A Robust Technique for the Detection and Quantification of Abdominal Aortic Calcification Using Dual Energy X-Ray Absorptiometry

Karima Mohamed Ali Elmasri

Cardiff University

School of Engineering

This thesis is submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

May 2018

Declaration and Statements

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed..... (candidate) Date .....

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ……(insert MCh, MD, MPhil, PhD etc, as appropriate).

Signed ..... (candidate) Date .....

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ..... (candidate) Date .....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate) Date .....

Abstract

Arterial calcification is a manifestation of atherosclerosis, which over the last two decades has become a recognised predictor of cardiovascular disease. Abdominal Aortic Calcification (AAC) and osteoporosis have been shown to coincide in older individuals. The accepted method of diagnosing osteoporosis is through the measurement of bone mineral density by dual energy x-ray absorptiometry (DXA). Vertebral fracture assessment (VFA) images obtained alongside BMD using DXA technology provide an inexpensive resource for AAC diagnosis.

Although several simple methods have been proposed for manual semi-quantitative scoring of AAC in x-ray images in the past, these methods have limitations in terms of capturing small changes in atherosclerosis progression and are time-consuming. Several automatic approaches have been proposed to measure AAC on radiographs. However, these methods have not been related to any accepted medical AAC scoring systems and thus are not likely to be adopted easily by the medical community. In addition, there has been no attempt to apply the proposed methods to VFA images.

The main focus of the research presented in this thesis is the automatic quantification of AAC in VFA images acquired in single energy mode. The thesis is divided into two main parts. In the first part, an automatic method for AAC detection and quantification in VFA images is proposed and evaluated on a large number of images. In the second part, the performance of both single and dual energy VFA imaging for the detection of uniformly distributed calcification is investigated.

The automatic method for AAC detection consists of two stages. In the first stage an active appearance model was employed for the purpose of segmentaion. In the second stage, adaptive thresholding techniques were used to detect AAC, whilst automatic

classification techniques were used to quantify the detected calcification. The performance of several classifiers were investigated, and the proposed method was evaluated against the manual AC-24 scoring method using several hundred images and two human readers. A thorough statistical analysis of the results showed that, overall, the SVM classifier gave the best results. Weighted accuracy, sensitivity, specificity assessed for 4 AAC categories were 89.2%, 78.5% and 92.3% respectively while the corresponding values for 3 AAC categories were 88.6%, 86%, 90.4%.

In the second part, a study using a tissue-mimicking physical phantom is described. The phantom consists of an aluminium strip within Perspex to simulate calcification and abdominal soft tissue respectively.

VFA images of different phantom configurations were acquired in single energy (SE) and dual energy (DE) modes. The minimum detectable aluminium thickness was assessed visually and related to contrast and contrast-to-noise ratio. Percentage coefficient of variation was used to quantify uniformity, repeatability and reproducibility with a Perspex width of 25 cm, the smallest thickness of aluminium that could be detected was 0.20-0.25 mm.

The initial results are promising, and the system proposed in this research can be used as an alternative method to the manual scoring system (AC-24) for a wide range of AAC. The principal conclusion from the phantom work is that under idealised imaging conditions, VFA images have the potential to be used for detecting small thicknesses of calcification with good linearity, repeatability and reproducibility in SE and DE modes for patients with a body width < 30 cm.

Acknowledgements

First, I am grateful to Allah (My Lord) the all high, the all great who made it possible for me to finish this work.

Second, I would like to thank my main supervisors, Doctor Yulia Hicks and Professor William Evans, not only for their constant assistance, excellent guidance and thoughtful advice, but also for their constant support and encouragement. The high standard of their research has always been an inspiration to me. I am also grateful to Doctor XinYang and Doctor Xianfang Sun for their advice.

I have been extremely fortunate to have supervisors who cared so much about my work, and who responded to my questions and queries so promptly.

I am grateful to the staff of the Department of Medical Physics at the University Hospital of Wales, especially Mrs Rebecca Pettit, for giving me the opportunity to carry out the practical aspects of this project.

Finally, I would like to express my sincere gratitude and appreciation to family: my sisters, Souad and Nadia, and my brothers, Amer, Mustafa, Doctor Aiad and Jamal, for their endless love, support and prayers. Thanks also to my beloved nephew Mohammed.

Karima Elmasri

May 2018

Table of contents

Declaration and Statements i

Abstract ii

Acknowledgements iv

List of Figures viii

List of Tables xi

List of Publication xiii

Abbreviations and Symbols xiv

Chapter: 1 Introduction 1

1.1 Motivation 1

1.2 Aims and Objectives 4

1.3 Thesis Outline 5

Chapter: 2 Abdominal Aortic Calcification 7

2.1 Introduction 7

2.2 Atherosclerosis 7

2.3 Abdominal Aortic Calcification (AAC) 10

2.4 Characteristics of Calcification 13

2.4.1 Calcification Chemical Composition 14

2.4.2 Calcification Thickness 14

2.5 Detection of Abdominal Aortic Calcification 15

2.5.1 Computed Tomography (CT) 16

2.5.2 Ultrasound 16

2.5.3 Radiography 17

2.5.4 Dual Energy X-ray Absorptiometry 18

2.6 Quantification of Abdominal Aortic Calcification 19

2.7 Summary 22

Chapter: 3 Medical Image Processing and Machine Learning 25

3.1 Introduction 25

3.2 Medical Image Processing 25

3.2.1 Image Enhancement Techniques 26

3.2.2 Image Segmentation 28

3.3 Statistical Machine Learning 37

3.3.1 Classification 38
3.3.2 Accuracy of Classification 40
3.4 Automatic Detection and Quantification of AAC 41
3.5 Summary 44
Chapter: 4 Dual Energy-X-ray Absorptiometry 45
4.1 Introduction 45
4.2 Clinical Application of Dual Energy-X-ray Absorptiometry 45
4.3 Physics of Absorptiometry 47
4.4 Dual Energy X-ray Absorptiometry Technology 53
4.4.1 Production of Dual Energy X-ray beam 54
4.4.2 DXA Detectors and Collimators 55
4.5 Hologic DXA 57
4.6 Quality Control 57
4.7 Vertebral Fracture Assessment (VFA) 59
4.8 Summary 63
Chapter: 5 Automatic Detection and Quantification of AAC 64
5.1 Introduction 64
5.2 Data set 65
5.3 AC-24 Manual Scoring System 68
5.4 AAC Categorisation using the manual AC-24 69
5.5 Segmentation of Lumbar Spine and Aorta 69
5.5.1 Statistical Shape Model 70
5.5.2 Applying statistical shape model to VFA images 72
5.5.3 Statistical model of texture 75
5.5.4 Combined Appearance Model 76
5.6 Active appearance model (AAM) searching algorithm 78
5.7 Extracting the aorta from the segmented image 78
5.8 Calcification Detection 80
5.9 Quantification of AAC 82
5.9.1 Feature Extraction 82
5.10 Automatic Calcification Classification 86
5.10.1 Feature selection 87
5.11 Comparison of Automatic and Manual Classification 88
5.12 Summary 90
Chapter: 6 Evaluation of Automatic Method for AAC Detection and Quantification 91
6.1 Introduction 91
6.2 Study Population 91
6.3 Variability of AAC Measurement in VFA Images 94
6.3.1 Intra-observer variability of calcification scores 96

6.3.2 Inter-observer Variability of Calcification Scores 97
6.4 Relationship between Patient Age and AAC 98
6.5 Evaluation of the Automatic method for AAC detection and Quantification 99
6.5.1 Manual Categorisation of AAC Severity 99
6.5.2 Classification Accuracy and Feature Selection 100
6.6 Summary 108
Chapter: 7 Phantom Design and construction 109
7.1 Introduction 109
7.2 Modelling of Abdomen Soft Tissue 110
7.2.1 Phantom Material 111
7.2.2 Phantom Dimensions 113
7.3 Modelling of Calcification 122
7.3.1 Phantom Material 122
7.3.2 Aluminium Strips Thickness Measurements 123
7.4 Summary 125
Chapter: 8 Performance evaluation of DXA for AAC Detection 126
8.1 Introduction 126
8.2 VFA Image Acquisition Protocol 127
8.3 Preparation of Images for Analysis 128
8.4 Uniformity and Noise 134
8.4.1 Method 134
8.4.2 Results 134
8.5 Detectability assessment 145
8.5.1 Contrast and Contrast to Noise Ratio (CNR) 145
8.5.2 Visual Assessment 151
8.6 Repeatability and Reproducibility of C and CNR 155
8.6.1 Results 156
8.7 Conclusion 161
8.8 Summary 162
Chapter: 9 Conclusions and Future work 164
9.1 Contributions 165
9.2 Conclusions 165
9.3 Limitations 166
9.4 Future Work 167

