachings. His passion for his work is contagious and a true inspiration.
- Supervisor, Prof Sathekge, and second co-supervisor, Prof Martinson, for their guidance and support.
- Ms Gopika Ramkilawon and Dr Tanita Botha (University of Pretoria) for their assistance with the statistical analysis required in this study.
- Ms Susan Scheepers (Information specialist at the University of Pretoria) and Ms Elaine Tam (Academic liaison librarian at the University of Sydney, Australia) for their contribution in the systematic review literature searches for Chapter 2.
- ICON Radiotherapy for contributing to the tuition fees for this study and for allowing me to do some parts of this study during office hours.
- Dr Andre Kyme (senior lecturer at the School of Biomedical Engineering at the University of Sydney) for setting up the virtual environment on the University of Sydney's super computers and for allowing me to use these computers for image segmentation.
- Ms Sarah Dixon (student at the School of Biomedical Engineering at the University of Sydney) for allowing me to use her automatic segmentation model as primary segmentation to mask the lungs on all chest x-rays.
- Perinatal HIV Research Unit (PHRU) for allowing me to retrospectively use their x-rays and data that was acquired exclusively for tuberculosis research purposes.
- Mr Floris Swanepoel (Data Coordinator at PHRU) for providing me with all the images and data, and for sorting all the relevant clinical data needed for this study.
- Dr Ziyaad Waja (Medical doctor at PHRU) for reporting on some of the x-rays in the dataset that had incomplete image reports.

## ABSTRACT

Pulmonary tuberculosis (PTB) is internationally one of the leading causes of death from a single infectious agent, and South Africa remains in the top 8 countries globally with the highest number of new infections. A chest x-ray (CXR) is still the most common radiological imaging procedure for PTB screening, diagnosis and monitoring, but it cannot be used as a standalone diagnostic tool due to the subjectivity associated with reporting. This can be addressed by quantifying digital CXR with tools such as radiomic feature extraction. Despite radiomics' increasing popularity, little evidence exists as to its application in non-neoplastic diseases such as PTB. A major limitation in the quantification of non-neoplastic diseases, especially from CXR, is the variability in disease segmentation which is the biggest cause of irreproducible radiomics outcomes.

Due to a lack of guidelines on interpolation and other post-acquisition pre-processing methods to use in radiomic studies carried out on planar projection x-ray images, a short comparative study was conducted in this thesis to determine its influence. It showed that image cropping was the only pre-processing step that significantly influenced the constructed radiomic signatures. Cropping was therefore carefully applied throughout this study.

In this thesis a unique sliding window segmentation method was developed to eliminate the difficult and time-consuming task of accurate PTB disease segmentation from planar images. It was applied as a secondary segmentation, superimposed on a primary automatic lung segmentation, that divided the entire lung region into uniform windows that overlapped while sliding over the CXR in both image dimensions. When radiomic features were extracted from each sliding window, it allowed the distribution of the features across the lung region to be evaluated.

The introduction of the secondary segmentation caused a significant increase to the dimensionality of the data which has not yet been encountered in any previous radiomic study. This necessitated thorough investigation of various statistical methods and dimensionality reduction and model construction algorithms. Correlation analysis (with a 0.8 cut-off value) proved to be the most robust method for dimensionality reduction in this study and it allowed the development of a reproducible PTB signature consisting of 11 texture features. This signature was used for the construction of a Random Forest radiomic model

which achieved good classification accuracy (AUC=0.9444 (95% CI, 0.8762; 0.9814)) for normal chest x-rays and PTB cavities.

To evaluate the change in the radiomic features over time, using serial images of the same patient, radiomic signature parameter maps were constructed using a unique approach of spatial mapping. For this the radiomic features extracted from each sliding window were re-distributed into the x- and y-coordinates in a two-dimensional matrix corresponding to the sliding window coordinates. These signature parameter maps accurately mimicked and highlighted the disease pathology on the radiological images when some image formatting methods were applied.

Further dimensionality reduction methods were applied to consolidate the signature parameter maps to radiomics scores. The change (baseline to first follow-up CXR) in these scores was compared to the change in its corresponding radiology- and clinical scores calculated using existing algorithms. A strong correlation (0.22 ( $p$ -value = 0.02)) was found between the radiological and the radiomic scores that were statistically significant. This showed that the developed radiomic scores were able to quantify the change in the disease characteristics as seen from digital CXR of patients diagnosed with PTB.

In this thesis three different outcomes were therefore achieved when radiomic feature extraction was applied to chest x-rays. Firstly a model was developed that can automatically differentiate normal CXR from CXR with PTB cavities, which could improve the accuracy of CXR reporting currently regaining prominence as a high-volume screening tool. Secondly, signature parameter maps that showed a strong correlation to the lung pathology were constructed. This might be valuable as a quantitative supplementary indicator in the management of PTB disease and further increase the acceptance of CXR as a tool for assessing the TB response in medical research and clinical practice. Finally, a radiomics score was constructed that was able to quantify the change in the disease characteristics as seen from digital CXR of patients diagnosed with PTB. This radiomic score analysis of serial x-rays taken while patients receive TB therapy has the potential to be a quantitative monitoring tool of response to therapy. Radiomics was therefore successfully applied in this study to quantify the characteristics of PTB from chest x-rays.

# TABLE OF CONTENTS

	Declaration	i
	Dedication	ii
	Acknowledgements	iii
	Abstract	iv
	Table of contents	vi
	List of tables, figures and abbreviations	viii
<b>Chapter 1</b>	<b>Introduction</b>	<b>1</b>
	1.1. Tuberculosis is an epidemic	1
	1.1.1. Tuberculosis diagnosis	2
	1.1.2. Tuberculosis prognosis and monitoring	2
	1.1.3. Imaging of pulmonary tuberculosis	3
	1.1.4. Radiological manifestations of pulmonary tuberculosis	3
	1.2. Physics of x-ray imaging	5
	1.2.1. Quantification of images	5
	1.2.2. Image formats	6
	1.3. Radiomics overview	7
	1.3.1. Radiomic libraries	9
	1.3.2. Limitations of radiomic feature extraction	10
	1.4. Description of study	11
	1.4.1. Patient selection	11
	1.4.2. Hypothesis	11
	1.4.3. Assumptions	11
	1.4.4. Aims and objectives	12
	1.5. Study overview	12
	1.6. References	15
<b>Chapter 2</b>	<b>Pulmonary tuberculosis diagnosis, differentiation and disease management: A review of radiomics applications</b>	<b>18</b>
<b>Chapter 3</b>	<b>The influence of post-acquisition image processing on a radiomic signature constructed from planar images</b>	<b>37</b>
<b>Chapter 4</b>	<b>Introducing a secondary segmentation to construct a radiomics model for pulmonary tuberculosis cavities</b>	<b>55</b>
<b>Chapter 5</b>	<b>Quantitative chest x-ray radiomics for therapy response monitoring in patients with pulmonary tuberculosis</b>	<b>75</b>

<b>Chapter 6</b>	Discussions	<b>98</b>
<b>Chapter 7</b>	Future perspectives	<b>106</b>
<b>Chapter 8</b>	Summary	<b>112</b>
<b>Annexures</b>		<b>113</b>
	<b>Annexure 1:</b> PhD committee approval letter	114
	<b>Annexure 2:</b> Research ethics committee approval letters	
	Ethics committee approval 2020	115
	Ethics committee approval 2021	116
	Ethics committee approval 2022	117
	<b>Annexure 3:</b> Flowcharts	
	Study overview: Flowchart of workflow followed in this study	118
	Chapter 3: Flowchart of workflow	119
	Chapter 3: Flowchart of image processing	120
	Chapter 4: Flowchart of workflow	121
	Chapter 5: Flowchart of workflow	122



# LIST OF TABLES, FIGURES AND ABBREVIATIONS

## List of tables

### Chapter 2

**Table 1:** Summary table that includes comparative, relevant information for the three studies reviewed in this article

**Table 2:** Summary of radiomic features used in each reviewed study

### Chapter 3

**Table 1:** Summary of features that displayed significant differences ( $p \leq 0.05$ ) when comparing the different interpolation methods with their respective individual group comparison results

**Table 2:** Summary of features that displayed significant differences ( $p \leq 0.05$ ) when comparing the various pre-processing methods with their respective individual group comparison results

**Table 3:** Summary of the signature features retained for each of the six datasets under consideration

### Chapter 4

**Table 1:** Details of the features retained for the two different consolidation measures in the cavity and normal datasets respectively when a 0.8 cut-off value was used in the Pearson's correlation analysis. The features not common to both datasets are highlighted in bold

**Table 2:** Performance comparison of the SD and variance radiomic signature-based models showing the AUC measures with the corresponding confidence intervals (95% CI), accuracy, sensitivity, specificity, and precision

### Chapter 5

**Table 1:** Clinical score chart used in this study with a maximum score of 12

**Table 2:** Radiological score conditions of the X-ray classification

**Table 3:** The following features were identified as signature features that highlight and quantify the disease and correlates to the lung pathology

**Table 4:** The correlation between the change in the TBscore, the RLscore and the RadScore as calculated using Spearman's correlation analysis with a p-value testing the significance of the correlation

## **Chapter 6**

**Table 1:** Performance indicators of the radiomic models or signatures developed in the literature study performed in Chapter 2 and compared to the results of this study (F)

**Table 2:** A summary (calculated as the average of all the normal and cavity CXR values) of the signature features for the standard deviation (SD) and variance models

## List of figures

### Chapter 1

**Figure 1:** Chest x-rays showing various radiological manifestations of pulmonary TB (a) Cavitory lesion on the upper left lung post-primary pulmonary TB (b) Dense opacity pleural effusion in the lower left lung of primary pulmonary TB (c) Bilateral diffuse miliary opacities of primary pulmonary TB (d) Ghon's complex of active TB (e) Dense homogenous opacity in right, middle and lower lobe (f) Bilateral hilar adenopathy of primary pulmonary TB.

**Figure 2:** Graphical representation of an original 4 x 4-pixel image with four grayscale intensities (0, 1, 2 and 3) indicating the relationship of pixels that will be considered to calculate each of the second order gray level matrixes. a) GLCM with co-occurrence in 1 direction, b) GLSZM with size zone of 4 pixels, c) GLRLM with run length of 3 pixels and d) NGTDM at 1 pixel distance with 8 neighbouring pixels.

**Figure 3:** Radiomics workflow and image processing scheme as suggested by the IBSI. (IH: intensity histogram; IVH: intensity-volume histogram; GLCM: grey level cooccurrence matrix; GLRLM: grey level run length matrix; GLSZM: grey level size zone matrix; NGTDM: neighbourhood grey tone difference matrix; NGLDM: Neighbouring grey level dependence matrix; GLDZM: grey level distance zone matrix.)

**Figure 4:** Schematic overview of the workflow followed in this study.

### Chapter 2

**Figure 1:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram for journal article screening and selection

### Chapter 3

**Figure 1:** Plot of the normalized means of the 93 features extracted using three different interpolation methods (Bilinear, Bicubic and Nearest Neighbour interpolation) to indicate the significant differences obtained in some of the radiomic features

**Figure 2:** Plot of the normalized means of the 93 features extracted using different image sizes (1024x1024 pixels and 256x256 pixels) to indicate the significant differences obtained in some of the radiomic features

**Figure 3:** Plot of the normalized means of the 93 features extracted using cropped and uncropped images to indicate the significant differences obtained in some of the radiomic features

**Figure 4:** Plot of the normalized means of the 93 features extracted using images segmented prior to interpolation and re-segmented images to indicate that no significant differences are seen in the extracted radiomic features

### **Chapter 4**

**Figure 1:** Output of the segmentation model (from left to right): The original image, the mask output (multiplied by 255 to be visible) and the mask superimposed with the image that was used to evaluate the segmentation accuracy

**Figure 2:** Above: Example of 5 sliding windows, sliding horizontally in the y axis (window coordinates  $[P_x, P_y] = [30,9], [30,10], [30,11], [30,12]$  and  $[30,13]$ ) superimposed on the lung mask and the CXR. Below: The same sliding windows, inverted to allow visualisation of the lung mask

**Figure 3:** The number of common features retained for the two different consolidation measures when different cut-off values were considered in the Pearson's correlation analysis

**Figure 4:** Receiver-operating characteristic (ROC) curves for the SD and variance signature-based random forest model using a random walk oversampling technique

### **Chapter 5**

**Figure 1:** A schematic overview of the process followed to develop a radiomics score (RadScore) in this study

**Figure 2:** An example of the output of the primary segmentation model, (a) The original image, (b) The mask (multiplied by 255 to be visually visible) and (c) The mask multiplied with the original image. This image was used to evaluate the accuracy of the primary segmentation model

**Figure 3:** First order feature parameter maps for a single baseline CXR of Patient A; (a)10<sup>th</sup> Percentile, (b)90<sup>th</sup> Percentile, (c)Energy, (d)Entropy, (e)Interquartile range, (f)Kurtosis, (g)Maximum, (h)Mean, (i)Mean-Absolute-Deviation, (j)Median, (k)Minimum, (l)Range, (m)Robust-Mean-Absolute-Deviation, (n)Root-Mean-Squared, (o)Skewness, (p)Total-Energy, (q)Uniformity and (r)Variance

**Figure 4:** Signature parameter map obtained for Patient A's baseline CXR

**Figure 5:** (a) Plot of the frequency proportions of a baseline CXR of Patient A indicating the four groups used to obtain a radiomics score and (b) Plot of the frequency proportions of the corresponding follow-up CXR of Patient A

**Figure 6:** Graph indicating the number of patients who showed a decline, no change or an improvement in the TBscore, RLscore and RadScore

**Figure 7:** Plot of the correlation between the TBscore and the RLscore

**Figure 8:** Plot of the correlation between the TBscore and the RadScore

**Figure 9:** Plot of the correlation between the RLscore and the RadScore

**Figure 10:** The radiomic signature parameter map (a) correlates strongly to the lung pathology on the CXR (b)

## **Chapter 7**

**Figure 1:** (a) A clinical example of an original unprocessed CXR (b) The same CXR with the Phillips Bone Suppression software applied that was able to successfully suppress all high-density anatomy

**Figure 2:** (a) An original CXR. (b) The same CXR, processed and inverted to be in the same format as the training set of the Dilated Conditional GAN bone suppression model

**Figure 3:** (a) Original input image of Patient A, (b) Output image of Patient A after the Dilated Conditional GAN bone suppression model was applied, (c) Original input image of Patient B and (d) Output image of Patient B after the Dilated Conditional GAN bone suppression model was applied

## List of abbreviations

AEC	Automatic-Exposure-Control
AI	Artificial intelligence
AUC	Area under the ROC (receiver operating characteristic) curve
CAD	Computer aided detection
CI	Confidence Interval
CT	Computed tomography
CXR	Chest x-ray(s)
DICOM	Digital Imaging and Communications in Medicine
DR	Digital radiography
DL	Deep learning
EMB	Ethambutol
GLCM	Gray Level Cooccurrence Matrix (GLCM)
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
HCC	Hepatocellular carcinoma
HIV	Human immunodeficiency virus
IBSI	Imaging Biomarker Standardization Initiative
ICC	Inter-class correlation
INH	Isoniazid
IQR	Inter quartile range
LASSO	Least shrinkage and selection operator
Lp/mm	Line pairs per millimeter
MAD	Mean absolute deviation
ML	Machine learning
MWMOT	Simulation and majority weighted oversampling technique
NGTDM	Neighbouring Gray Tone Difference Matrix
OPI	Opportunistic pulmonary infections

PA	Posterior-anterior
PARKS	Picture archiving and radiomics knowledge system
PCR	Polymerase chain reaction
PET	Positron emission tomography
PTB	Pulmonary tuberculosis
PZA	Pyrazinamide
RadScore	Radiomics score
RALE	Radiographic Assessment of Lung Edema
RIF	Rifampicin
RLscore	Radiology score
ROC	Receiver operating characteristics
ROI	Region of interest
SD	Standard deviation
SMOTE	Synthetic minority sampling technique
TB	Tuberculosis
TBscore	TB score
VOI	Volume of interest
WHO	World Health Organization
2D	2 dimensional
3D	3 dimensional

# CHAPTER 1

## Introduction

Images are more than pictures, they are data. Digital images are two- or three-dimensional matrixes consisting of hundreds, to thousands, of discrete spatial elements called pixels or voxels. In planar images each pixel contains quantitative information regarding the visual intensities present during acquisition of the image. When the inter-pixel relationships of the quantitative pixel data are studied an entire new dimension of information is uncovered. This exposes image information invisible to the human eye which is extremely advantageous in medical image interpretation and processing. Digital imaging and the current developments in computer science allow radiological image interpretation to evolve from subjective reporting to an objective science, and from being primarily a diagnostic tool to being a significant role player in personalized precision medicine.

This study utilized these developments and combined them with a unique segmentation method in radiomic feature extraction to quantify the characteristics of pulmonary tuberculosis as imaged using digital planar chest x-rays.

### 1.1 Tuberculosis is an epidemic

According to the World Health Organization (WHO) tuberculosis (TB) is one of the top ten causes of death worldwide, and the second leading cause of death, after COVID-19, from a single infectious agent.<sup>1</sup> Early and accurate diagnosis and proper disease management is key to fighting this global epidemic.

The World Health Organization launched an initiative called “The End TB Strategy” by 2035.<sup>2</sup> One pillar identified to achieve this strategy, is to intensify research and innovation for the rapid development of new high-impact tools that can dramatically change TB prevention and care.<sup>2</sup> Unfortunately TB is not usually classified as an international focus area for medical research as this disease is not common in the world’s most powerful countries (e.g. In 2021 only 2.9% of all global TB incidences were reported in the America’s and 2.9% in Europe).<sup>1</sup>



### **1.1.1 Tuberculosis diagnosis**

Diagnosis of TB disease is usually done when a person presents with symptoms suggestive of TB using a combination of biological and radiological methods.<sup>3</sup> Biological methods, polymerase chain reaction (PCR) technologies and culture conversions, are considered the gold standard for diagnosing pulmonary TB (PTB).<sup>3-4</sup> Digital radiography is a radiological tool used for PTB diagnosis,<sup>3</sup> with a posteroanterior (PA) chest x-ray (CXR) as standard orientation.<sup>4</sup> Although CXR has long been accepted as a definite indicator of TB disease,<sup>5</sup> it must still be used in combination with other diagnostic tools as the radiological abnormalities in PTB is not definitive. Many other diseases may mimic radiological PTB findings and this makes diagnosis from CXR alone impossible.<sup>6</sup>

### **1.1.2 Tuberculosis prognosis and monitoring**

Patients on PTB treatment are monitored clinically, bacteriologically, and often radiologically. Biological methods are currently used as the gold standard biomarker for PTB, with sputum culture-confirmed smear conversion (positive to negative) the most widely accepted biomarker used to predict a relapse-free cure of PTB.<sup>7</sup> Negative smear results allow patients to change onto the continuation phase of treatment, else another 2 months of the intensive treatment phase is followed.<sup>3</sup> The only other biomarker endorsed by the WHO is the detection of DNA in sputum, but this is only for limited use.<sup>8</sup> TB biomarker research remains a high priority and an active research field.<sup>8</sup>

The search for trusted radiological biomarkers is also yet to be achieved. CXR are currently used as a subjective and qualitative tool to assist in the management of PTB. CXR are typically reported on as either “better” or “worse” than the previous CXR. Literature suggests that a scheduled follow up CXR be taken after the 2 month intensive phase of treatment and visually compared to the initial CXR taken at diagnosis.<sup>4</sup> With an improved CXR and a negative culture at 2 months into treatment, patients may be allowed to start with the continuation phase of treatment.<sup>4</sup> It is also recommended that a CXR be taken at completion of treatment to establish a new baseline.<sup>4</sup>

### 1.1.3 Imaging of pulmonary tuberculosis

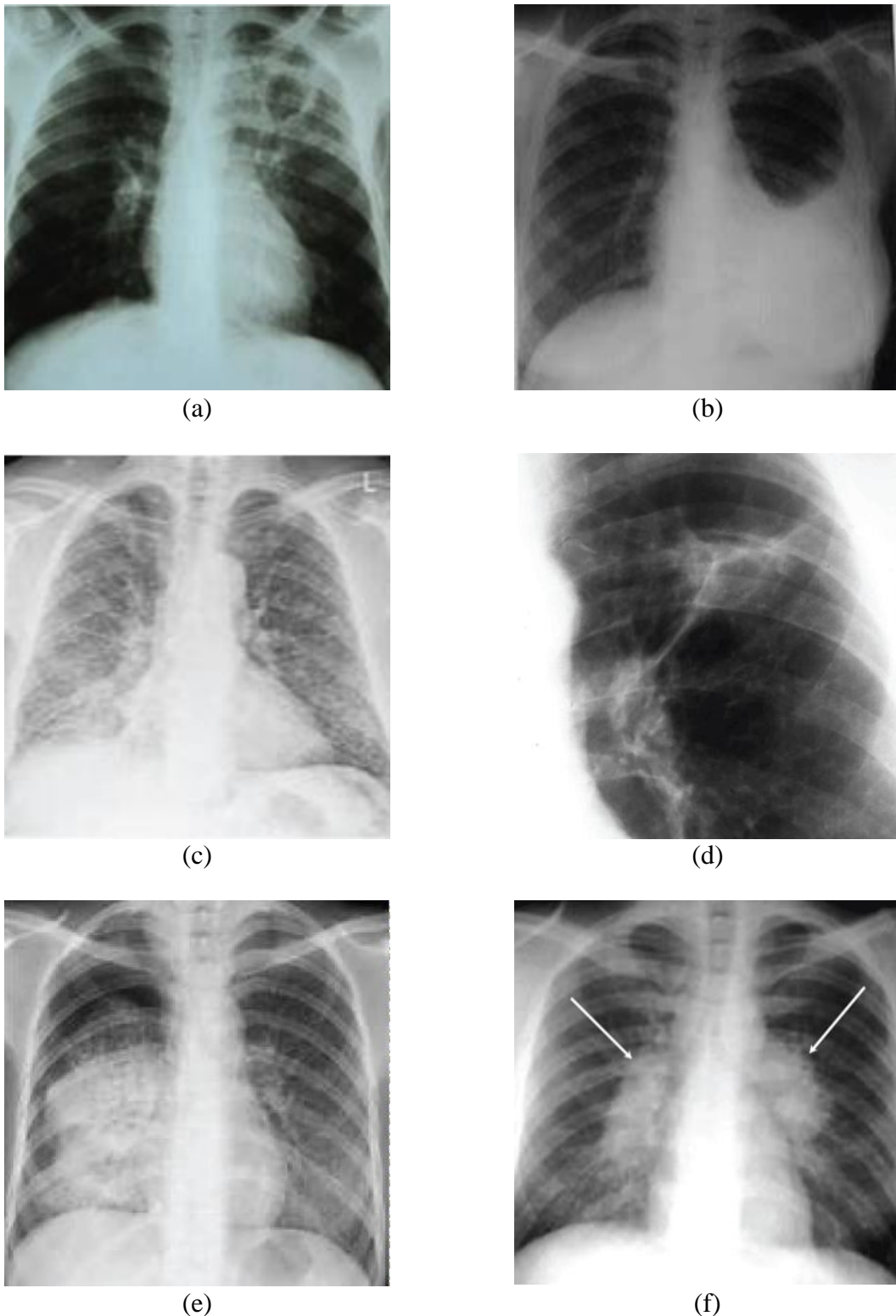
CXR remain the most common radiological imaging modality, even though more sophisticated 3-dimensional (3D) imaging modalities are available.<sup>9-11</sup> CT and PET/CT do provide superior clinical information to digital radiography (DR) because CT produces 3D images with no superimposed anatomy and PET images provide additional metabolic information. CT has a sensitivity of almost double that of CXR in detecting cavities, and is often used as complimentary imaging where resources are available.<sup>4</sup> The radiological manifestation of typical TB disease is however somewhat different on CT than on CXR.<sup>12</sup> Unfortunately 3D imaging resources are not widely available in developing countries and are therefore not considered as standard of care for TB diagnosis in these countries. Research has however proven that CXR are a very effective diagnostic modality as it yields high sensitivity and moderate to high specificity.<sup>9</sup> But since the abnormalities noted in TB imaging are not definitive, it reduces the specificity in all imaging modalities.<sup>2</sup>

Fixed bucky x-ray units should be the modality of choice when acquiring a CXR for PTB diagnosis. Some instances where this is not possible, e.g., in remote areas where bucky units are not available, or when a patient is confined to the ICU, mobile x-ray units with somewhat inferior image quality, are used.

### 1.1.4 Radiological manifestations of PTB

Pulmonary TB causes a wide variety of pathological changes in the lungs which lead to protean radiological manifestations. Primary PTB generally demonstrates radiologically as cavitory lesions (or cavitation) (figure 1a), pleural effusion (figure 1b), pericardial effusion, Miliary opacities (figure 1c), Ghon's complex (figure 1d), consolidation or Hilar and mediastinal lymphadenopathy (figure 1f).<sup>4,6</sup> The last being the most common radiological manifestation as seen in 83%-96% of paediatric and 10%-43% of adult primary PTB cases.<sup>4</sup> Various other radiological findings on CXR, such as infiltrate, may also indicate active PTB in patients who are not immunosuppressed (with high CD4 counts).<sup>4,6</sup> The radiological manifestations are however more complex in patients who are HIV-positive because the normal immunological response to TB is not present in these patients who are immunosuppressed.<sup>13</sup> The initial radiological features of postprimary PTB are broadly categorized as parenchymal disease with cavitation, consolidation, airway involvement,

pleural extension and other complications.<sup>13-14</sup> Inactive TB again may have some other radiological presentations.



**Figure 1:** Chest x-rays showing various radiological manifestations of pulmonary TB (a) Cavitory lesion on the upper left lung post-primary pulmonary TB (b) Dense opacity pleural effusion in the lower left lung of primary pulmonary TB (c) Bilateral diffuse miliary opacities of primary pulmonary TB (d) Ghon's complex of active TB (e) Dense homogenous opacity in right, middle and lower lobe (f) Bilateral hilar adenopathy of primary pulmonary TB.<sup>14</sup>

## 1.2 Physics of x-ray imaging

Radiological imaging is a non-invasive technique used to obtain anatomical information from inside the human body. Projection imaging is a two-dimensional (2D) superimposed representation of the 3-dimensional (3D) tissues lying between the x-ray source and the detector.<sup>15</sup> This leads to significant loss of image contrast, loss of all depth information and leads to ambiguity of relevant sizes of objects at different depths in the image.<sup>15</sup> There is therefore some uncertainty as to the exact spatial position of the anatomical information captured in each pixel, but this can be minimized by using large SIDs (source-image-distances) which reduces the relative magnification artifacts, and by using standard orientation and patient positioning procedures for all chest imaging of all patients. Another disadvantage is that objects of different depths are directly overlaid and it can be challenging to distinguish one from the other.<sup>15</sup> Some of the advantages of x-rays are that it is relatively inexpensive, fast and a good indicator of severity.<sup>5</sup> Digital x-rays also allow the radiologist to interpret the disease while the patient is onsite and can immediately refer the patient for further consultation and disease management as necessary.<sup>5</sup>

### 1.2.1 Quantification of images

In digital projection imaging, images are acquired by passing low energy ionizing radiation (x-rays) through a patient and detecting the transmitted energy on a solid-state detector after it exits the patient.<sup>15</sup> The intensity of the energy that exists the patient is stored to a 2D image matrix of pixels as grayscale values,<sup>16</sup> and is an indication of how much was differentially absorbed in the patient. It thus carries information about the attenuation of the original x-ray beam, and this can be related to the anatomical structures and pathologies within the patient. Pixels are therefore the discrete spatial elements (usually square elements) that are displayed as an image and stored as an image matrix. The size of the matrix is defined by the number of columns ( $M$ ) and rows ( $N$ ) of pixels and can mathematically be expressed by Equation 1 where the matrix is a function of  $f(i,j)$  where  $i$  and  $j$  indicates the spatial position of the elements.

$$f_{i,j} = \begin{bmatrix} f_{0,0} & \cdots & f_{0,M-1} \\ \vdots & \ddots & \vdots \\ f_{N-1,0} & \cdots & f_{N-1,M-1} \end{bmatrix} \quad [\text{Eq. 1}]$$

The spatial resolution of an image depends on the size of the pixels and is defined by the number of *line pairs per millimetre (lp/mm)*.<sup>17</sup> An approximate resolution for digital radiography is 4 lp/mm with 2.5 lp/mm the limiting spatial resolution.<sup>17</sup>

The value of the grayscale level (e.g., shades of gray) is calculated using Equation 2 where  $n$  is the bit depth. For example if an image has a pixel bit depth of 8, the grayscale values will range between 0 and 255 since each pixel is represented by a one 8-bit byte ( $2^8 = 256$ ) in the computer's memory.<sup>16</sup> In this case pixels with values close to 0 represents structures with less attenuation to the beam (e.g. lung) and pixels with values close to 255 represents structures with more attenuation to the beam (e.g. calcifications).<sup>16</sup> These numerical grayscale values are being used to quantify images and form the basis of all quantitative image processing studies. Planar DR images are typically produced with bit-depths of 10 or 12.<sup>15</sup>

$$\text{Grayscale level} = 2^n \quad [\text{Eq. 2}]$$

To enhance certain features on images and to assist humans with interpretation, many image post-processing algorithms have been developed.<sup>15</sup> Some of these processing algorithms include image filtering, noise reduction, and edge/feature enhancement within an image.<sup>15</sup> It is important to note that image processing cannot increase the amount, or quality, of data available in the image, but it can only use mathematical algorithms to remove information.<sup>15</sup> Post-processing is therefore always at the cost of loss of information, but is considered advantageous as the enhanced features are more easily visible and contrast is improved, allowing certain objects of interest to be more easily identified.

### 1.2.2 Image formats

The data stored within the image matrix of medical images can be stored in various image formats, e.g., jpeg, png, nrrt, tiff and DICOM (Digital Imaging and Communications in Medicine), but DICOM is the medical imaging standard.

Most image formats, e.g., png, store data in an RGB-mode. RGB-mode has 3 channels per pixel, one for red, one for green and one for blue, and thus each pixel takes 3 bytes of storage. Images can also be palletised (have a P-mode), where only the index into the fixed colour palette is saved. For example, an 8-bit image therefore have only 256 different colours and each pixel only requires 1 byte of storage. The advantage is that these pictures only require 1/3<sup>rd</sup> of memory space, compared to RGB-mode, but the limited number of unique colours can cause artefacts or banding. An L-mode image (or scalar image) is a single channel image

where only the luminance is stored and is normally interpreted as a grayscale. DICOM medical images are scalar images.

The biggest difference between DICOM and most other image formats is that a single DICOM file contains both a header as well as all the image data in various dimensions.<sup>18</sup> Other image formats save the header and the data in separate files. When DICOM images are used for data processing or image quantification it is essential to change the corresponding header group element with the image data. This dual processing makes DICOM images an unpopular format for image processing. Currently png or jpeg formats are the most popular image formats for image processing and for developing machine or deep learning models.

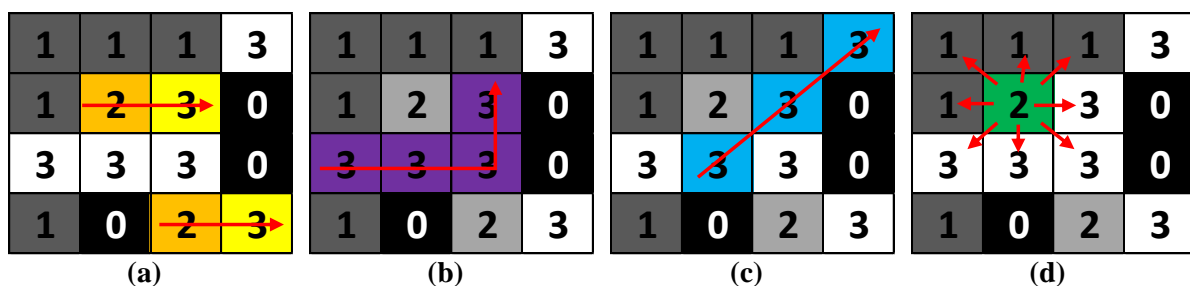
### 1.3 Radiomics Overview

The term ‘*radiomics*’ was first introduced in medicine in 2012.<sup>19</sup> Radiomics (as applied to radiology) is a field of medical study that has the potential to extract large amounts of quantitative features from medical images.<sup>20</sup> This multi-step process makes use of statistically based imaging analysis algorithms to identify image features which can be used to quantify disease characteristics.<sup>20</sup> This feature extraction tool therefore allows medical images to be converted into minable high-dimensional data.<sup>21</sup> Radiomics had its origin in the medical field of oncology,<sup>21</sup> and numerous articles have been published on the use of radiomics in tumours, but very few on the application in non-neoplastic diseases such as TB.<sup>22</sup> Chapter 2 is a thorough literature review on the literature currently available on the application of radiomics explicitly for the purpose of diagnosis, differentiation from other pulmonary diseases and disease management of PTB.

Radiomic features can be categorized into subgroups namely; shape-based features, first order statistical features, second order statistical features and higher order statistical features.<sup>23</sup> Shape-based features (morphological 2D and 3D features) describes the geometrical properties and shape of the ROI.<sup>23</sup> First order statistics describe the value and distribution of a group of pixels without concern for spatial relationships.<sup>23</sup> Second order statistics describe the textural features and are calculated by the statistical inter-relationship between the pixels in the ROI.<sup>24</sup> Second order statistics include five feature classes namely, Gray Level Cooccurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM) and Gray Level Dependence Matrix (GLDM).<sup>24</sup> These second order features are derived by

first applying one of the aforementioned algorithms to the original image matrix before applying first order statistics to the modified image matrix. These features provide a measure of the spatial arrangements of the pixel intensities and are influenced by various variables such as the pixel size and image orientation.<sup>23</sup>

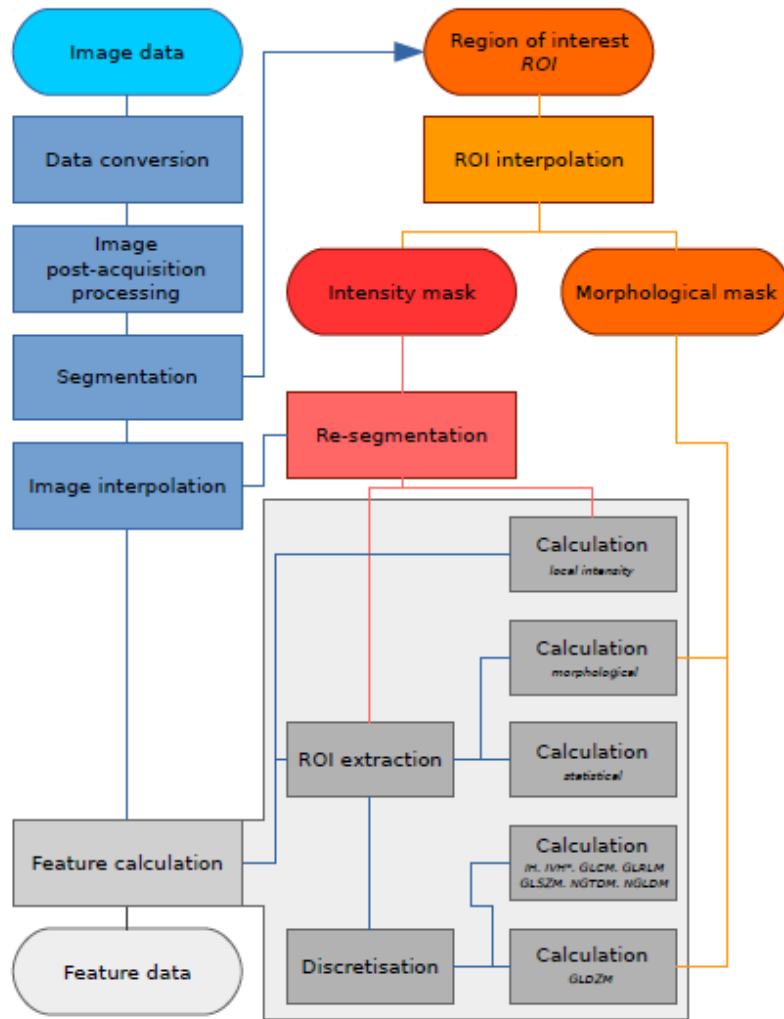
For example, GLCM features quantify the incidence of pixels with same intensities at a predetermined distance along a fixed direction. GLRLM features quantify consecutive pixels with the same intensity along fixed directions.<sup>23</sup> GLSZM features quantify gray levels zones, which is the number of connected pixels that share the same gray level intensity, in an image.<sup>25</sup> NGTDM features quantify the difference between a gray value and the average gray value of its neighbours within a specified distance.<sup>25</sup> GLDM features quantifies the number of connected pixels within a specified distance that are dependent on the centre pixel.<sup>25</sup> Figure 2 is a graphical representation of the how the second order image matrixes are derived.