List of Figures

Figure 2-1: Diagram showing the effects of atherosclerosis on an artery. 10
Figure 2-2: Abdominal aorta bifurcation. (Openstax, Circulatory Pathways, 2015, https://anatomychartee.co/show/branch-of-abdominal-aorta.html 12
Figure 4-1: Lumbar spine DXA image showing analysis regions and an example of BMD, BMC, and BA results summary. 52
Figure 4-2: DXA scanner components AP scan. (https://ufhealthjax.org/womens-imaging/images/dxa-procedure.jpg) 53
Figure 4-3: Hologic Discovery DXA scanner in position for a lateral spine scan. 54
Figure 4-4: Quality control plot for L1 to L4 BMD of the Hologic spine phantom between 2012 and 2015 59
Figure 4-5: Hologic Discovery DXA system in lateral scan. (http://www.hologic.com/products/imaging/skeletal-health/discovery-dxa-system 62
Figure 4-6: Lateral VFA image acquired by Hologic Discovery DXA, (a) in single energy mode, b) in dual energy mode. 62
Figure 5-1: Examples of the data retrieved from the Medical Physics archive at the University Hospital of Wales in Cardiff, UK, VFA images enhanced for contrast (A) 91- year-old female, (B) an 83-year-old female, (C) 79- year-old female 67
Figure 5-2: Cropped and labelled VFA images of female patient with, (A) severe AAC, (B) moderate AAC, (C) mild AAC 67
Figure 5-3: Labelled image with AC-24 score points. 68
Figure 5-4: Example of an annotated VFA image with 56 landmarks (shown as black dots): (A) original image; (B) cropped image with landmarks. 73
Figure 5-5: (A) First three mode of shape variation in turn $\pm 3SD$, (B) Shape variation. 74
Figure 5-6: Appearance mean and first three eigenvectors. 77
Figure 5-7: The original texture and original texture described by combined model 77
Figure 5-8: (A) Input image; (B) finding the best fit, initial position (red points) the best fit (blue points); (C) segmented spine and aorta 78
Figure 5-9: (A) Segmented calcified aorta produced from AAM; (B)- biggest region area (aorta). 79
Figure 5-10: Examples of segmented aortas with different degree of calcification and different brightness 80
Figure 5-11: Plots of PSNR vs. number of thresholds for three different calcified aortas: moderate AAC; low AAC; high AAC. 81
Figure 5-12: (A) original aorta image; (B) 2-level thresholding; (C) 4-level; (D) 6-level; (E) 7-level; (F) 14-level. 82

Figure 5-13: (A) Original image; (A) segmented aorta,(C) segmented aorta quantised using 7 threshold levels; (D) binary images after thresholding. 85
Figure 5-14: (A) Segmented aorta quantised using 7 threshold levels; (B), (C) binary images .. 85
Figure 5-15: Variation of CCR with feature number 87
Figure 5-16: (A) accuracy of K-nn and SVM classifiers for AAC classes 1, 2 and 3; (B) AUC for each classifier and AAC class. 89
Figure 6-1:.Examples of VFA images obtained by DXA scanner; (A) a 91- year-old patient with highly calcified aorta (AC-24 =19), (B) a 74-year-old-patient with moderate calcified aortic walls (AC-24= 8), (C) an 82 year-old- patient with mild calcified aorta (AC-24= 3), (D) a 38 year-old-patient with no-aortic calcification (AC-24=0). 93
Figure 6-2: Examples of VFA images excluded from the dataset due to;(A) inadequate space anterior to the aorta; (B) anatomical artefacts, (C) artefacts due to patient's movement. 94
Figure 6-3: The correlation of AC-24 points between two readers 97
Figure 6-4: Scatter plot and linear regressions with 95% confidence intervals for the correlation 99
Figure 6-5: Data distribution of the grouped diagnosis in AC-24 scoring by Reader 1 system. 100
Figure 6-6: Average classification accuracy achieved for 4 AAC categories by three classifiers with the number of features used. 101
Figure 6-8: Average classification accuracy achieved for 3 AAC categories by three classifiers with the number of features used. 106
Figure 7-1: (A) abdominal region (B) ellipse shape. 114
Figure 7-2:Perspex Phantom and aluminium strip 118
Figure 7-3: Experimental set up for lateral VFA scan of Hologic spine phantom 121
Figure 7-4: Lateral VFA image of Hologic spine phantom with: a)17 cm scan length, b) 30 cm scan length. 122
Figure 7-5: Digimatic screw gage micrometre. 124
Figure 8-1: Experimental setup for VFA imaging 128
Figure 8-2: Original phantom VFA images acquired by Hologic Horizon scanner for 15 cm Perspex and 0.5 mm Al, (a) single energy mode, (b) dual energy mode. 129
Figure 8-3: Original VFA images acquired by DXA Hologic Discovery for 15 cm Perspex and 0.5 mm Al, (a) single energy mode, (b) dual energy mode. 129
Figure 8-4: ROIs on a phantom images of 15 cm Perspex (left, L and right, R) and 0.5 mm aluminium (centre, C) obtained by Horizon scanner, (a) single energy mode, (b) dual energy mode. 131
Figure 8-5: ROIs on a phantom images of 15 cm Perspex (left, L and right, R) and 0.5 mm aluminium (centre, C) obtained by Discovery scanner, (a) single energy mode, (b) dual energy mode. 131
Figure 8-6: VFA image profile showing the aluminium strip edge effect for 15 cm Perspex combined with 2 mm Aluminium obtained by Horizon scanner in (a) single energy mode, (b) dual energy mode. 132

Figure 8-7: VFA image profile showing the aluminium strip edge effect for 15 cm Perspex combined with 2 mm Aluminium obtained by Discovery scanner in (a) single energy mode, (b) dual energy mode. 133
Figure 8-8: Image profile showing the Perspex with 15 cm, 25 cm and 40 cm in single energy mode for Horizon scanner, (a) in X-direction (left), (b) Y-direction(right). 137
Figure 8-9: Image profile showing the Perspex with 15 cm, 25 cm and 40 cm in dual energy mode for Horizon scanner, (a) in X-direction(left), (b) Y-direction(right). 138
Figure 8-10: Image profile showing the Perspex with 15 cm, 25 cm and 40 cm in single energy mode for Discovery scanner, (a) in X-direction, (b) Y-direction. 139
Figure 8-11: Image profile showing the Perspex with 15 cm, 25 cm and 40 cm in dual energy mode for Discovery scanner, (a) in X-direction, (b) in Y-direction 140
Figure 8-12: Variation of the mean of pixels values in VFA images obtained by DXA Horizon with Perspex width between 10 and 40 cm in (a) single energy mode (b) dual energy mode.. 141
Figure 8-13: Variation of the mean of pixels values in VFA images obtained by DXA Discovery with Perspex width between 10 and 40 cm in (a) single energy mode (b) dual energy mode.. 142
Figure 8-14: The Contrast as a function of the thickness of Al for different Perspex width in SE mode images performed by Horizon scanner. 147
Figure 8-15: The CNR as a function of the thickness of Al for different Perspex width in DE mode images obtained by Horizon scanner. 147
Figure 8-16: The Contrast as a function of the thickness of Al for different Perspex width in DE mode images obtained by Horizon scanner 148
Figure 8-17: CNR as a function of the thickness of AL for different Perspex width in DE mode images obtained by Horizon scanner. 148
Figure 8-18: The Contrast as a function of the thickness of Al for different Perspex width in SE mode images obtained by Discovery scanner. 149
Figure 8-19: The CNR as a function of the thickness of AL for different Perspex width In SE mode images obtained by Discovery scanner. 149
Figure 8-20: The contrast as a function of the thickness of Al for different Perspex width in DE mode images obtained by Discovery scanner. 150
Figure 8-21: CNR as a function of the thickness of Al for different Perspex width in SE mode images performed by Discovery 150
Figure 8-22: VFA images of a Perspex 25 cm combined with (a) 0.15 mm of AL (not visible), (b) 0.2 mm of AL (not visible), (c) 0.25 mm of AL in SE mode (visible) and (d) 0.15 mm (not visible) (e) 0.2 mm (not visible), (f) 0.25 mm (visible) in DE mode performed by DXA Horizon scanner 153
Figure 8-23: VFA images of a Perspex 25 cm combined with (a) 0.15 mm of AL (not visible), (b) 0.2 mm of AL (not visible), (c) 0.25 mm of AL in SE mode (visible) and (d) 0.15 mm (not visible) (e) 0.2 mm (not visible), (f) 0.25 mm (visible) in DE mode performed by Discovery. 154

List of Tables

Table 4-1: QC results of three DXA scanners used in clinical VFA scans. 58
Table 4-2: QC results of three DXA scanners used experimental phantom scan. 58
Table 5-1: An example of AC24 scoring for one VFA scan of the aorta 69
Table 6-1: characteristics of study data set. 92
Table 6-2: Intra- observer obtained for two Readers for AC-24 and AAC categories with 95% CI. 96
Table 6-3: Inter-observer agreement measured for two Readers for AC-24 and AAC categories with 95% CI. 97
Table 6-4: Accuracy, sensitivity, specificity and SD of three classifiers for 4 AAC categories.. 103
Table 6-5: Confusion matrix obtained using SVM classifier for 4-AAC categories. 104
Table 6-6: Confusion matrix obtained using K-nn classifier for 4-AAC categories. 104
Table 6-7: Confusion matrix obtained using RF classifier for 4-AAC categories. 104
Table 6-8: Accuracy, sensitivity, specificity and SD of three classifiers for 4 AAC categories.. 106
Table 6-9: Confusion matrix obtained using SVM classifier for 3-AAC categories. 107
Table 6-10: Confusion matrix obtained using K-nn classifier for 3-AAC categories. 107
Table 6-11: Confusion matrix obtained using RF classifier for 3-AAC categories. 107
Table 7-1: Elemental composition of water, Perspex and soft tissue (Mihailescu and Borcia 2006; Singh et al. 2014). 111
Table 7-2: Properties of soft tissue, Perspex and water (Lewis et al. 2012) 113
Table 7-3: Estimated body width for men. 116
Table 7-4: Estimated body width for women. 117
Table 7-5: Maximum and minimum body width estimated for men and women. 118
Table 7-6: Properties Hydroxyapatite and aluminium energies of 40 and 70 kev (R. Jonson, 1993). 123
Table 7-7: Thickness measurements for an aluminium strip with nominal thickness 0.5mm ... 124
Table 7-8: Thickness measurements for an aluminium strip with nominal thickness 0.05,0.1,0.2,0.3,0.5 and 1 mm. 125
Table 8-1: The mean, SD and % CV of pixel values on images acquired by DXA Horizon in single energy mode. 143
Table 8-2: The mean, SD and %CV of pixel values on images obtained by DXA Horizon in dual energy mode. 143
Table 8-3: The mean, SD and %CV of pixel values on images obtained by DXA Discovery in single energy mode. 144
Table 8-4: The mean, SD and %CV of pixel values on images obtained by DXA Discovery in dual energy mode. 144

Table 8-5: The minimum detectable thickness assessed visually on the raw VFA images obtained by DXA Horizon and Discovery and the CNR threshold of detectability 152
Table 8-6: %CV of the mean pixel values in The Perspex ROI and %CV of C and CNR for both repeatability and reproducibility for Horizon in SE and DE modes for a phantom configuration of 15 cm of Perspex with 0.1, 0.5 and 1 mm of Al obtained by DXA Horizon 158
Table 8-7: %CV of the mean pixel values in The Perspex ROI and %CV of C and CNR for both repeatability and reproducibility for Horizon in SE and DE modes for a phantom configuration of 25 cm of Perspex with 0.1, 0.5 and 1 mm of Al obtained by DXA Horizon 159
Table 8-8: %CV of the mean pixel values in The Perspex ROI and %CV of C and CNR for both repeatability and reproducibility for Horizon in SE and DE modes for a phantom configuration of 40 cm of Perspex with 0.1, 0.5 and 1 mm of Al obtained by DXA Horizon 160

List of Publication

Conference proceedings:

Karima Elmasri, Yulia Hicks, William D Evans, Xin Yang, Xianfang Sun, Rebecca Pettit. A robust technique for detecting abdominal aortic calcification using dual energy x-ray absorptiometry. All Wales Summer Meeting Medical Physics and Clinical Engineering, Abstracts. August (2016), pp.9–11.

Karima Elmasri, Yulia Hicks, William Evans, Xin Yang, Xianfang Sun, Rebecca Pettit. Evaluation of vertebral fracture assessment images for the detection of abdominal aortic calcification. All Wales Summer Meeting Medical Physics and Clinical Engineering, Abstracts. June (2017), pp.3–5.

Karima Elmasri, Yulia Hicks, William Evans, Xin Yang, Xianfang Sun, Rebecca Pettit. (2016). Automatic Detection and Quantification of Abdominal Aortic Calcification in Dual Energy x-ray Absorptiometry. Procedia Computer Science, 96, pp.1011–1021.

Journal articles under preparation:

Karima Elmasri, Yulia Hicks, William Evans, Xin Yang, Xianfang Sun, Rebecca Pettit (2018). Evaluation of automatic method for detection and quantification of Abdominal Aortic Calcification using vertebral Fracture Assessment Images. Will be submitted to Radiology Journal.

Karima Elmasri, Yulia Hicks, William Evans, Xin Yang, Xianfang Sun, Rebecca Pettit (2018). Performance evaluation of DXA for the detection of Calcification. Will be submitted to Biomedical Physics & Engineering journal.

Abbreviations and Symbols

The following are abbreviations and symbols used in this thesis:

AAC	Abdominal aortic calcification
ABQ	Algorithm-based qualitative method
AC-8	8-points semi-quantitative scoring of AAC
AC-24	24-points semi-quantitative scoring of AAC
AAM	Active appearance model
AP	Anterior-posterior
ASM	Active shape model
BA	Bone area
BC	Bone mineral content
BMD	Bone mineral density
BMI	Bone mass index
C	Contrast
CC	Calcium carbonate
CAC	Coronary artery calcium
CAD	Computer-aided diagnosis

CCR	Correct classification rate
CHD	Coronary heart disease
CNR	Contrast to noise ratio
CKD	Chronic kidney disease
CO	Calcium oxalate
CP	Calcium phosphate
CPP	Calcium pyrophosphate
CT	Computed tomography
CV	Coefficient of variation
CVD	Cardiovascular disease
DE	Dual energy
DXA	Dual energy x-ray absorptiometry
FCM	Fuzzy C-means
FN	False negative
FP	False positive
FRAX	Fracture risk assessment tool
GADOX	gadolinium sulfoxylate scintillator
GLCM	Grey-level co-occurrence matrix
HA	Hydroxyapatite

HU	Hounsfield Unit
IVUS	Intravascular US
K-nn	K-nearest neighbour
L1	1st lumbar vertebra
L4	4th lumbar vertebra
LDL	Low density lipoprotein
LOOCV	Leave-one-out-cross validation
MRI	Magnetic resonance imaging
MRF	Markov Random Fields
μSv	Micro sievert
PET	Positron emission tomography
PD	Paget's disease
RF	Random forest
RMSE	Root mean square error
ROI	Region of interest
SD	Standard deviation
SPECT	Single photon emission computed tomography
SVM	Support vector machine
TN	True negative

TP	True positive
US	Ultrasound
VFA	Vertebral fracture assessment

Chapter: 1

Introduction

1.1 Motivation

Cardiovascular disease (CVD) is a major cause of mortality and the main cause of morbidity worldwide. The most prevalent form of CVD is a condition called atherosclerosis, in which the arteries become narrowed and hardened due to an excessive build-up of plaque on the inner artery wall (Sanz and Fayad 2008; Golestani et al. 2010). More than 50% of patients with atherosclerosis die without prior clinical symptoms (Barascuk et al. 2011). According to the British Heart Foundation, in the United Kingdom CVD is the second highest cause of death, with a total number of deaths of about 155,000 per year. In 2014, CVD caused 27% of all deaths (Nick et al. 2015).

Atherosclerosis is an inflammatory disease of the large and medium-sized arteries, including the aorta, the carotid arteries, coronary arteries and muscular arteries, leading to ischemia of the heart and brain and possibly resulting in infarction (Lorkowski and Cullen 2007; Tuttolomondo et al. 2012). Heart attack, stroke, aortic aneurysms and peripheral vascular disease may occur due to atherosclerosis.

Arterial calcification, especially abdominal aortic calcification (AAC), is a manifestation of atherosclerosis and a predictor of CVD (Cecelja et al. 2013). The presence of AAC means that is it very likely that there is some atherosclerosis in the largest artery and elsewhere in the arterial system. Moreover, an inverse relationship between bone mineral density (BMD) and measures of AAC has been reported (Grant et al. 2017). Atherosclerosis increases with advancing age, as the plaque develops and the size of

calcified regions increases to be visible on radiographs of the thoracic as well as the abdominal region (Wilson et al. 2001; Okuno et al. 2007).