**Figure 2:** Graphical representation of an original 4 x 4-pixel image with four grayscale intensities (0, 1, 2 and 3) indicating the relationship of pixels that will be considered to calculate each of the second order gray level matrixes. a) GLCM with co-occurrence in 1 direction, b) GLSZM with size zone of 4 pixels, c) GLRLM with run length of 3 pixels and d) NGTDM at 1 pixel distance with 8 neighbouring pixels.

Higher order statistics are obtained by applying filters or mathematical transforms to the image first before applying statistical algorithms. The purpose of these mathematical transforms is to identify repetitive- or non-repetitive patterns, to suppress noise or to highlight features.<sup>23</sup> Details on the specific features and the mathematical and statistical equations of these different feature classes can be found in literature.<sup>24</sup>

Figure 3 is the recommended radiomics workflow and image processing scheme as suggested by the Imaging Biomarker Standardization Initiative (IBSI).<sup>26</sup> Depending on the specific imaging modality and purpose, some steps may be omitted, e.g. *data conversion* is not required for planar images. The region of interest (ROI) is explicitly split into two masks, namely an intensity and morphological mask.<sup>26</sup> Feature calculation is expanded to show the different feature groups with specific pre-processing required.



**Figure 3:** Radiomics workflow and image processing scheme as suggested by the IBSI. (IH: intensity histogram; IVH: intensity-volume histogram; GLCM: grey level cooccurrence matrix; GLRLM: grey level run length matrix; GLSZM: grey level size zone matrix; NGTDM: neighbourhood grey tone difference matrix; NGLDM: Neighbouring grey level dependence matrix; GLDZM: grey level distance zone matrix.)

Every step in a the multi-step workflow of radiomics can influence the results and reproducibility.<sup>27</sup> To address this, the IBSI aims to standardize image biomarker nomenclature and definitions to standardize image biomarkers extraction.<sup>26</sup> The IBSI guidelines were adhered to where possible during this study.

### 1.3.1 Radiomic libraries

Many open source and commercial radiomic libraries are available for feature extraction for example; Matlab, MaZda, LIFEx, PyRadiomics, and IBEX.<sup>28</sup> PyRadiomics is one of the most popular libraries,<sup>29</sup> and will also be used for feature extraction in this study. It is an open



source radiomics library/toolbox developed by a team from Harvard Medical School. It is written in Python which is a common language in scientific computing. PyRadiomics is divided into two sections; Engineered features and Deep learning methods.<sup>24</sup> In this study only the engineered features will be used for the quantification of PTB features from x-rays.

Pyradiomics can extract up to 1500 features per images. These consists of shape-based descriptors and other features extracted from original and derived images. Derived images can be obtained by applying various Laplacian of Gaussian filters with 5 sigma levels with 1 level of Wavelet decompositions yielding 8 derived images.<sup>24</sup> Images can be derived using Square, Square Root, Logarithm and Exponential filters or Gradient and Local Binary Pattern (2D or 3D).<sup>24</sup> When no filters are used, the number of available features are reduced by approximately 1/15<sup>th</sup>.

Since naming standardization is still lacking in radiomics, the feature descriptions in this study all correspond to the features available for extraction in Pyradiomics, as suggested by the IBSI.<sup>26</sup>

### **1.3.2 Limitations of radiomic feature extraction**

Quantitative radiomics feature extraction poses many challenges and limitations. Many of these challenges will be discussed in the chapters to follow, but possibly the biggest challenge is that the number of possible features available to extract are often more than the number of patient images. This requires careful feature selection and dimensionality reduction to avoid overfitting.<sup>30</sup> Another challenge is standardization in radiomic studies as there are many different open source radiomic libraries available, all with slightly different functionalities.<sup>30</sup> The IBSI addresses many of these issues and compliance to these standards are of utmost importance to develop reproducible and robust radiomic signatures and models.<sup>26</sup> Lately research also focuses on the standardization of radiomic features not only on developing radiomics nomograms and models.<sup>31</sup>

In theory radiomics can be applied to any field of medicine that can be imaged, but planar images, with its own set of limitations, are still an uncommon imaging modality for radiomics studies. A few research groups for example do focus on the quantitative assessment of breast cancer from Mammograms using radiomics.<sup>32</sup> Mammograms are also 2D medical images and are subjected to the same superimposed limitations as CXR, with the exception that breast cancer presents as masses, while PTB is a non-neoplastic disease. Planar radiomic feature

extraction was also applied in a few recent studies published on the use of radiomic feature extraction from CXR in COVID-19 diagnosis.<sup>33-34</sup> This proved that, even though still uncommon, radiomic feature extraction can be successfully applied to study non-neoplastic diseases from CXR.

Many of these limitations and challenges still need to be addressed before radiomic feature extraction can be considered as standalone tool in the clinical environment. Regardless of this, image-derived information is expected to soon have a significant contribution as an integrated component in personalized medical treatment.<sup>30</sup>

## **1.4 Description of study**

### **1.4.1 Patient selection**

This is a retrospective study and used images, with their corresponding clinical data and radiological data, collected by researchers at the Perinatal HIV Research Unit (PHRU) based at the Chris Hani Baragwanath Academic Hospital, Soweto, South Africa, for other clinical TB research studies.

### **1.4.2 Hypothesis**

First order and textural radiomics features can be used to develop a radiomics signature from digital chest x-rays to quantitatively determine the textural changes and locality of the radiological manifestations of PTB. These features can then be used, in combination with the clinical data and radiological reports, in longitudinal studies to quantitatively measure disease status over time.

### **1.4.3 Assumptions**

For this study, the following assumptions were made:

- i. All radiological CXR reports were accurate
- ii. All CXR received were in the original DICOM format without any post-processing applied
- iii. All CXR were acquired with standard imaging protocols using an SID of 180 cm with the patient positioned close to the detector

#### 1.4.4 Aims and Objectives

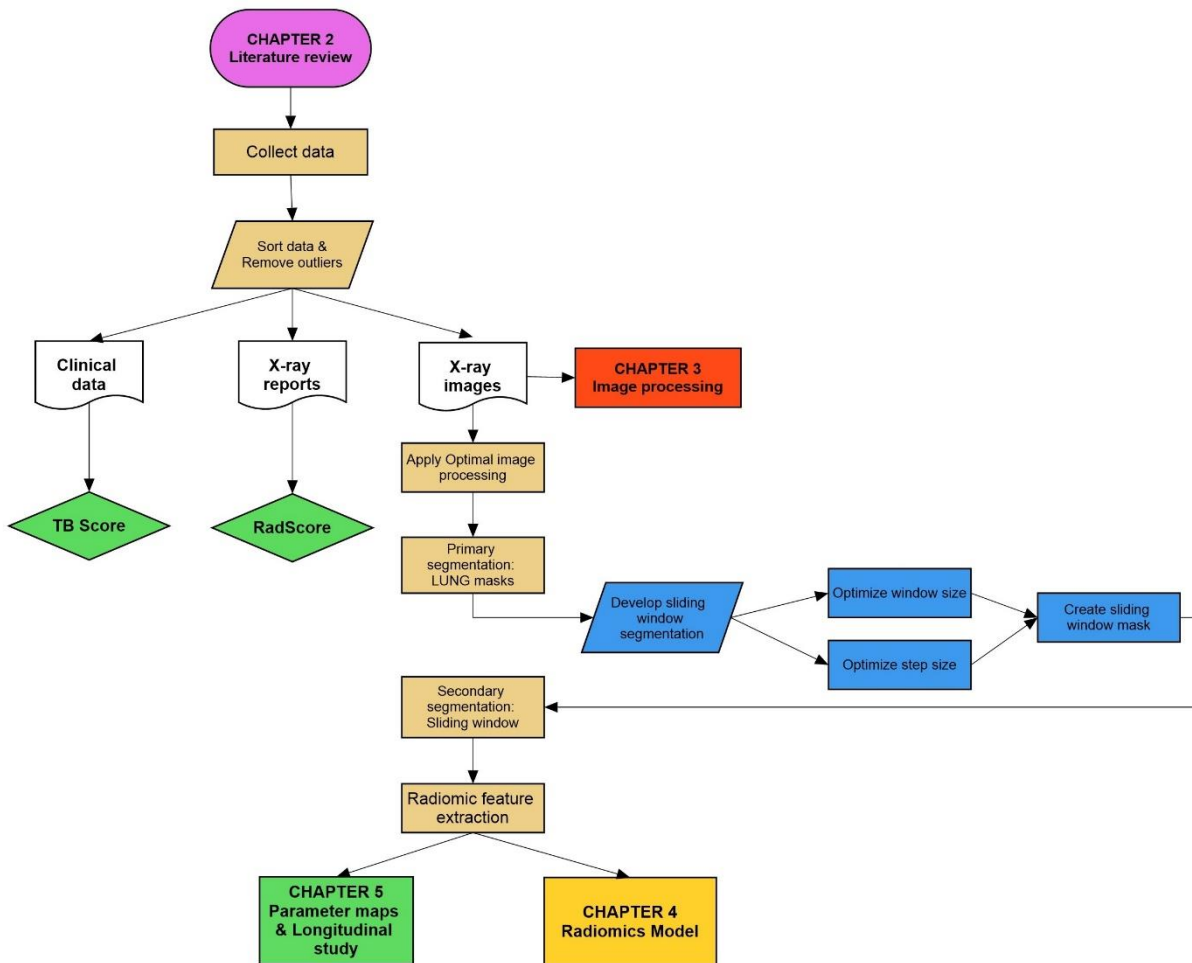
The aim of this study was to quantify clinical image characteristics of PTB from digital chest x-rays and to evaluate the time-course changes of these quantified features.

This study comprised of 4 objectives:

- i. Describe and perform quantitative feature extraction of clinical TB characteristics from chest x-rays using a novel secondary segmentation process and radiomic algorithms.
- ii. Perform dimensionality reduction processes of the extracted radiomics features to develop a radiomics signature that describes PTB characteristics from digital CXR.
- iii. Build a radiomics model that can automatically differentiate CXR with PTB disease expressions from normal CXR.
- iv. Do time-course analysis, from baseline and follow-up CXR, to test the signature's relevance and application in the monitoring of TB disease response and to evaluate the probability of using the developed radiomics signature as an imaging biomarker.

#### 1.5 Study overview

An apparent need has been identified in the literature for a tool that can quantitatively monitor TB treatment response. Many studies focus on the development of biological biomarkers to evaluate this treatment response,<sup>35-36</sup> but robust and reliable radiological biomarkers will be equally valuable. Chest x-rays are routinely acquired as part of TB diagnosis and in follow-up visits, and these large x-ray databases are therefore available at no extra cost. To utilize the wealth of information available in these databases the x-rays must be quantified. One such tool to quantify medical images is radiomic feature extraction. From these statistically based imaging analysis algorithms signatures and models can be developed that may address unmet clinical needs or compare it's performance to that of radiologists.<sup>37</sup> Figure 4 is a schematic overview of the study and the detailed workflow charts can be found in Annexure 3.



**Figure 4:** Schematic overview of the workflow followed in this study.

In chapter 2 of this thesis a review study, that followed a formal systematic review strategy, was performed to identify the research that has been done to date on the use of radiomic feature extraction on PTB diagnosis and disease management. It revealed a large gap that exists in current research studies to apply radiomic feature extraction to quantify TB characteristics. It also provided evidence that feature extraction is primarily applied to 3D images, and mostly to segmented tumours. This explained why all available image post-acquisition processing guidelines were only applicable to 3D images. Chapter 3 was therefore dedicated to determining the influence of image pre-processing and interpolation on the radiomic features and signatures extracted from 2D images, such as x-rays. The optimal pre-processing methods for x-rays that was confirmed in chapter 3 was then applied to all images in this study going forward.

Chapters 4 and 5 are original research that focus on the development of radiomic signatures, models and scores that reveal meaningful quantitative TB characteristics from digital planar medical images. A novel secondary segmentation method was developed in chapter 4 that

eliminates the need for precise disease segmentation and that allowed evaluation of the spread of radiomic features across the entire lung. Following this segmentation and dimensionality reduction, two radiomic models were constructed that performed very well to identify PTB cavities from CXR. This secondary segmentation however caused an immense increase in the dimensionality of the data. Chapter 4 briefly describes the mathematical and statistical processes followed during dimensionality reduction and model construction to deal with this increased dimensionality of the data.

A radiomics score was developed in chapter 5 to address the need identified for a quantitative tool that can assist in the management of PTB. In this chapter radiomic signature parameter maps were also created that showed excellent qualitative correlations to the images. These parameter maps can be used, not only as a visual tool enhancing radiological features, but also as a quantitative tool to assist in CXR reporting and especially in disease management. This can further increase the acceptance of chest x-rays as a tool for assessing the TB response in medical research and clinical practice.

In this thesis various endpoints, signatures, models, scores, and signature parameter maps, were successfully achieved by quantifying PTB characteristics from digital CXR using radiomics.

## 1.6 References

1. WHO. Global tuberculosis report 2022. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>: World Health Organization, 2022.
2. WHO. The end tb strategy. <https://www.who.int/publications/i/item/WHO-HTM-TB-2015.19>: World Health Organization, 2016.
3. Department\_of\_Health. National tuberculosis magement guidelines. Republic of South Africa: 2014 Contract No.: 1.
4. Nachiappan AC, Rahbar K, Shi X, Guy ES, Mortani Barbosa EJ, Shroff GS, et al. Pulmonary tuberculosis: Role of radiology in diagnosis and management. *RadioGraphics*. 2017; 37(1):52-72. doi:10.1148/rg.2017160032
5. Karargyris A, Siegelman J, Tzortzis D, Jaeger S, Candemir S, Xue Z, et al. Combination of texture and shape features to detect pulmonary abnormalities in digital chest x-rays. *International Journal of Computer Assisted Radiology and Surgery : A journal for interdisciplinary research, development and applications of image guided diagnosis and therapy*. 2016; 11(1):99-106. doi:10.1007/s11548-015-1242-x
6. Harisinghani MG, McCloud TC, Shepard J-AO, Ko JP, Shroff MM, Mueller PR. Tuberculosis from head to toe. *RadioGraphics*. 2000; 20(2):449-70. doi:10.1148/radiographics.20.2.g00mc12449
7. Chen RY, Dodd LE, Lee M, Paripati P, Hammoud DA, Mountz JM, et al. Pet/ct imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Science translational medicine*. 2014; 6(265):265ra166. doi:10.1126/scitranslmed.3009501
8. MacLean E, Broger T, Yerlikaya S, Fernandez-Carballo BL, Pai M, Denkinger CM. A systematic review of biomarkers to detect active tuberculosis. *Nature microbiology*. 2019; 4(5):748-58. doi:10.1038/s41564-019-0380-2
9. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Ayles H, Sánchez CI. On combining multiple-instance learning and active learning for computer-aided detection of tuberculosis. *IEEE Transactions on Medical Imaging*. 2016; 35(4):1013-24. doi:10.1109/TMI.2015.2505672
10. Santosh KC, Antani S. Automated chest x-ray screening: Can lung region symmetry help detect pulmonary abnormalities? *IEEE transactions on medical imaging*. 2018; 37(5):1168-77. doi:10.1109/TMI.2017.2775636
11. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *European Journal of Radiology*. 2020; 123 doi:10.1016/j.ejrad.2019.108774
12. Müller NL, Silva CIS Imaging of the chest. [Internet] Philadelphia: Saunders/Elsevier; 2008 [cited Access 2008 Access Date]. Available from: <http://www.clinicalkey.com.au/dura/browse/bookChapter/3-s2.0-C20091572048>.
13. Padyana M, Bhat RV, Dinesha M, Nawaz A. Hiv-tuberculosis: A study of chest x-ray patterns in relation to cd4 count. *North American journal of medical sciences*. 2012; 4(5):221-5. doi:10.4103/1947-2714.95904

14. Basem Abbas Al U. The radiological diagnosis of pulmonary tuberculosis (tb) in primary care. *Journal of Family Medicine and Disease Prevention*. 2018; 4(1) doi:10.23937/2469-5793/1510073
15. Dance DR, Christofides S, Maidment ADA, McLean ID, Ng K-H, International Atomic Energy A Diagnostic radiology physics : A handbook for teachers and students. [Internet] Vienna: International Atomic Energy Agency; 2014 [cited Access 2014 Access Date]. Available from: <http://www-pub.iaea.org/MTCD/Publications/PDF/Pub1564web-82725456.pdf>.
16. Tan L, Jiang J Digital signal processing : Fundamentals and applications. [Internet] London, United Kingdom: Academic Press, an imprint of Elsevier; 2019 [cited Access 2019 Access Date]. Available from: <https://www.sciencedirect.com/science/book/9780128150719>.
17. Williams MB, Krupinski EA, Strauss KJ, Breeden WK, Rzeszutarski MS, Applegate K, et al. Digital radiography image quality: Image acquisition. *Journal of the American College of Radiology*. 2007; 4(6):371-88. doi:10.1016/j.jacr.2007.02.002
18. Digital imaging and communications in medicine : (dicom). Washington, DC: National Electric Manufacturers; 1993.
19. 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. *European Journal of Cancer*. 2012; 48(4):441-6. doi:10.1016/j.ejca.2011.11.036
20. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magnetic resonance imaging*. 2012; 30(9):1234-48. doi:10.1016/j.mri.2012.06.010
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016; 278(2):563-77. doi:10.1148/radiol.2015151169
22. Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. *BMC Medical Imaging* [Internet]. 2019; 19(1):1-11. [cited Access Year Access Date]]. Available from: URL
23. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: The facts and the challenges of image analysis. *European Radiology Experimental*. 2018; 2(1):1-8. doi:10.1186/s41747-018-0068-z
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer research*. 2017; 77(21):e104-e7. doi:10.1158/0008-5472.CAN-17-0339
25. Han Y, Chen C, Tang L, Lin M, Jaiswal A, Ding Y, et al. Using radiomics as prior knowledge for abnormality classification and localization in chest x-rays. 2020.
26. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020; 295(2):328-38. doi:10.1148/radiol.2020191145

27. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *La radiologia medica*. 2021; 126(10):1296-311. doi:10.1007/s11547-021-01389-x
28. Koçak B, Durmaz EŞ, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: A practical guide for beginners. *Diagnostic and interventional radiology (Ankara, Turkey)*. 2019; 25(6):485-95. doi:10.5152/dir.2019.19321
29. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging*. 2020; 11(1):91. doi:10.1186/s13244-020-00887-2
30. Larue RT, Defraene G, De Ruyscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. *The British journal of radiology*. 2017; 90(1070):20160665. doi:10.1259/bjr.20160665
31. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, et al. Standardization of imaging features for radiomics analysis. *The Journal of Medical Investigation*. 2019; 66(1.2):35-7. doi:10.2152/jmi.66.35
32. Zhang HX, Sun ZQ, Cheng YG, Mao GQ. A pilot study of radiomics technology based on x-ray mammography in patients with triple-negative breast cancer. *Journal of X-ray science and technology*. 2019; 27(3):485-92. doi:10.3233/XST-180488
33. Hu Z, Yang Z, Lafata KJ, Yin F-F, Wang C. A radiomics-boosted deep-learning model for covid-19 and non-covid-19 pneumonia classification using chest x-ray images. *Medical physics*. 2022; 49(5):3213-22. doi:10.1002/mp.15582
34. Giraudo C, Frattin G, Fichera G, Motta R, Stramare R. A practical integrated radiomics model predicting intensive care hospitalization in covid-19. *Critical care (London, England)*. 2021; 25(1):145. doi:10.1186/s13054-021-03564-y
35. Drain PK, Gardiner J, Hannah H, Broger T, Dheda K, Fielding K, et al. Guidance for studies evaluating the accuracy of biomarker-based nonsputum tests to diagnose tuberculosis. *Journal of Infectious Diseases*. 2019; 220
36. Delia G, Elisa P, Simone AJ, Tom HMO. Tuberculosis biomarkers: From diagnosis to protection. *Infectious Disease Reports [Internet]*. 2016; 8. [cited Access Year Access Date]]. Available from: URL
37. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging*. 2020; 20(1):33. doi:10.1186/s40644-020-00311-4



## CHAPTER 2

### **Pulmonary tuberculosis diagnosis, differentiation and disease management: A review of radiomics applications**

*Tamarisk du Plessis<sup>1</sup>, William Ian Duncombe Rae<sup>2</sup>, Mike Michael Sathekge<sup>1</sup>*

1. Department of Nuclear Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa.
2. Discipline of Medical Imaging Sciences, University of Sydney, Sydney, Australia.

*Polish J. Medical Phys. Eng. 2021;27(4):251-259*

*doi: 10.2478/pjmpe-2021-0030*

**TITLE:**

Pulmonary tuberculosis diagnosis, differentiation and disease management: A review of radiomics applications

**SHORT TITLE:**

A review of radiomics in pulmonary TB

**ABSTRACT:**

Pulmonary tuberculosis is a worldwide epidemic that can only be fought effectively with early and accurate diagnosis and proper disease management. The means of diagnosis and disease management should be easily accessible, cost effective and be readily available in the high tuberculosis burdened countries where it is most needed. Fortunately, the fast development of computer science in recent years has ensured that medical images can accurately be quantified. Radiomics is one such tool that can be used to quantify medical images. This review article focuses on the literature currently available on the application of radiomics explicitly for the purpose of diagnosis, differentiation from other pulmonary diseases and disease management of pulmonary tuberculosis. Despite using a formal search strategy, only five articles could be found on the application of radiomics to pulmonary tuberculosis. In all five articles reviewed, radiomic feature extraction was successfully used to quantify digital medical images for the purpose of comparing, or differentiating, pulmonary tuberculosis from other pulmonary diseases. This demonstrates that the use of radiomics for the purpose of tuberculosis disease management and diagnosis remains a valuable data mining opportunity not yet realized.

## INTRODUCTION:

Pulmonary tuberculosis (PTB) is an ongoing global epidemic and has been identified as a research priority by the World Health Organization (WHO) and various countries' national health departments.<sup>1</sup> Early and accurate diagnosis and proper disease management is key to fighting this epidemic. To diagnose and manage PTB clinicians generally use a combination of biological methods (such as the Mantoux tuberculin skin test, the tuberculosis (TB) blood test or sputum smear tests),<sup>2</sup> demographic data and radiological methods.<sup>3-4</sup> Biological methods are used as PTB biomarkers, with a sputum smear culture conversion (smear-positive to smear-negative status) currently the most widely accepted biomarker to predict a relapse-free cure of PTB.<sup>5</sup> A meta-analysis study showed that this is not a reliable biomarker with a sensitivity of only 40% (95% CI 25–56) and specificity of 85% (95% CI 77–91) for predicting relapse.<sup>5</sup> Researchers therefore still face the challenge of identifying more sensitive and specific biomarkers that can be used to quantitatively evaluate TB disease progression and response to treatment.<sup>5-7</sup> Another challenge in fighting this global epidemic is the burden of multi-drug resistant TB, where India (27%), China (14%) and The Russian Federation (8%) carry the highest number of cases globally.<sup>1</sup> To ensure a reduction in incidences and deaths, this epidemic needs to be addressed universally.

The top 8 highest TB burdened countries, who accounts for two thirds of the global TB incidence, are all developing countries, with 44% of all global cases in the South-East Asian region.<sup>1</sup> In these countries chest x-ray (CXR) remains the most common radiological imaging modality for PTB screening, as access to more sophisticated three-dimensional (3D) imaging modalities are limited.<sup>8-9</sup> Even in countries where patients have easier access to 3D modalities, it was shown that CXR remains the foundation for imaging certain radiological expressions of PTB, e.g. parenchymal disease.<sup>10</sup> But CT is more sensitive in detecting many other radiological expressions, e.g. lymphadenopathy and early bronchogenic spread in post-primary TB.<sup>10</sup> Not only is CXR the most widely accessible imaging modality,<sup>11</sup> but by using projection imaging, radiation doses to the patients are kept to a minimum.<sup>12</sup> Radiation dose and long term radiation effects become a considering factor when screening large cohorts or when multiple follow-up images are acquired. Research has showed that CXR is a very effective diagnostic modality as it yields high sensitivity ( $\kappa$  0.78, 95%, CI 0.73-0.82) and moderate to high specificity (75.7%).<sup>8,13</sup> But to utilize these properties expert readers are needed to interpret these images, and these experts are often scarce in resource-limited countries.<sup>3</sup>

With the fast development of computer science in recent years various research studies have been conducted on the use of artificial intelligence (AI) applications (Machine Learning (ML) and Deep Learning (DL)) to assist with, among other things, medical image processing and interpretation.<sup>11</sup> ML is a subset of AI that makes use of statistical algorithms that enable machines to improve with experience, while DL is a subset of ML and is based on artificial neural networks that enable algorithms to train themselves. Lately ML and DL have allowed radiological image interpretation to evolve from subjective reporting to an objective science,<sup>11</sup> and from being a primarily diagnostic tool, to being a central role player in personalized precision medicine.<sup>14</sup>

PTB causes a wide variety of pathological changes in the lungs which lead to many different radiological manifestations, and AI systems that are developed must have the ability to adapt to and manage these distinct morphological patterns.<sup>15</sup> Pathological changes visible on CXR include changes in the lung shape, size and context (texture) which in turn influence the lung-field symmetry.<sup>9</sup> When radiologists perform bi-lateral comparisons on CXR, the differences in corresponding regions between the left and the right lung assist greatly in the detection of these abnormalities.<sup>16</sup> This same principle is mimicked when automatic TB detection systems use feature vectors of various sizes and combinations,<sup>9,17-18</sup> with different classification algorithms to extract information from radiological images.<sup>9,19</sup>

To date AI applications are mainly used for the diagnosis of TB, but a few other studies have been done, for example on the automatic differentiation of drug-sensitive TB from drug-resistant TB from CXR.<sup>19</sup> This study tested various classifiers by using them in combination with a set of shape and texture features. It achieved the best performance, with an area under the receiver operating characteristic curve (AUC) of 66%, when using a traditional artificial neural network (ANN).<sup>19</sup> Many studies have also been done on differentiating PTB from other pulmonary diseases, often lung cancer, using both CXR and CT scans. The reason for this being that PTB mimics various other pulmonary diseases, and diagnostic imaging of PTB remains challenging.<sup>10</sup>

Automatic PTB screening systems are extremely useful, efficient and can be a low cost mass screening tool which can be well utilized in resource-constrained countries.<sup>20</sup> DL methods for some diseases do not yet perform to the same accuracy as radiologists, but it has been demonstrated that automatic PTB detection from CXR have reached radiologist level performance.<sup>21</sup> However research into automated radiological image analysis for the purpose

of disease management and prognosis remains limited.<sup>21</sup> For quantitative disease management, the automatic algorithms need to produce measurable outputs. One such method to quantify medical images is by using radiomic feature extraction. Although recently developed DL networks have led to more robust models for radiomics, and this has allowed the high-throughput extraction of quantitative features from radiological images, this has not been widely utilised as quantitative outputs need to be extracted from the networks and are generally hidden within the layers of the DL network or within the last layer of the algorithm.<sup>22</sup>

Radiomics is based on the hypothesis that the correct combination of these features, together with the clinical data, can identify significant tissue properties useful in the management of a disease.<sup>23</sup> Radiomics differs from traditional Computer-aided-detection (CAD) systems in the sense that CAD systems were designed to only diagnose or detect a disease.<sup>24</sup> Radiomic feature extraction can also be used to quantify disease characteristics and progression from medical images, as it makes use of statistically based imaging analysis algorithms to act as quantitative biomarkers for the identification of radiological features.<sup>24-25</sup> These can be used to quantify change and categorisation and not only identify and categorise predicted outcomes or disease states. This feature extraction tool allows medical images to be converted into minable multi-dimensional statistical data sets which characterise the relationship between the high dimensional data of the images.<sup>24</sup>

Radiomics is a complex, multi-step field of study that includes the following identifiable steps: image acquisition, image segmentation, feature extraction and qualification, analysis and database development.<sup>25</sup> Database development includes developing a radiomics signature or nomogram. A radiomics signature is a computational model built to meet specific clinical needs,<sup>26</sup> whereas a radiomics nomogram integrates a radiomics signature with the clinical data to evaluate parameters such as prognosis or disease management.<sup>27</sup> To development accurate, robust and reproducible radiomics signatures, the knowledge and skills of qualified and experienced researchers are crucial. They need to understand the influence of exposure parameters, image pre- and post-processing, image segmentation and mask modelling. Experienced researchers also play a vital role in optimisation and standardisation of image acquisition protocols, modelling, developing algorithms and the statistical analysis of high dimensional data.

Extracted features can be sub-categorized into first order statistical features, shape-based features (morphological 2D and 3D), textural features (or second order statistical features) and higher-order statistical features.<sup>28</sup> Higher order statistics are obtained by applying filters or mathematical transforms to the image before applying statistical algorithms.<sup>28</sup> Each one of the steps in radiomics poses its own set of challenges. Rizzo et al. explained it well in a narrative review article; “Radiomics: the facts and the challenges of image analysis”.<sup>23</sup>

Radiomics had its origin in the medical field of oncology,<sup>24</sup> and numerous articles have been published on the use of radiomics in tumours, but very few on the application in non-neoplastic diseases such as TB.<sup>29</sup> The purpose of this review article is to determine what research has already been done on the application of radiomics explicitly for the diagnosis and management of PTB, or on the differentiation of PTB from other pulmonary diseases. This will also reveal areas not yet addressed in the available literature and the potential opportunities for future research. Even though this is a narrative review, a formal systematic research strategy was followed to ensure that all published journal articles on this topic were included.

## **METHODS:**

### ***Review search strategy***

A comprehensive search for relevant literature was done by two independent information specialists at two separate universities. The following data bases were included in the search: PubMed (US National Library of Medicine), OVID Medline, Scopus, OVID Embase, CINAHL and IEEE Xplore.

The search strategy that was applied to text-words can be summarized as follows: ((radiomics OR radiomic\*) OR ('imaging biomarker\*' OR 'radiological biomarker\*' OR 'texture feature\*' OR 'texture analysis')) AND (tuberculosis OR TB). The year limit that was used was 2000 to May 2021 (week 2) when the search was concluded. The term 'radiomics' was first introduced in medicine in 2012,<sup>30</sup> and the year 2000 was an arbitrarily selected date prior to 2012.

### ***Study selection criteria***

All journal articles were considered, but only studies that met the following inclusion criteria were included: 1) Original studies with full-text articles published in English. 2) Studies with

a human study population who were diagnosed with pulmonary tuberculosis. 3) Articles that included medical images (from any imaging modality: CT, MRI, PET/CT and x-rays) as datasets. 4) Studies that used radiomics or texture feature extraction methods. 5) Articles with the aim to develop imaging- or radiological biomarkers for PTB.

Exclusion criteria: 1) Articles that did not extract quantitative information from medical images. 2) Non-peer reviewed academic journal articles, including conference proceedings.

### ***Search results and data extraction***

The search strategies returned a total of 66 journal articles (n = 18 in the PubMed database, n = 16 in the OVID Medline database, n = 5 in Scopus database, n = 6 in OVID Embase, n = 5 in CINAHL and n = 16 IEEE Xplore database). After duplicates were removed, n = 49 remained. Since a systematic review search strategy was followed, the search results were recorded and summarized accordingly in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram (figure 1).<sup>31</sup>

## **RESULTS:**

Despite the thorough search strategy that was followed to ensure complete inclusion of all available published data, only 5 articles could be found that met all the inclusion criteria on the application of radiomics to PTB. The relevant detail for each of the studies, such as radiomics libraries, models, algorithms, dimensionality reduction methods, clinical features, etc. used, is summarized in Table 1.