AAC is an established risk factor for CVD (Doherty et al. 2004; Chow et al. 2008; Golestani et al. 2010; Wong et al. 2011; Ganz et al. 2012; Bastos Gonçalves et al. 2012; Kim et al. 2013; Maroules et al. 2013; Tatami et al. 2015). In general, AAC is common in both CVD and chronic kidney disease (CKD) (Leckstroem et al. 2014).

Several non-invasive modalities are employed for AAC detection. These include multi-slice computed tomography (CT), radiography and ultrasound (US). Currently, CT is the gold standard technique to measure AAC but it is restricted due to high radiation exposure (Cecelja et al. 2013; Chuang et al. 2014; Leckstroem et al. 2014). The simplest method of AAC detection is radiography. Calcified lesions can be visualised on radiographs of the thorax and abdomen (Wilson et al. 2001).

Dual energy x-ray absorptiometry (DXA) is a standard diagnostic technique widely used to detect osteoporosis and vertebral fractures. Vertebral fracture assessment (VFA) has become a valuable and significantly used method to diagnose patients with osteoporosis due to the presence of vertebral fractures. During a VFA scan, the soft tissue anterior to the lumbar spine is incidentally included. This allows for the detection of calcified plaques in the abdominal aorta.

VFA images performed by DXA in single energy mode are used for the detection of AAC (Schousboe et al. 2006; Schousboe et al. 2007; Cecelja et al. 2013; Netto et al. 2013). Several recent studies (Lewis et al. 2016; Grant et al. 2017; Schousboe et al. 2017) used VFA methods as a reasonable substitute for radiography to measure AAC.

VFA images have several advantages over radiographs. These comprise simplicity, KeV, quick acquisition of images and lower cost. Another advantage is the low radiation dose required, 3-40 μSv, to perform a VFA image, whereas the radiation exposure for a radiograph is approximately 600 μSv (Schousboe et al. 2010; Szulc et al. 2013).

Furthermore, VFA uses a fan-beam of x-rays to scan the spine whereas in radiography, images are performed using a cone-beam, which produces parallax distortion within vertebrae located above or below the central point of the beam. Nonetheless, images obtained by radiography provide better visibility with less noise and higher spatial resolution than VFA images by DXA, with the result that cortical edges and endplates can be identified properly (Schousboe et al. 2010).

Although several simple methods have been proposed for manual semi-quantitative scoring of AAC, these methods have limitations, in particular in terms of capturing small changes in atherosclerosis progression. Moreover, the techniques rely on input from the observer/investigator and are time-consuming. The most popular method is the established and validated visual approach of Kauppila et al. (1997), in which calcification in the walls of the abdominal aorta is scored on a scale ranging from 0 to 24. This method is known as the AC-24 or 24 point system; it was applied by the Framingham study group (Schousboe et al. 2006; Nielsen et al. 2010; Cecelja et al. 2013; Mohammad et al. 2014) to estimate CVD risk from lumbar aortic radiographs.

With regard to the performance of DXA compared with other imaging modalities for AAC measurement using the AC-24 system, agreement was good to very good with radiography and moderate to good with CT (Schousboe et al. 2006; Toussaint et al. 2009).

In recent years, several automatic approaches have been proposed to measure AAC on radiographs. However, none of them were applied to VFA images, which have lower resolution and higher noise. In addition, the proposed methods have not been related to existing manual AAC scoring techniques and have not been properly evaluated from a medical point of view.

VFA images obtained in single energy mode alongside BMD measurement using DXA can be used to assess risk factors for CVD, although currently this kind of assessment

is not normally included in the BMD report. In principle medical intervention could be provided to individuals with AAC detected during osteoporosis screening in order to prevent the progression of CVD.

In addition, introducing computer-based assessment algorithms to measure AAC in VFA images could facilitate the detection of this risk factor for CVD.

1.2 Aims and Objectives

The goal of the research presented in this thesis is to assess the suitability of the DXA modality for the detection of AAC in VFA images with the purpose of finding an additional low-cost method of diagnosing CVD.

The aims of this project are: (1) development of a new method for the automated detection and quantification of AAC on VFA images relating this to existing AAC scoring techniques; (2) investigation of the limitations of VFA images for AAC detection using a physical phantom.

The specific objectives necessary to achieve the aims in the first part are identified as:

1. Creating an automatic method for AAC detection and quantification and assessing it against existing manual scoring technique in order to assist clinicians in the detection of AAC at all stages of development;

2. Validating the developed automatic system for all possible AAC categories on a large dataset consisting of several hundred images.

The sensitivity of the VFA medical imaging technique in detecting calcification in the patient's aorta is an important factor for achieving accurate AAC evaluation. Thus, the assessment of the capability of VFA to image AAC under various scan acquisition conditions is the second aim in this research. To achieve this aim, a phantom simulation

study was conducted. This is the first physical phantom study to address this issue, and the corresponding research objectives were identified as follows:

1. Designing and constructing a physical phantom to simulate soft tissue in the abdomen and calcification within this region with the capability to cover a range of abdominal width as well as a range of calcification thickness.

2. Conducting a study with this phantom to examine the sensitivity of VFA for AAC detection including:

2.1 investigating the performance of VFA with different phantom configurations and image acquisition conditions;

2.2 Investigating the minimum detectable thickness of uniformly distributed AAC.

1.3 Thesis Outline

This thesis is organised into the following structure:

- Chapter 1 provides an introduction to the work.

- Chapter 2 presents a review of background information on cardiovascular diseases and their association with calcification in the abdominal aorta, and discusses related work in this field. It also reviews available scoring systems for calcification in the abdominal aorta.

- Chapter 3 reviews techniques relevant to the automatic assessment of AAC, such as automatic image segmentation and data classification methods.

- Chapter 4 reviews the fundamentals of the DXA modality including VFA.

- Chapter 5 introduces a new automatic technique for AAC detection and quantification.
- Chapter 6 presents a validation of the new method using a large dataset.
- Chapter 7 describes phantom design and construction.
- Chapter 8 presents a novel phantom study involving two DXA scanners.
- Chapter 9 highlights the contributions, limitations and conclusions of this thesis, and proposes further work.

Chapter: 2

Abdominal Aortic Calcification

2.1 Introduction

This chapter surveys previous studies relevant to the detection and quantification of AAC presented in this thesis. The chapter provides general background information related to atherosclerosis and AAC and presents the common modalities used to detect AAC. The chapter also discusses the physical characteristics of calcification such as the size and chemical composition. Existing medical visual assessment systems for AAC quantification are also presented.

2.2 Atherosclerosis

Atherosclerosis or arteriosclerotic vascular disease is a systemic disease characterised by thickening and hardening of blood vessel walls to form atherosclerotic plaque (Sanz and Fayad 2008; Golestani et al. 2010). The disease is identified by the formation of fatty lines, fibrous plaques and calcified plaques in the inner layer of the arterial wall. The calcified plaques affect the shape and mechanical properties of the arteries. Atherosclerosis occurs mostly in large and medium-sized arteries and can lead to ischemia of the heart and brain, possibly resulting in infarction (Lorkowski and Cullen 2007; Tuttolomondo et al. 2012).

The plaque is divided into two categories: stable and unstable (vulnerable) (Fishbein 2010). Atherosclerotic plaque pathobiology is very complicated; in general, stable atherosclerotic plaques are rich in extracellular matrix and smooth muscle cells. Vulnerable plaques have also been called “high-risk” or “thrombosis-prone” plaques (Bentzon et al. 2014).

The presence of calcium is not a marker of stable or vulnerable plaques, but there is an association between coronary atherosclerotic plaques and the extent of atherosclerosis and this allows to atherosclerosis to be categorised (Oei et al. 2002).

Cholesterol deposition and thrombosis inflammation play an important role in plaque formation (Kim et al. 2000). As the plaque develops, it projects into the space inside the artery where the blood is flowing. Subsequently, calcium is deposited and collagen fibre production by smooth muscle cells declines. A complicated lesion forms when the fibrous plaque breaks and opens, revealing the cholesterol and connective tissue underneath combined with a blood clot.

An increasing blood concentration of low-density lipoprotein (LDL) in combination with other risk factors such as smoking, hypertension, diabetes mellitus, gender, and a complex genetic susceptibility to the disease can be sufficient to form atherosclerosis. Typically, the abdominal aorta, coronary arteries, iliofemoral arteries and carotid bifurcations are the most affected locations (Bentzon et al. 2014).

Arteries consist of three layers: tunica adventitia, tunica media and tunica intima. Tunica adventitia components are type I collagen networks of elastic fibres and proteoglycans that form loose connective tissue. Scattered fibroblasts, adipocytes, rarely macrophages, lymphocytes and mast cells are the components within the adventitia. The tunica media is composed of smooth muscle cell sub-layers arranged diagonally and separated by wavy sheets of elastic membranes. Collagen type III fibres, elastin, proteoglycans and glycoproteins are its other major components. The tunica intima is made up of one layer of endothelial cells. Within the sub-endothelial space that forms the border between intima and media, there are fibrils of elastin and extracellular matrix (Bukowska et al. 2014). Figure 2-1 shows a normal an artery and artery narrowed by atherosclerotic plaque.