*Study A:* In the first article by Bei et al. [2019] radiomic features extracted from CT images were used to create a radiomics signature capable of distinguishing primary progressive PTB from community-acquired pneumonia in children.<sup>29</sup> Manual segmentation was performed by a radiologist with more than 10 years of experience and verified by a second radiologist.<sup>29</sup> The radiologist delineated the margins of pulmonary consolidation as the first region-of-interest (ROI) and the mediastinal lymph nodes as the second ROI.<sup>29</sup> In this study they developed two radiomics signatures, one from each ROI, and constructed a radiomics model by combining these two signatures.<sup>29</sup> Finally they used the least shrinkage and selection operator (LASSO) algorithm to build a predictive nomogram by combining the radiomics model with the clinical data.<sup>29</sup> The predictive nomogram's classification outperformed the

senior radiologist's clinical judgement (AUC=0.971, 95% CI:0.912-1 vs. AUC=0.832, 95% CI:0.677-0.987).<sup>29</sup>

*Study B:* In the second article Shi et al. [2019], used radiomic features to assist with the identification of opportunistic pulmonary infections (OPIs) misdiagnosed as lung cancers in patients with human immunodeficiency virus (HIV).<sup>32</sup> In this study 76.2% of the OPI cohort had PTB and 23.8% other pulmonary infections.<sup>32</sup> Semi-automatic segmentation of the lesion was done using inhouse software.<sup>32</sup> The ROI borders were then manually adjusted by a radiologist to ensure that the lesion boundary was entirely included and that the bronchi and vessels were excluded.<sup>32</sup> The morphological CT features, clinical data and radiomic features were statistically compared between the two disease groups. They concluded that radiomics might assist with the identification of OPIs mimicking lung cancers for central-type lesions. Four radiomic features in these lesions were significantly different (large dependence high gray level emphasis (LDHGLE) (P=0.008), skewness (P=0.017), inverse difference normalized (IDN) (P=0.017) and kurtosis (P=0.017).<sup>32</sup> But they found that radiomics features of the peripheral-type lesions might not be useful for differentiating the diseases.<sup>32</sup>

*Study C:* In the third article Feng et al. [2020] used radiomic features to differentiate between lung tuberculoma and adenocarcinoma presenting as solitary pulmonary solid nodules.<sup>33</sup> They used a U-net based volume-of-interest (VOI) segmentation method to automatically delineate the boundary of the lesions.<sup>33</sup> A radiomics signature consisting of 6 features was identified and combined with clinical data to build a predictive radiomics nomogram using LASSO logistic regression.<sup>33</sup> The signature showed improved diagnostic accuracy compared to any single model (AUC=0.9064, 95% CI:0.9390-0.9931).<sup>33</sup>

*Study D:* In this article Cui et al. [2020] developed a radiomics nomogram model, using LASSO algorithms, to differentiate TB from lung cancer from pre-operative lung CT data.<sup>34</sup> They have also evaluated different radial dilation distances outside the lesion to determine the best performance.<sup>34</sup> The nomogram showed good discriminative performance in distinguishing TB from lung cancer (AUC=0.914, sensitivity=0.788, specificity=0.907).<sup>34</sup> This quantitative study again shows improved detection and discrimination performance of medical images when using the radiomics nomogram compared to decisions made by radiologists alone.<sup>34</sup>

*Study E:* In the final article Du et al. [2021] developed CT, PET and PET/CT radiomics signatures. These signatures were combined with semantic features to develop radiomics



nomograms to differentiate between active pulmonary TB and lung cancer.<sup>35</sup> Nine CT-based semantic features (maximal tumor diameter, tumor location, cavitation, vacuole, spiculation, vessel convergence, lobulation, pleural indentation and air bronchogram) and two PET-based semantic features (radionecrosis and metabolic activity greater than the adjacent mediastinal blood pool) were included.<sup>35</sup> They showed that the performance of the CT signature was superior to that of the PET signature (AUC=0.86 vs. 0.79,  $p=0.1585$ ), and that the PET/CT signature improved diagnostic performance even further compared to CT alone (AUC=0.91 vs. 0.86,  $p=0.0247$ ).<sup>35</sup> They concluded that PET and CT radiomic features can offer complementary diagnostic value when used in combination with the semantic features defined by radiologists.<sup>35</sup>

## DISCUSSION:

In this review study only five articles could be identified that met the inclusion criteria. In all 5 articles radiomic feature extraction was successfully used to quantify images for the purpose of comparing or differentiating pulmonary tuberculosis from other pulmonary diseases (pneumonia<sup>29</sup>, lung cancers<sup>32,34-35</sup> and adenocarcinomas<sup>33</sup>). All imaging modalities (CT, MRI, Molecular imaging and Planar imaging) were included in the search, but only CT and PET/CT (study E only) scans were used for feature extraction. It is also interesting to note that all five studies were carried out in China. This might be because China is the third highest TB burdened country (accounting for 8.4% of all global instances),<sup>1</sup> they are technologically very advanced and most patients have access to 3D imaging resources.

Four of the reviewed studies<sup>29,33-35</sup> built radiomics nomograms, while one study<sup>32</sup> only developed a radiomics signature. Table 2 is a summary of the significant radiomic features organized according to the feature groups. There was minimal correlation between the features selected as significant in the five studies. The only two features that were identified by more than one study to be significant, were Kurtosis<sup>32-33</sup> and 10<sup>th</sup> Percentile<sup>34-35</sup>. Kurtosis is a measure of the sharpness of the peak of the distribution of the values in the region of interest,<sup>28</sup> and 10<sup>th</sup> percentile of the N<sup>th</sup> voxel in the ROI is a first order statistical feature.<sup>28</sup> Not one study identified any higher-order statistical features as significant.

There were some weaknesses noticed in the articles reviewed. The training and validation cohorts of most studies (all except for study C which had both an internal and external validation cohort) were from the same centers. It is however preferable to have external

validation cohorts when DL methods are used to reveal possible overfitting of the training data sets. All studies were also retrospective studies, which is the only way to obtain a reasonable sample size in a new field of study. But the downside to this is that the researchers had no control on the acquisition parameters and the management of the data sets.

Unfortunately acquisition and post processing play a vital role in high throughput quantitative image analysis.

The studies identified between 4 and 11 (mean  $7.6 \pm 2.4$ ) radiomics features as significant from 99 to 3556 (mean  $1415.8 \pm 1239.3$ ) features extracted.<sup>29,32-33</sup> Small data sets are generally a limitation of radiomic studies. Thousands of radiomic features are available for extraction, but datasets are often smaller than the number of possible features to mine. This was also the case in the reviewed articles where the primary cohort sample sizes ranged from 24 to 319 (mean  $134.8 \pm 98.9$ ) patients, while the number of features extracted ranged from 99 to 3556 (mean  $1415.8 \pm 1239.3$ ).<sup>32-34,36</sup> The number of features extracted are almost 10 times more than the number of patients. A prospective multicenter study with larger cohorts is necessary to confirm the results and improve reliability of all studies. This suggestion was also acknowledged by most authors from the reviewed studies. From the limited number of studies available no preliminary conclusions can be made regarding which features or feature groups are likely to produce robust and reliable quantitative image information from PTB radiological images.

Radiomics is intrinsically a complex multi-step process, but to further complicate this is a lack of standardized guidelines and definitions. A lack of consensus on general definitions makes study intercomparison difficult. Since reproducibility and robustness are vital in radiomics studies, all details regarding the image processing and feature extraction should be clearly recorded and disclosed. The Image Biomarker Standardization Initiative (IBSI) was published in 2019 and hope to address various issues that will improve the validation and reproducibility of radiomics studies.<sup>37</sup> All five of the reviewed articles did well to comply with these guidelines, with the most recent<sup>35</sup> conforming the best.

The low number of studies included in this review article might seem like unconvincing evidence to address the purpose of this study, but this does indicate two things. Firstly, that high-throughput quantitative image analysis, especially in non-neoplastic deceases, is a new, but rapidly growing, field of study.<sup>37</sup> This is evident from the publication dates of the 5 articles that qualified to be included in this review study (2019 to 2021). Secondly, it shows

that the possibility of using radiomics for quantitative image analysis for the purpose of PTB disease management and differentiation from other pulmonary diseases is still an under investigated field of study.

## **CONCLUSIONS:**

Six data bases were searched and only five articles could be found detailing the application of radiomics to pulmonary tuberculosis. In all five articles reviewed, radiomic feature extractions from CT or PET/CT images were used effectively to quantify digital medical images for the purpose of comparing or differentiating tuberculosis from another pulmonary disease.

The outcome of this study evidently raises two questions.

1. Firstly, why has no attempt yet been made to use radiomics for the quantitative management and prognosis of pulmonary tuberculosis? Particularly while researchers still have a major challenge to identify more sensitive and specific biomarkers that can be used to identify the different stages of tuberculosis and to quantitatively evaluate disease progression or response to treatment.
2. Secondly, why are all current radiomics studies on pulmonary tuberculosis performed from three-dimensional imaging modalities (CT or PET/CT scans) when patients in countries where pulmonary tuberculosis is most prevalent have very limited access to these modalities?

It can be concluded from this review that the application of radiomics feature extraction and analysis of pulmonary tuberculosis, for the purpose of quantitative disease management and prognostication from chest x-rays, is a valuable data mining opportunity yet to be realized. This might seem challenging at this point in time, but with the fast development of computer science, and creative application of established mathematical solutions, this might be realized in the near future.

## REFERENCES:

1. WHO. Global tuberculosis report 2020. Report. Geneva: World Health Organization, 2020.
2. Prevention CfDCa [Internet]. Testing for tuberculosis (tb). Available from: [https://www.cdc.gov/tb/publications/factsheets/testing/tb\\_testing.htm](https://www.cdc.gov/tb/publications/factsheets/testing/tb_testing.htm).
3. Tan JH, Acharya UR, Tan C, Abraham KT, Lim CM. Computer-assisted diagnosis of tuberculosis: A first order statistical approach to chest radiograph. *Journal of Medical Systems*. 2012; 36(5):2751-9. doi:10.1007/s10916-011-9751-9
4. Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, et al. Official american thoracic society/infectious diseases society of america/centers for disease control and prevention clinical practice guidelines: Diagnosis of tuberculosis in adults and children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2017; 64(2):111-5. doi:10.1093/cid/ciw778
5. Chen RY, Dodd LE, Lee M, Paripati P, Hammoud DA, Mountz JM, et al. Pet/ct imaging correlates with treatment outcome in patients with multidrug-resistant tuberculosis. *Science translational medicine*. 2014; 6(265):265ra166. doi:10.1126/scitranslmed.3009501
6. Drain PK, Gardiner J, Hannah H, Broger T, Dheda K, Fielding K, et al. Guidance for studies evaluating the accuracy of biomarker-based nonsputum tests to diagnose tuberculosis. *Journal of Infectious Diseases*. 2019; 220
7. Delia G, Elisa P, Simone AJ, Tom HMO. Tuberculosis biomarkers: From diagnosis to protection. *Infectious Disease Reports [Internet]*. 2016; 8. [cited Access Year Access Date]]. Available from: URL
8. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Ayles H, Sánchez CI. On combining multiple-instance learning and active learning for computer-aided detection of tuberculosis. *IEEE Transactions on Medical Imaging*. 2016; 35(4):1013-24. doi:10.1109/TMI.2015.2505672
9. Santosh KC, Antani S. Automated chest x-ray screening: Can lung region symmetry help detect pulmonary abnormalities? *IEEE transactions on medical imaging*. 2018; 37(5):1168-77. doi:10.1109/TMI.2017.2775636
10. Skoura E, Zumla A, Bomanji J. Imaging in tuberculosis. *International Journal of Infectious Diseases*. 2015; 32:87-93. doi:<https://doi.org/10.1016/j.ijid.2014.12.007>
11. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *European Journal of Radiology*. 2020; 123 doi:10.1016/j.ejrad.2019.108774
12. Mettler FA, Jr., Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology*. 2008; 248(1):254-63. doi:10.1148/radiol.2481071451

13. Hoog AH, Meme Hk Fau - van Deutekom H, van Deutekom H Fau - Mithika AM, Mithika Am Fau - Olunga C, Olunga C Fau - Onyino F, Onyino F Fau - Borgdorff MW, et al. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. (1815-7920 (Electronic))
14. Lambin P, Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, et al. Radiomics: The bridge between medical imaging and personalized medicine. 2017.
15. Hogeweg L, Sánchez CI, Maduskar P, Philipsen R, Story A, Dawson R, et al. Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis. *IEEE Transactions on Medical Imaging*. 2015; 34(12):2429-42. doi:10.1109/TMI.2015.2405761
16. Ginneken Bv, Katsuragawa S, Romeny BMtH, Kunio D, Viergever MA. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging*. 2002; 21(2):139-49.
17. Shen R, Cheng I, Basu A. A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. *IEEE Transactions on Biomedical Engineering*. 2010; 57(11):2646-56. doi:10.1109/TBME.2010.2057509
18. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Reither K, Breuninger M, et al. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays. *IEEE Transactions on Medical Imaging*. 2015; 34(1):179-92. doi:10.1109/TMI.2014.2350539
19. Jaeger S, Juarez-Espinosa OH, Candemir S, Poostchi M, Yang F, Kim L, et al. Detecting drug-resistant tuberculosis in chest radiographs. *International journal of computer assisted radiology and surgery*. 2018; 13(12):1915-25. doi:10.1007/s11548-018-1857-9
20. Abideen ZU, Ghafoor M, Munir K, Saqib M, Ullah A, Zia T, et al. Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks. *IEEE Access*. 2020; 8:22812-25. doi:10.1109/ACCESS.2020.2970023
21. Summers RM. Are we at a crossroads or a plateau? Radiomics and machine learning in abdominal oncology imaging. *Abdominal Radiology*. 2019; 44(6):1985-9. doi:10.1007/s00261-018-1613-1
22. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: Threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European Radiology Experimental*. 2018; 2(1):1-10. doi:10.1186/s41747-018-0061-6
23. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: The facts and the challenges of image analysis. *European Radiology Experimental*. 2018; 2(1):1-8. doi:10.1186/s41747-018-0068-z
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016; 278(2):563-77. doi:10.1148/radiol.2015151169

25. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The process and the challenges. *Magnetic resonance imaging*. 2012; 30(9):1234-48. doi:10.1016/j.mri.2012.06.010
26. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging*. 2020; 20(1):33. doi:10.1186/s40644-020-00311-4
27. Liu Q, Li J, Liu F, Yang W, Ding J, Chen W, et al. A radiomics nomogram for the prediction of overall survival in patients with hepatocellular carcinoma after hepatectomy. *Cancer Imaging*. 2020; 20(1):82. doi:10.1186/s40644-020-00360-9
28. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer research*. 2017; 77(21):e104-e7. doi:10.1158/0008-5472.CAN-17-0339
29. Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. *BMC Medical Imaging [Internet]*. 2019; 19(1):1-11. [cited Access Year Access Date]]. Available from: URL
30. 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. *European Journal of Cancer*. 2012; 48(4):441-6. doi:10.1016/j.ejca.2011.11.036
31. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLOS Medicine*. 2009; 6(7):e1000097. doi:10.1371/journal.pmed.1000097
32. Shi W, Zhou L, Peng X, Ren H, Wang Q, Shan F, et al. Hiv-infected patients with opportunistic pulmonary infections misdiagnosed as lung cancers: The clinicoradiologic features and initial application of ct radiomics. *Journal of thoracic disease*. 2019; 11(6):2274-86. doi:10.21037/jtd.2019.06.22
33. Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *European Journal of Radiology*. 2020; 128 doi:10.1016/j.ejrad.2020.109022
34. Cui EN, Yu T, Shang S-J, Wang X-Y, Jin Y-L, Dong Y, et al. Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans. *World Journal of Clinical Cases*. 2020; 8(21):5203-12. doi:10.12998/wjcc.v8.i21.5203
35. Du D, Gu J, Chen X, Lv W, Feng Q, Rahmim A, et al. Integration of pet/ct radiomics and semantic features for differentiation between active pulmonary tuberculosis and lung cancer. *Molecular Imaging & Biology*. 2021; 23(2)
36. Cui EN, Yu T, Shang SJ, Wang XY, Jin YL, Dong Y, et al. Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans. *World journal of clinical cases*. 2020; 8(21):5203-12. doi:10.12998/wjcc.v8.i21.5203

37. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020; 295(2):328-38.

## TABLES:

**Table 1:** Summary table that includes comparative, relevant information for the three studies reviewed in this article

Study	A	B	C	D	E
<b>Author</b>	<b>Bei et al.</b>	<b>Shi et al.</b>	<b>Feng et al.</b>	<b>Cui et al.</b>	<b>Du et al.</b>
<b>Year of publication</b>	<b>2019</b>	<b>2019</b>	<b>2020</b>	<b>2020</b>	<b>2021</b>
<b>Imaging modality</b>	CT	CT	CT	CT	PET & CT & PET/CT
<b>Institution where data sets were acquired</b>	Beijing Childrens Hospital, Beijing, China	Shanghai Public Health Clinical Center, Shanghai, China	2 Unknown Centres in China	Liaoning Cancer Hospital in China	1 Unknown Hospital in China
<b>Date data sets were obtained</b>	Jan 2011 - Jan 2018	Jun 2013 - Feb 2018	Jan 2014 - Dec 2018	Jan 2012 - Oct 2018	Jan 2013 - March 2019
<b>Sample size</b>	115 (53 TB, 62 CAP)	73 (24 OPI & 49 LungCA)	426	478 (244 TB & 234 LungCA)	174 (77 TB & 97 LungCA)
<b>Primary cohort</b>	86	24 (19 TB, 5 other OPI)	123	319	122
<b>Validation cohort</b>	29	49	303 (121 - Internal & 182 - External)	159	52
<b>Significant clinical features used</b>	Duration of fever	Age, Smoking, HAART duration, CD4 <sup>+</sup> T cell count, CD4 <sup>+</sup> /CD8 <sup>+</sup> ratio	Gender, Age, Lesion size, Location, Lesion margin, Lobulated sharp, Spiculation sign	No clinical features included	9 CT based & 2 PET based semantic features
<b>Segmentation method</b>	Manual segmentation by radiologist (>10yrs experience). Second radiologist verification.	Semi-automatic inhouse software. (Manual adjustment by 2 radiologists)	U-net based Deep Learning model (Python)	Manual segmentation by 2 radiologist (12yrs & 14yrs experience)	Manual segmentation by nuclear physician (>3yrs experience). Second physician verification.
<b>Feature extraction software</b>	Matlab	PyRadiomics	Matlab	PyRadiomics	Matlab (SERA software)
<b>Number of features extracted</b>	970	99	3556	1967	487
<b>Type/Category of features extracted</b>	Gray intensity (First order statistics), Shape & Size, Texture, Wavelet	First order statistics, Shape, Texture	First order statistics, Shape & Size, Texture, Wavelet	First order statistics, Shape and Texture	First order, Morphological, Intensity based, Higher order statistics
<b>Dimensionality reduction method</b>	LASSO	n/a	LASSO	LASSO	Pearson correlation analysis & LASSO



Study	A	B	C	D	E
<b>Classifier used to build predictive nanogram</b>	Linear SVM	n/a	Not specified	Multivariable Logistic Regression analysis ('rms' package in R)	Multivariable Logistic Regression analysis
<b>Number of radiomic features</b>	11	4	6	8	PET: 3, CT: 5, PET/CT: 9
<b>Features selected to build radiomics nomogram or signature</b>	X7_fos_maximum	LDHGLE	ZSV_GLSZM_0.5_0.67_Equal_8	lbp_2D_firstorder_Entropy	stat_p10
	X0_GLCM_maximum_probability	Inverse difference normalized (IDN)	Kurtosis_Global_1_1.5_Equal_8	lbp_3D_k_firstorder_10Percentile	morph_asphericity
	X6_GLCM_IMC1	Skewness	complexity_NGTD_M_1_1.2_Lloyd_16	log_sigma_3_0_m_3D_glc_m_ldn	cm_info_corr1_2D_avg
	X1_GLRLM_LRLGLE	Kurtosis	HGZE_GLSZM_1_0.67_Lloyd_8	log_sigma_5_0_m_3D_RunLength-NonUniformity	szm_size_3D
	X1_GLRLM_LRE	---	SZHGE_GLSZM_1_0.67_Lloyd_8	squareroot_gldm-Dependence-NonUniformity	ngl_lde_3D
	Max3D		SRHGE_GLRLM_1_1.5_0.67_Lloyd_64	wavelet_HLH_glc_m_ldn	dzm_zdnu_3D
	Sph_dis		---	wavelet_HLL_glc_m_ldn	morph_a_dens_mvee
	Compactness1			wavelet_LLL_glc_m_ldmn	cm_clust_shade_2D_mrg
	Surface_to_volume_ratio			---	PET_cm_info_corr1_2D_mrg
	X2_fos_minimum				PET_cm_info_corr2_2D_mrg
	X0_GLRLM_LRHGLE				PET_stat_p10
	---				PET_morph_asphericity
					CT_szm_size_3D
					CT_ngl_lde_3D
					CT_morph_a_dens_mvee
					CT_cm_clust_shade_2D_mrg
					CT_dzm_zdnu_3D

(Abbreviations: TB = Tuberculosis, CAP = Community Acquired Pneumonia, OPI = Opportunistic Pulmonary Infections, LungCA = Lung Cancer, HAART = Highly Active Antiretroviral Therapy, LASSO = Least absolute shrinkage and selection operator, SVM = Support Vector Machine, GLCM = Gray Level Cooccurrence Matrix, GLSZM = Gray Level Size Zone Matrix, GLRLM = Gray Level Run Length Matrix, GLDM = Gray Level Dependence Matrix, NGTDM = Neighboring Gray Tone Difference Matrix LDHGLE = Large dependence high gray level emphasis)

**Table 2:** Summary of radiomic features used in each reviewed study

Study	A	B	C	D	E*
	Bei et al.	Shi et al.	Feng et al.	Cui et al.	Du et al.
<b>1st order features</b>	X7_fos_maximum	Skewness	Kurtosis_Global_1_1.5_Equal_8	lbp_2D_firstorder_Entropy	PET_stat_p10
	X2_fos_minimum	Kurtosis		lbp_3D_k_firstorder_10Percentile	
<b>Shape</b>	Max3D	---	---	log_sigma_3_0_mm_3D_glcml_dn	PET_morph_sphericity
	Sph_dis				
	Compactness1			log_sigma_5_0_mm_3D_RunLength-NonUniformity	CT_morph_area_mvee
	Surface_to_volume_ratio				
<b>GLCM</b>	X0_GLCM_maximum_probability	Inverse difference normalized (IDN)	---	wavelet_HLH_glcml_dn	CT_cm_clust_shade_2D_mrg
	X6_GLCM_IMC1			wavelet_HLL_glcml_dn	PET_cm_info_corr1_2D_mrg
				wavelet_LLL_glcml_dmn	PET_cm_info_corr2_2D_mrg
<b>GLRLM</b>	X1_GLRLM_LRE	---	SRHGE_GLRLM_1.5_0.67_Lloyd_64	---	---
	X0_GLRLM_LRHGLE				
	X1_GLRLM_LRLGLE				
<b>GLDM</b>	---	LDHGLE	---	squareroot_gldm_Dependence-NonUniformity	---
<b>GLSZM</b>	---	---	HGZE_GLSZM_1_0.67_Lloyd_8	---	CT_szm_size_3D
			SZHGE_GLSZM_1_0.67_Lloyd_8		
			ZSV_GLSZM_0.5_0.67_Equal_8		
<b>GLDZM</b>	---	---	---	---	CT_dzm_zdnu_3D
<b>NGTDM</b>	---	---	complexity_NGTDM_1_1.2_Lloyd_16	---	---
<b>NGLDM</b>	---	---	---	---	CT_ngl_lde_3D

(Abbreviations: GLCM = Gray Level Cooccurrence Matrix, GLRLM = Gray Level Run Length Matrix, GLDM = Gray Level Dependence Matrix, GLSZM = Gray Level Size Zone Matrix, GLDZM = Gray Level Distance Zone Matrix, NGTDM = Neighboring Gray Tone Difference Matrix, NGLDM = Neighboring Grey Level Dependence Matrix)

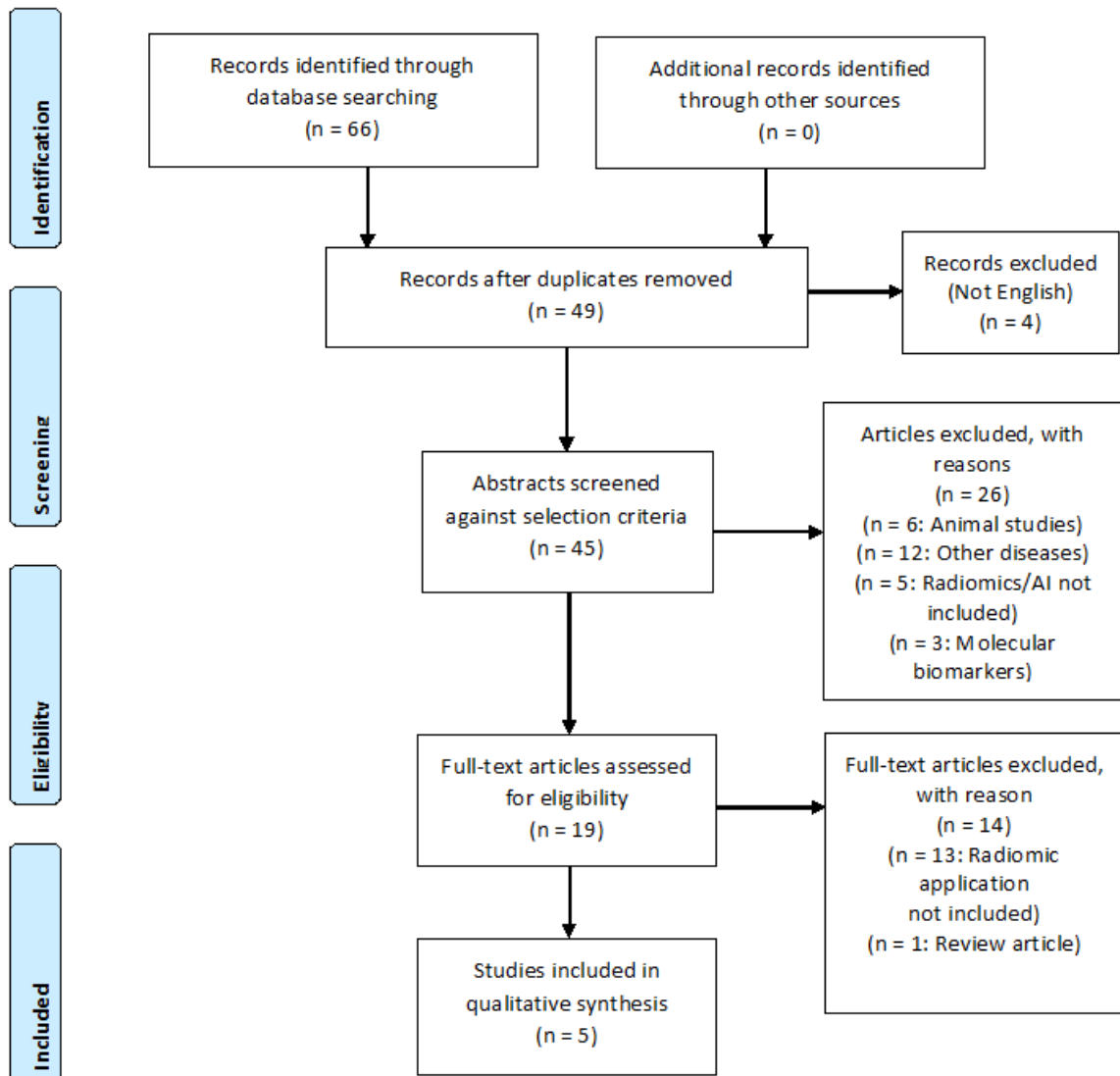
\*Since the PET/CT signature is simply a combination of the CT and PET signatures, only the PET/CT features are listed to avoid duplication.

**FIGURES:**

**Figure 1:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram for journal article screening and selection



**PRISMA Flow Diagram**



## CHAPTER 3

# The Influence of Post-Acquisition Image Processing on a Radiomic Signature Constructed from Planar Images

*Tamarisk du Plessis<sup>1</sup>, Gopika Ramkilawon<sup>2</sup>, Christophe Van de Wiele<sup>3</sup>, Mike M. Sathekge<sup>1</sup>*

1. Department of Nuclear Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa.
2. Department of Statistics, University of Pretoria, Pretoria, South Africa.
3. Department of Diagnostic Sciences, University of Ghent, Ghent, Belgium.

*Int. J. Med. Imaging. 2023;11(2):34-41*

*doi: 10.11648/j.ijmi.20231102.13*

## Abstract

For radiomics to be accepted as a definite tool in medicine, the outputs must be robust, repeatable and reliable. Image processing alters the quality of the input data which might have an impact on the values of the extracted features and ultimately the signatures developed. This study evaluated the magnitude of the influence of various interpolation and post-acquisition processing methods on the radiomic feature values extracted from planar images and radiomic signatures. Three different interpolation methods were applied to a chest x-ray dataset before 2-dimensional (2D) radiomic features were extracted using Pyradiomics. The influence of image size, cropping and re-segmentation were also evaluated by changing the respective variable before applying bilinear interpolation and extracting 2D features. ANOVA and post-hoc Bonferroni corrections were used to assess the differences in the radiomic feature values. Of the 93 first order- and texture- features extracted, 42 texture features (56.8%) proved to be significantly influenced ( $p \leq 0.05$ ) by the interpolation method. Only 2 first order features (10.5%) were significantly influenced ( $p \leq 0.05$ ) by the image size and 62 texture features (83.8%) by the other pre-processing methods evaluated. Pearson's Correlation Analysis was then applied to develop a separate radiomics signature from each of the six image processing datasets under consideration. Five identical signatures were developed, with only the uncropped dataset that resulted in a unique signature. This study showed that the interpolation algorithms and other processing applied to planar images do have a noticeable influence on most radiomic feature values extracted. But regardless of the differences seen in the feature values, the radiomic signatures were reproducible for most datasets using different image processing methods.

## Keywords

Radiomics, Image Processing, Interpolation, Chest X-rays, Radiomic Signatures

## 1. Introduction

Radiomics is a field of study that has the potential to extract large amounts of quantitative features from medical images [1]. It makes use of statistically based imaging analysis algorithms to identify image features which can be used to quantify disease characteristics [2]. Radiomic feature extraction has the hypothesis that the correct combination of these algorithms, together with the clinical data, can express meaningful tissue properties useful in the management of a disease [3]. Numerous articles have been published on the use of radiomics in tumours [4-7], and a few on the application in non-neoplastic diseases such as pulmonary tuberculosis and Covid-19 [8-10]. In recent years this data mining tool advanced to the point where it incorporates machine learning and deep learning approaches to build state of the art radiomic signatures and models [11]. However, radiomic studies can only be used as a definite tool in medicine once the outcomes are reliable, repeatable, robust and validated.

The complex multi-step process of radiomics includes; Image acquisition, image post-acquisition processing, image segmentation, feature extraction, dimensionality reduction, association analysis, model construction and database development [2].

The foundation of this multi-step process however remains the input image, either two- (2D) or three dimensional (3D). But with the fast development of quantitative imaging methods the focus on this foundational step has been lost. It was shown that the quality of the input data has a considerable impact on the value of the extracted features [3]. And that variables such as a variety of acquisition, reconstruction and post processing parameters influence the image texture and noise and consequently the value of the extracted features [3]. The robustness of radiomic features therefore depend significantly on the image post-processing applied [12].

The Imaging Biomarker Standardization Initiative (IBSI) was published in 2019 with the aim to standardize image biomarker nomenclature and definitions, to suggest tools for verifying radiomics software implementations and to standardize reporting guidelines [13]. The IBSI gives useful suggestions on the radiomics workflow with detailed technical instructions regarding the image post-acquisition processing workflow required [13]. It recommends that the following steps should be followed before feature extraction algorithms are applied; dataset classification, data conversion, post-acquisition processing, segmentation, interpolation and re-segmentation [13].

Data classification refers to sorting the images in the dataset to only include images of the same modality, patient orientation and photometric interpretation. Data conversion of raw image data into more meaningful data is only required for certain image modalities without definite grayscale values, e.g. PET [11]. To homogenize datasets further various post-acquisition processing steps are required; image format conversion, normalization and discretization of the grayscale intensities and image interpolation [13].

Interpolation algorithms translate image intensities from the original image grid to an interpolation grid. Pixels are spatially represented by their centres in such grids [11, 13]. Isotropic pixel spacing is required for texture features to be rotationally invariant, and to allow comparison between different datasets [13]. There are currently no clear recommendations on whether up-sampling or down-sampling should be the preferred interpolation method for radiomic studies [12]. Only a recommendation that a calculated decision should be made regarding this, as up-sampling to the smallest pixel dimension can introduce artificial information, while down-sampling to the largest pixel dimension can result in information loss [13].

Various interpolation algorithms are commonly used for volumetric image pre-processing in medicine, e.g. nearest neighbour, trilinear, tricubic convolution and tricubic spline interpolation [13]. The 2D equivalents of these popular interpolation methods are 2D-nearest neighbour-, bilinear- and bicubic interpolation, but no literature could be found on the optimal choice for interpolation of planar medical images. Nearest neighbour interpolation adapts the intensity of the nearest neighbouring pixel without regard to the intensities of other neighbouring pixels [14]. This is the simplest interpolation method but may result in blocky images. Bilinear interpolation takes the intensity of 4 neighbouring pixels into account and applies two linear interpolations to obtain a new pixel intensity [14]. Bicubic interpolation results in the smoothest interpolation as it uses 16 pixels and applies a third order polynomial function to interpolate the new pixel intensity [14].

In the recommended workflow discussed above only 3D images were mentioned, which leaves the question as to whether the same labour intensive steps are required for planar images. Also no literature could be found that quantitatively shows the influence that each of these post-acquisition processing steps have on the feature values extracted or on the radiomic signatures being developed.

This study was performed with the aim to evaluate the scale of influence that different interpolation methods and other common post-acquisition processing applied to planar images will have on the extracted radiomic features. It also evaluated which radiomic features are most sensitive to these post-acquisition processing methods and how these feature values influence radiomics signatures. A chest x-ray (CXR) dataset was selected for this study since these planar images are still clinically used in many countries for diagnosis and disease management [15].

## **2. Methods**

### **2.1. Dataset**

This is a retrospective study consisting of 103 posteroanterior (PA) CXR of patients diagnosed with active pulmonary tuberculosis. Ethical clearance was granted by the researchers' tertiary institution to use these images for this study.

### **2.2. Image pre-processing**

The original dataset consisted of DICOM images of various sizes and width-to-height ratios. To remove as many variables as possible that might have an underlying influence on the interpolation, all images were cropped to a square (equal width-to-height ratio). Only one dimension of the images was cropped to preserve the maximum dimensions of the CXR. Unfortunately automatic batch cropping was unsuccessful as most CXR were acquired at asymmetrical patient positions and automatic cropping of these images removed lung pixels. An expensive added layer of manual cropping of all images was therefore performed using commercial image processing software, Total Image Converter (by CoolUtils.com file converters) version 8.2.0.237. To eliminate the dual processing required in DICOM images, where the corresponding header group element must be changed with the image data, images were converted to another format. The same software was used to convert all images to PNG format and to correct the unconventional photometric interpretation of Monochrome1. During the conversion from DICOM to PNG-format, the bit depth was changed from 14 or 12 bits to 8 bits to obtain a uniform dataset and the scalar DICOM type was converted to conventional RGB type for PNG formatted images. Image size, pixel size and image resolution were



preserved. Since the radiomics library used can only apply the feature algorithms to scalar images, the images were finally converted to L-mode with the ITU-R 601-2 luma transform.

### **2.3. Image interpolation**

All images in the original DICOM dataset had dimensions equal to or larger than 1024x1024 pixels after being cropped. These large images ensured that no image had to be extrapolated. All cropped images were therefore down sampled to 1024x1024 pixels using 3 basic interpolation algorithms commonly used in image processing; bilinear interpolation, bicubic interpolation and 2D nearest neighbour interpolation. This was done using the `cv2.resize()` function (`cv2.INTER_LINEAR`, `cv2.INTER_CUBIC` and `cv2.INTER_NEAREST`) in Python version 3.7.6.

### **2.4. Segmentation and feature extraction**

A fully automatic in-house U-net based segmentation model was used to segment the lung region-of-interest (ROI) [16]. The segmentation model resizes images to 256x256 pixels (using bilinear interpolation) before segmenting the lungs as a 256x256 pixel mask. All masks therefore had to be extrapolated to 1024x1024 pixels to have similar dimensions to its corresponding image. Extrapolation of the masks were done using nearest neighbour interpolation. The extrapolation method will have no influence on the segmentation, or the results of this study, as a mask is simply a binary matrix. Each image therefore had 3 versions (one version for each interpolation method) associated to a single mask.

The Pyradiomics library has 103 2D features available for extraction (10 2D shape features, 19 first order- and 74 texture features). Pyradiomics (version 3.0) in Python (version 3.7.6) was used to extract a total of 93 2D features (first order- and texture features only) from each image version. No shape-based features were evaluated as these are calculated using the shape of the ROI defined by the mask. Since the same mask is used for all 3 image versions this will naturally result in identical features. Pyradiomics have five different texture feature groups namely; GLCM = Gray Level Co-occurrence Matrix, GLDM = Gray Level Dependence Matrix, GLRLM = Gray Level Run Length Matrix, GLSZM = Gray Level Size Zone Matrix and NGTDM = Neighbouring Gray Tone Difference Matrix [1]. All variables in the Pyradiomics library were left at default and no filters were applied.

## 2.5. Evaluating the influence of image size, cropping and re-segmentation

The influence of 3 other common post-acquisition image processing techniques was also evaluated. To do this the bilinearly interpolated dataset that was discussed in the above methodology (section 2.3), was used as baseline. Mask segmentation and feature extraction was then repeated three times while only 1 post-acquisition processing variable was changed at a time to obtain a second, third and fourth feature set for comparison.

Firstly, the influence of image dimension was studied by interpolating the baseline dataset to 256x256 pixels instead of 1024x1024 pixels with bilinear interpolation. This dataset was again segmented and radiomic features extracted to obtain the second set of features called *Size*. Secondly, the influence of image cropping before applying the segmentation model and radiomic algorithms was evaluated. This dataset was left uncropped before all other image pre-processing steps were performed. The masks were segmented and radiomic features extracted to obtain the third set of features called *Uncropped*. Lastly, the influence of re-segmenting the images after image interpolation was evaluated. The fourth set of features called *Re-segmented* was therefore obtained by using masks that were re-segmented after image interpolation.

## 2.6. Statistical analysis

Statistical analysis for both image processing sections of this study was performed using SPSS 28.0. ANOVA and post-hoc Bonferroni corrections were used to assess the differences between features obtained for the different variables using the different image post-processing methodologies under study.

## 2.7. Dimensionality reduction

A Shapiro-Wilk test was used to ensure no normality assumption violations in the features. This dataset does not fully adhere to the assumption of normality, but since this sample size is sufficiently large (larger than 100), Pearson Correlations can still be applied as it is robust against such violations [17]. Pearson Correlation Analysis was therefore applied to obtain a radiomics signature for each of the six above-mentioned datasets. Feature pairs with absolute correlations greater than 0.8 were removed.

### 3. Results

#### 3.1. Section 1: Influence of the interpolation algorithm

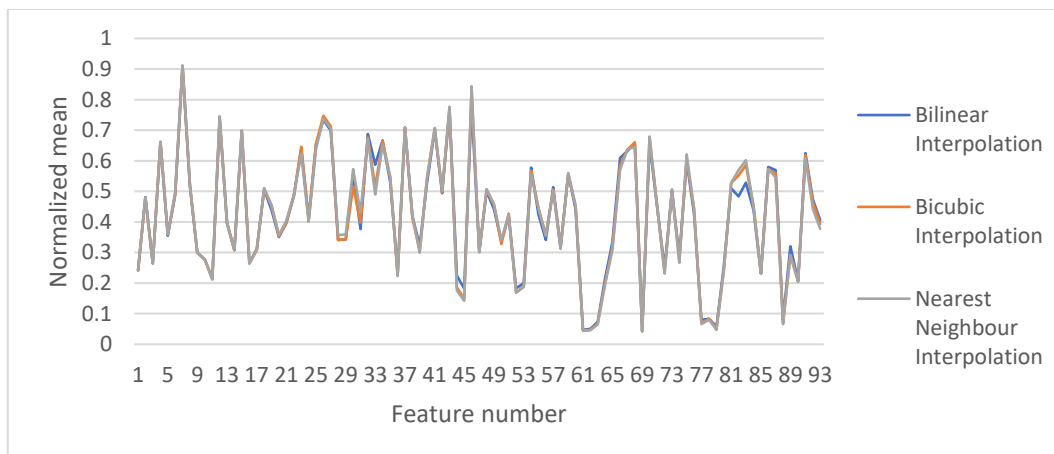
When comparing bilinear, bicubic and nearest neighbour interpolation, significant results (F-test) between these three methodologies were found for 42 of the 93 features extracted. The features with significant p-values ( $p \leq 0.05$ ) are summarized in table 1. To evaluate which interpolation method contributed to the significantly lower or higher feature values, individual group comparisons were performed by comparing the mean values of each feature to the feature group mean. The results are also summarized in table 1 as either < or > to indicate a significantly lower mean (<) or a significantly higher mean (>) respectively.

**Table 1:** Summary of features that displayed significant differences ( $p \leq 0.05$ ) when comparing the different interpolation methods with their respective individual group comparison results

FEATURE NO	FEATURE NAME	SIGNIFICANT (F-TEST)	BILINEAR	BICUBIC	NEAREST NEIGHBOUR
23	glcm_Contrast	0.0001	<	>	>
24	glcm_Correlation	0.0001	>		
25	glcm_DifferenceAverage	0.0001	<		>
26	glcm_DifferenceEntropy	0.0001	<		>
27	glcm_DifferenceVariance	0.0001	<	>	>
28	glcm_Id	0.0001	>		<
29	glcm_Idm	0.0001	>		<
30	glcm_Idmn	0.0001	>		
31	glcm_Idn	0.0001	>		
32	glcm_Imc1	0.0001	<		>
33	glcm_Imc2	0.0001	>		
34	glcm_InverseVariance	0.0001	<	>	>
37	glcm_JointEntropy	0.0320	<		>
38	glcm_MCC	0.0001	>		
43	gldm_DependenceEntropy	0.0290	<		>
44	gldm_DependenceNonUniformity	0.0050	>	<	<
45	gldm_DependenceNonUniformityNormalized	0.0080	>		<
46	gldm_DependenceVariance	0.0001	<	>	>
50	gldm_LargeDependenceEmphasis	0.0001	>		<
51	gldm_LargeDependenceHighGrayLevelEmphasis	0.0130	>	<	<
54	gldm_SmallDependenceEmphasis	0.0001	<		>
55	gldm_SmallDependenceHighGrayLevelEmphasis	0.0001	<		>
56	gldm_SmallDependenceLowGrayLevelEmphasis	0.0010	<		

57	glrlm_GrayLevelNonUniformity	0.0001	<	>	>
65	glrlm_RunEntropy	0.0001	>		
66	glrlm_RunLengthNonUniformity	0.0001	<		>
67	glrlm_RunLengthNonUniformityNormalized	0.0001	<		
68	glrlm_RunPercentage	0.0001	<	>	>
70	glrlm_ShortRunEmphasis	0.0001	<	>	>
71	glrlm_ShortRunHighGrayLevelEmphasis	0.0020	<	>	>
73	glszm_GrayLevelNonUniformity	0.0001	<		>
78	glszm_LargeAreaHighGrayLevelEmphasis	0.0290	>		<
81	glszm_SizeZoneNonUniformity	0.0001	<	>	>
82	glszm_SizeZoneNonUniformityNormalized	0.0001	<	>	
83	glszm_SmallAreaEmphasis	0.0001	<	>	
84	glszm_SmallAreaHighGrayLevelEmphasis	0.0500	<		
86	glszm_ZoneEntropy	0.0001	>		
87	glszm_ZonePercentage	0.0001	<		>
89	ngtdm_Busyness	0.0001	<		>
91	ngtdm_Complexity	0.0001	<	>	>
92	ngtdm_Contrast	0.0001	<	>	>
93	ngtdm_Strength	0.0001	>		<

The significant differences can be appreciated by observing Figure 1 where the normalized means of the 93 features extracted using the different interpolation methods were plotted. Before plotting the graphs, the means were normalized with min-max normalization to compensate for the scale variety of the radiomic features.



**Figure 1:** Plot of the normalized means of the 93 features extracted using three different interpolation methods (Bilinear, Bicubic and Nearest Neighbour interpolation) to indicate the significant differences obtained in some of the radiomic features

### 3.2. Section 2: Influence of image size, image cropping and re-segmentation

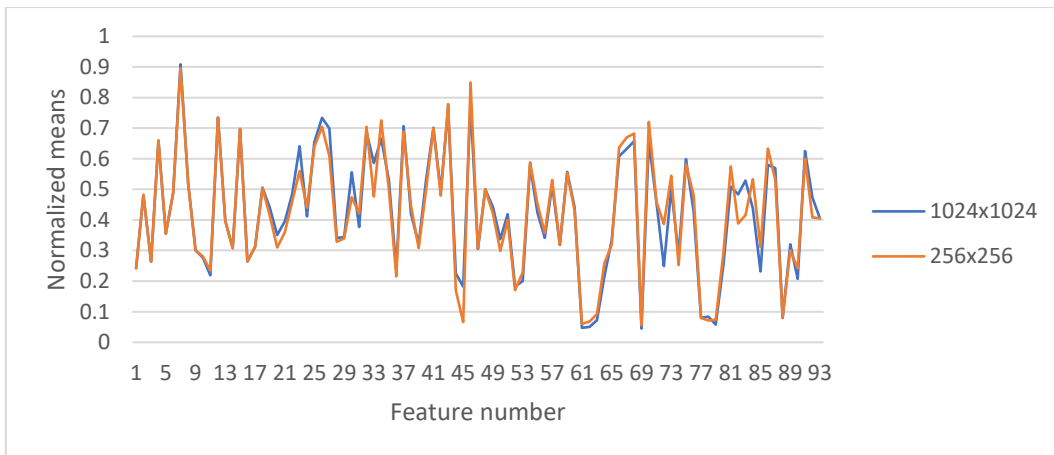
When comparing the baseline, Size, Uncropped and Re-segmented datasets, significant differences (F-test) were found for 63 of the 93 features extracted. The features with significant p-values ( $p \leq 0.05$ ) are summarized in table 2. Individual group comparisons were performed to evaluate which pre-processing methods contributed to the significantly lower or higher feature values by comparing the means of each feature to the feature group mean. The results are also indicated in table 2 as either  $<$  or  $>$  to indicate a significantly lower mean ( $<$ ) or a significantly higher mean ( $>$ ) respectively.

**Table 2:** Summary of features that displayed significant differences ( $p \leq 0.05$ ) when comparing the various pre-processing methods with their respective individual group comparison results

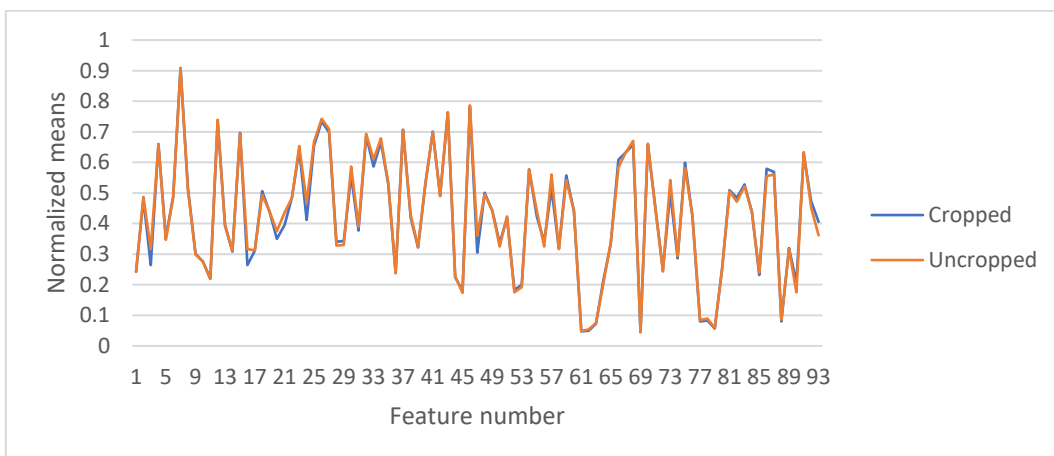
FEATURE NO	FEATURE NAME	SIGNIFICANT (F-TEST)	BASELINE	SIZE	UNCROPPED	RE-SEGMENTED
3	firstorder_Energy	0.0001		<		
16	firstorder_TotalEnergy	0.0001		<		
19	glcm_Autocorrelation	0.0250		<		
23	glcm_Contrast	0.0001		>		
24	glcm_Correlation	0.0001		>		
25	glcm_DifferenceAverage	0.0001		>		
26	glcm_DifferenceEntropy	0.0001		>		
27	glcm_DifferenceVariance	0.0001		>		
28	glcm_Id	0.0001		<		
29	glcm_Idm	0.0001		<		
30	glcm_Idmn	0.0001		<		
31	glcm_Idn	0.0001		<		
32	glcm_Imc1	0.0001		>		
33	glcm_Imc2	0.0001		<		
34	glcm_InverseVariance	0.0001		>		
35	glcm_JointAverage	0.0390		<		
36	glcm_JointEnergy	0.0001		<		
37	glcm_JointEntropy	0.0001		>		
38	glcm_MCC	0.0001		<		
39	glcm_MaximumProbability	0.0001		<		
40	glcm_SumAverage	0.0390		<		
41	glcm_SumEntropy	0.0050		>		
43	gldm_DependenceEntropy	0.0001		>		
44	gldm_DependenceNonUniformity	0.0001		<	<	
45	gldm_DependenceNonUniformityNormalized	0.0001		<		
46	gldm_DependenceVariance	0.0001		>		

47	gldm_GrayLevelNonUniformity	0.0001	<
50	gldm_LargeDependenceEmphasis	0.0001	<
51	gldm_LargeDependenceHighGrayLevelEmphasis	0.0001	<
54	gldm_SmallDependenceEmphasis	0.0001	>
55	gldm_SmallDependenceHighGrayLevelEmphasis	0.0001	>
56	gldm_SmallDependenceLowGrayLevelEmphasis	0.0001	>
57	glrlm_GrayLevelNonUniformity	0.0001	<
60	glrlm_HighGrayLevelRunEmphasis	0.0310	<
61	glrlm_LongRunEmphasis	0.0001	< <
62	glrlm_LongRunHighGrayLevelEmphasis	0.0001	<
63	glrlm_LongRunLowGrayLevelEmphasis	0.0030	<
65	glrlm_RunEntropy	0.0001	<
66	glrlm_RunLengthNonUniformity	0.0001	<
67	glrlm_RunLengthNonUniformityNormalized	0.0001	>
68	glrlm_RunPercentage	0.0001	<
69	glrlm_RunVariance	0.0010	<
70	glrlm_ShortRunEmphasis	0.0001	<
71	glrlm_ShortRunHighGrayLevelEmphasis	0.0001	>
72	glrlm_ShortRunLowGrayLevelEmphasis	0.0001	>
73	glszm_GrayLevelNonUniformity	0.0001	<
74	glszm_GrayLevelNonUniformityNormalized	0.0010	<
75	glszm_GrayLevelVariance	0.0001	>
77	glszm_LargeAreaEmphasis	0.0001	< <
78	glszm_LargeAreaHighGrayLevelEmphasis	0.0001	> <
79	glszm_LargeAreaLowGrayLevelEmphasis	0.0001	< <
80	glszm_LowGrayLevelZoneEmphasis	0.0410	>
81	glszm_SizeZoneNonUniformity	0.0001	<
82	glszm_SizeZoneNonUniformityNormalized	0.0001	<
83	glszm_SmallAreaEmphasis	0.0001	<
86	glszm_ZoneEntropy	0.0170	>
87	glszm_ZonePercentage	0.0001	>
88	glszm_ZoneVariance	0.0001	< <
89	ngtdm_Busyness	0.0001	<
90	ngtdm_Coarseness	0.0001	>
91	ngtdm_Complexity	0.0001	>
92	ngtdm_Contrast	0.0001	>
93	ngtdm_Strength	0.0001	>

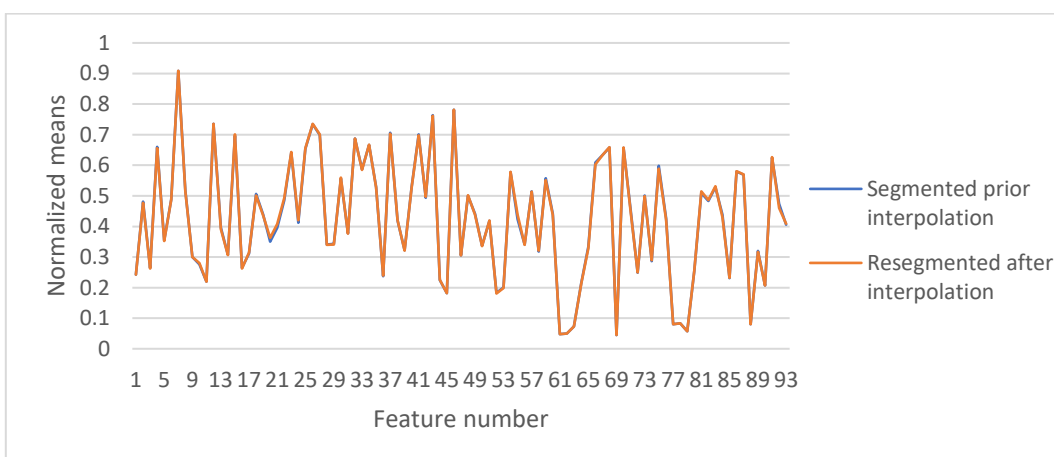
The individual group comparisons are graphically presented in Figures 2, 3 and 4. The normalized mean values of the baseline feature set were plotted against the *Size* features (Figure 2), *Uncropped* features (Figure 3) and the *Re-segmented* features (Figure 4) respectively.



**Figure 2:** Plot of the normalized means of the 93 features extracted using different image sizes (1024x1024 pixels and 256x256 pixels) to indicate the significant differences obtained in some of the radiomic features



**Figure 3:** Plot of the normalized means of the 93 features extracted using cropped and uncropped images to indicate the significant differences obtained in some of the radiomic features



**Figure 4:** Plot of the normalized means of the 93 features extracted using images segmented prior to interpolation and re-segmented images to indicate that no significant differences are seen in the extracted radiomic features

### 3.3. Section 3: Influence of Radiomics signatures

Six datasets were under consideration in this study, three evaluating the influence of the interpolation algorithm (*Bilinear*, *Bicubic* and *Nearest Neighbour*) and three evaluating the influence of other image post processing methods (*Size*, *Uncropped* and *Re-segmented*).

Table 3 summarizes the signatures obtained for each of these datasets.

**Table 3:** Summary of the signature features retained for each of the six datasets under consideration

DATASET	# FEATURES REMOVED	SIGNATURE FEATURES RETAINED
<b>BI-LINEAR</b>	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
<b>BI-CUBIC</b>	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
<b>NEAREST NEIGHBOUR</b>	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
<b>SIZE</b>	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
<b>UNCROPPED</b>	89	glcm_Idmn, gldm_GrayLevelNonUniformity, ngtdm_Busyness, ngtdm_Strength
<b>RE-SEGMENTED</b>	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength

## 4. Discussions

In radiomics first order statistics uses basic statistical algorithms to describe the value and distribution of a group of pixels without concern for spatial relationships [3]. Second order statistics describe the textural features and are calculated by the statistical inter-relationship between the pixels in the ROI [1]. This was confirmed by the results of this study where 42 (56.8%) and 62 (83.8%) out of the 74 texture features extracted were significantly influenced by the interpolation method and other post-acquisition image processing respectively. The only 2 first order feature values influenced by any image processing were Energy ( $p=0.001$ ) and Total Energy ( $p=0.001$ ).



#### 4.1. Section 1: Interpolation

Since no ground truth exists regarding the correct interpolation algorithm to apply in radiomic studies, the group mean for each feature was considered as the baseline for individual comparisons. When individual group comparisons were performed, no significant differences were seen in any first order feature values. However significant differences in 56.8% of the texture features amongst all three interpolation methods were observed. The differences were random without an obvious trend or pattern that can be identified. The mean feature values for bicubic interpolation, which is the smoothest interpolation, do however have outputs closest to the group mean with only 15 features that differ considerably from the group mean. Another observation is that bilinear and nearest neighbour interpolation were always distributed to opposite sides of the group mean in the 42 features that differed significantly. When selecting an interpolation algorithm all factors, such as the quality of the images, available computational power and study outcomes, should be considered. By definition bilinear interpolation is considered as the conservative choice in image interpolation as it takes the intensity of 4 neighbouring pixels into account, compared to 16 in bicubic interpolation and only 1 in nearest neighbour interpolation [13]. Without a ground truth, the above results highlight the importance of consistency in radiomics studies. Regardless of the method selected only a single interpolation algorithm must be applied across an entire study and in comparative studies.

#### 4.2. Section 2: Size, cropping and re-segmentation

In this part of the study all four feature sets (*Baseline, Size, Uncropped and Re-segment*) were interpolated with bilinear interpolation. When the *Size* CXR dataset was down sampled to a quarter (256x256 pixels) of the baseline images' dimensions, 83.8% of the texture feature values extracted were significantly influenced. This can be accounted for by the statistical inter-relationship of the pixels in the secondary matrixes that would have changed by down sampling the images. Some 3D radiomic studies also mentioned the influence of image size on the outcome of their studies [18, 19].

Only 6 out of 74 (8.1%) texture features extracted were significantly influenced by not cropping the images into a square before applying the interpolation algorithm. This can also be accounted for by the image matrix that would have changed when the interpolation algorithm had unevenly down sampled the pixels of the uncropped images in the two

dimensions. It is possible that the influence of uncropped images will increase as the width-to-height ratios of the uncropped images increase.

By re-segmenting the masks after image interpolation were applied, no significant influences were observed in any features. This might hold true for this study only as the segmentation model used also applied bilinear interpolation to the images before segmenting the masks. In essence the masks were always re-segmented without doing it intentionally. Even being study specific, these results are valuable as it proves good repeatability of the interpolation algorithm, the segmentation model and the radiomics library used.

### **4.3. Section 3: Radiomics signature**

Five out of the six radiomic signatures that were constructed from each dataset were identical consisting of 3 texture features; `glcm_Idmn`, `ngtdm_Busyness` and `ngtdm_Strength`. Only the *Uncropped* dataset signature had an additional feature, `gldm_GrayLevelNonUniformity`. This reproducible result is seen regardless of the significant differences found in the extracted feature values caused by the different interpolation methods, sizes and re-segmentations applied. Also the unique signature obtained when the images were left uncropped indicates the labour intensive process of cropping images to squares (or equal width-to-height ratios for the entire dataset) is imperative.

## **5. Conclusions**

This study showed that first order feature values are not significantly impacted by the interpolation algorithms and other image processing methods applied, but that it does significantly influence most texture feature values extracted from planar images. It also showed that regardless of the significant differences seen in the extracted feature values, caused by most post-acquisition image processing methods, the outcome of the radiomics signatures remains reproducible. The only image post-acquisition processing step that resulted in a different signature was image cropping and it must therefore strongly be considered in all planar image studies.

Larger sample sizes are required to verify this study, but it preliminarily shows that image post-processing, except cropping, does not considerably influence the outcome of radiomic signatures. Focus should rather be placed on applying the correct dimensionality algorithms

that are robust against any instabilities caused by image post-processing. It is however imperative to be consistent with all image processing steps applied across an entire radiomics study. Thorough reporting of all image processing applied in radiomic studies is also crucial to increase the reproducibility and validity of this field of study.

### **Conflict of interest**

The authors declare that no funding, grants, or other financial support were received for the preparation of this manuscript. The authors also have no competing interests to declare that are relevant to the content of this article.

## References

1. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer research* 2017;77:e104-e107
2. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging* 2012;30:1234-1248
3. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *European Radiology Experimental* 2018;2:1-8
4. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR. American journal of neuroradiology* 2018;39:208-216
5. Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *European Journal of Radiology* 2020;128
6. Feng B, Chen X, Chen Y, Lu S, Liu K, Li K, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. *European Radiology* 2020
7. Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of 18F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. *European Journal of Nuclear Medicine and Molecular Imaging* 2020
8. Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. In: *BMC Medical Imaging*, 2019; 1-11
9. Giraud C, Frattin G, Fichera G, Motta R, Stramare R. A practical integrated radiomics model predicting intensive care hospitalization in COVID-19. *Critical care (London, England)* 2021;25:145
10. Han Y, Chen C, Tang L, Lin M, Jaiswal A, Ding Y, et al. Using Radiomics as Prior Knowledge for Abnormality Classification and Localization in Chest X-rays. In, 2020
11. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Medical physics* 2020;47:e185-e202
12. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging* 2020;11:91
13. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020;295:328-338

14. Acharya T, Tsai P-S. Computational foundations of image interpolation algorithms. *Ubiquity* 2007;2007:1-17
15. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *European Journal of Radiology* 2020;123
16. Dixon SAP. Using Deep Learning to Segment Chest X-Rays for the Analysis of Pneumonia. In:*School of Biomedical Engineering*: University of Sydney, 2019
17. Havlicek LL, Peterson NL. Robustness of the Pearson Correlation against Violations of Assumptions. *Perceptual and Motor Skills* 1976;43:1319-1334
18. Altazi BA, Zhang GG, Fernandez DC, Montejo ME, Hunt D, Werner J, et al. Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. *Journal of Applied Clinical Medical Physics* 2017;18:32-48
19. Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C, et al. Revisiting the Robustness of PET-Based Textural Features in the Context of Multi-Centric Trials. In:*PLOS ONE*, 2016

## CHAPTER 4

### **Introducing a secondary segmentation to construct a radiomics model for pulmonary tuberculosis cavities**

*Tamarisk du Plessis<sup>1</sup>, Gopika Ramkilawon<sup>2</sup>, William Ian Duncombe Rae<sup>3</sup>, Tanita Botha<sup>2</sup>, Neil Alexander Martinson<sup>4,5</sup>, Sarah Alice Parry Dixon<sup>6</sup>, Andre Kyme<sup>6</sup>, Mike Michael Sathekge<sup>1</sup>*

1. Department of Nuclear Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
2. Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, South Africa
3. Medical Imaging Department, Prince of Wales Hospital, Sydney, Australia
4. Perinatal HIV Research Unit (PHRU), University of the Witwatersrand, Johannesburg, South Africa
5. Johns Hopkins University Centre for TB Research, Baltimore, MD, USA
6. School of Biomedical Engineering, University of Sydney, Sydney, Australia

*Radiol. Med. 2023;128(9):1093-1102*

*doi: 10.1007/s11547-023-01681-y*

## Abstract

**Purpose:** Accurate segmentation (separating diseased portions of the lung from normal appearing lung) is a challenge in radiomic studies of non-neoplastic diseases, such as pulmonary tuberculosis (PTB). In this study we developed a segmentation method, applicable to chest X-rays (CXR), that can eliminate the need for precise disease delineation, and that is effective for constructing radiomic models for automatic PTB cavity classification.

**Methods:** This retrospective study used a dataset of 266 posteroanterior CXR of patients diagnosed with laboratory confirmed PTB. The lungs were segmented using a U-net based in-house automatic segmentation model. A secondary segmentation was developed using a sliding window, superimposed on the primary lung segmentation. Pyradiomics was used for feature extraction from every window which increased the dimensionality of the data, but this allowed us to accurately capture the spread of the features across the lung. Two separate measures (standard-deviation and variance) were used to consolidate the features. Pearson's correlation analysis (with a 0.8 cut-off value) was then applied for dimensionality reduction followed by the construction of Random Forest radiomic models.

**Results:** Two almost identical radiomic signatures consisting of 10 texture features each (9 were the same plus 1 other feature), were identified using the two separate consolidation measures. Two well performing random forest models were constructed from these signatures. The standard-deviation model (AUC=0.9444 (95% CI, 0.8762; 0.9814)) performed marginally better than the variance model (AUC=0.9288 (95% CI, 0.9046; 0.9843)).

**Conclusion:** The introduction of the secondary sliding window segmentation on CXR could eliminate the need for disease delineation in pulmonary radiomic studies, and it could improve the accuracy of CXR reporting currently regaining prominence as a high-volume screening tool as the developed radiomic models correctly classifies cavities from normal CXR.

## Keywords:

Radiomics, Segmentation, Pulmonary Tuberculosis, Sliding window

## Introduction

Tuberculosis (TB) is one of the top ten causes of death worldwide according to the World Health Organization [1]. However an estimated 66 million lives were saved in the past two decades through TB diagnosis and treatment [1]. Early and accurate diagnosis is vital in fighting this global battle against TB spread and infections. Planar chest X-rays (CXR), in combination with biological methods, are commonly used to screen for or diagnose pulmonary TB (PTB) in patients at high risk of TB disease. CXR is the most widely accessible imaging modality in high TB burdened countries, and is regaining prominence as a high volume screening modality [2]. Advantages of CXR include that it is relatively inexpensive, fast, non-invasive and a good indicator of the extent of disease in the lungs [3]. Some dis-advantages are that expert x-ray interpreters are often scarce in resource-limited countries [4], and results are influenced by intra-observer subjectivity [5]. To lower the subjectivity associated with x-ray interpretation data science research has focussed on quantifying and analysing features on CXR [6,7].

Radiomic feature extraction is one such tool that can be used to quantify disease characteristics, or features, and assess progression from serial medical images in the same patient, as it makes use of statistically based imaging analysis algorithms to convert medical images into mineable high-dimensional data [8,9]. Radiomics can extract relevant image information that can comprehensively assess the entire two dimensional landscape in the region-of-interest (ROI) [10]. Radiomic libraries can extract hundreds to thousands of features per image. As an image mining tool generating such big data, radiomics naturally lends itself to the application of machine learning or deep learning approaches for developing signatures or advanced model building [10,11].

Radiomics is a trending research technique in oncology imaging, but it is less studied in non-neoplastic pathologies such as PTB [12,13]. A recent systematic review showed that radiomic feature extraction, for the purpose of PTB diagnosis or differentiation from other pulmonary pathology, has only been applied in five studies [12]. In all five studies CT or PET/CT scans were used as the input imaging modality [12]. The review highlighted the need, and the challenges, of applying feature extraction to chest x-rays [12]. One challenge is that PTB has diverse radiological presentations: cavities, adenopathy, infiltrates and plural effusions, miliary pattern with the disease spreading across either a relatively small proportion of a single lung, or with extensive bilateral disease. Accurate segmentation of these diverse



disease presentations is difficult and time consuming and not always feasible with large data sets [14], and can result in significant observer-bias [6]. This is a major limitation in the quantification of non-neoplastic diseases, because variability in segmentation is the biggest cause of irreproducible radiomics outcomes [10].

Several radiomic features are interpreted differently when subjected to inter- and intra-observer assessments in delimiting ROIs [14]. Some articles use manual segmentation by expert readers as the ground truth for segmentation [15], but both manual and semi-automatic segmentation have limitations, while fully automatic segmentation models are fast and have good reproducibility. Many segmentation algorithms have already been trained with deep-learning methods to perform automatic segmentation tasks for various imaging modalities, including CXR [14]. These models are used primarily for organ segmentation but cannot yet identify the pathology [16], especially in non-neoplastic pulmonary diseases. The principal aim of our study was to develop a segmentation method, applicable to chest x-rays, that could eliminate the need for precise disease delineation in the lungs. This segmentation method will be applicable to any CXR quantification study, but we developed it specifically for radiomic feature extractions.

In recent years radiomics has gained increasing popularity due to its ability to quantify medical images and for the construction of radiomic signatures, nomograms, machine learning classifiers and models to assist in disease diagnosis, prediction of disease status, response to treatment and disease prognosis [17]. Radiomics improve discrimination performance and detection of medical images compared with those made by radiologists alone [9,17,18]. Our study used radiomics to develop a model to automatically differentiate normal CXR from CXR with cavities, to assist clinicians with improved and faster PTB diagnosis.

PTB has many radiological presentations, but thick-walled cavities are generally an excellent radiological indicator of active PTB that with treatment and time resolve into thin-walled smooth cavities in treated TB [19]. These cavities cause textural changes in the lung that are visually apparent. Cavities were therefore selected as the radiological TB expression under investigation for this study.

## Methods

Every step in a the multi-step workflow of radiomics can influence the results and reproducibility [20]. To address this, the Imaging Biomarker Standardization Initiative (IBSI), published in 2019, aims to standardize image biomarker nomenclature and definitions to standardize image biomarkers extraction [21]. The IBSI guidelines were adhered to where possible during this study.

### Patient selection

This is a retrospective study that used a dataset consisting of 266 posteroanterior (PA) CXR of patients diagnosed with laboratory-confirmed PTB between August 2013 and July 2018. The CXR were radiologically reported on by clinicians who were part of the initial study. Additionally for this study a single experienced TB clinician was asked to retrospectively review the X-rays individually, blinded to the previous reports. The second observer confirmed the CXR classifications as either normal ( $n=71$ ) or with the presence of cavities ( $n=195$ ). All CXR with discordant or indeterminate classifications were removed from this analysis. In this retrospective dataset the acquisition equipment was not recorded, but it can be assumed that various imaging units were used.

### X-ray pre-processing

The original dataset included images in DICOM format which were acquired using non-standardised patient positions, image sizes, orientations, photometric interpretations and bit depths using a range of different imaging units. Total Image Converter version 8.2.0.237 (by CoolUtils.com file converters) was used for initial pre-processing to ensure a uniform dataset. The following pre-processing steps were applied; Manually cropped all images to square dimensions, corrected unconventional photometric interpretations on some images and converted DICOM images to PNG format. Python version 3.7.6 was used to interpolate all images to  $256 \times 256$  pixels with bilinear interpolation and to convert the conventional RGB type for PNG format to scalar type as required by the radiomics library.

### Primary Image segmentation

A fully automatic in-house segmentation model was used to segment the lung fields [22]. This U-net based model was trained and validated on a publicly available Chest X-ray 14 Dataset (CX14) [22]. It was then tested on an unseen publicly available dataset, the JSRT

dataset, and achieved a maximum Intersection over Union (IoU) of 0.8301, 0.9210, and 0.7791 for the heart, lungs and clavicles, respectively [22]. The segmentation model resizes images to  $256 \times 256$  pixels with bilinear interpolation before segmenting the lungs as a  $256 \times 256$  pixel mask output. Because all images had been previously interpolated to the same dimensions as the masks, they could simply be multiplied with the masks to visually evaluate the accuracy of the segmentation model on our unseen dataset (Figure 1). All CXR were correctly segmented, and no manual corrections were needed.



**Figure 1:** Output of the segmentation model (from left to right): The original image, the mask output (multiplied by 255 to be visible) and the mask superimposed with the image that was used to evaluate the segmentation accuracy

### Secondary segmentation

To create the secondary sliding window segmentation a square mask of  $n \times n$  pixels was created and called the sampling window ( $w$ ). This sampling window was selected to be large enough for the enclosed region to exhibit similar characteristics to those of the underlying region and at the same time to be as small as possible to enable the accurate detection of borders between adjacent textural regions. This window will slide over the image in both vertical and horizontal dimensions with a predetermined window step size ( $w_{step}$ ).  $w_{step}$  therefore determines the number of pixels that the sampling window slides across at each step and determines how well boundaries between features are resolved. There is a trade-off: if  $w_{step}$  is too large boundaries will be unresolved, while if  $w_{step}$  is too small then extended times are devoted to computation, and the window would place bounds on many of the features and increase their variability.

The sliding window masks were created in Python (version 3.7.6) using `Numpy.array()` and `PIL.Image()` functions. A square window ( $w$ ) of  $16 \times 16$  pixels and a window step size ( $w_{step}$ ) of 4 pixels was selected. The window will move, or slide, from one side of the CXR to the other in both x- and y-dimensions to create a window matrix of  $61 \times 61$  windows. The

number of windows in the matrix can be calculated using equation 1 where  $[P_x, P_y]$  is the dimensions of the window matrix,  $[n_x, n_y]$  is the dimensions of the image in pixels,  $w$  is the window size and  $w_{step}$  is the window step size.