Atherosclerosis causes CVD such as heart attack, stroke, aortic aneurysm, and peripheral vascular disease, which represent the greatest cause of death worldwide. This disease can be diagnosed only at advanced stages (Sanz and Fayad 2008). Symptoms of atherosclerosis depend on the location of the affected artery and whether the walls narrowed gradually or become blocked suddenly. A heart attack occurs when coronary arteries are blocked suddenly; stroke occurs when the arteries supplying the blood to the brain are blocked. If the arteries supplying blood to the leg are blocked, this may cause gangrene of a toe, foot or leg. Angina occurs when heart muscles affected and leg pain while walking if the leg muscles are affected.

Calcification detected by lateral lumbar radiographs is an indicator of subclinical atherosclerotic disease and an independent predictor of CVD (Bots et al. 1993; Lillemark 2010). CVD mortality in a large population sample (1049 men and 1466 women) was studied by Wilson et al. (2001). The study proved that calcification seen on lumbar radiographs is a mark of subclinical atherosclerosis and vascular morbidity and mortality.

Studies relating to the coronary arteries have also proved an association between calcification and cardiovascular events, in particular, myocardial infarction (Alexopoulos et al. 2003; Doherty et al. 2004; Shakeri et al. 2013).

The association between atherosclerosis and osteoporosis has been proven in many studies; they are correlated with age as both conditions are common in the elderly (Chow et al. 2008). Generally, atherosclerosis increases and advances with age, as the plaque develops and the size of calcification increases to be visible on radiographs of the abdominal and thoracic regions (Wilson et al. 2001; Okuno et al. 2007).

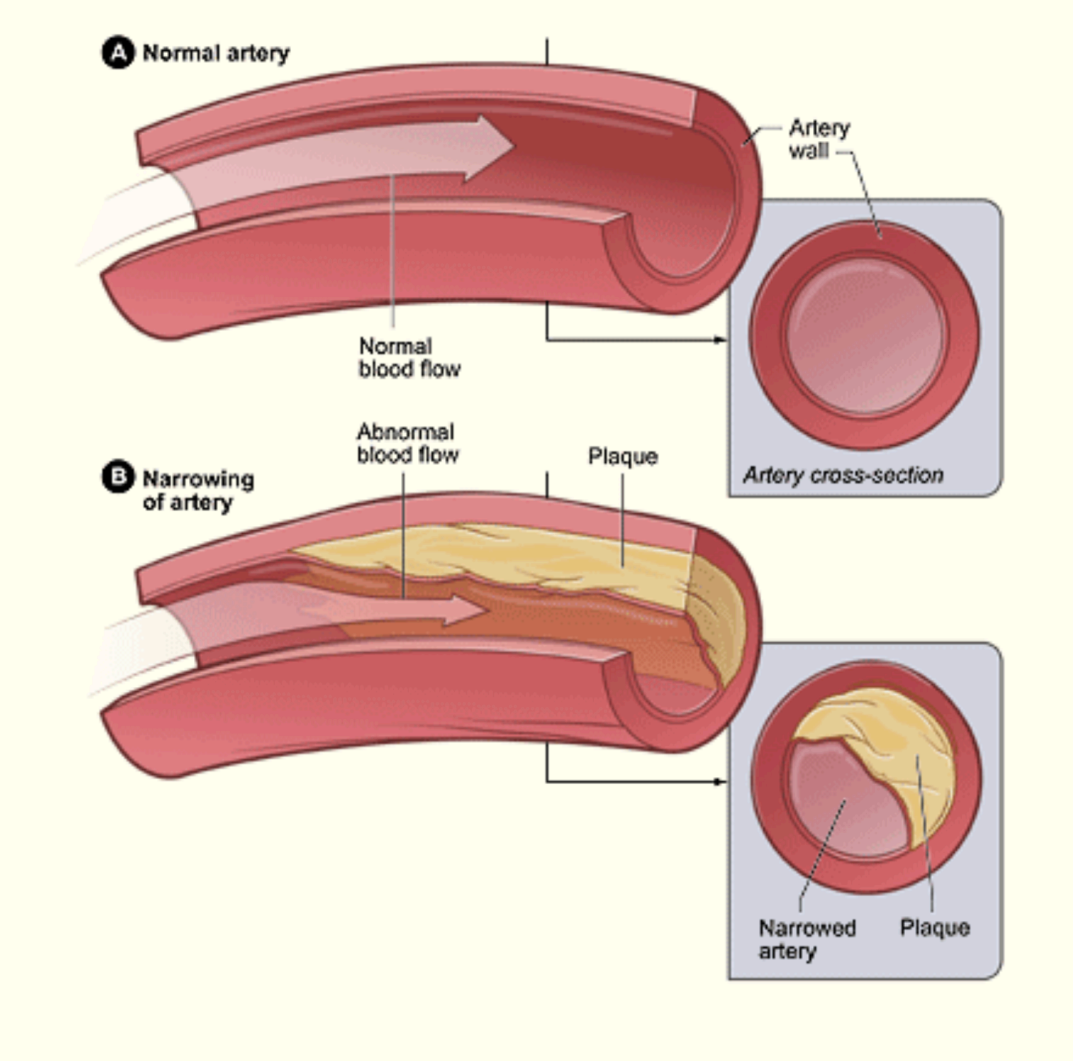


Figure 2-1: Diagram showing the effects of atherosclerosis on an artery (https://en.wikipedia.org/wiki/Artery)

2.3 Abdominal Aortic Calcification (AAC)

The abdominal aorta (Van Metter et al. 2000a) is the biggest blood vessel in the body; it is imbedded in adipose tissue behind the peritoneal cavity. It starts at the aortic hiatus of the diaphragm, anterior to the inferior border of the 12th thoracic vertebra, and it descends anteriorly to the end at the lower border of the 4th lumbar vertebra (Songur et

al. 2010). It bifurcates at lumbar vertebra L4 into the common iliac arteries. Abdominal aortic bifurcation is shown in Figure 2-2.

The normal diameter of the abdominal aorta is considered to be less than 30 mm (Van Metter et al. 2000b; Erbel and Eggebrecht 2006). The mean (SD) aortic diameter was reported as 18.8 (3.0) mm at the proximal bifurcation by Laughlin et al. (2011) and as 18.6 (2.2) mm by Allison et al. (2008). The normal arterial vessel wall thickness was reported to be < 1 mm in a study by Bentzon et al. (2014).

The normal abdominal aorta cannot be distinguished from surrounding soft tissue by any x-ray imaging technique because its composition is similar to soft tissue. However, the abdominal aorta can be captured on radiographs if it is calcified. AAC is identified visually as either a diffuse white stippling of the aorta extending out to the aortic wall or as white linear calcification of the anterior and/or posterior aortic walls (Schousboe et al. 2017).

There are two categories of arterial calcification, intimal and medial. Both types can be seen in the abdominal aorta; however, most of the available studies have concentrated on intimal calcification because of its strong connection with atherosclerosis (Jayalath et al. 2005). Intimal calcification can be seen in patients with renal failure and diabetes mellitus. Medial calcification can affect the abdominal aorta but it is rarely detected in the coronary arteries. Discrimination between these two types of calcification is not possible and it seems to be a limitation of medical imaging techniques (Jayalath et al. 2005).

There is a strong association between factors such as age, sex, BMI, and the presence and extent of calcified atherosclerosis (Allison et al. 2008). Calcification and stenosis generally affect people older than 65 years of age. Calcification strongly correlates with coronary artery events (Okuno et al. 2007).