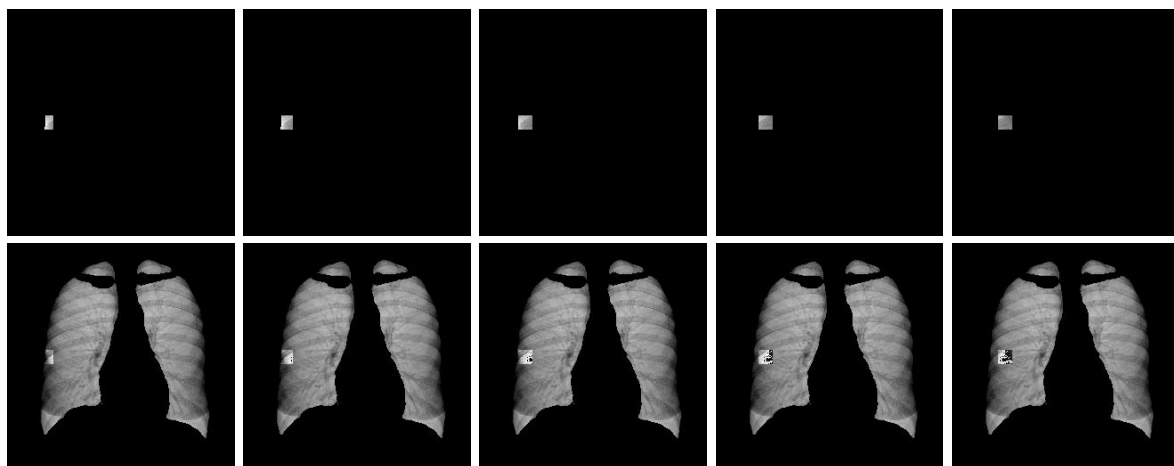
$$P_x = \frac{(n_x - w)}{w_{step}} + 1 \quad \text{AND} \quad P_y = \frac{(n_y - w)}{w_{step}} + 1 \quad [\text{Eq.1}]$$

The sliding windows will therefore cause the effective dimensionality of each image's features to increase by a factor 3721 ( $61 \times 61$ ). This resolution is adequate to resolve the change in the radiomic features across the lung, within an acceptable computational time.

### Radiomic feature extraction

The sliding window masks were superimposed on the primary segmented lung mask of each CXR (see Figure 2). Radiomic features were extracted from each window in the window matrix if the window was not masked off by the lung segmentation.

The Pyradiomics library (version 3.0) was used to extract 93 two dimensional (2D) first order- and texture-features from each sliding-window on each CXR. Shape-based features will be meaningless for the purpose of this study, as these features use the masked ROI for calculating the values and was omitted.



**Figure 2:** Above: Example of 5 sliding windows, sliding horizontally in the y axis (window coordinates  $[P_x, P_y] = [30,9], [30,10], [30,11], [30,12]$  and  $[30,13]$ ) superimposed on the lung mask and the CXR. Below: The same sliding windows, inverted to allow visualisation of the lung mask

### Dimensionality reduction

Statistical analysis was performed using R Software (version 4.1.3; <http://www.r-project.org/>). The secondary segmentation caused an approximately 3000 fold increase in the dimensionality of the features extracted. Before traditional dimensionality reduction methods

could be applied, additional measures were introduced, namely standard deviation (SD) and variance, to quantify and capture the change in features over the lung region [23].

These 2 measures were calculated for each of the 93 features in the cavity dataset (195 CXR) and the normal dataset (71 CXR) respectively. Using the resultant expressions of the features in these 2 measures formed the first step in the feature selection process to limit the complications of dimensionality which arise from the over-abundance of features. For dimensionality reduction, Pearson's correlation coefficient  $\rho$  was used to identify the uncorrelated features [24,25].

The Pearson correlation coefficients between 0 and 1 indicate a positive correlation, correlations equal to zero indicate no correlation and correlations between -1 and 0 indicate a negative correlation. These correlation coefficients were calculated for each feature pair. Feature pairs with absolute correlations ( $|\rho|$ ) greater than a pre-determined cut-off value were removed. Three cut-off values were considered for this study, namely 0.7, 0.8 and 0.9, and later examined to decide which was the most appropriate for the purpose of dimensionality reduction. In addition to the removal of highly correlated features, only features common in both the cavity- and normal datasets were retained.

### **Model development**

To apply this developed radiomic signature in a meaningful manner, a random forest model was constructed to differentiate cavities seen on CXR of people suffering from PTB and normal CXR [26]. This random forest model was used due to its attractive computational features and classification performance as it is robust to overfitting data by design [26]. Due to the imbalance between the cavity and normal samples, a random walk oversampling technique was applied to improve the model's performance [27,28]. Four other sampling strategies were also considered and recorded in the results section. After adjustments to the data were made to ensure equal representation, the entire dataset was split into the training and testing sets with a 70/30% split respectively.

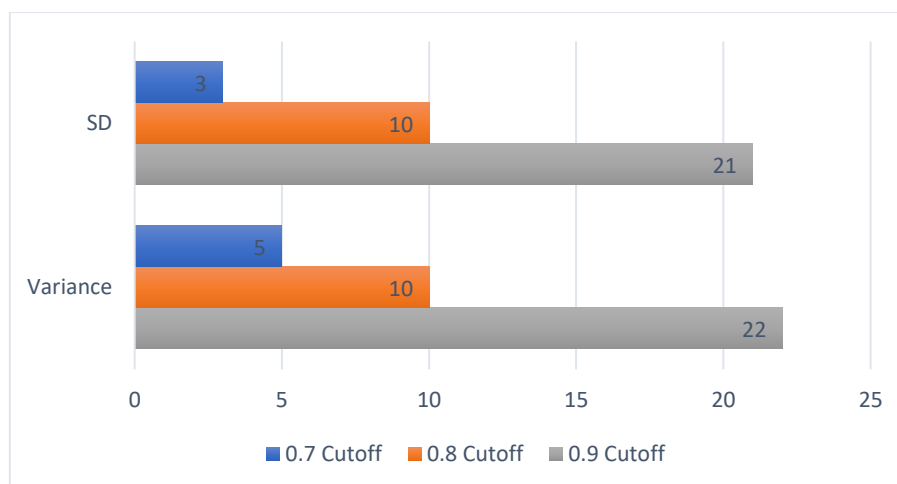
Two separate random forest models were built, using the same CXR set on which the SD- and variance signatures were developed respectively [29]. For both random forest models, a grid search cross-validation method was employed to determine the ideal number of variables to try at each tree split with 15 folds and 5 repeats to further limit overfitting. Across both random forest models, the grid search cross-validation indicated that the ideal number of

variables to try at each tree split was 1. The performance of the models was validated using the testing set, AUC measure, accuracy, sensitivity, and specificity.

## Results

### Signature results

The first step in finding the signature was to determine the optimal cut-off value for the Pearson's correlation. To evaluate this, only the number of features common to both the cavity- and normal datasets were considered. Figure 3 indicates the number of features retained for the two different consolidation measures when different cut-off values were considered in the correlation analysis.



**Figure 3:** The number of common features retained for the two different consolidation measures when different cut-off values were considered in the Pearson's correlation analysis

The results shown in Figure 3 will be conversed in the discussions section, but due to the number of features retained, 0.8 was considered as the optimal cut-off value for dimensionality reduction in this study. When the cavity and normal datasets were considered separately, the number of features retained with a 0.8 cut-off value in the two different consolidation measures were 11 and 12 in the SD measure and 12 and 12 in the variance measure, respectively. For both measures only 10 features were common to the normal and cavity dataset (as seen in Figure 3), with either 1 or 2 additional unique features retained. The features retained are recorded in Table 1 with the unique features emphasized in bold.

**Table 1:** Details of the features retained for the two different consolidation measures in the cavity and normal datasets respectively when a 0.8 cut-off value was used in the Pearson’s correlation analysis. The features not common to both datasets are highlighted in bold.

MEASURE	CAVITY CXR DATASET	NORMAL CXR DATASET
SD	glcm_Correlation	glcm_Correlation
	gldm_DependenceEntropy	<b>glcm_DifferenceEntropy</b>
	gldm_DependenceNonUniformityNormalized	gldm_DependenceEntropy
	gldm_DependenceVariance	gldm_DependenceNonUniformityNormalized
	gldm_SmallDependenceLowGrayLevelEmphasis	gldm_DependenceVariance
	glrlm_RunEntropy	<b>gldm_LargeDependenceHighGrayLevelEmphasis</b>
	<b>glrlm_RunLengthNonUniformityNormalized</b>	gldm_SmallDependenceLowGrayLevelEmphasis
	glrlm_ShortRunLowGrayLevelEmphasis	glrlm_RunEntropy
	glszm_ZoneEntropy	glrlm_ShortRunLowGrayLevelEmphasis
	ngtdm_Busyness	glszm_ZoneEntropy
ngtdm_Contrast	ngtdm_Busyness	
		ngtdm_Contrast
VARIANCE	glcm_Correlation	glcm_Correlation
	gldm_DependenceEntropy	glcm_DifferenceEntropy
	gldm_DependenceNonUniformityNormalized	<b>gldm_DependenceEntropy</b>
	gldm_DependenceVariance	gldm_DependenceNonUniformityNormalized
	gldm_SmallDependenceLowGrayLevelEmphasis	gldm_DependenceVariance
	<b>glrlm_RunEntropy</b>	<b>gldm_LargeDependenceHighGrayLevelEmphasis</b>
	<b>glrlm_RunLengthNonUniformityNormalized</b>	gldm_SmallDependenceLowGrayLevelEmphasis
	glrlm_ShortRunLowGrayLevelEmphasis	glrlm_ShortRunLowGrayLevelEmphasis
	glszm_GrayLevelNonUniformity	glszm_GrayLevelNonUniformity
	glszm_ZoneEntropy	glszm_ZoneEntropy
ngtdm_Busyness	ngtdm_Busyness	
ngtdm_Contrast	ngtdm_Contrast	

glcm = gray level cooccurrence matrix, gldm = gray level dependance matrix, glrlm = gray level run length matrix, glszm = gray level size zone matrix, ngtdm = neighbouring gray tone difference matrix

## Model results

The radiomic signatures obtained for both the SD and variance measures contained the same number of features (10), but the features included in their respective signatures differed (see Table 1). Two separate random forest models were therefore built, using a random walk oversampling technique, to model each of the two radiomic signatures, respectively. We note similar model performance for both models across all performance metrics (see Table 2).

**Table 2:** Performance comparison of the SD and variance radiomic signature-based models showing the AUC measures with the corresponding confidence intervals (95% CI), accuracy, sensitivity, specificity, and precision

	SD	Variance
<b>AUC (95% CI)</b>	0.9444 (0.8762; 0.9814)	0.9288 (0.9046; 0.9843)
<b>Accuracy</b>	0.8333	0.8750
<b>Sensitivity</b>	0.7708	0.8542
<b>Specificity</b>	0.8958	0.8958
<b>Precision</b>	0.8810	0.8913

## Discussion

The prompt diagnosis of PTB is vital for providing timely and accurate treatment, as a delay in treatment can lead to poor outcomes [17]. Biological methods are the gold standard for TB diagnosis, but culture or smear analysis takes time [30]. CXR are immediately available but cannot be used as a standalone tool for diagnosis. It has been shown that radiomics can improve discrimination performance and detection of medical images compared with those made by radiologists alone [9,17,18]. In this study we developed a well performing radiomic model that could assist clinicians with the diagnosis of cavities due to TB on CXR. When added to the clinical signs and symptoms, this might reduce requirements for laboratory results and shorten time to treatment. It can also improve the accuracy of CXR reporting currently regaining prominence as a high-volume screening tool. The radiomic model that can detect cavities will also be useful in future PTB management studies when serial CXR, with their corresponding models, are studied.

For model construction, we first had to address the challenge of PTB disease segmentation that is required when quantifying X-rays using radiomic feature extraction. We developed a sliding window segmentation that allowed the extracted radiomic features to mimic the textural changes across the lung region caused by the disease. Inspiration for this newly developed segmentation method was obtained from a previous study that used deep learning approaches to sub-divide the lung region on a CXR into multiple stationary blocks [31]. They then used multi-instance learning (MIL) to classify each block as either a normal or a TB-manifestation class for TB diagnosis [31]. Instead of the stationary blocks, we introduced a sliding window approach to ensure that the boundaries between feature windows are well resolved.

An advantage of the secondary segmentation is that it eliminates the need to accurately delineate the diseased ROI which is time consuming and difficult in non-neoplastic diseases.



It is also completely automated which eliminates observer-bias and increases reproducibility. The disadvantage of the secondary segmentation is that it increases the dimensionality of the data significantly, but this was addressed by introducing two different consolidation measures before performing traditional dimensionality reduction and model construction.

In previous studies on radiomic signature or nomogram construction for PTB, feature extraction was used to quantify digital medical images for the purpose of comparing, or differentiating, PTB from other pulmonary diseases, mainly tumors [13,32-34,24]. To the best of our knowledge, radiomics has not yet been used for the purpose of PTB diagnosis or disease management. The previously mentioned radiomics studies were performed from CT or PET/CT images [13,32-34,24], which seems redundant when patients in countries where PTB is most prevalent have very limited access to three-dimensional imaging modalities [35,36]. For this reason, we used relevant 2D CXR for segmentation and feature extraction.

Although planar images are an unpopular modality for radiomic studies, a previous study was found where they applied a unique segmentation using a deep learning approach to train a model to automatically identify the thoracic disease in the lung and to generate bounding boxes around it [37]. In this study radiomics features were used to create heat maps to assist the model in identifying the disease, rather than to quantify disease characteristics [37].

### **Signature development**

To develop a radiomic signature for PTB from CXR, dimensionality reduction was required to highlight the most important features and to remove redundant features. Pearson correlations analysis used in this study assumes that the data is normally distributed. We noted that most of the variables being analysed are normally distributed and only a small proportion of the variables violate the normality assumption which was tested using the Shapiro Wilk test. The correlation coefficients should therefore be largely unbiased and unaffected. Furthermore, various studies have indicated that Pearson correlations are robust to violations in the underlying assumptions [38], particularly when the normality assumption is violated. This then eliminated the consideration to use non-parametric alternatives to calculate the correlations between each feature given the few violations of the underlying normality assumption. Other less successful dimensionality reduction models considered for this study were: Inter-class correlation (ICC), Lasso regression, Factor analysis, Standardizing and Mean-absolute-deviation (MAD).

No recommendation on an optimal cutoff value for Pearson Correlations dimensionality reduction could be found in the literature. One study did however mention using 0.8, without validation [24]. We therefore evaluated three different cut-off values, 0.7, 0.8 and 0.9, in this study. When a cut-off value of 0.7 was used in the correlation analysis it retained very few common features (3 and 5 features: 3.2% and 5.3%) in the different measures. This was found to be too conservative and eliminated some features that might be useful. A 0.9 cutoff value retained the most features (21 and 22 features; 22.5% and 23.6%), but this is too liberal and not useful in the context of dimension reduction. It was decided that 0.8 is therefore a balanced cut-off value to be applied.

Two general statistical methods were considered to quantify and consolidate the 3721 windows' extracted features for each CXR. By statistical definition variance and standard deviation gives an indication of how much each data entry in a group differs from the mean of the group [23]. Average, median and IQR measures were also initially considered, but by definition they all average out the data and give no indication of the spread in the data [23]. From their statistical definition, these three related measures should produce results similar to when the secondary segmentation would have been disregarded. These three consolidation methods are therefore meaningless to achieve the study's aim to evaluate the spread in the radiomic features across the lung region and were ignored.

Two separate signatures for the SD and variance measures were developed by only including the features that were common to both the normal and cavity dataset for each measure, respectively. Each signature consists of 10 features, 9 common and 1 different feature (see Table 1). No first order statistical features were included in this signature as these features use basic statistical algorithms to describe the value and distribution of the pixels in the ROI [15], and has no concern for spatial relationships [39]. Texture features are calculated by using the statistical inter-relationship between the pixels in the ROI [39].

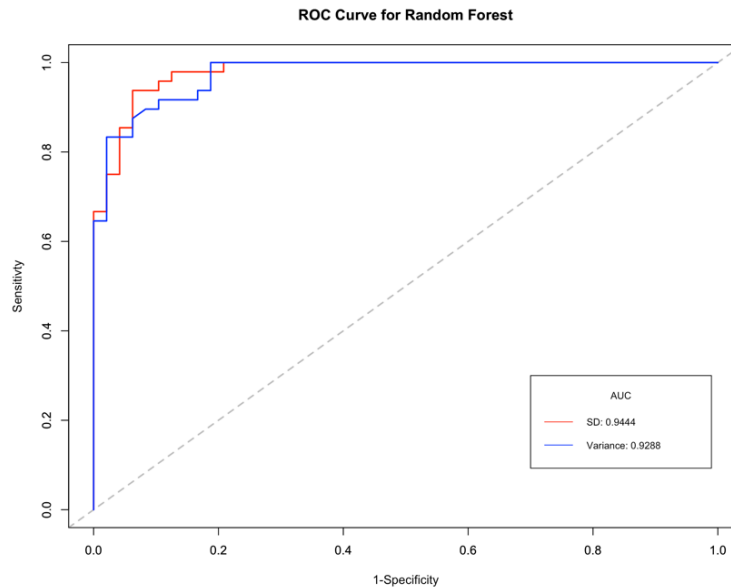
### **Model construction**

The objective of model construction was to develop a noninvasive tool which can automatically differentiate cavities seen on CXR of people suffering from PTB and normal CXR to further assist with PTB diagnosis. As an additional benefit, a successful model will also prove the effectiveness and accuracy of the secondary segmentation introduced. The classification results (Table 2) of both models developed using the SD and variance radiomic signatures showed strong diagnostic power across most measures.

Theoretically, machine learning algorithms are most suitable for samples with uniform distributions in the model training process [40]. For this reason the data was adjusted using the random walk oversampling technique which is an algorithm that generates synthetic instances so that the mean and SD of the numerical attributes remain close to the original data [27,28]. This technique did correct the imbalance between the cavity and normal groups in the sampling distribution and therefore improved the classification performance of the model. Four other sampling strategies were also considered but performed less convincingly: oversampling, synthetic minority sampling technique (SMOTE), simulation and majority weighted oversampling technique (MWMOT).

To construct the random forest models, a grid search cross-validation method was employed to determine the ideal number of variables to try at each tree split with 15 folds and 5 repeats to further limit overfitting. The grid search method is an exhaustive method commonly used to find the optimal parameter value by considering all possible combinations of these values for the model so that the classifier can more accurately predict the unlabelled or testing data [29]. Across both random forest models, the grid search cross-validation method indicated that the ideal number of variables to try at each tree split was 1.

As a result, the SD model performed marginally better than the variance model having a higher AUC value of 0.9444 (95% CI, 0.8762; 0.9814), which is larger than the variance model's AUC value of 0.9288 (95% CI, 0.9046; 0.9843), but the 95% CI for the variance model is narrower which means there is less range in this estimate. This is visually supported by Figure 4. The variance model had a better classification accuracy than the SD model with 87.50% and 83.33% respectively, indicating that the variance model correctly predicted more cavities to the total observations in the data than the SD model. The variance model once again had a better model sensitivity than the SD model with a measure of 85.42% and 77.08% respectively which indicates that the variance model correctly identifies 85.42% of all cavity CXR. Interestingly, both models have a specificity measure of 89.58% which indicates that both models will identify 89.58% of patients who do not have cavities, i.e., have normal CXR. The variance model once again had a higher precision value than the SD model with a precision of 89.13% and 88.1% respectively. This indicates that when the variance model predicts a cavity, it is correct 88.1% of the time.



**Figure 4:** Receiver-operating characteristic (ROC) curves for the SD and variance signature-based random forest model using a random walk oversampling technique

Three other popular machine learning algorithms were also considered for model construction: Logistic regression and Lasso Regression (with a minimum error and one standard error away). All three performed poorly compared to the random forest model in correctly classifying the cavities due to the complex nature of the data and the ability of the random forest model to classify less distinctive groups with overlapping information for the classes.

Although we were able to develop robust radiomics models, there are limitations to this study. One major restriction is caused by the intrinsically superimposed nature of CXR images. Ribs and other higher density overlying structures cause noise in the lungs that is detected in the radiomic features. Currently there are some attempts to develop bone suppression software that can retrospectively remove the ribs from CXR [41], but these models were not yet matured enough to apply to our unseen dataset. The successful removal of all superimposed high-density structures might further improve the performance of our model. Cavities are a single representation of PTB, but it is also a common challenge in image-based studies that multiple radiological TB expressions (e.g., adenopathy, infiltrates, and plural effusions) are present on a single x-ray. It is important to note that these other radiological expressions that might be present on the cavity CXR dataset can cause some subjectivity to the outcome of this study. Our sliding window segmentation method was only tested on a single representation of PTB. Future perspectives will extend this technique to other presentations and clinical models.

## Conclusion

In this study two separate radiomic models were constructed, both of which achieved good classification accuracy for normal chest X-rays and cavities on X-rays of people suffering from pulmonary TB. This was achieved by the introduction of a secondary sliding window segmentation that was superimposed on a conventional automatic lung segmentation. This reproducible automatic segmentation method eliminates the difficult and labor-intensive manual disease delineation task, and it alleviates the subjectivity introduced by human judgement on X-rays. The well performing radiomic model could assist clinicians with the prompt diagnosis of pulmonary TB using digital chest X-rays. Accompanied with clinical signs and symptoms, it might aid diagnosis and commencement of pulmonary TB treatment and improve the accuracy of high volume X-ray TB screening or surveillance programs.

## Statements and declarations

**Funding:** The authors declare that no funding, grants, or other financial support were received for the preparation of this manuscript.

**Conflict of interest:** The authors have no competing interests to declare that are relevant to the content of this article.

**Author contributions:** The following authors contributed to the study conception and design; Tamarisk du Plessis, William Rae and Neil Martinson. Material preparation and data collection were performed by Tamarisk du Plessis. Data analysis was performed by Tamarisk du Plessis, Gopika Ramkilawon, William Rae and Tanita Botha with the assistance of Sarah Dixon and Andre Kyme. The first draft of the manuscript was written by Tamarisk du Plessis and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Compliance with ethical standards:** This retrospective study, involving the chest x-rays of human participants, was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Research Ethics Committee of The Faculty of Health Sciences, University of Pretoria, approved this study (Ethics Ref No: 473/2020).

## References

1. WHO (2022) Global Tuberculosis Report 2022. ISBN 978-92-4-006172-9. World Health Organization, <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>
2. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP (2020) Artificial intelligence applications for thoracic imaging. *European Journal of Radiology* 123 <https://doi.org/10.1016/j.ejrad.2019.108774>
3. Karargyris A, Siegelman J, Tzortzis D, Jaeger S, Candemir S, Xue Z, Santosh KC, Vajda Sr, Antani S, Folio L, Thoma GR (2016) Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. *International Journal of Computer Assisted Radiology and Surgery : A journal for interdisciplinary research, development and applications of image guided diagnosis and therapy* 11 (1):99-106 <https://doi.org/10.1007/s11548-015-1242-x>
4. Tan JH, Acharya UR, Tan C, Abraham KT, Lim CM (2012) Computer-Assisted Diagnosis of Tuberculosis: A First Order Statistical Approach to Chest Radiograph. *Journal of Medical Systems* 36 (5):2751-2759 <https://doi.org/10.1007/s10916-011-9751-9>
5. Harisinghani MG, McLoud TC, Shepard J-AO, Ko JP, Shroff MM, Mueller PR (2000) Tuberculosis from Head to Toe. *RadioGraphics* 20 (2):449-470 <https://doi.org/10.1148/radiographics.20.2.g00mc12449>
6. Stephen SFY, Hugo JWLA (2016) Applications and limitations of radiomics. *Physics in Medicine and Biology* 61 (13):R150-R166 <https://doi.org/10.1088/0031-9155/61/13/R150>
7. Brady A, Laoide RÓ, McCarthy P, McDermott R (2012) Discrepancy and error in radiology: concepts, causes and consequences. *The Ulster medical journal* 81 (1):3-9
8. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ (2012) Radiomics: the process and the challenges. *Magnetic resonance imaging* 30 (9):1234-1248 <https://doi.org/10.1016/j.mri.2012.06.010>
9. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 278 (2):563-577 <https://doi.org/10.1148/radiol.2015151169>
10. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, Mattonen SA, El Naqa I (2020) Machine and deep learning methods for radiomics. *Medical physics* 47 (5):e185-e202 <https://doi.org/10.1002/mp.13678>
11. Papanikolaou N, Matos C, Koh DM (2020) How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 20 (1):33 <https://doi.org/10.1186/s40644-020-00311-4>

12. Du Plessis T, Rae WID, Sathekge MM (2021) Pulmonary tuberculosis diagnosis, differentiation and disease management: A review of radiomics applications. *Polish Journal of Medical Physics and Engineering* 27 (4):251-259 <https://doi.org/doi:10.2478/pjmpe-2021-0030>
13. Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, Zhimin L, Tong Y, Jie T, Di D, Yun P (2019) Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. *BMC Medical Imaging* 19 (1):1-11. doi:<https://doi.org/10.1186/s12880-019-0355-z>
14. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B (2020) Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging* 11 (1):91 <https://doi.org/10.1186/s13244-020-00887-2>
15. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M (2018) Radiomics: the facts and the challenges of image analysis. *European Radiology Experimental* 2 (1):1-8 <https://doi.org/10.1186/s41747-018-0068-z>
16. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K (2021) Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis* 72:102125 <https://doi.org/https://doi.org/10.1016/j.media.2021.102125>
17. Cui EN, Yu T, Shang S-J, Wang X-Y, Jin Y-L, Dong Y, Zhao H, Luo Y-H, Jiang X-R (2020) Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans. *World Journal of Clinical Cases* 8 (21):5203-5212 <https://doi.org/10.12998/wjcc.v8.i21.5203>
18. Lohmann P, Bousabarah K, Hoevels M, Treuer H (2020) Radiomics in radiation oncology-basics, methods, and limitations. *Strahlentherapie und Onkologie : Organ der Deutschen Röntgengesellschaft [et al]* 196 (10):848-855 <https://doi.org/10.1007/s00066-020-01663-3>
19. Skoura E, Zumla A, Bomanji J (2015) Imaging in tuberculosis. *International Journal of Infectious Diseases* 32:87-93 <https://doi.org/10.1016/j.ijid.2014.12.007>
20. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E (2021) A deep look into radiomics. *La radiologia medica* 126 (10):1296-1311 <https://doi.org/10.1007/s11547-021-01389-x>
21. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat In, Cook GJR, Davatzikos C, Depeursinge A, Desseroit MC, Dinapoli N, Dinh CV, Echegaray S (2020) The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295 (2):328-338 <https://doi.org/10.1148/radiol.2020191145>
22. Dixon SAP (2019) Using Deep Learning to Segment Chest X-Rays for the Analysis of Pneumonia. University of Sydney,

23. Bain LJ, Engelhardt M (1992) Introduction to probability and mathematical statistics. The Duxbury advanced series in statistics and decision sciences, 2nd ed. edn. Duxbury Press, Belmont, Calif.
24. Du D, Gu J, Chen X, Lv W, Feng Q, Rahmim A, Wu H, Lu L (2021) Integration of PET/CT Radiomics and Semantic Features for Differentiation between Active Pulmonary Tuberculosis and Lung Cancer. *Molecular Imaging & Biology* 23 (2)  
<https://doi.org/10.1007/s11307-020-01550-4>
25. Devore JL, Berk KN (2012) Modern mathematical statistics with applications. Springer texts in statistics, 2nd ed. edn. Springer, New York ;. doi:<https://doi.org/10.1007/978-1-4614-0391-3>
26. Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning : data mining, inference, and prediction. Springer series in statistics, 0172-7397, Second edition. edn. Springer, New York
27. Cordon I, Garcia S, Fernandez A, Herrera F (2018) Imbalance: Oversampling algorithms for imbalanced classification in R. *Knowledge-Based Systems* 161:329-341  
<https://doi.org/10.1016/j.knosys.2018.07.035>
28. Zhang H, Li M (2014) RWO-Sampling: A random walk over-sampling approach to imbalanced data classification. *Information Fusion* 20 (1):99-116  
<https://doi.org/10.1016/j.inffus.2013.12.003>
29. Ramadhan MM, Sitanggang IS, Nasution FR, Ghifari A (2017) Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency. *DEStech Transactions on Computer Science and Engineering (cece)*  
<https://doi.org/10.12783/dtcse/cece2017/14611>
30. Che-Engku-Chik CEN, Yusof NA, Abdullah J, Othman SS, Mat Said MH, Wasoh H (2016) Detection of tuberculosis (TB) using gold standard method, direct sputum smears microscopy, PCR, qPCR and electrochemical DNA sensor: a mini review. Hibiscus Publisher,
31. Khatibi T, Shahsavari A, Farahani A (2021) Proposing a novel multi-instance learning model for tuberculosis recognition from chest X-ray images based on CNNs, complex networks and stacked ensemble. *Physical and Engineering Sciences in Medicine* 44 (1):291-311 <https://doi.org/10.1007/s13246-021-00980-w>
32. Shi W, Zhou L, Peng X, Ren H, Wang Q, Shan F, Zhang Z, Liu L, Shi Y (2019) HIV-infected patients with opportunistic pulmonary infections misdiagnosed as lung cancers: the clinicoradiologic features and initial application of CT radiomics. *Journal of thoracic disease* 11 (6):2274-2286 <https://doi.org/10.21037/jtd.2019.06.22>
33. Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, Yao N, Li Z, Li R, Zhang C, Ji J, Long W (2020) Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *European Journal of Radiology* 128  
<https://doi.org/10.1016/j.ejrad.2020.109022>



34. Cui EN, Yu T, Shang SJ, Wang XY, Jin YL, Dong Y, Zhao H, Luo YH, Jiang XR (2020) Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans. *World journal of clinical cases* 8 (21):5203-5212  
<https://doi.org/10.12998/wjcc.v8.i21.5203>
35. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Ayles H, Sánchez CI (2016) On Combining Multiple-Instance Learning and Active Learning for Computer-Aided Detection of Tuberculosis. *IEEE Transactions on Medical Imaging* 35 (4):1013-1024  
<https://doi.org/10.1109/TMI.2015.2505672>
36. Santosh KC, Antani S (2018) Automated Chest X-Ray Screening: Can Lung Region Symmetry Help Detect Pulmonary Abnormalities? *IEEE transactions on medical imaging* 37 (5):1168-1177 <https://doi.org/10.1109/TMI.2017.2775636>
37. Han Y, Chen C, Tang L, Lin M, Jaiswal A, Ding Y, Peng Y (2020) Using Radiomics as Prior Knowledge for Abnormality Classification and Localization in Chest X-rays.
38. Havlicek LL, Peterson NL (1976) Robustness of the Pearson Correlation against Violations of Assumptions. *Perceptual and Motor Skills* 43 (3\_suppl):1319-1334  
<https://doi.org/10.2466/pms.1976.43.3f.1319>
39. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H (2017) Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer research* 77 (21):e104-e107  
<https://doi.org/10.1158/0008-5472.CAN-17-0339>
40. Wang X-H, Long L-H, Cui Y, Jia AY, Zhu X-G, Wang H-Z, Wang Z-H, Wang W-H (2019) A MRI-based radiomics model for preoperative prediction of five-year survival status in hepatocellular carcinoma. *Journal of Clinical Oncology* 37 (15\_suppl):e15596  
[https://doi.org/10.1200/JCO.2019.37.15\\_suppl.e15596](https://doi.org/10.1200/JCO.2019.37.15_suppl.e15596)
41. Zhou Z, Zhou L, Shen K (2020) Dilated conditional GAN for bone suppression in chest radiographs with enforced semantic features. *Medical physics* 47 (12):6207-6215  
<https://doi.org/10.1002/mp.14371>

## CHAPTER 5

### **Quantitative chest x-ray radiomics for therapy response monitoring in patients with pulmonary tuberculosis**

*Tamarisk du Plessis<sup>1\*</sup>, William I.D. Rae<sup>2</sup>, Gopika Ramkilawon<sup>3</sup>, Neil A. Martinson<sup>4,5</sup>, Mike M. Sathekge<sup>1</sup>*

1. Department of Nuclear Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, 0001, South Africa
2. Medical Imaging Department, Prince of Wales Hospital, Sydney, NSW 2031, Australia
3. Department of Statistics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Pretoria, 0081, South Africa
4. Perinatal HIV Research Unit (PHRU), University of the Witwatersrand, Johannesburg, 1862, South Africa
5. Centre for Tuberculosis Research, Johns Hopkins University, Baltimore, MD, 21205, USA

\*Correspondence: [tamarisk.duplessis@gmail.com](mailto:tamarisk.duplessis@gmail.com)

*Diagnostics. 2023;13(17):2842-2856*

*doi: 10.3390/diagnostics1317284*

## Abstract

Tuberculosis (TB) remains the second leading cause of death globally from a single infectious agent, and there is a critical need to develop improved imaging biomarkers and aid rapid assessments of responses to therapy. We aimed to utilize radiomics, a rapidly developing image analysis tool, to develop a scoring system for this purpose. A chest X-ray radiomics score (RadScore) was developed by implementing a unique segmentation method, followed by feature extraction and parameter map construction. Signature parameter maps that showed a high correlation to lung pathology were consolidated into four frequency bins to obtain the RadScore. A clinical score (TBscore) and a radiological score (RLscore) were also developed based on existing scoring algorithms. The correlation between the change in the three scores, calculated from serial X-rays taken while patients received TB therapy, was evaluated using Spearman's correlation. Poor correlations were observed between the changes in the TBscore and the RLscore (0.09 ( $p$ -value = 0.36)) and the TBscore and the RadScore (0.02 ( $p$ -value = 0.86)). The changes in the RLscore and the RadScore had a much stronger correlation of 0.22, which is statistically significant ( $p$ -value = 0.02). This shows that the developed RadScore has the potential to be a quantitative monitoring tool for responses to therapy.

## Keywords:

radiomics; tuberculosis; segmentation; feature extraction; chest x-rays; radiomics score

## 1. Introduction

Tuberculosis (TB) is a leading cause of death globally [1]. Planar chest X-rays (CXRs) are a method commonly used for early pulmonary TB (PTB) diagnosis in patients with clinical symptoms suggestive of TB. CXRs are used rather than more sophisticated three-dimensional modalities because they are the most widely accessible and inexpensive imaging modality in countries where TB is most prevalent [2-4]. Follow-up CXRs are also widely used in the management of PTB to evaluate clinical responses to treatment, though there are no objective tests for improvement. Intra- and interobserver discrepancies are common in X-ray reporting [5], and one study found an error rate of up to 23% when experienced radiologists reported on a collection of normal and abnormal X-rays [6].

For an assessment of responses to PTB treatment, in some settings, a follow-up CXR taken at the completion of the 2-month intensive four-drug treatment phase is often visually compared to the CXRs taken at the initiation of TB treatment [7]. With a subjective improvement in CXRs and symptoms, patients proceed to the two-drug continuation phase of TB treatment [7]. Currently, CXRs are only used for qualitative visual assessments in PTB management and are typically reported as either “improved” or “worse” compared to previous CXRs. However, if PTB characteristics can be quantified from chest X-rays, these characteristics might be used to evaluate disease response and could assist with identifying those who do not respond to TB treatment earlier in their treatment course.