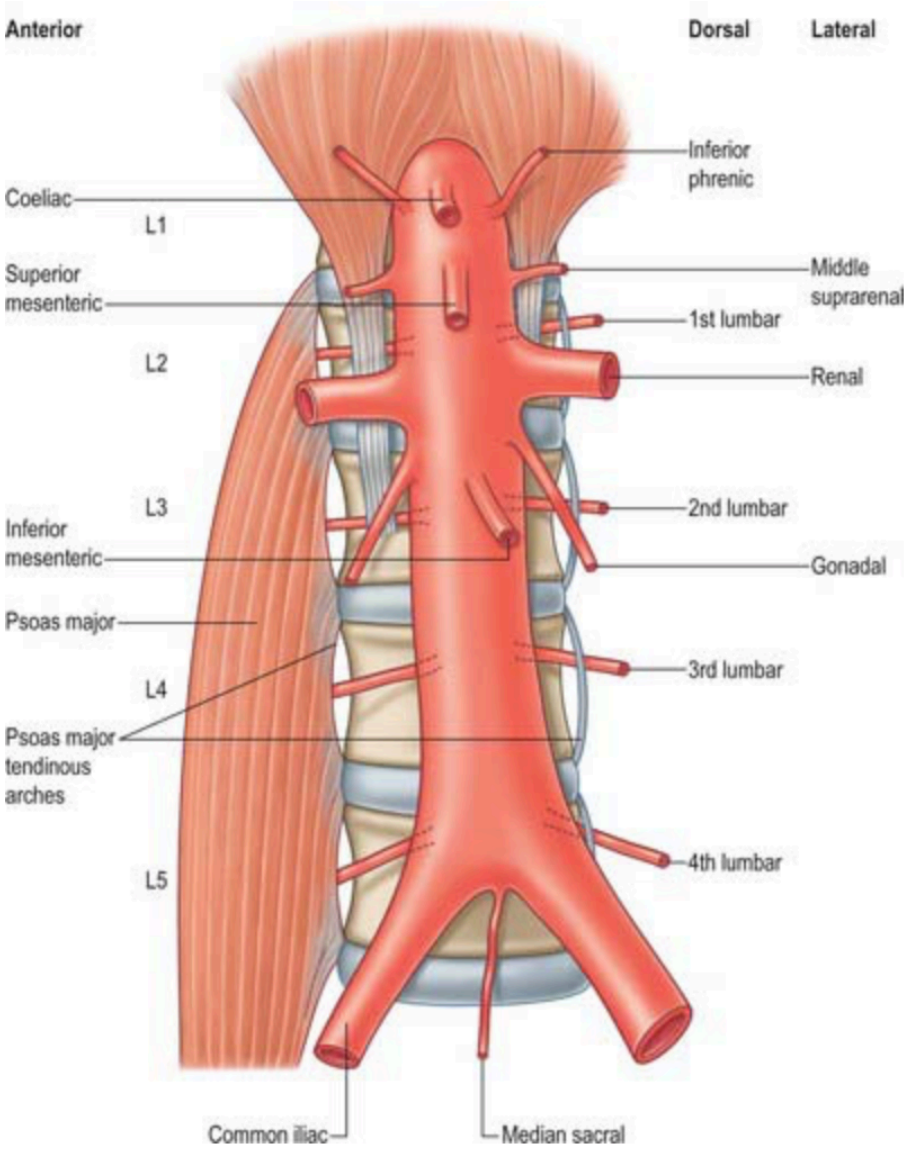


Figure 2-2: Abdominal aorta bifurcation. (Openstax, Circulatory Pathways, 2015, https://anatomychartee.co/show/branch-of-abdominal-aorta.html

A significant correlation between aortic calcification, aortic wall thickness and aortic extensibility and coronary artery disease was reported by Jang et al. (2012). Generally, AAC is a strong predictor of the CVD (Doherty et al. 2004; Chow et al. 2008; Golestani et al. 2010; Wong et al. 2011; Ganz et al. 2012; Bastos Gonçalves et al. 2012; Kim et al. 2013; Maroules et al. 2013; Tatami et al. 2015).

The prevalence of calcification is higher in the abdominal aorta compared to the thoracic aorta and it is correlated with fat deposition plaques (Benvenuti et al. 2005; Cecelja et al. 2013). There is consistency in many studies regarding the association of AAC with age in the middle-age group (Kim et al. 2013). Moreover, the progression rate of AAC is correlated significantly with chronic kidney disease CKD (Yamamoto et al. 2016).

The Framingham Heart Study revealed that abdominal aortic calcification is a heritable atherosclerotic feature (O'Donnell et al. 2002). Risk factors such as smoking, age, physical activity, and diabetes mellitus are common in both atherosclerosis and AAC (Golestani et al. 2010; Bolland et al. 2010).

CVD mortality risk can be predicted by the distribution, size, and shape of aortic calcification, in addition to its extent (Nielsen et al. 2010).

In general, AAC is considered as a strong predictor of CVD. Many studies have proven a strong correlation between AAC and coronary artery calcification, kidney disease and osteoporosis (Doherty et al. 2004; Chow et al. 2008; Golestani et al. 2010; Wong et al. 2011; Ganz et al. 2012; Bastos Gonçalves et al. 2012; Kim et al. 2013; Maroules et al. 2013; Tatami et al. 2015).

2.4 Characteristics of Calcification

Calcification associated with atherosclerosis results from highly regulated and organised active cellular processes. The mechanism of arterial calcification is not completely understood. However, it has been suggested that arterial wall cells, such as vascular smooth muscle cells, may produce a mineralised matrix (Lee et al. 2012). There is considerable interest in generating further understanding of the mechanisms of calcification and its implications in CVD. Calcification formation and composition have been intensively investigated with different methods in both in vitro and in vivo studies.

This section discusses two characteristics of calcification: its chemical composition and its thickness.

2.4.1 **Calcification Chemical Composition**

The chemical composition of calcification has been investigated by various methods, including x-ray diffraction, Fourier transform infrared spectroscopy and chemical analysis (Liao et al. 2009). The calcification of human artery plaque was examined by Fitzpatrick et al. (1994). This study proved that the chemical composition of calcification was identical to hydroxyapatite ($Ca_{10}[PO_4]_6\,[OH]_2$), which is the major inorganic component of bone (Lee et al. 2012). Calcified deposits appear to be similar to bone mineral, chemically and morphologically (Dmitrovsky and Boskey 1985). Higgins et al. (2005) reported that calcium hydroxyapatite is one of the forms of calcification throughout the vasculature. Calcification in the aortic valves was analysed by Ortlepp et al. (2004) and they proved that it consists of Ca-deficient hydroxyapatite.

Some studies have reported that breast calcification consists of calcium oxalate (Warren et al. 2013). Micro-calcifications have been separated into two types: type I, calcium oxalate and type II calcium hydroxyapatite (Haka et al. 2002; Prieto et al. 2011). According to the same study, discrimination between these two types is not possible. Similar were published by Lee et al. (2012); in cardiovascular tissues calcium deposits comprised different calcium salts, such as calcium carbonate (CC), calcium oxalate (CO), calcium phosphate (CP), calcium pyrophosphate (CPP), and calcium hydroxyapatite (HA).

2.4.2 **Calcification Thickness**

There is little information about the size and distribution of calcified plaques within the aorta. To some extent, this is difficult to obtain during in vivo research. The thickness of arteriosclerotic calcification was measured using a multi- metric ruler in a study by

Laroche and Delmotte (2005); the length of calcification was also measured.The study aimed to quantify the thickness and length of vascular calcification in patients with Paget's disease (PD) and compare them with those in a control group. Calcification thickness was scored as follows: 1, between 1 and 2 mm; 2, between 2 and 3 mm; 3, more than 3 mm. The mean length of calcification was 1.93 ± 2.85 cm in PD cases and 0.84 ± 1.69 cm in controls ($p = 0.04$). Calcification thickness was 1.24 ± 1.30 mm in PD patients, and 0.56 ± 0.94 mm ($p = 0.01$) in controls.

Babiarz et al. (2003) scored the thickness of calcification in the cavernous carotid arteries based on computed tomography (CT) as follows: 0, 1 for 1 mm, 2 for 2 mm, 3 for 3 mm and 4 for > 3 mm. In a study by Moreau et al. (1994), the thickness of tissue and bone within atherosclerotic plaque was studied using a calibration phantom that had similar geometry to that of calcified plaque. Lucite and an epoxy-based bone mimicking-material (SB3) were chosen with thicknesses of 15 mm and 5 mm respectively. The authors reported that the reason for choosing this large range was that in heavily calcified arteries, the thickness can be as large as 3-4 mm for same type of calcification.

2.5 **Detection of Abdominal Aortic Calcification**

Arterial calcification measurements are the most effective approaches to predict clinical events associated with the abdominal aorta. CT, radiography and ultrasound (US) are the most common modalities used for the assessment of atherosclerotic plaque (Jayalath et al. 2005).

Numerous studies have measured and validated AAC on lateral spine radiographs and on CT scans (Lewis et al. 2016; Schousboe et al. 2017). CT is presently used as the gold standard measure of artery calcification, but is limited by high radiation exposure (Cecelja et al. 2013; Chuang et al. 2014; Leckstroem et al. 2014; Lee et al. 2017).

The most popular techniques used for AAC detection are presented briefly in this section.

2.5.1 **Computed Tomography (CT)**

CT is a diagnostic imaging procedure that uses x-rays to build cross-sectional images ("slices") of the body. Cross-sections are reconstructed from measurements of attenuation coefficients in the volume of the object studied.

Multislice computed tomography is an important tool for non-invasive imaging of cardiovascular structures (Schmitz et al. 2005; Schousboe et al. 2017). In addition, CT easily permits complete quantification of calcification in arterial walls.

Coronary calcium scores were determined by CT in a study by Bolland et al. (2010). Scoring was performed using a 64 slice CT scanner. In another study, abdominal CT scans of 380 patients were analysed to investigate the correlation of AAC with life- style and risk factors for CVD (Kim et al. 2013). The findings of this research were that AAC was related to gender, age, presence of dyslipidemia, exercising, smoking in male subjects and the presence of diabetes mellitus and hypertension in female subjects.

The association of AAC with CVD in 1974 men and women aged 45 to 84 years who had complete AAC and coronary artery calcium data from CT scans was investigated in a study by Criqui et al. (2014). They found that both AAC and CAC could predict heart disease and CVD events independently of one another and that AAC was more strongly correlated with total mortality than CAC.

2.5.2 **Ultrasound**

Ultrasound is non-invasive tomographic technique, which has the capability to immediately image vessel walls. The tomographic orientation of ultrasound can visualise the entire circumference of the vessel wall and its surfaces. The main disadvantage of US is that it is operator- dependent and has low reproducibility (Hartshorne et al. 2011). US may be used to visualise the of intima-media thickness of carotid arteries as they are superficial structures. This can provide additional information about myocardial infarction

or stroke (Fayad and Fuster 2001; Sanz and Fayad 2008; Ganz et al. 2012). Ultrasonic contrast agents have been introduced to improve image resolution and specificity.

Intravascular ultrasound (IVUS) is another diagnostic modality that combines anatomic and physiologic measurements in real time. This technique is a catheter-based, invasive modality that allows direct imaging of atheroma and offers a cross- sectional, tomographic perspective of the vessel (Fayad and Fuster 2001; Ricotta et al. 2008; Sanz and Fayad 2008). Artefacts adversely affect the images; these include artefacts generated by acoustic oscillations, which makes the catheter appear larger in size than its physical size (Nissen and Yock 2001). IVUS cannot define calcification thickness as ultrasound cannot penetrate calcium, but it can detect the arc of calcium layers (Dey and Roy 2012).

2.5.3 **Radiography**

Radiography is an imaging modality that uses a beam of x-rays projected toward the object to view its internal structure. A certain amount of radiation is absorbed by the object, dependent on its density, composition and thickness. Plain radiography is the simplest method of abdominal aortic calcification detection.

Aortic calcification has been diagnosed by radiography in many studies (Walsh et al. 2002; Hollander et al. 2003; Jayalath et al. 2005; Ganz et al. 2012). This is because the calcified lesions have high density and are usually visible on standard radiographs of the thorax and abdomen (Wilson et al. 2001; O'Donnell et al. 2002).

Lateral lumbar radiographs were used to predict CHD, CVD, and CVD mortality in 1049 men and 1466 women in a study by Kauppila et al. (2001). Jayalath et al. (2005) presented a comprehensive review of aortic calcification, the techniques used for measurement, the associated risk factors and the outcomes for these with AAC. In another published work by Honkanen et al. (2008), severe AAC was recorded using lateral lumbar radiographs; patients with stage 5 chronic kidney disease (CKD) were

included. The method that uses for AAC assessment was recommended as a part of pre-transplant work-up and cardiovascular risk factor determination. Radiography is widely used but it is less sensitive in detecting atherosclerotic lesions than new methods such as US and CT (Schousboe et al. 2017).

2.5.4 **Dual Energy X-ray Absorptiometry**

DXA is a standard diagnostic technique widely used to detect osteoporosis and vertebral fractures. VFA has become a valuable method that is used to identify patients with osteoporosis because of the presence of vertebral fractures (Golestani et al. 2010; Schousboe et al. 2006; Schousboe et al. 2007; Cecelja et al. 2013; Grant et al. 2017; Schousboe et al. 2017). Prevalent vertebral fractures and predict future fractures independently of other risk factors such as age and BMD (Wilson 2006).

VFA has the potential to detect AAC with low radiation exposure and is a useful alternative to CT in this regard AAC (Cecelja et al. 2013; Elmasri et al. 2016; Lewis et al. 2016; Grant et al. 2017). For example, the modality was also employed for the detection of vascular calcification disease in dialysis patients in a study by Toussaint et al. (2009). Screening of vascular calcification using VFA images is similar to radiography and may provide a dual advantage in CKD because it can measure BMD and simultaneously detect AAC with good sensitivity and specificity.

VFA images have several advantages over lateral spine radiographs. These include simplicity, speed of image acquisition and low-cost. Another advantage is the low radiation exposure from VFA compared with radiographs (Schousboe et al. 2010; Szulc et al. 2013).

In addition, VFA imaging uses a fan-beam to scan the spine, whereas radiography is performed using a cone-beam, which produces parallax distortion within vertebrae located above or below the central point of the beam (Schousboe et al. 2010). VFA images have low resolution compared with radiographs; these images may fail to identify

small amounts of AAC that may be visible on radiographs. At the same time, both VFA images and radiographs may miss smaller calcifications that are detected by CT (Schousboe et al. 2017).

2.6 Quantification of Abdominal Aortic Calcification

There are different manual methods for AAC measurement depending on the modalities utilised for this task. A brief review of the methods associated with each modality described in Section 2.5 is presented here.

AAC has been measured using CT in several studies (Ichii et al. 2013; Cecelja et al. 2013; Leckstroem et al. 2014; Lee et al. 2017). Calcification measurements on CT scans are often performed using the Agatston score method. The Agatston score is defined by the area of a calcified plaque multiplied by a scaling cofactor that represents peak attenuation of this calcification (Hong and Pilgram 2003; Lee et al. 2017). A constant threshold of attenuation 130 Hounsfield units (HU) is used to distinguish between non-calcified and calcified areas.

AAC and BMD were assessed in a study by Chow et al. (2008) to define the association between vascular calcification and bone mass and structure. The Agatston scoring method was used to quantify AAC for 693 subjects. The study also showed that osteoporosis and atherosclerosis are age-dependent.

In another study, females aged 52-80 years from the UK twins registry years underwent non-contrast CT and lateral DXA imaging of the abdominal aorta at lumbar vertebrae L1–L4 (Cecelja et al. 2013). Every cross-sectional slice in the CT image was analysed separately, and calcification in the aorta was expressed as the area that in this region > mm^2 with attenuation ≥130 HU. Calcification was measured in the unit of mm^3 using the volume score calculated as the product of voxel volume and the number of voxels for

each cross-sectional slice; by summing the score for all cross-sectional images, the total volume score was obtained.

Leckstroem et al. (2014) measured AAC on CT scans of 93 living kidney donors. AAC was derived using a manual scoring system to calculate total aortic calcium severity score. Each CT slice was scored individually for circumference calcified between the coeliac axis and the aortic bifurcation calcification was classified as mild, moderate or severe, where mild AAC extended through 45^0 of the circumference, moderate 46^0–180^0 and severe 180^0. Chuang et al. (2014) also determined AAC by a modified Agatston score in 100 Framingham Heart Study participants using CT. They found that AAC was common in middle aged and older adults and that AAC distribution was not uniform in the aorta.

In the Framingham Heart Study, abdominal aortic wall calcification seen on standard lateral spine radiograph was measured using what has become the most widely accepted semi-quantitative technique (Kauppila et al. 1997; Wilson et al. 2001; Schousboe et al. 2008). This visual scoring system quantified calcific deposits in the region of the aorta at the level of L1 through L4 and is known as the AC-24 method.

To date, the AC-24 method has been widely used to assess AAC (Ganz et al. 2012; Mohammad et al. 2014; Lewis et al. 2016; Grant et al. 2017; Schousboe et al. 2017). In the AC-24 system, calcified deposits are considered to be present if calcification is visible in an area parallel to the lumbar spine and anterior to the lower part of the spine. Calcification of the anterior and posterior aortic walls is assessed separately; and graded as follows: 0, no calcification; 1, only small scattered calcified deposits occupying one third or less of the longitudinal wall of the aorta; 2, one thirds but less than two thirds of the longitudinal aortic wall calcified; 3, two thirds or more of the longitudinal wall of the aorta calcified. Finally, the AAC scores in the anterior and posterior walls are summed to give an overall severity score in the range 0-24 (Wilson et al. 2001).

AAC may also be quantified using the AC-8 system, which is a simplification of AC-24 method. The AC-8 system is described as follows. For each aortic wall, the total length of calcification is scored between 0 and 4 relative to vertebral body height. A segment with no calcification is scored as 0 if the aggregate length of calcification is one vertebral body height or less, the score is 1. The score is 2 when the aggregate length of calcification is greater than one and less than or equal to 2 vertebral heights, and so on to the maximum score of 4. Scores for each wall are summed individually to give the total AC-8 score, which ranges between 0 and 8 (Bolland et al. 2010).