Laboratory results are considered as ground truth for PTB management to determine responses to treatment [8]. However, in cases where laboratory results are not available, clinical signs and symptoms related to PTB is also a good indicator of treatment response. Scoring systems are tools designed to predict outcomes, assist in clinical decision-making, support treatment options and manage clinical risk [9]. Various clinical scoring algorithms for TB exist, and a systematic review study summarized the sensitivity and specificity of some of them [10]. The TBscore is one such system that was developed as a rapid and inexpensive tool to monitor TB patients in their treatment and to assess clinical outcomes [11]. This scoring system included five symptoms that were self-reported (cough, dyspnoea, night sweats, haemoptysis and chest pain) and six signs that were assessed by a trained nurse (anaemia, tachycardia, positive finding on lung auscultation, fever, body mass index, and mid-upper arm circumference) [11]. This clinical score was also useful for TB diagnosis among adults living with HIV who presented to healthcare institutions with signs and

symptoms suggestive of TB with high sensitivity (95.5%) and a specificity of 36.9% when a cut-off value of two was implemented [12]. In our study, we used this TB scoring system and modified it slightly for the population group under consideration.

Several radiological chest X-ray scoring systems have been reported, for example the RALE (Radiographic Assessment of Lung Edema) score and the BRIXIA Chest X-ray Severity Scoring System [13]. Both these scoring systems have lately been re-purposed, for COVID-19 research [13-16]. In the RALE scoring system, each lung (left and right, respectively) is scored from 0 to 4, with 0 for no disease involvement and 4 for more than 75% involvement, with a total score out of 8 [13]. The BRIXIA score is used to grade lung abnormalities caused by disease on an 18-point severity scale [17]. This scoring system divides the lungs into 6 regions (3 regions for each lung), and each region is given a score (from 0 to 3) based on the extent of the lung abnormalities detected [17]. For our retrospective study, we developed a TB radiological scoring system based on the aforementioned scoring systems that could be used to determine the correlation of newly developed scores.

Studies on the use of computer science from medical images for the diagnosis of PTB or on the differentiation of TB from another pulmonary disease have been reported [18-20], but very little evidence exists on the use of radiological image analysis for monitoring treatment response and prognosis [21]. As radiomics is gaining popularity in the medical field, it may be beneficial to use this to assess responses to TB therapy.

The motivation for our study was based on a recent study that developed a radiomics score (rad score) from CT scans which acted as a potential prognostic imaging feature for postoperative survival in solitary hepatocellular carcinoma (HCC) patients [22]. They identified six signature features and then applied the least absolute shrinkage and selection operator (LASSO) logistic model to develop the optimal rad-score algorithm [22]. This algorithm ensured that each feature contributed equally to the rad-score since the dimensionality of features could differ significantly. This, in essence, is just a different approach to normalization. This study concluded that their rad-score might be complementary to the current tumour staging systems [22].

Most radiomic studies are performed using data from CT scans because of simplified disease segmentation and exposed (non-superimposed) anatomy. If PTB is the disease under investigation, CXRs would be more practical as they are an inexpensive, widely accessible

imaging modality for disease diagnosis and management [4]. Our study, therefore, used a chest X-ray dataset of patients diagnosed with active PTB.

Some restrictions exist on performing radiomic feature extraction from planar chest X-rays, for example, the superimposed chest anatomy overlying the disease of interest and accurate disease segmentation to select the ROI for feature extraction [23]. To address this, a recent study developed a sliding window segmentation that was applied as a secondary segmentation to the entire segmented lung to include the normal tissue and the disease [23]. This article shows that the sliding window segmentation can eliminate the need for accurate disease segmentation on planar images, as it produces accurate and reproducible radiomic signatures and models (AUC = 0.9444 (95% CI [0.8762, 0.9814])) [23]. This sliding window segmentation method was applied to our study, even though the traditional consequent steps of dimensionality reduction, signature development and model construction were not followed through.

In this study, a unique approach to dimensionality reduction and signature construction was applied to assess changes in CXRs that could be linked to other objective measures of TB treatment response using a novel CXR radiomics score. The correlation between the radiomics score and several calculated clinical- and radiological scores was evaluated.

## **2. Methodology**

### **2.1. Patient selection**

This was a secondary data analysis on the clinical- and radiological data of 111 patients collected for an observational study between August 2013 and July 2018. These patients had a laboratory-confirmed diagnosis of PTB with clinical data recorded and a CXR acquired at the baseline- and first follow-up clinic visit. The mean time lapse between the baseline and the first follow-up visit was 54.39 days (a range of 50–64 days), with all X-rays taken within one week of the date of the clinical visit. The CXRs were all reported by radiologists or attending doctors with extensive experience in PTB X-ray reporting.

### **2.2. Clinical score (TBscore)**

The TBscore algorithm was used to calculate a clinical score for each patient [12]; it was minimally altered (see Table 1) to only include those signs and symptoms that were included

in the initial clinical study database. Tolerances for BMI and pulse rate were intentionally altered by expert TB clinicians to have a similar quantitative contribution as the original TBscore. In the original TBscore algorithm a BMI < 18 kg/m<sup>2</sup> scored 1, a BMI < 16 kg/m<sup>2</sup> scored 2 and a pulse rate of >90 bpm scored 1 with no value for score 2 [12].

**Table 1:** Clinical score chart used in this study with a maximum score of 12

Sign or Symptom	CRITERIA		
	0	1	2
<b>How do you feel?</b>	7 - 10 (Good)	4 - 6 (Okay)	≤3 (Awful)
<b>Fever (°C)</b>	≤37.5 (NO)		>37.5deg (YES)
<b>Pulse rate (bpm)</b>	< 90	90 - 100	> 100
<b>Cough (days)</b>	No cough	< 60 or No cough at prior visit	≥60 or No cough at prior visit
<b>BMI (kg/m<sup>2</sup>)</b>	>20	18 - 20	<18
<b>Nights sweats (days)</b>	0	< 60 days	≥ 60days

°C = Degrees Celsius, bpm = beats per minute, kg/m<sup>2</sup> = kilogram per meter squared

In this study four self-reported symptoms were included; general well-being (recorded as a self-reported answer to “how does the patient feel?” as either a score out of 10 (with 10 being “excellent” and 1 “awful”), or on a qualitative scale as “Good”, “Okay” or “Awful”), cough (duration in days coughing was experienced), BMI (body-mass-index in kg/m<sup>2</sup>) and night sweats (number of days night sweats were experienced to date). Two clinical signs were included; temperature (oral or axillary temperature ≥ 37.5 °C) and pulse rate (in beats per minute) [12]. The symptoms and signs all had a possible score of 0, 1 or 2 resulting in a maximum clinical score out of 12 (see Table 1).

### 2.3. Radiological score (RLscore)

The radiological dataset had the following 4 data entries recorded that were applicable to this study: (1) The CXRs’ acquisition date. (2) The presence of 4 radiological TB expressions (cavities, infiltrates, adenopathy and pleural effusion) in the left or right lung recorded separately. This was recorded as either ‘YES’ or ‘NO’, where ‘YES’ indicated disease presence and ‘NO’ indicated absent disease. For scoring purposes, binary values were assigned to the conditions in this study as YES = 1 and NO = 0. This added up to a maximum score out of 8 if all 4 expressions were present in both lungs. (3) The CXRs’ cavitation classification (recorded as 1, 2 or 3) were according to the conditions set out in Table 2. (4) Additionally, the extent of disease classification (recorded as A, B or C) was according to the

conditions set out in Table 2. For scoring purposes, numerical values were assigned to each condition: A = 1, B = 2 and C = 3. Normal X-rays = 0.

**Table 2:** Radiological score conditions of the X-ray classification

<b>X-ray classification: Cavitation</b>		<b>X-ray classification: Extent of disease</b>	
1	Absent, as seen on posteroanterior (PA) or anteroposterior (AP) CXR view.	A (1)	Limited: Lesion(s) involving a total lung area less than one-quarter the area of the entire thoracic cavity, as seen on PA or AP view.
2	Single or multiple cavities with diameter < 4 cm in aggregate (for each cavity, measure at point of maximum diameter) for PA or AP CXR view.	B (2)	Moderate: Lesion(s) of greater than A, but, even if bilateral, involve a total lung area of less than one-half the area of the entire thoracic cavity, as seen on PA or AP view.
3	Single or multiple cavities with diameter ≥ 4 cm in aggregate (for each cavity, measure at point of maximum diameter) for PA or AP CXR view.	C (3)	Extensive: Lesion(s) involving a total lung area equal to, or more than half the area, of the entire thoracic cavity, as seen on PA or AP view.

A radiological score (RLscore) was obtained for each x-ray by including the recorded TB expressions (possible 8 points), cavitation classification (possible 3 points) and the extend of the disease (possible 3 points). It was calculated according to the algorithm in Equation 1 for a maximum score out of 14.

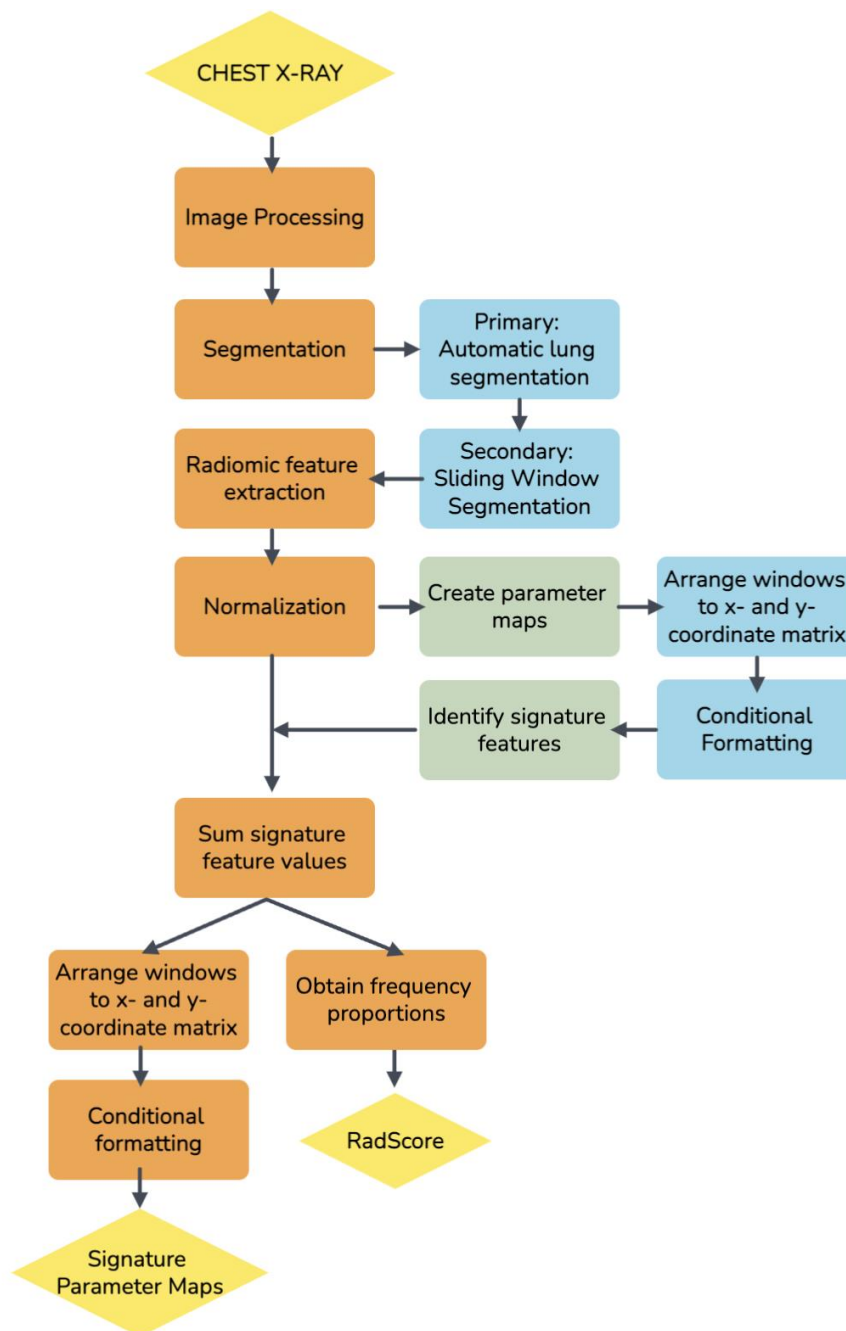
$$\begin{aligned}
 RLScore = & \text{Sum}(\text{Radiological TB expressions}) + (\text{Extend of disease}) \\
 & + (\text{Cavitation classification})
 \end{aligned}$$

[Eq. 1]

#### 2.4. Radiomics Score (RadScore)

A non-traditional method was followed to develop the radiomics score (RadScore) in this study when a sliding window segmentation was introduced [23]. Figure 1 is a schematic overview of the approach followed.





**Figure 1:** A schematic overview of the process followed to develop a radiomics score (RadScore) in this study

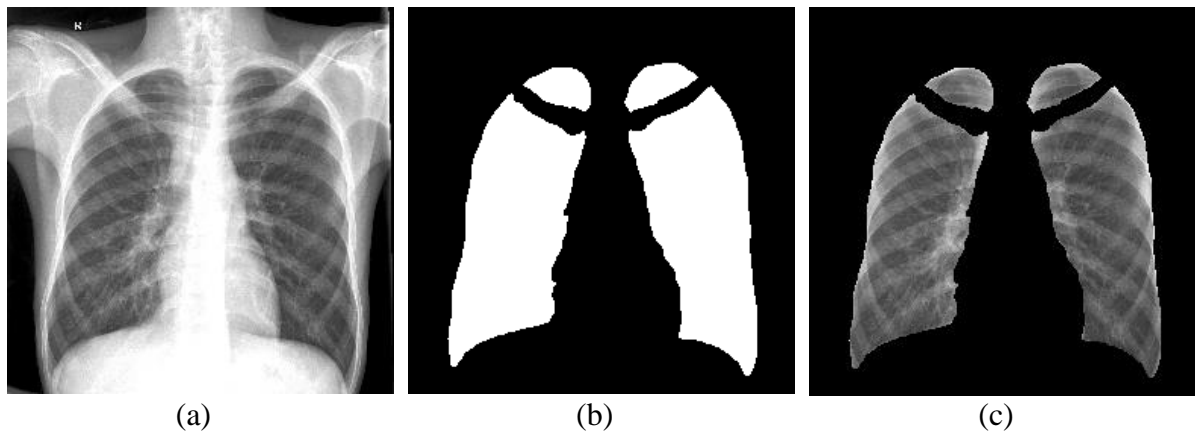
### Image Processing:

The Total Image Converter (version 8.2.0.237) and Python (version 3.7.6) were used for initial pre-processing to ensure that a uniform CXR dataset was used as a cohort and that it was in the required format for the radiomics library. Image processing included manually cropping all images to square dimensions, correcting unconventional photometric interpretations on some images, converting DICOM images to the PNG format, interpolating

all images to  $256 \times 256$  pixels with bilinear interpolation, and converting the conventional RGB type of PNG images to the scalar type

### Primary Segmentation:

A fully automatic in-house segmentation model was used to segment the lungs as the primary segmentation [24]. The segmentation model resized images to  $256 \times 256$  pixels with bilinear interpolation before segmenting the lungs as a  $256 \times 256$  pixel mask output. Figure 2 is an example of the lung segmentation achieved with the clavicles removed.



**Figure 2:** An example of the output of the primary segmentation model, (a) The original image, (b) The mask (multiplied by 255 to be visually visible) and (c) The mask multiplied with the original image. This image was used to evaluate the accuracy of the primary segmentation model

### Secondary Segmentation & Radiomic Feature Extraction

Sliding window masks were created in Python (version 3.7.6) using `Numpy.array()` and `PIL.Image()` functions. A square sampling window mask ( $w$ ) with  $16 \times 16$  pixels and a window step size ( $w_{step}$ ) of 4 pixels was selected as has been previously described [23]. The window moved, or slid, from one side of the CXR to the other in both  $x$ - and  $y$ -dimensions to create a new window matrix of  $61 \times 61$  windows. The number of windows in the matrix can be calculated using Equation 2 where  $[P_x, P_y]$  refers to the dimensions of the window matrix,  $[n_x, n_y]$  is the dimensions of the image in pixels,  $w$  is the window size and  $w_{step}$  is the window step size [23].

$$P_x = \frac{(n_x - w)}{w_{step}} + 1 \quad \text{AND} \quad P_y = \frac{(n_y - w)}{w_{step}} + 1 \quad [\text{Eq.2}]$$

The sliding windows, therefore, cause the effective dimensionality of each image's features to increase by a factor of 3721 ( $61 \times 61$ ). This resolution was thought to be adequate to resolve the change in radiomic features across the lung within an acceptable computational time.

The sliding window masks were superimposed on the primary segmented lung mask of each CXR. Radiomic features were extracted from each window in the window matrix if the window was not masked off by the lung segmentation.

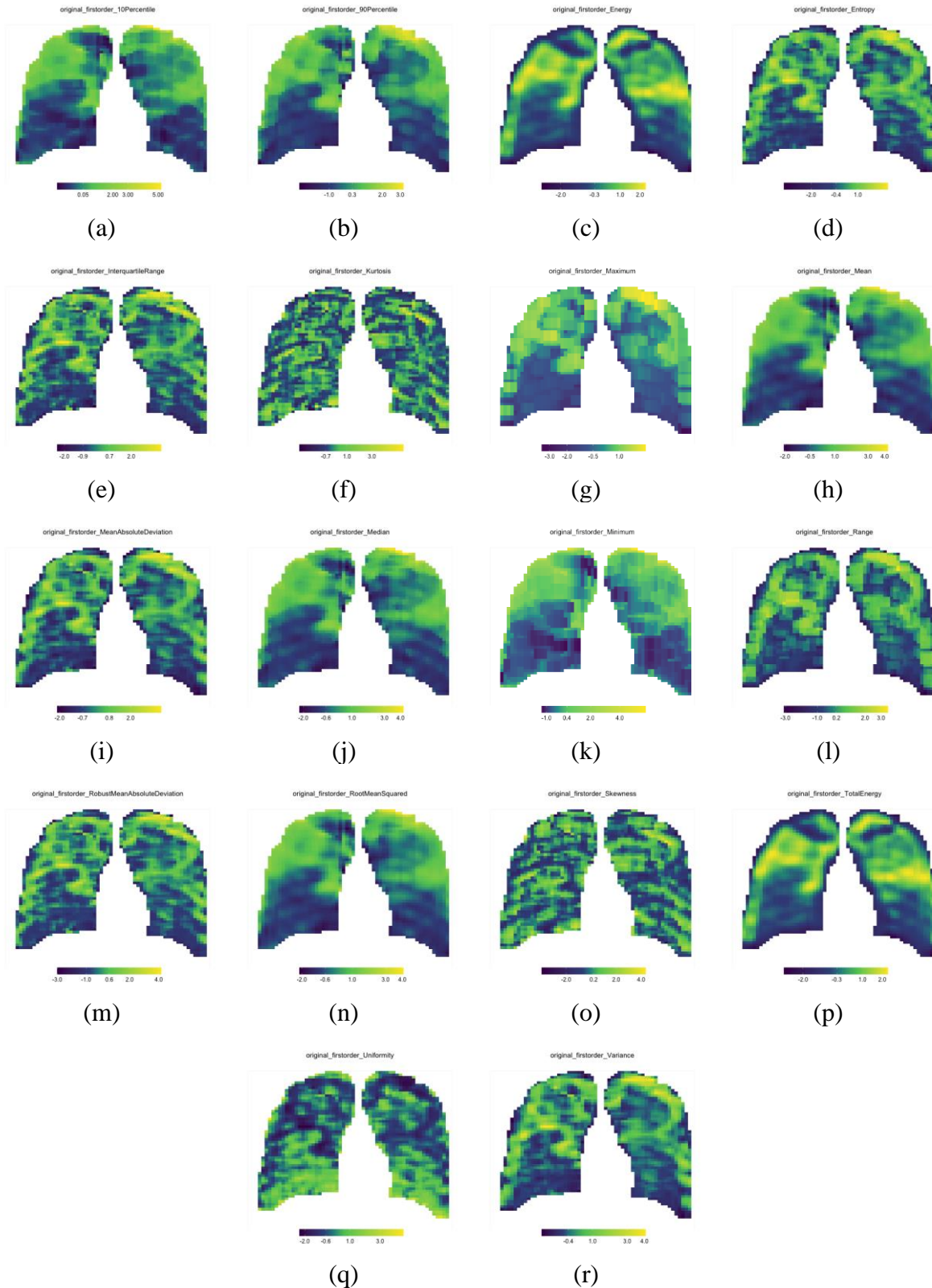
The Pyradiomics library (version 3.0) was used to extract 93 two-dimensional (2D) first-order and texture features from each sliding window on each CXR. Shape-based features were omitted, as these features used the ROI defined by the mask to calculate the values, which is meaningless for the purpose of this study.

### Creating Parameter Maps

All data processing was performed using R Software (version 4.1.3). From Equation 2, it can be seen that 3271 window features were extracted per feature from each CXR. To account for the dimensional variability in these features, all windows' features were normalized with z-score normalization. All feature values were redistributed to the correct  $x$ - and  $y$ -coordinates in the sliding window matrix. This was conducted for each of the 93 features extracted on each of the 111 CXRs. Heatmaps were applied, and a parameter map was printed for each feature (see Figure 3).

### Signature Features & Signature Parameter Map

A visual inspection of the parameter maps was the only method to identify the features that could result in a robust and reproducible radiomics signature to collate with the extent of the disease. Due to the enormous number of maps under investigation, it was decided to sample a simple random sample comprising more than 10% of the parameter maps. All parameter maps from this sample were visually inspected and compared to the original X-ray to identify the features that represented the disease. Six features were identified on 100% of the randomized samples' CXRs, and these six features were acknowledged as the signature. The 6 signature feature values were summed to obtain a single signature value for each window in the parameter map out of a maximum of 6 (the sum of 6 normalized features). Since we noticed that these features all increased with an increase in disease status, we expected that the summed features would further highlight the diseased and the higher-density areas in the lung. When the 3721 values of the signature features were again redistributed to the correct  $x$ - and  $y$ -coordinates, a single signature parameter map of  $61 \times 61$  windows for each CXR was developed.



**Figure 3:** First order feature parameter maps for a single baseline CXR of Patient A; (a)10<sup>th</sup> Percentile, (b)90<sup>th</sup> Percentile, (c)Energy, (d)Entropy, (e)Interquartile range, (f)Kurtosis, (g)Maximum, (h)Mean, (i)Mean-Absolute-Deviation, (j)Median, (k)Minimum, (l)Range, (m)Robust-Mean-Absolute-Deviation, (n)Root-Mean-Squared, (o)Skewness, (p)Total-Energy, (q)Uniformity and (r)Variance

### Developing the Radiomics Score (RadScore)

To develop a radiomics score that can be compared to a single integer radiology and clinical score, all the signature parameter maps were statistically summarized into a single value. The summed features for each CXR, with a numerical range from 0 to 6, were consolidated into four frequency bins with a bin width of 1.5. These 4 groups were [0.0; 1.5], (1.5; 3.0], (3.0; 4.5] and (4.5, 6.0]. The corresponding frequencies of the summed features for each group were calculated to obtain crucial information about where most feature information lies. Since the signature feature values increased with increased disease status, we could assume that a higher score indicated a worse state of PTB and would likely fall in the (3.0; 4.5] or (4.5, 6.0] frequency group with healthier lung tissues falling in the [0.0; 1.5] and (1.5; 3.0] groups. Due to dimensionality issues caused by the image segmentation techniques applied to each CXR, the summed features did not all contain information at the exact same windows from baseline to follow-up. The proportions were, therefore, used for each of the four groups to circumvent spurious results and to improve accuracy in comparisons between the baseline and follow-up CXRs.

### **2.5. Longitudinal Change**

To evaluate the change in the RadScore, RLscore and the TBscore, the follow-up visit's score was subtracted from the baseline visit's score according to Equation 3.

$$\text{Change}(\text{Score}) = \text{First visit}(\text{Score}) - \text{Second visit}(\text{Score}) \quad [\text{Eq.3}]$$

A positive change, therefore, indicates a radiological or clinical improvement, a negative change indicates worsening, and 0 indicates no change.

Spearman's correlation analysis (the non-parametric equivalent of Pearson's correlation analysis) was performed with a test of significance to evaluate the RadScore, TBscore and RLscore correlations. Spearman's test was used since the clinical score was found to not be normally distributed at a 5% level with a  $p$ -value of 0.01 using the Shapiro–Wilk test of normality.

### 3. Results

#### 3.1. Parameter Maps

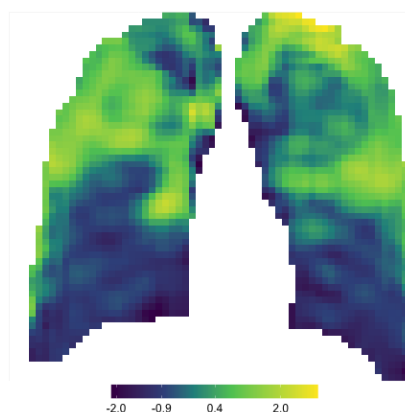
Parameter maps are the normalized features extracted from each CXR's sliding window mask and re-distributed into the correct spatial dimensions with conditional formatting applied. Figure 3 is only the 18 first-order features' parameter maps of a single CXR (Patient A: Baseline CXR).

Visual evaluation was used to identify the features whose parameter maps correlated with disease pathology in the lungs. Table 3 provides a list of the six features identified as signature features.

**Table 3:** The following features were identified as signature features that highlight and quantify the disease and correlate to the lung pathology

Feature number	Feature name
1	First order – 90 <sup>th</sup> percentiles
2	First order – Median
3	First order – Mean
4	First order – Energy
5	First order – Root mean square
6	First order – Total Energy

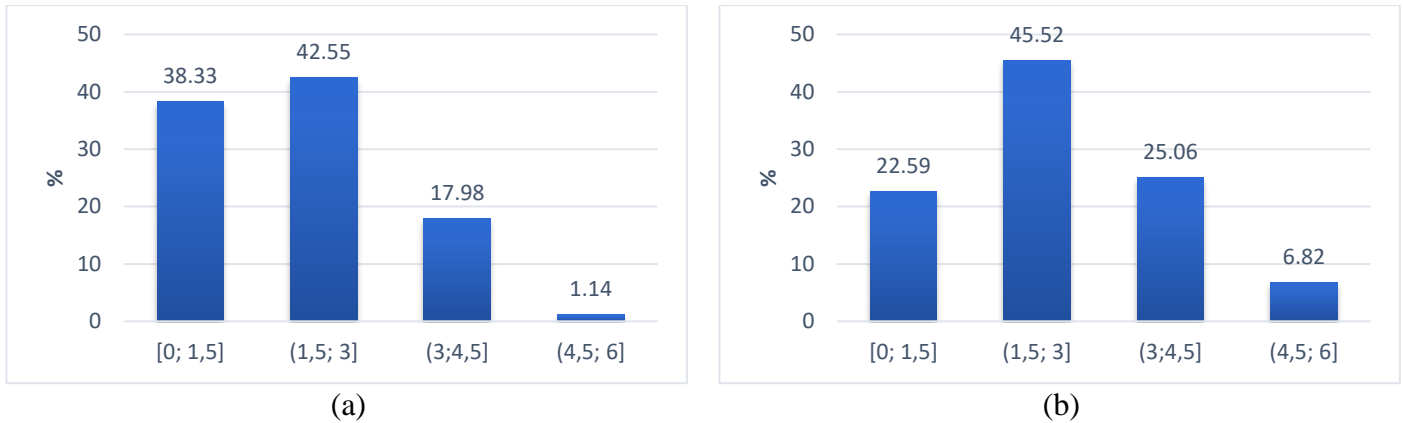
To obtain signature parameter maps, the six signature features' values were summed to obtain a signature value for every window out of six. These values were again re-distributed to the correct  $x$ - and  $y$ -coordinate matrix, and conditional formatting was applied to obtain a single signature parameter map for each CXR. Figure 4 is an example of the signature parameter map of Patient A's baseline CXR (the same patient used in Figure 3).



**Figure 4:** Signature parameter map obtained for Patient A's baseline CXR

### 3.2. Developing a RadScore

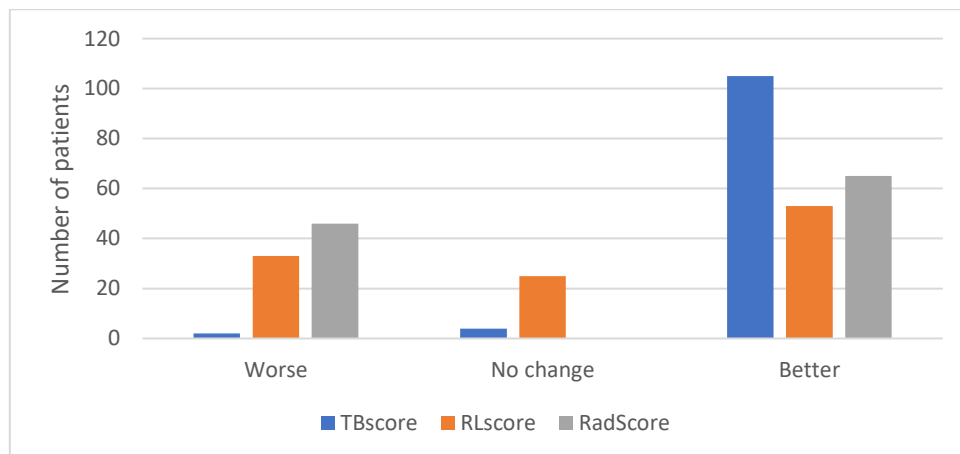
Frequency bins were used to develop a RadScore for each patient at each clinical visit. Figure 5 shows the frequency bins of Patient A at baseline and first follow-up CXR.



**Figure 5:** (a) Plot of the frequency proportions of a baseline CXR of Patient A indicating the four groups used to obtain a radiomics score and (b) Plot of the frequency proportions of the corresponding follow-up CXR of Patient A

### 3.3. Change in RLscore, TBscore & RadScore

Equation 3 was used to calculate the longitudinal change in the RLscore, TBscore and the RadScore with the results shown in Figure 6.

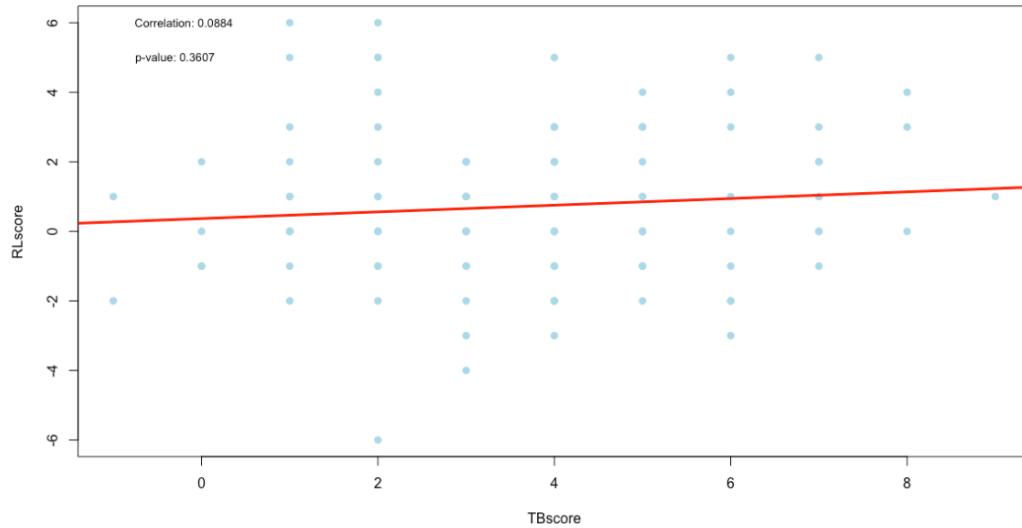


**Figure 6:** Graph indicating the number of patients who showed a decline, no change or an improvement in the TBscore, RLscore and RadScore

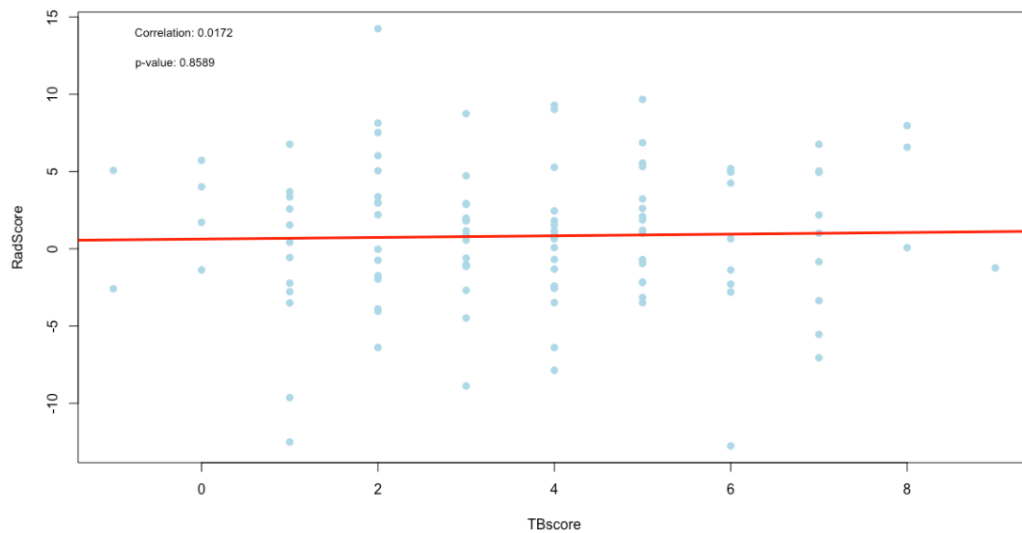
The correlation between the change in the scores were evaluated using Spearman's correlation, with a test of significance for the correlation. The results are recorded in Table 4, Figure 7 (TBscore vs. RLscore), Figure 8 (TBscore vs. RadScore) and Figure 9 (RLscore vs. RadScore).

**Table 4:** The correlation between the change in the TBscore, the RLscore and the RadScore as calculated using Spearman’s correlation analysis with a  $p$ -value testing the significance of the correlation

Scores Being Compared	Correlation Value	p-value
TBscore vs. RLscore	0.0884	0.3607
TBscore vs. RadScore (prop. 3 to 6)	0.0172	0.8589
RLscore vs. RadScore (prop. 3 to 6)	0.2178	0.0216

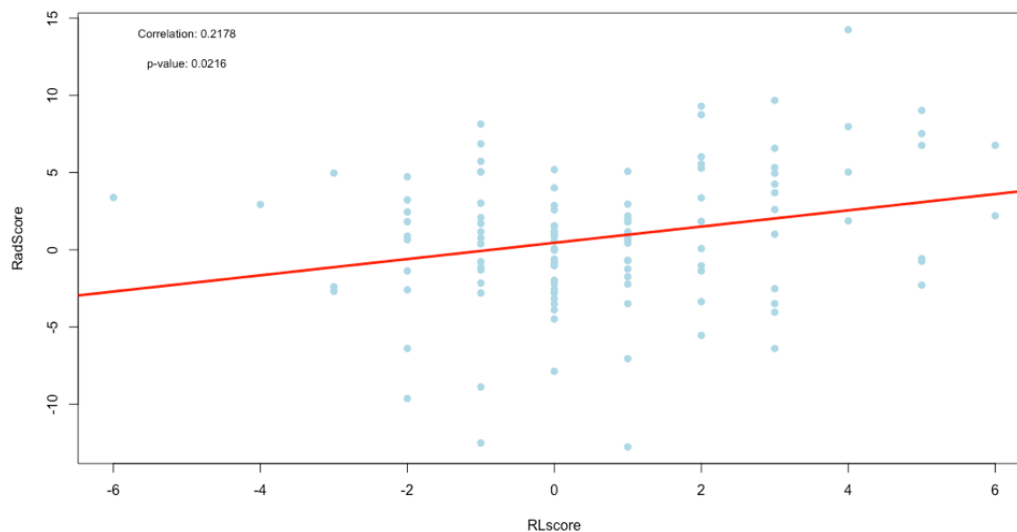


**Figure 7:** Plot of the correlation between the TBscore and the RLscore



**Figure 8:** Plot of the correlation between the TBscore and the RadScore





**Figure 9:** Plot of the correlation between the RLscore and the RadScore

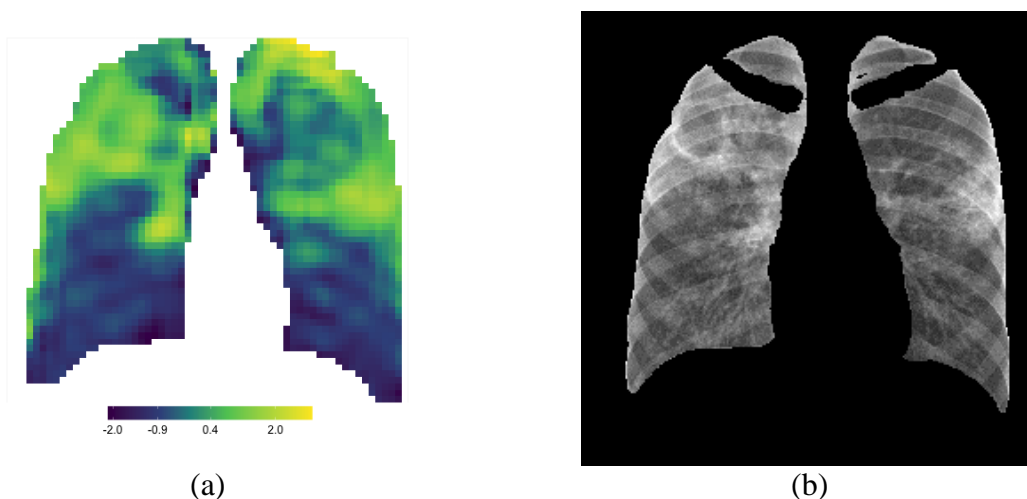
#### 4. Discussions

In this study, we developed a TB radiomics score (RadScore) from chest X-rays without any need for disease segmentation. In traditional radiomic studies, the disease under investigation (mostly tumours) is precisely delineated, either manually or with an automatic or semi-automatic model [25, 26]. The radiomic features are then extracted in the segmented ROI only, and a signature is developed that describes the properties of the delineated disease. However, the precise delineation of a non-neoplastic disease such as TB is very difficult. The sliding window method that was introduced as a secondary segmentation allowed us to evaluate the change in radiomic features across the entire lung region. By re-distributing every window's extracted feature into the corresponding  $x$ - and  $y$ -coordinates, parameter maps could be developed. These parameter maps are, in essence, a quantified interpretation, either as a first- or second-order (texture features) statistical algorithm of the original image. We noticed that only first-order features' parameter maps corresponded to lung pathology. This was expected as the texture features derived from secondary matrixes are too far removed from the original grayscale intensities of the CXR.

The parameter maps appeared as smoothed image versions because the sliding windows in the secondary segmentation reduced the resolution of the image. This segmentation allows for some overlap of the pixels from one window to the next, which ensures that no image information is lost. When the window size of the secondary segmentation is reduced, the signature parameter map seems less smooth, but this is at the cost of additional computational

time and resources. The boundaries between the features in this study are well resolved as we used a window step size of only 4 pixels [23].

Signature parameter maps were developed in this study by adding the six features that individually corresponded to lung pathology. Other mathematical and statistical methods were also considered to develop a single signature value for each window, but it was found that the simple sum of these features was the optimal choice to highlight lung pathology compared to the standalone features' parameter maps. Figure 10 shows the strong correlation between the radiomic signature map (a) and the original CXR (b). Since each window in the signature parameter map had a normalized quantitative signature value that was directly proportional to the colour scale, this parameter map could be used not only as a visual tool enhancing radiological features but also as a quantitative tool to assist in CXR reporting, especially in disease management. When the parameter map of a baseline CXR is compared to that of a follow-up CXR, uncertainties in visual assessment can be eliminated using quantitative comparisons of the lung ROI.



**Figure 10:** The radiomic signature parameter map (a) correlates strongly to the lung pathology on the CXR (b)

If image registration was included as an initial step in this study, it might have been possible to use image subtraction of the baseline and follow-up CXR to determine differences in the signature maps. These differences would have clearly indicated the change in the extent of the disease (better or worse) and the position, visually and quantitatively. However, in this retrospective study, image registration was not considered. For individual cases, manual quantitative comparisons of the lung region of interest could be made, but this is too time-consuming to incorporate as a standard practice in the clinical environment.

The aim of our study was ultimately to develop a radiomics score that could automatically assign a single numerical value to each CXR and contribute to quantitative X-ray reporting, which is essential but still lacking for evaluating TB responses to treatment. Various consolidation methods were evaluated; however, we concluded that by using the change in the proportion of the windows (with feature values between 3.0 and 6.0), the most reliable score could be achieved. To evaluate the relevance of the developed radiomics score (RadScore), its correlation to a developed clinical- and radiological score was tested.

Figures 7-9 are the plots of the correlations, and visually they seem to indicate the presence of outliers. The presence of outliers is known to be problematic, particularly within correlation analysis [27]. To this end, all outliers were tested to see if they are statistically significant using Grubbs Test [28]. Furthermore, the practical and clinical importance of these outliers were carefully considered as well. In all cases, the outliers were statistically insignificant and were therefore not removed.

The TBscore and the RLscore indicated a poor correlation of 0.09 with no statistical significance ( $p$ -value = 0.36). This might be accounted for by the slower improvement seen in lung pathology on CXR compared to a faster response to treatment to the clinical signs and symptoms of PTB. Additionally, the lower sensitivity of X-rays to detect minor changes in lung pathology can be a contributing factor [29]. This could also explain the even poorer correlation of 0.02 between the TBscore and the RadScore, with no statistical significance ( $p$ -value=0.86). It is known that CXR plays a vital role in the management of PTB, but that it cannot be used as a standalone tool [30]. The poor correlations between the clinical signs and symptoms and the RL- and the RadScore also show that they cannot be used in isolation from laboratory-confirmed results in the management of PTB.

The RLscore and the RadScore had a much stronger correlation of 0.22, which was statistically significant ( $p$ -value = 0.02). This shows that the RadScore accurately quantifies the subjective radiological reports. The RadScore might, therefore, have the potential to eliminate the need for CXR reporting in TB management programs, which is currently only qualitative. This could benefit some of the world's highest TB-burdened countries that have limited resources and a shortage of clinicians experienced in CXR reporting. If the RadScore is applied and supported with the developed signature parameter maps as a quantitative objective interpretation of CXR, it could improve the accuracy of X-ray reporting and strengthen the current role of CXRs in TB management programs.

Since the x-rays were used as foundational input for the RadScore construction an excellent correlation between the RL- and the RadScore was expected if the dimensionality reduction and signature development were accurately applied, which we ensured in this study. The likely cause of the lower-than-expected correlation is the high-density ribs that were superimposed on the lung anatomy in the lung ROI. These higher density overlying structures cause noise that is detected in the radiomic features. This reduced the accuracy of the features extracted, especially in the lateral lung regions as seen in Figure 4. We attempted to address this issue by applying a bone suppression model, which was proven to outperform most other available models at the time [31], to all CXRs in this study before the primary lung segmentation. This was, however, unsuccessful as this model was still too immature at the time of this study due to the small training cohort, with a limited dynamic range, that was used to train this model. When a matured and accurate bone suppression model is developed that can retrospectively remove the ribs, most of the noise that influences the radiomic features will be removed. The radiomics scoring method proposed in this study would then most likely result in a RadScore that has a very high correlation to the RLscore.

The methodology introduced in this study is labour-intensive and expensive in terms of knowledge and computational resources. A future perspective is to develop this into software with a user-friendly interphase where a CXR image can easily be uploaded and the clinical data entered. The software can then automatically perform the image post-acquisition processing, apply the segmentation model and radiomic feature extraction library to the CXR, and complete the rapid parameter map and score constructions. If this is developed as an application rather than server-based software, it should be accessible from any computer, regardless of its processing power. This eliminates the restriction of computational power, which might be a limiting factor in many high-TB-burdened countries. Following this, RadScore should be useful as a quantitative tool to evaluate the changes in PTB disease characteristics as seen from CXRs, which could greatly assist clinicians, especially in resource-limited countries, in the management of PTB.

## 5. Conclusions

In this study, a radiomics score (RadScore) was developed from chest X-rays that showed a good correlation, with statistical significance, to a developed radiological score. It was shown that the RadScore can be used to quantify the change in the disease characteristics seen on X-rays of patients diagnosed with pulmonary TB. Integrating the RadScore as a quantitative objective interpretation of X-rays could improve the accuracy of X-ray reporting and strengthen its role in TB management programs. As part of the RadScore construction, signature parameter maps were created that showed excellent qualitative correlations that could further increase the acceptance of chest X-rays as a quantitative tool for assessing TB response in medical research and clinical practice.

**Author Contributions:** Conceptualization, T.dP., W.R. and N.M.; Methodology, T.dP. and W.R.; Software, T.dP.; Validation, T.dP. and G.R.; Formal analysis, T.dP. and G.R.; Investigation, T.dP.; Resources, N.M.; Writing—original draft preparation, T.dP.; Writing—review and editing, W.R., N.M. and M.M.; Supervision, M.M.; All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by the Department of Nuclear Medicine, University of Pretoria, South Africa.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of The Faculty of Health Sciences, University of Pretoria (Ethics Ref No: 473/2020) on 13 August 2020.

**Informed Consent Statement:** Informed patient consent was waived due to the retrospective nature of this study.

**Data Availability Statement:** Data is available on requests from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## References

1. WHO. Global Tuberculosis Report 2021. In. Geneva: World Health Organization, 2021
2. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Ayles H, Sánchez CI. On Combining Multiple-Instance Learning and Active Learning for Computer-Aided Detection of Tuberculosis. *IEEE Transactions on Medical Imaging* 2016;35:1013-1024. doi:10.1109/TMI.2015.2505672
3. Santosh KC, Antani S. Automated Chest X-Ray Screening: Can Lung Region Symmetry Help Detect Pulmonary Abnormalities? *IEEE transactions on medical imaging* 2018;37:1168-1177. doi:10.1109/TMI.2017.2775636
4. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *European Journal of Radiology* 2020;123. doi:10.1016/j.ejrad.2019.108774
5. Stephen SFY, Hugo JWLA. Applications and limitations of radiomics. *Physics in Medicine and Biology* 2016;61:R150-R166. doi:10.1088/0031-9155/61/13/R150
6. Brady A, Laoide RÓ Fau - McCarthy P, McCarthy P Fau - McDermott R, McDermott R. Discrepancy and error in radiology: concepts, causes and consequences. *The Ulster medical journal* 2012;81:3-9
7. Nachiappan AC, Rahbar K, Shi X, Guy ES, Mortani Barbosa EJ, Shroff GS, et al. Pulmonary Tuberculosis: Role of Radiology in Diagnosis and Management. *RadioGraphics* 2017;37:52-72. doi:10.1148/rg.2017160032
8. Che-Engku-Chik CEN, Yusof NA, Abdullah J, Othman SS, Mat Said MH, Wasoh H. Detection of tuberculosis (TB) using gold standard method, direct sputum smears microscopy, PCR, qPCR and electrochemical DNA sensor: a mini review. *Journal of Biochemistry, Microbiology and Biotechnology* 2016;4:16-21
9. Willis BH, Coomar D, Baragilly M. Clinical scores in primary care. *British Journal of General Practice* 2020;70:1-279. doi:10.3399/bjgp20X709985
10. Jensen SB, Rudolf F, Wejse C. Utility of a clinical scoring system in prioritizing TB investigations - a systematic review. *Expert Review of Anti-infective Therapy* 2019;17:475-488. doi:10.1080/14787210.2019.1625770
11. Wejse C, Gustafson P, Nielsen J, Gomes VF, Aaby P, Andersen PL, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. *Scandinavian Journal of Infectious Diseases* 2008;40:111-120
12. Aunborg JW, Hønge BL, Jespersen S, Rudolf F, Medina C, Correia FG, et al. A clinical score has utility in tuberculosis case-finding among patients with HIV: A feasibility study from Bissau. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2020;92S:S78-S84. doi:10.1016/j.ijid.2020.03.012
13. Setiawati RA-O, Widyoningroem AA-O, Handarini TA-O, Hayati FA-O, Basja AA-O, Putri AA-O, et al. Modified Chest X-Ray Scoring System in Evaluating Severity of COVID-19 Patient in Dr. Soetomo General Hospital Surabaya, Indonesia.

- International Journal of General Medicine* 2021;9:2407-2412.  
doi:10.2147/IJGM.S310577
14. Nishant A, Samruddhi Dhanaji C, Prashant J, Shivakumar I, John D. Brixia Chest X-ray Scoring System in Critically Ill Patients with COVID-19 Pneumonia for Determining Outcomes. In:*Journal of Clinical and Diagnostic Research*; OC15-OC17 doi:10.7860/JCDR/2021/48844.15197
  15. Sofic A, Cizmic M, Beslagic E, Becirevic M, Mujakovic A, Husic-Selimovic A, et al. Brixia Chest X-ray Severity Scoring System is in Relation with C-reactive Protein and D-dimer Values in Patients with COVID-19. *Diagnostics* 2023;13:2122. doi:10.3390/diagnostics13122122
  16. Borghesi A, Zigliani A, Golemi S, Carapella N, Maculotti P, Farina D, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. *International Journal of Infectious Diseases* 2020;96:291-293. doi:10.1016/j.ijid.2020.05.021
  17. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *La radiologia medica* 2020;125:509-513. doi:10.1007/s11547-020-01200-3
  18. Jaeger S, Juarez-Espinosa OH, Candemir S, Poostchi M, Yang F, Kim L, et al. Detecting drug-resistant tuberculosis in chest radiographs. *International journal of computer assisted radiology and surgery* 2018;13:1915-1925. doi:10.1007/s11548-018-1857-9
  19. Shen R, Cheng I, Basu A. A Hybrid Knowledge-Guided Detection Technique for Screening of Infectious Pulmonary Tuberculosis From Chest Radiographs. *IEEE Transactions on Biomedical Engineering* 2010;57:2646-2656. doi:10.1109/TBME.2010.2057509
  20. Melendez J, Ginneken Bv, Maduskar P, Philipsen RHHM, Reither K, Breuninger M, et al. A Novel Multiple-Instance Learning-Based Approach to Computer-Aided Detection of Tuberculosis on Chest X-Rays. *IEEE Transactions on Medical Imaging* 2015;34:179-192. doi:10.1109/TMI.2014.2350539
  21. Du Plessis T, Rae WID, Sathekge MM. Pulmonary tuberculosis diagnosis, differentiation and disease management: A review of radiomics applications. *Polish Journal of Medical Physics and Engineering* 2021;27:251-259. doi:10.2478/pjmpe-2021-0030
  22. Zheng B-H, Liu L-Z, Zhang Z-Z, Shi J-Y, Dong L-Q, Tian L-Y, et al. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC Cancer* 2018;18. doi:10.1186/s12885-018-5024-z
  23. Du Plessis T, Ramkilawon G, Rae WID, Botha T, Martinson NA, Dixon SAP, et al. Introducing a secondary segmentation to construct a radiomics model for pulmonary tuberculosis cavities. *La radiologia medica* 2023;128. doi:10.1007/s11547-023-01681-y
  24. Dixon SAP. Using Deep Learning to Segment Chest X-Rays for the Analysis of Pneumonia. In:*School of Biomedical Engineering*: University of Sydney, 2019

25. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *La radiologia medica* 2021;126:1296-1311. doi:10.1007/s11547-021-01389-x
26. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging* 2020;11:91. doi:10.1186/s13244-020-00887-2
27. Jason WO, Amy O. The power of outliers (and why researchers should ALWAYS check for them). In: *Practical Assessment, Research & Evaluation*; 1-8
28. Urvoy M, Atrousseau F. *Application of Grubbs' test for outliers to the detection of watermarks*. In. *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*, 2014:49-60
29. Hoog AH, Meme Hk Fau - van Deutekom H, van Deutekom H Fau - Mithika AM, Mithika Am Fau - Olunga C, Olunga C Fau - Onyino F, Onyino F Fau - Borgdorff MW, et al. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. *International Journal of Tuberculosis and Lung Disease* 2011;15:1308-1314. doi:10.5588/ijtld.11.0004
30. Harisinghani MG, McLoud TC, Shepard J-AO, Ko JP, Shroff MM, Mueller PR. Tuberculosis from Head to Toe. *RadioGraphics* 2000;20:449-470. doi:10.1148/radiographics.20.2.g00mc12449
31. Zhou Z, Zhou L, Shen K. Dilated conditional GAN for bone suppression in chest radiographs with enforced semantic features. *Medical physics* 2020;47:6207-6215. doi:10.1002/mp.14371



## CHAPTER 6

### Discussions

Pulmonary tuberculosis (PTB) remains one of the leading causes of death globally from a single infection agent. Chest x-rays (CXR) are the most frequently used imaging procedure for diagnosis and management of this disease and is also gaining prominence as a high-volume screening tool. But image reporting continues to be a challenge of this non-neoplastic pulmonary disease, due to a lack of expert readers and the inherent superimposed nature of x-rays. This can be addressed by quantifying digital x-rays with tools such as radiomic feature extraction. In recent years radiomics has gained increasing popularity due to its ability to quantify medical images and for the construction of radiomic signatures, nomograms, machine learning classifiers and models to assist in disease diagnosis, prediction of disease status, response to treatment and disease prognosis.

A systematic review on the existing applications of radiomics to PTB, for the purpose of diagnosis or differentiation from other pulmonary pathology, was performed in Chapter 2 of this thesis. It showed that, at the time of this review in 2021, radiomic feature extraction has only been applied in five studies, all of which used CT or PET/CT images as the input imaging modality. This review highlighted many disparities and uncertainties that remain in the field of radiomic studies applied to PTB. One key finding was the lack of radiomic research using planar images. However, in 2022 additional papers were published on the application of radiomics in PTB by applying standard computer vision techniques and one of these was briefly discussed in Chapter 4.<sup>1</sup> This study applied a unique segmentation using a deep learning approach to train a model to automatically identify the thoracic disease in the lung and to generate bounding boxes around it. In this study radiomic features were used to create heat maps to assist the model in identifying the disease, rather than to quantify disease characteristics.<sup>1</sup> This study achieved an AUC of 0.843 and localization of 0.679 in  $T(\text{IoU}) = 0.1$  in disease detection, but did not mention which radiomic features were utilised.<sup>1</sup> The models developed in this thesis performed much better than the one mentioned above with the standard-deviation model achieving an AUC of 0.9444 (95% CI, 0.8762; 0.9814) and the variance model an AUC of 0.9288 (95% CI, 0.9046; 0.9843). Refer to Table 2 in Chapter 4 for the complete performance indicators.

Even though planar radiomic feature extraction studies remain uncommon for PTB it is gaining some popularity for other diseases such as COVID-19. A few recent studies were published on the use of radiomic feature extraction from CXR for COVID-19 diagnosis and management.<sup>2-3</sup> One study developed a CXR-based radiomics integrated model to assess the predictive role of such models for ICU (intensive care unit) hospitalization and the overall outcome of COVID-19 patients.<sup>3</sup> Their preliminary results demonstrated that a rapid model based on two radiomic features (inverse variance, run length non-uniformity normalized) and a basic inflammatory index collected at admission, can predict ICU hospitalization.<sup>3</sup> In 2022 another study was published that implemented a 2D sliding kernel to map the impulse response of radiomic features throughout the entire CXR image.<sup>2</sup> They investigated three deep neural network architectures that showed an improved sensitivity, specificity, accuracy and ROC AUC results with the inclusion of radiomic analysis in deep learning model designs.<sup>2</sup> Different radiomic features were identified as important in the three different neural networks; GLCOM entropy and GLRLM short-run-emphasis were selected in the VGG-16 model, GLCOM entropy and GLCOM sum-entropy were selected in the VGG-19 model and GLCOM sum-average and GLRLM short-run-high-gray-level-emphasis were selected in the DenseNet-121 model.<sup>2</sup>

Chapter 2 of this thesis is a literature review paper that was published without adding the performance information of the methods used. The table below can be viewed as an addition to Tabel 1 (page 33) where the reviewed studies' results are compared to the results of the radiomic models developed in this thesis (indicated as Study F). In instances where multiple radiomic models were developed in the reviewed studies, only the best performing model or signature was recorded. From Table 1 below it can be seen that the radiomic models developed in this thesis (Study F) outperforms all other models except for the model developed in Study A (Bei et al.).

**Table 1:** Performance indicators of the TB radiomic models or signatures developed in the literature study performed in Chapter 2 and compared to the results of this study (F).

Study	Author	Cohort	AUC	95% CI	Accuracy	Sensitivity	Specificity
A	Bei et al. <sup>4</sup>	Training	0.970	0.939-1.000	93.02%	0.950	0.913
		Validation	0.957	0.899-1.000	93.10%	0.923	0.938
B	Shi et al. <sup>5</sup>	Training	0.93	0.84-1.00	*	*	*
		Validation	*	*	*	*	*
C	Feng et al. <sup>6</sup>	Training	0.933	0.892-0.974	0.878	0.831	0.931
		Validation	0.874	0.813-0.934	0.846	0.868	0.763
D	Cui et al. <sup>7</sup>	Training	0.914	*	*	0.890	0.796
		Validation	0.900	*	*	0.788	0.907
E	Du et al. <sup>8</sup>	Training	0.91	0.84-0.95	0.84	0.74	0.96
		Validation	0.91	0.80-0.96	0.83	0.86	0.78
F	Du Plessis et al. <sup>9</sup>	SD Model	0.944	0.876-0.981	0.833	0.771	0.896
		Variance Model	0.929	0.905-0.974	0.875	0.854	0.896

\* These values were not recorded in the literature

Since no scientific literature could be found on the optimal post-acquisition image processing steps to apply to planar images in radiomic studies, a comparative study was performed on this in Chapter 3. For this study one third of the total dataset was randomly selected and consisted of 103 PA CXR images with sizes ranging between 1479 x 1629 pixels and 3520 x 4280 pixels (mean: 1812 x 2020 pixels). For clarity on the workflow followed in this chapter a comprehensive workflow flowchart is available in Annexure 3. A second flowchart that includes images was also added to Annexure 3 to illustrate the image processing workflow. This chapter showed that the interpolation algorithm and most standard post-acquisition processing applied to planar images does not significantly influence the radiomic signatures obtained. Only cropping images to squares (equal width-to-height pixel ratios) influenced the constructed signatures. It was therefore recommended that the focus in radiomic studies from planar images should rather be on selecting the correct dimensionality reduction methods that are robust against the influence of possible variations in image quality that might be emphasised during image processing.

Chapter 3 also showed that radiomic feature extraction was reproducible when the various image processing and interpolation methods were applied to the same input image. If the CXR however would have been repeated a few minutes apart there would have been some changes to the disease characteristics on the CXR. However, the image protocols used for CXR acquisition in this study do attempt to minimize these changes. The CXR images were all acquired using standard protocols with large SIDs (Source Image Distances = 180cm), PA orientation and with the patient positioned close to the detector. This minimizes the

magnification artefacts and the size distortion. The changes to the intensity matrix would therefore be small and repeatable outcomes should be achieved. Furthermore, to ensure that the models constructed in this thesis are mature enough to deliver reproducible signatures even with repeated CXR, adequately large sample sizes were used.

PTB has diverse radiological presentations. Accurate segmentation of these disease presentations is difficult and time consuming and not always feasible with large data sets. This can result in significant inter- and intra- observer-bias, and variability in segmentation is the biggest cause of irreproducible radiomics outcomes in non-neoplastic diseases. To eliminate the need for disease delineation a novel sliding window segmentation, that allowed the extracted radiomic features to mimic the textural changes across the lung region caused by the disease, was developed in Chapter 4. This segmentation was applied as a secondary segmentation superimposed on a primary automatic lung segmentation mask. Following these segmentations, two radiomics signatures consisting of 10 secondary features were developed. It was shown in a recent study that radiomic feature selection is highly dependent on the feature selection methods used, but that features selected by statistically similar models are similar.<sup>10</sup> The relevance of the features used in radiomic models therefore depends on the model used.<sup>10</sup> In this thesis we have used six different feature selection methods and evaluated four different models, not only on their predictive performance, but also on their selected features. Only second order features were included in the models. These texture features give insight into the spatial arrangement of the pixels and hence the intra-lesion heterogeneity. Entropy, from various feature groups, were included in the signature features, which is an indication of the degree of disorder and uncertainty found in the image matrixes. A possible reason for the exclusion of any first order features are the shared pixels in the sliding windows. Since each window consists of 16 x 16 pixels and the windows step size is 4 pixels, a quarter of each window is shared with its neighbouring windows. The change in the basic first order feature values from one window to the next might therefore not be sufficiently strong to be detected as significant by the models.

After thorough evaluation of the feature selection methods and model algorithms, two well performing radiomic models were constructed that could assist clinicians with the diagnosis of PTB cavities from digital CXR. When added to the clinical signs and symptoms, this might reduce requirements for laboratory results and shorten time to treatment. It can also improve the accuracy of CXR reporting currently regaining prominence as a high-volume

screening tool. This radiomic model that can detect cavities will also be useful in future PTB management studies when serial CXR, with their corresponding models, are studied.

The disadvantage of the developed secondary segmentation is that it significantly increased the dimensionality of the data by more than 3500 folds. This additional dimensionality has not yet been encountered in any previous radiomic study and necessitated thorough investigation of various statistical methods and dimensionality reduction and model construction algorithms. It is fundamental to use the correct algorithms for dimensionality reduction and model construction that are robust against any variations that may occur in the input data. The various statistical methods considered were briefly mentioned in Chapter 4. Table 2 is a summary (calculated as the average of all the CXR feature values) of the actual feature values that was obtained for the standard deviation and the variance consolidation methods. These values were however not used to develop the models, but each CXR's value were considered separately. Table 2 does however give an indication of the dimensionality differences that are present in radiomic features.

**Table 2:** A summary (calculated as the average of all the normal and cavity CXR values) of the signature features for the standard deviation (SD) and variance models.

Signature feature	Standard deviation		Variance	
	Cavity	Normal	Cavity	Normal
glcm_Correlation	0.2464	0.2471	0.0608	0.0612
glcm_DifferenceEntropy	0.2032	0.2015	0.0417	0.0409
gldm_DependenceEntropy	0.6744	0.6679	0.4570	0.4474
gldm_DependenceNonUniformityNormalized	0.0718	0.0706	0.0053	0.0051
gldm_DependenceVariance	1.1150	1.1181	1.2480	1.2552
gldm_LargeDependenceHighGrayLevelEmphasis	109.1003	109.0483	12119.1640	12007.1294
gldm_SmallDependenceLowGrayLevelEmphasis	0.0332	0.0319	0.0011	0.0010
glrlm_RunEntropy	0.6400	0.6489	0.4115	0.4229
glrlm_RunLengthNonUniformityNormalized	0.1157	0.1165	0.0134	0.0136
glrlm_ShortRunLowGrayLevelEmphasis	0.0916	0.0919	0.0084	0.0085
glszm_GrayLevelNonUniformity	1.7593	1.7146	3.1369	2.9771
glszm_ZoneEntropy	0.6618	0.6402	0.4403	0.4112
ngtdm_Busyness	20.6448	26.6745	1735.3986	2128.3086
ngtdm_Contrast	0.0345	0.0326	0.0015	0.0012

Studies on the use of computer science from medical images for the diagnosis of PTB or on the differentiation of TB from another pulmonary disease have been reported on, but little evidence exists on the use of radiological image analysis for monitoring treatment response and prognosis. To evaluate the change in the radiomic features over time, using serial images of the same patient, a unique approach to radiomic signature construction was taken in

Chapter 5. The radiomic features extracted from each sliding window, when the secondary segmentation was applied to the CXR of each PTB patient, were re-distributed into the x- and y-coordinates in a two-dimensional matrix, corresponding to the sliding window coordinates, to create feature parameter maps. Feature parameter maps showing a strong correlation to the lung pathology was summed to construct a signature parameter map for each CXR. The signature parameter maps were an accurate visual resemblance to the radiological images when some image formatting methods were applied. Since the signatures in this chapter was developed by visual assessment, it is expected that only first order features are included in the signature. Texture features derived from secondary matrixes are too far removed from the original grayscale intensities of the CXR and would not produce meaningful visual results. Early radiomic research was dominated by the analysis of semantic, radiologist-defined features that carried qualitative real-world meaning.<sup>11</sup> But lately, with the rapid development of machine learning approaches, the emphasis is placed on high-throughput agnostic analyses.<sup>11</sup> This took the focus toward an increase in predictive power and further away from a biological and mathematical understanding of the findings.<sup>11</sup> The signatures developed in this chapter aimed to re-connect the radiomic features to the image characteristics and disease status as was done in initial radiomic studies. After additional processing, feature parameter maps were developed that might be valuable as a quantitative supplementary indicator in the management of PTB disease and further increase the acceptance of CXR as a tool for assessing the TB response in medical research and clinical practice.

In Chapter 5 further dimensionality reduction methods were applied to consolidate the signature parameter maps to radiomics scores. A radiology- and clinical score for each corresponding patient was also developed using existing scoring algorithms. When the 3 scores were compared a strong correlation was found between the radiological and the radiomics score that was statistically significant. This showed that the developed radiomics score was able to quantify the change in the disease characteristics as seen from digital CXR of patients diagnosed with PTB. Radiomic analysis of serial x-rays taken while patients receive TB therapy has the potential be a quantitative monitoring tool of response to therapy.

In this study three different outcomes were therefore successfully achieved when radiomic feature extraction was used to quantify PTB characteristics from chest x-rays; A radiomics model was constructed that achieved good classification accuracy for normal chest x-rays and cavities on x-rays of people suffering from PTB, signature parameter maps were developed that quantitatively represents the lung pathology, and a radiomics score was built that could

be used as a quantitative monitoring tool for response to therapy. These outcomes could increase the accuracy of CXR reporting, which is currently regaining prominence as a high-volume screening tool, by alleviating the subjectivity introduced by human judgement on x-rays. Accompanied with clinical signs and symptoms, it might also aid diagnosis and commencement of TB treatment. Quantitative CXR interpretation can further increase the acceptance of CXR as a tool for assessing the TB response in medical research and clinical practice.

The successful quantification of PTB characteristics from chest x-rays using radiomics, as achieved in this study, has great potential to improve diagnosis and disease management and to being a significant role player in personalized precision medicine.

## References

1. Han Y, Chen C, Tang L, Lin M, Jaiswal A, Ding Y, et al. Using radiomics as prior knowledge for abnormality classification and localization in chest x-rays. 2020.
2. Hu Z, Yang Z, Lafata KJ, Yin F-F, Wang C. A radiomics-boosted deep-learning model for covid-19 and non-covid-19 pneumonia classification using chest x-ray images. *Medical physics*. 2022; 49(5):3213-22. doi:10.1002/mp.15582
3. Giraud C, Frattin G, Fichera G, Motta R, Stramare R. A practical integrated radiomics model predicting intensive care hospitalization in covid-19. *Critical care (London, England)*. 2021; 25(1):145. doi:10.1186/s13054-021-03564-y
4. Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. *BMC Medical Imaging [Internet]*. 2019; 19(1):1-11. [cited Access Year Access Date]]. Available from: URL
5. Shi W, Zhou L, Peng X, Ren H, Wang Q, Shan F, et al. Hiv-infected patients with opportunistic pulmonary infections misdiagnosed as lung cancers: The clinicoradiologic features and initial application of ct radiomics. *Journal of thoracic disease*. 2019; 11(6):2274-86. doi:10.21037/jtd.2019.06.22
6. Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *European Journal of Radiology*. 2020; 128 doi:10.1016/j.ejrad.2020.109022
7. Cui EN, Yu T, Shang SJ, Wang XY, Jin YL, Dong Y, et al. Radiomics model for distinguishing tuberculosis and lung cancer on computed tomography scans. *World journal of clinical cases*. 2020; 8(21):5203-12. doi:10.12998/wjcc.v8.i21.5203
8. Du D, Gu J, Chen X, Lv W, Feng Q, Rahmim A, et al. Integration of pet/ct radiomics and semantic features for differentiation between active pulmonary tuberculosis and lung cancer. *Molecular Imaging & Biology*. 2021; 23(2) doi:10.1007/s11307-020-01550-4
9. Du Plessis T, Ramkilawon G, Rae WID, Botha T, Martinson NA, Dixon SAP, et al. Introducing a secondary segmentation to construct a radiomics model for pulmonary tuberculosis cavities. *La radiologia medica*. 2023; 128 doi:10.1007/s11547-023-01681-y
10. Demircioğlu A. Evaluation of the dependence of radiomic features on the machine learning model. *Insights into Imaging*. 2022; 13(1):28. doi:10.1186/s13244-022-01170-2
11. Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology*. 2021; 298(3):505-16. doi:10.1148/radiol.2021202553



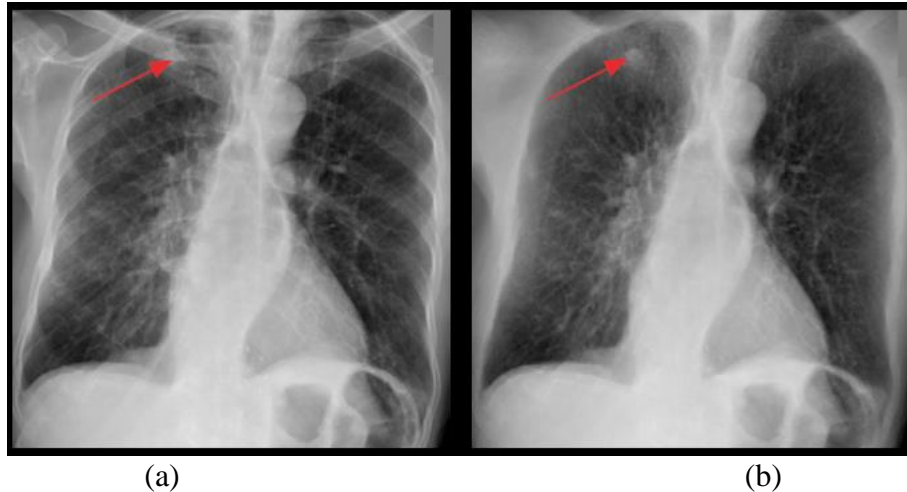
## CHAPTER 7

### Future Perspectives

The most pressing challenges and limitations of radiomic studies applied to planar images have been discussed in this thesis and include the inherently superimposed nature of planar images, small datasets, limited computational resources and a lack of standardization. All four of these limitations and challenges should further be addressed in future studies.

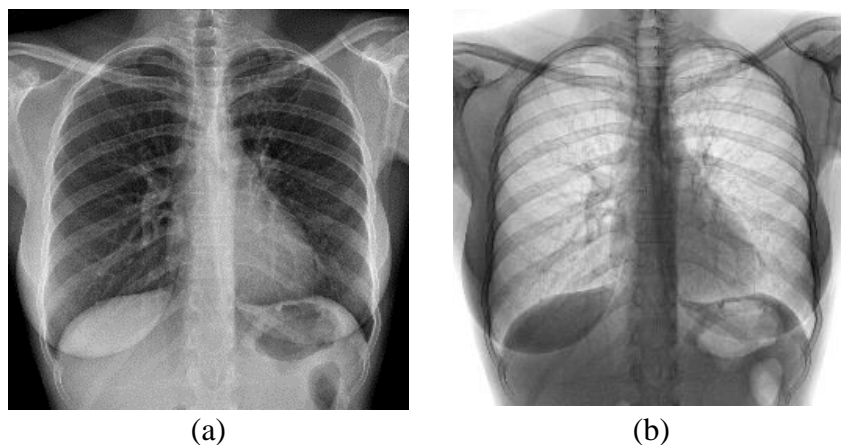
This study did attempt to address the limitations caused by the inherent superposed nature of two-dimensional (2D) images by applying bone suppression software to the chest x-rays (CXR) before extracting radiomic features. Since CXR are a two-dimensional superimposed representations of the three-dimensional anatomy in the chest, the rib and clavicle bones are overlaying the lung tissue of interest. These denser structures in the ROI will cause the quantified grey-level intensities to represent the lung pathology less accurately. By removing these denser structures from the superimposed CXR, the PTB expressions in the lung will be more obvious.

There are two approaches to bone suppression software applications. The retrospective approach is to apply image processing algorithms that can train a model to remove the high-density anatomical structures. Or prospectively a second x-ray with a different energy can be acquired immediately after the first and Dual-Energy subtraction can be applied.<sup>1</sup> Some commercial bone suppression software packages are readily available and clinically in use, such as the Phillips- and the Carestream Bone Suppression Software packages (see Figure 1). Both have prospective approaches with the software packages integrated in the x-ray acquisition workflow and applied to the images before they are sent to the PACS.<sup>1</sup> These would therefore not be relevant for this retrospective study.



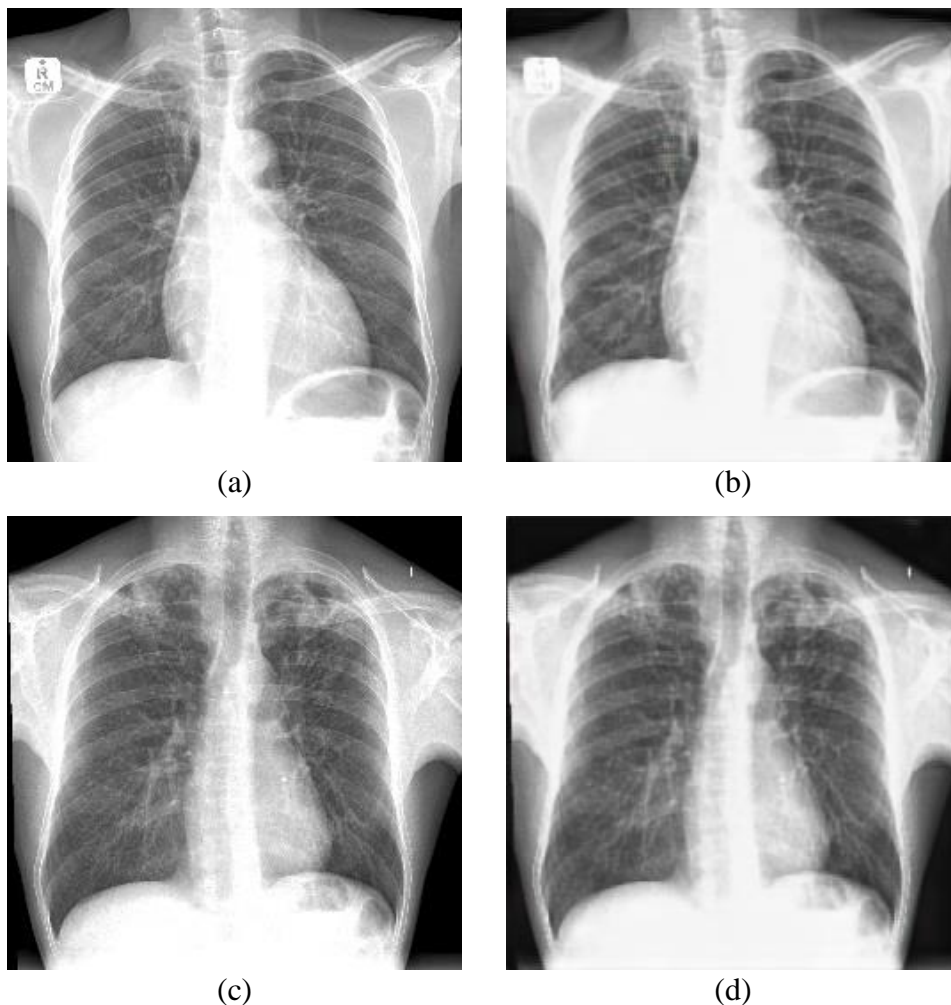
**Figure 1:** (a) A clinical example of an original unprocessed CXR (b) The same CXR with the Phillips Bone Suppression software applied that was able to successfully suppress all high density anatomy<sup>1</sup>

Some retrospective bone suppression studies on x-rays have already been done, with different levels of accuracy, using various deep learning models.<sup>2</sup> One study used Dilated Conditional GAN (Generative Adversarial Network) for bone suppression and proved that this model “outperforms current state-of-the-art bone suppression methods using X-ray images”.<sup>2</sup> With the developer’s permission, the Dilated Conditional GAN bone suppression model was obtained and applied to this thesis’ dataset. Some image processing was required to get this dataset in the same format as the training dataset of the bone suppression model. It included: Image format conversion from DICOM to PNG, interpolation to match the training images’ size, and image inversion to match the training set intensities (e.g., the maximum and minimum pixel values were swapped around). Figure 2 is an example of one original CXR with its corresponding processed and inverted image.



**Figure 2:** (a) An original CXR. (b) The same CXR, processed and inverted to be in the same format as the training set of the Dilated Conditional GAN bone suppression model

After image processing the bone suppression model was applied with the hope of removing the ribs and clavicles to evaluate the influence it has on this study's outcomes. Figure 3 are examples of two patients' outcomes. Unfortunately, the bone suppression model gave very poor results. The bones were not suppressed, but instead the image quality was compromised. This is most likely due to the immaturity of the model, as it was only trained using a single publicly available dataset of 272 images.<sup>2</sup> Another explanation might be the limitation of deep neural networks that losses performance with a new 'domain', i.e. if the data distribution, or the dynamic range, on which the model was trained and/or tested differs significantly from the one it is applied to.



**Figure 3:** (a) Original input image of Patient A, (b) Output image of Patient A after the Dilated Conditional GAN bone suppression model was applied, (c) Original input image of Patient B and (d) Output image of Patient B after the Dilated Conditional GAN bone suppression model was applied

Since this immature model had already been compared to, and outperformed, all other known models at the time,<sup>2</sup> no other model was considered for this study. The results presented in this thesis will most likely improve significantly when a mature bone suppression model can

be developed and applied to the CXR dataset before radiomic feature extraction. The radiomic features will then not be influenced by the high-density bony anatomy overlying the lung and TB disease expressions.

Possibly the biggest challenge in radiomic studies remains the limited number of medical images available for feature extraction. The number of possible features available to extract are often more than the number of images. Even more so, in this retrospective study where the dimensionality of each x-ray's features was increased by more than 3500 folds. This sample size limitation is not only apparent in radiomic feature extraction, but in all computer science studies involving medical image datasets. The rule of thumb for data sample size in computer vision for machine learning is 1000 samples per class and in deep learning 4000 samples per class.<sup>3</sup> It is unlikely to ever obtain this quantity of samples in medical image studies. However, by re-training models with an unseen dataset, regardless of its size, the accuracy and reliability of the models will always improve. An increased sample size will naturally follow when unseen datasets are used for model construction in all future studies that follow from this one.

To extract the radiomic features and to construct the radiomic signatures and model in this study was a labour intensive process that involved plenty of hard coding and a thorough understanding of radiomics and image processing. More work is needed to make the developed tools available to any medical professional who does not have pre-existing knowledge in the field of computer science or radiomics. This can be achieved by developing a user friendly interface where an CXR image can easily be uploaded and incorporated to the segmentation model, feature extraction library and model algorithms used and developed in this study. If this is developed as an application, rather than server based software, this should be accessible from any computer, regardless of its processing power. This will eliminate the restriction of computational power which might be a limiting factor in many high TB burdened countries. The radiomics- and radiological score developed in Chapter 5 can also be incorporated into this application to assist clinicians with the quantitative interpretation of CXR in follow-up appointments. Once this tool is developed and made readily available to assist in the management of TB, it can allow the treatment of PTB across the world to be more patient specific.

The rapid growth in the radiomics field has contributed to a lack of standardization of both its scientific integrity and its clinical relevance.<sup>4</sup> This restriction was attended to as best possible

in this study, but it will be beneficial to pay further attention to it in follow-up studies when the plea for standardization was addressed. A published suggestion to address the issue is for all researchers to contribute to a Picture archiving and radiomics knowledge system (PARKS) where quantitative image features are shared, compared and co-analyzed.<sup>4</sup> This will assist with the improvement of standardization of image acquisition, analysis and reporting in radiomics. An intrinsic limitation of radiomics is that the feature algorithms summarise information over a region of the image and lose context and detail in many cases. This might be an area that needs to be addressed in all future radiomic studies.

In this thesis it was seen that different radiomic features were included in the signatures developed in chapter 4 and 5 respectively. Chapter 2 also showed that many different features are included in signatures and nomograms developed in different studies even when the same image modality and anatomical ROI are considered. A future perspective for all radiomic studies would be to explore the reasons to why the features included in the final signatures work? What do they mean physically? And why some texture features are superior to others?

Chapter 4 briefly discusses the various statistical algorithms and machine learning models tested and considered for dimensionality reductions and model construction purposes in this study. A final future perspective will be to thoroughly describe, formalize and record all the statistical work completed in this study.

## References

1. Freedman MT, Lo S-CB, Seibel JC, Bromley CM. Lung nodules: Improved detection with software that suppresses the rib and clavicle on chest radiographs. *Radiology*. 2011; 260(1):265-73. doi:10.1148/radiol.11100153
2. Zhou Z, Zhou L, Shen K. Dilated conditional gan for bone suppression in chest radiographs with enforced semantic features. *Medical physics*. 2020; 47(12):6207-15. doi:10.1002/mp.14371
3. Cho J, Lee K, Shin E, Choy G, Do S. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? 2015.
4. Lambin P, Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, et al. Radiomics: The bridge between medical imaging and personalized medicine. 2017.

## CHAPTER 8

### Conclusion

Pulmonary tuberculosis (PTB) is a worldwide epidemic that can only be fought effectively with early and accurate diagnosis and proper disease management. The means of diagnosis and disease management should be easily accessible, cost effective and readily available in the high tuberculosis (TB) burdened countries where it is most needed. Currently digital chest x-ray imaging (CXR) remains the most widely accessible imaging modality in high TB burdened countries, and is regaining prominence as a high-volume screening tool, hence it was used in this thesis to quantify PTB characteristics using radiomics.

The robust radiomic signatures, models and scores that were developed in this thesis could be used as a valid and readily available biomarker, when the future perspectives are completed. This might not only have an enormous impact on the clinical care provided to patients suffering from PTB but will also greatly assist towards the World Health Organization's "The End TB Strategy". This study fits in with the one pillar identified in this strategy to intensify research and innovation for the rapid development of new high-impact tools that can dramatically change TB prevention and care. The radiomic signatures, models and scores, together with the sliding window segmentation developed in this thesis, will also address a few other challenges in TB diagnosis and monitoring such as a shortage of expert readers and human subjectivity. By statistical quantification of imaging features subjectivity can be eliminated and reproducibility can be achieved. Quantitative analysis might also reveal underlying biological characteristics of TB that are not visible to the human eye. Additionally, these models and scores could assist clinical trials of novel agents being evaluated for efficacy against TB by identifying early in the process those agents or regimens that have the greatest effect against PTB and will result in fewer treatment failures.

Radiomics is a high-impact tool that was successfully utilized in this thesis to quantify the characteristic of PTB using planar chest x-rays, and that has the potential to be used as an imaging biomarker in the near future.

# ANNEXURES





29 June 2020

Prof M Sathekge  
Department of Medical Nuclear Sciences  
Faculty of Health Sciences

Dear Prof M Sathekge

**STUDENT: DU PLESSIS T (PhD MEDICAL NUCLEAR SCIENCES)**

**TITLE: Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics**

The above-mentioned student's protocol has been approved by the PhD committee.

We wish the student all the best with her studies.

Kind regards



PROF V STEENKAMP  
CHAIR: PhD COMMITTEE

**Institution:** The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

13 August 2020

**Approval Certificate  
New Application**

**Ethics Reference No.: 473/2020**

**Title: Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics**

**Dear Mrs T du Plessis**

The **New Application** as supported by documents received between 2020-07-09 and 2020-08-12 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2020-08-12 as resolved by its quorate meeting.

Please note the following about your ethics approval:

- Ethics Approval is valid for 1 year and needs to be renewed annually by 2021-08-13.
- Please remember to use your protocol number (473/2020 ) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

**Ethics approval is subject to the following:**

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

**Yours sincerely**



**Dr R Sommers**

MBChB MMed (Int) MPharmMed PhD

**Deputy Chairperson** of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)



**Institution:** The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 22 May 2002 and Expires 03/20/2022.
- IORG #: IORG0001762 OMB No. 0990-0279 Approved for use through February 28, 2022 and Expires: 03/04/2023.

## Faculty of Health Sciences Research Ethics Committee

14 October 2021

### Approval Certificate Annual Renewal

Dear Mrs T du Plessis

**Ethics Reference No.: 473/2020**

**Title: Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics**

The **Annual Renewal** as supported by documents received between 2021-09-22 and 2021-10-13 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2021-10-13 as resolved by its quorate meeting.

Please note the following about your ethics approval:

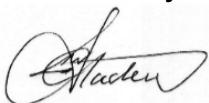
- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2022-10-14.
- Please remember to use your protocol number (473/2020 ) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

**Ethics approval is subject to the following:**

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely



**On behalf of the FHS REC, Professor Werdie (CW) Van Staden**

MBChB, MMed(Psych), MD, FCPsych(SA), FTCL, UPLM

**Chairperson: Faculty of Health Sciences Research Ethics Committee**

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)



**Institution:** The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

- FWA 00002567, Approved dd 18 March 2022 and Expires 18 March 2027.
- IORG #: IORG0001762 OMB No. 0990-0278 Approved for use through August 31, 2023.

## Faculty of Health Sciences Research Ethics Committee

11 November 2022

### Approval Certificate Annual Renewal

Dear Mrs T du Plessis,

**Ethics Reference No.: 473/2020 – Line 2**

**Title: Quantification of pulmonary tuberculosis characteristics from digital chest x-rays using radiomics**

The **Annual Renewal** as supported by documents received between 2022-10-26 and 2022-11-09 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2022-11-09 as resolved by its quorate meeting.

Please note the following about your ethics approval:

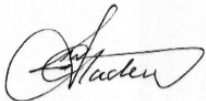
- Renewal of ethics approval is valid for 1 year, subsequent annual renewal will become due on 2023-11-11.
- Please remember to use your protocol number (473/2020) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

#### **Ethics approval is subject to the following:**

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

We wish you the best with your research.

Yours sincerely



**On behalf of the FHS REC, Professor Werdie (CW) Van Staden**

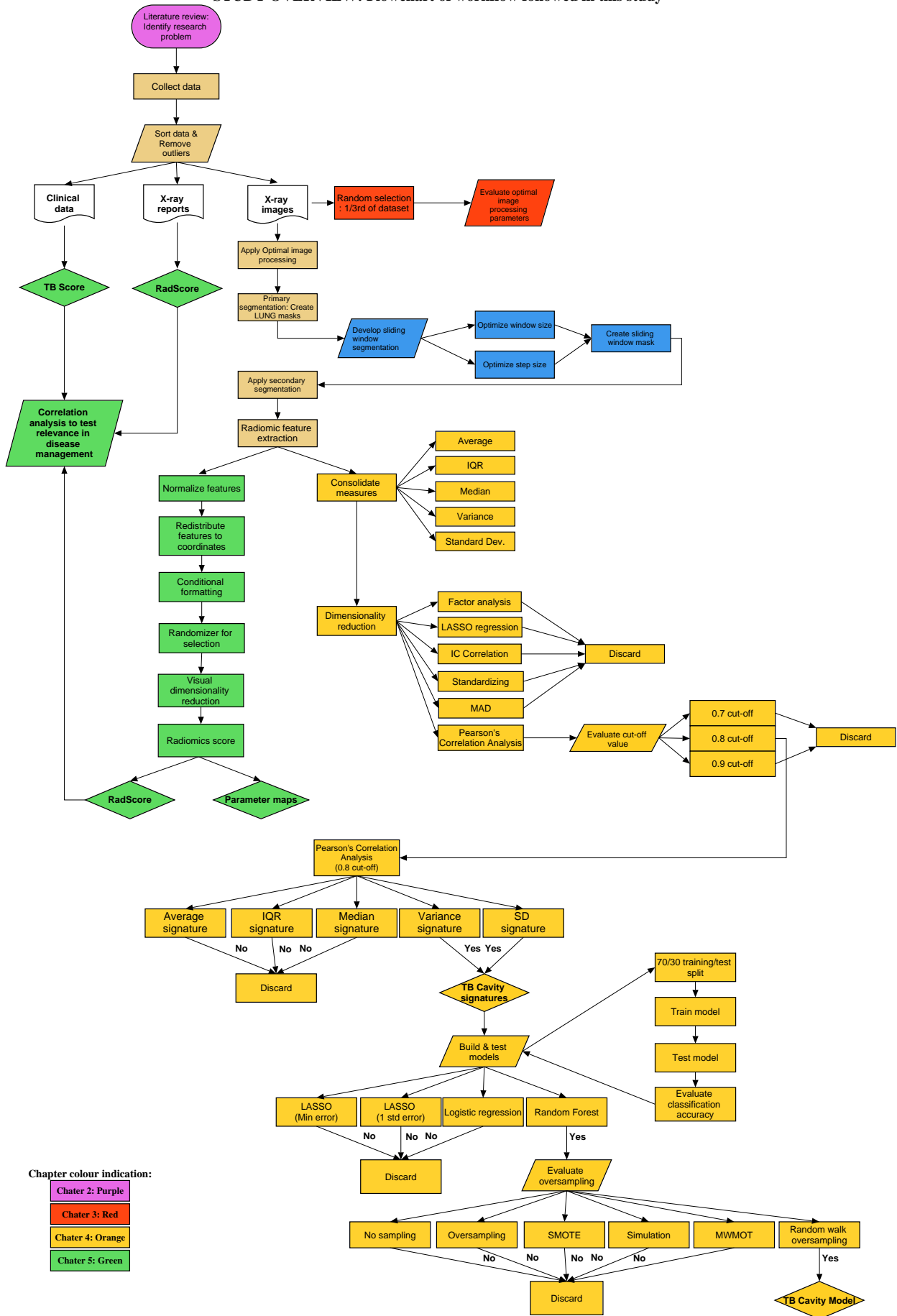
MBChB, MMed(Psych), MD, FCPsych(SA), FTCL, UPLM

**Chairperson: Faculty of Health Sciences Research Ethics Committee**

*The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2015 (Department of Health)*

# ANNEXURE 3

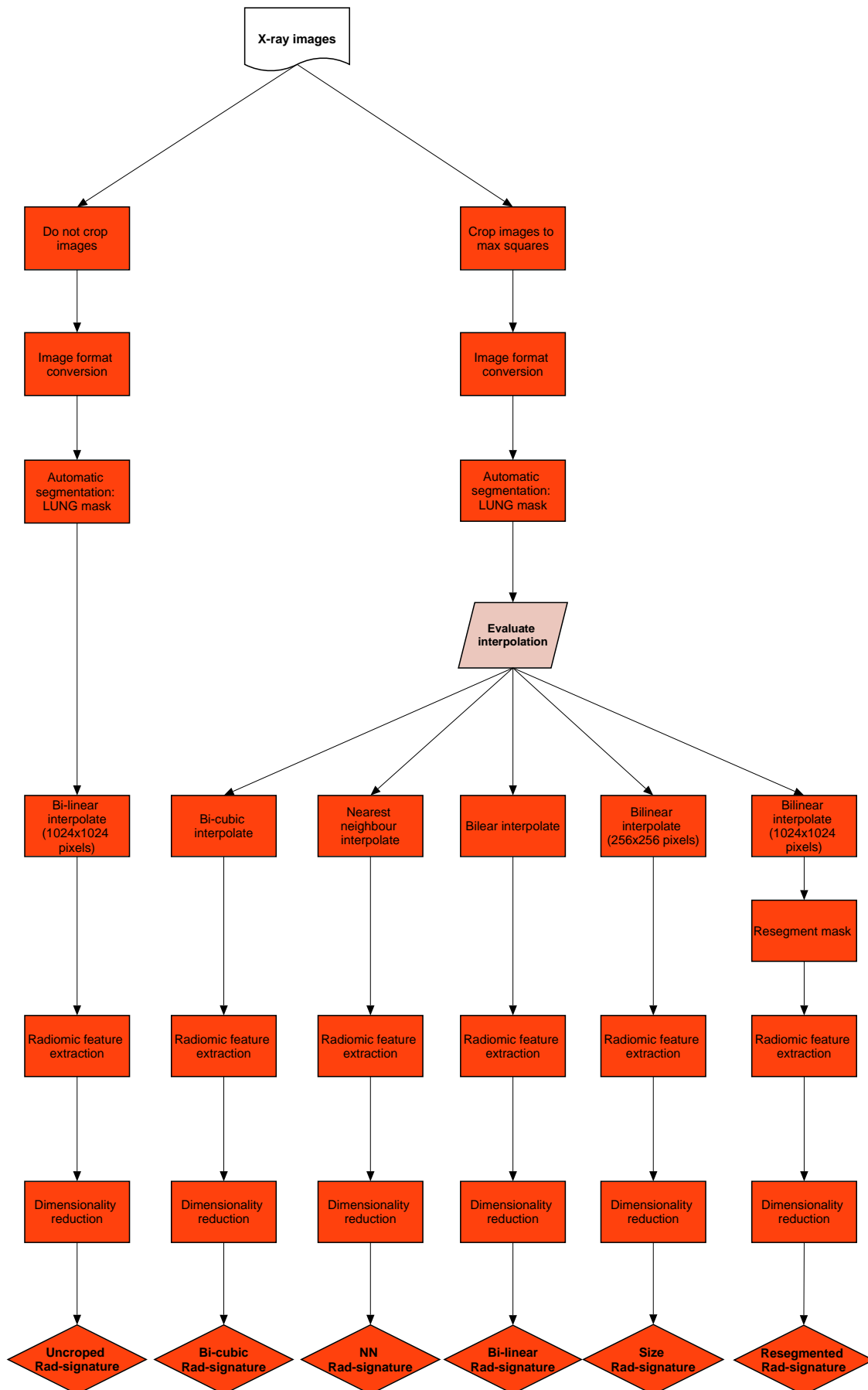
STUDY OVERVIEW: Flowchart of workflow followed in this study



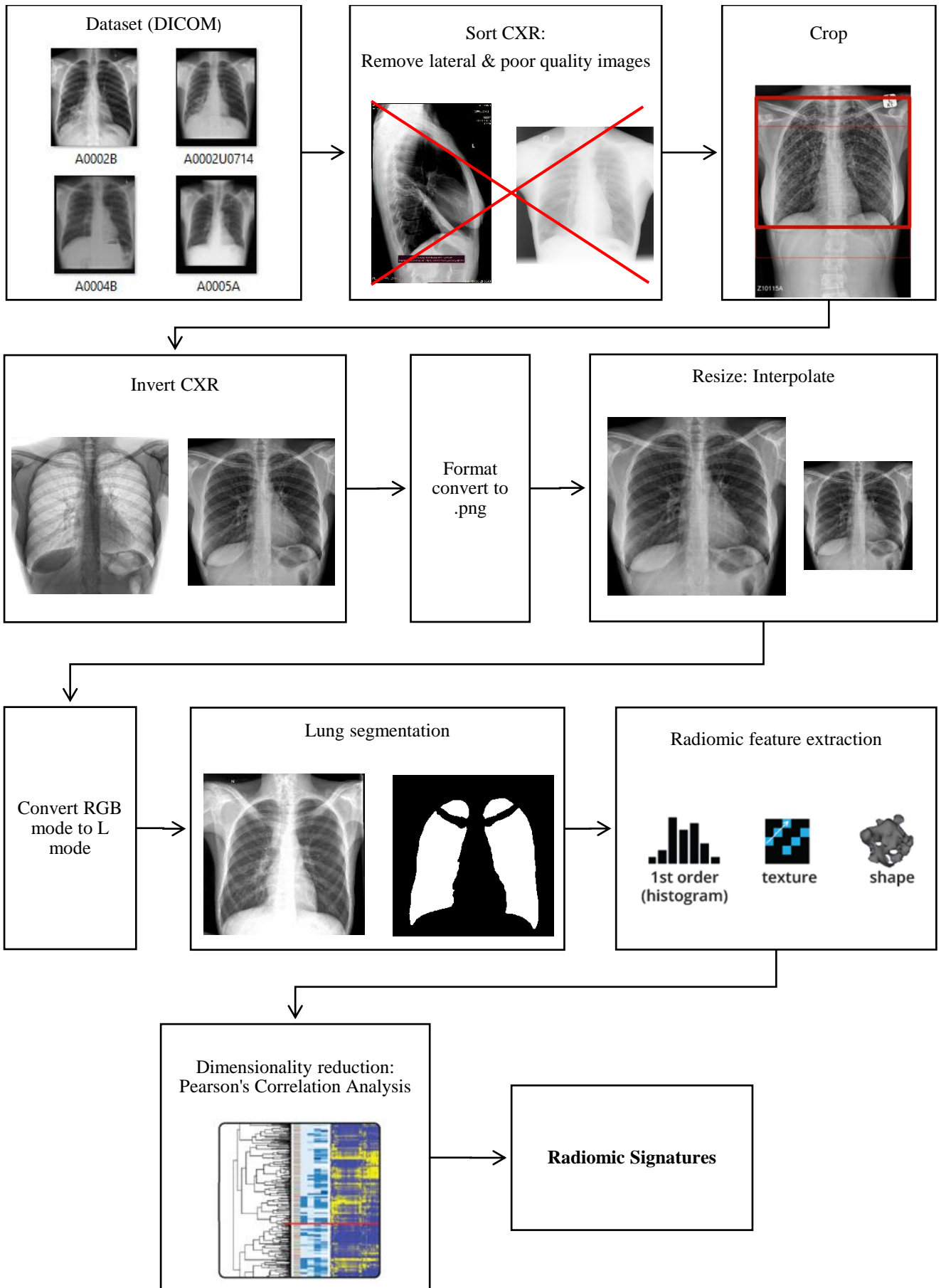
Chapter colour indication:

- Chapter 2: Purple
- Chapter 3: Red
- Chapter 4: Orange
- Chapter 5: Green

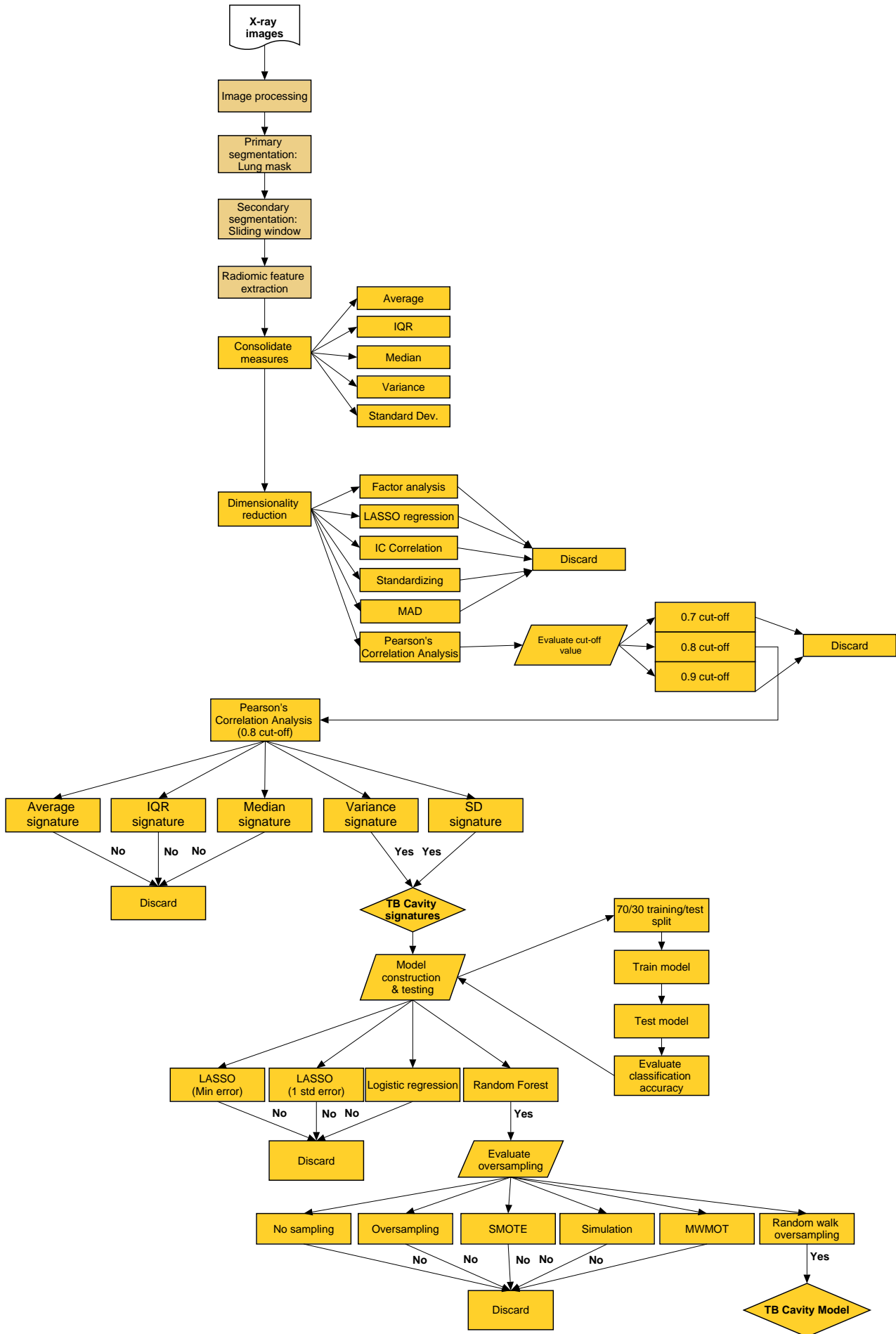
### CHAPTER 3: Flowchart of workflow



### CHAPTER 3: Flowchart of image processing

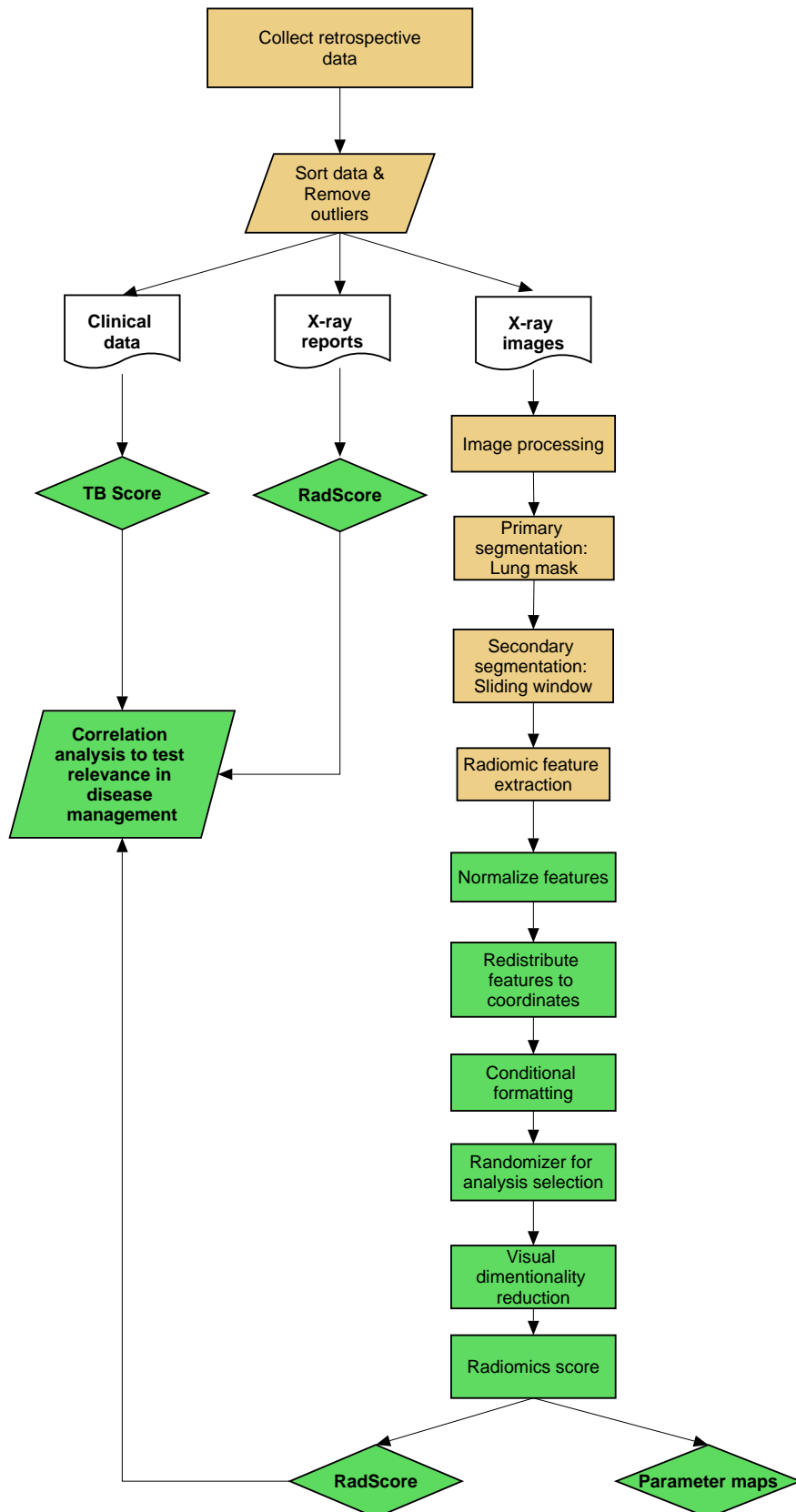


CHAPTER 4: Flowchart of workflow





CHAPTER 5: Flowchart of workflow



- End -