The size, shape and the distribution of AAC were investigated by Ganz et al. (2012) for 308 postmenopausal women. In this study, the number of calcified deposits was quantified as well as the percentage of the abdominal aorta occupied by the lesions based on their area, thickness, wall coverage and length. AC-24 was used for AAC quantification and three radiologists annotated the same images manually. Statistical agreement tests were conducted in this study. Schousboe et al. (2006) used the AC-24 system to quantify AAC manually on 57 VFA images of women who had participated in a previous study for the detection of vertebral fracture.

In a study by Lewis et al. (2016), VFA images of 892 elderly women aged > 70 years were used to investigate whether AAC measures were related to atherosclerosis. AAC scores were measured using AC-24. VFA images of 2,500 subjects were evaluated for AAC detection and quantification by Golestani et al. (2010) using AC-8. VFA assessment was validated against radiography in 53 subjects. Subjects with identified AAC were classified into two groups: low score (1-3) and high score (>3).

The severity of calcification in eight anterior and posterior aortic segments was assessed by the AC-24 system in a study conducted by Szulc et al. (2013). Here, AAC also was categorised into three groups: score 0, score 1-2 and score 3-6.

In the work of Mohammad et al. (2014), two radiologists scored AAC on VFA images using the AC-24 scoring system described above. In this study, the severity of AAC was classified categorically. Three categories of calcification scores were described: mild (1–4), moderate (5–12), and severe (>12).

Recently, Schousboe et al. (2017) reported that the severity of AAC on radiographs and VFA images should be categorised as follows: low (AC-24 =1or 2, AC-8 =1); moderate (AC-24 =3 to 5, AC-8=2 or 3); and severe (AC-24 6 or more; A-C8 =4 or more). A limitation of this scoring system is the time and cost of assessing the lateral spine images.

Several other recent studies (Lewis et al. 2016; Grant et al. 2017) have used VFA images to measure AAC. Grant et al. (2017) scored AAC using a modified Framingham score in 100 patients who had osteoporosis examinations.

The accuracy and variability of calcification measurements are important considerations in clinical and research applications of AAC scoring. With regards to AC-24 scores, radiographs recorded a very high inter-observer and intra-observer reliability, with intra-class correlation coefficients (ICC) consistently reported as 0.90 or higher (Kauppila et al. 1997). In another study, the intra-observer intra-class correlation coefficient was 0.93 and inter-observer intra-class correlation coefficient was 0.90 (Pariente-rodrigo et al. 2016). Several studies have shown that AC-24 and AC-8 can discriminate between those with and those without a high level of AAC (AAC-24 scale score ≥5) on radiographs with high accuracy (Bolland et al. 2010; Mohammad et al. 2014; Schousboe et al. 2017).

2.7 **Summary**

This chapter has provided background information on abdominal aortic calcification (AAC) and its correlation with atherosclerosis, as well as information on medical imaging techniques used for AAC detection. Evidence has been presented that VFA images

acquired as a part of osteoporosis diagnosis by DXA are a suitable alternative to CT scans in the detection of AAC due to their simplicity, low cost and low radiation exposure has been presented.

Available manual scoring systems used as assessment tools for measuring AAC by different modalities have been reviewed. These include the AC-24 points system that has been employed in many studies on radiographic images. It has been shown that the same system can be used for AAC measurement in VFA images with high accuracy and reliability.

The main findings of the literature review are as follows:

1. Atherosclerosis is a common cause of CVD, such as aortic aneurysm, peripheral vascular disease, heart attack and stroke. Studies related to CVD have proved that the presence of calcification in the aorta is correlated with the presence of aortic atherosclerosis.

2. Only few prospective studies have examined how well the detection of AAC on lateral spine images can predict incident CVD. However, most of them have shown that the presence and severity of AAC is correlated to presence of CVD.

3. DXA is a standard diagnostic technique widely used to detect osteoporosis; if it has VFA capability, it may also be used to identify vertebral fracture. VFA also represents a new method for the detection of abdominal aortic calcification with low radiation exposure and cost. DXA with VFA can provide two diagnostic examinations at the same time, namely atherosclerosis and osteoporosis.

4. Although the presence of calcification in the abdominal aorta is commonly noted incidentally in VFA images obtained using DXA, normally it is not included in the DXA report. At the same time, there is potential that identifying AAC during osteoporosis screening using VFA images can help in diagnosis of CVD.

5. Further studies are needed to determine whether AAC identified in VFA images improves the prediction of CVD risk. In addition, an investigation of whether AAC detection with VFA images contributes to treatment or alters healthcare provision is also needed.

6. The manual AC-24 points system is the only current accepted method for AAC scoring on both radiographs and VFA images. The limitation of the current assessment of AAC is its time and cost.

7. Several studies have proposed points scales to classify AAC into mild, moderate and severe. Recently, an AC-24 score greater than 6 has been reported as an indication of severe AAC.

The next chapter surveys medical image processing, automatic classification techniques and reviews available automatic methods for AAC assessment.

Medical Image Processing and Machine Learning

3.1 Introduction

This chapter reviews popular medical image processing techniques relevant to the task at hand; relevant machine learning methods are also discussed in this Chapter. This Chapter also surveys previously proposed automatic methods for AAC assessment using a variety of imaging modalities.

3.2 Medical Image Processing

Medical images are very important for the diagnosis and treatment of wide range of medical conditions. Medical image processing and analysis is helpful in transforming raw images into a quantifiable images suitable for searching and mining, and extracting from them significant quantitative information to aid medical diagnosis (Norouzi, 2014).

A variety of imaging modalities have been developed over the last several decades. Examples include CT, US, Magnetic Resonance Imaging (MRI), other tomographic modalities such as single photon emission computed tomography (SPECT) and positron emission tomography (PET). Images obtained by different medical imaging modalities are subject to degradations and/or deformations during the process of acquisition. Consequently, to extract information for quantification, at least some preliminary image processing is needed.

3.2.1 **Image Enhancement Techniques**

Image enhancement is an important step in medical image processing; its purpose is to improve the visual quality of images for human viewing or to provide better images for the steps involved in automated processing. These could include image segmentation, object detection and feature extraction (Maini and Aggarwal 2010; Bedi and Khandelwal 2013; Sasi and Jayasree 2013). Medical and other images may suffer from poor contrast, limited resolution and high noise. Removing blurring, noise and increasing contrast, are examples of enhancement operations that are done to reveal hidden detail.

Image enhancement does not increase the information content of the data, but it highlights the main features of interest in order to detect objects in a simple and efficient way (Vasuki and Devi 2017). Available techniques of image enhancement include spatial domain methods and frequency domain methods. The choice of technique is highly dependent on the imaging modality, image processing task and viewing conditions.

Spatial domain methods work on the image pixels directly; they comprise point arithmetic operations and neighbourhood enhancement algorithms. Point arithmetic operations include grey-level transformation and histogram equalisation (Vasuki and Devi 2017). For example, the contrast enhancement operation, often known as windowing, utilises a contrast stretching technique to map the pixel values of an original image into new values, which have increased contrast (Gonzalez and Woods 2010). Histogram equalisation, (HE) is one of the popular and widely used contrast enhancement techniques. HE changes the spatial histogram of an image into an approximately uniform distribution. It cannot highlight image edges and cannot be utilised for detection applications (Kumar and Rana 2016; Vasuki and Devi 2017). Neighbourhood enhancement algorithms include image smoothing and image sharpening. Image smoothing is used to remove noise; a good image smoothing algorithm should be able to deal with different types of noise. The mean filter, Wiener filter, Gaussian filter and

median filter are used for effective noise removal (Vasuki and Devi 2017). A comprehensive review of image smoothing techniques has been presented in many studies (Goyal et al. 2012; Kaur and Singh 2015; Vasuki and Devi 2017).

In frequency domain methods, images transformed into the frequency domain using low-pass filtering, high-pass filtering and homomorphic filtering techniques. Fourier transformation and Gaussian low pass are two common low pass filtering techniques (Makandar 2015).

High pass filtering allows high frequencies to pass but eliminates frequencies lower than the cut-off frequency. Sharpening is a high pass operation in the frequency domain; it uses filters such as Butterworth and Gaussian. These methods are time consuming, dependent on cut-off frequency and also need expert knowledge to select the right process (Wang and Tan 2011). In addition, they often cannot meet medical image processing task requirements because they process the entire image. Thus, they may lose both partial and specific information and also may be easily affected by noise.

Morphological image processing is an image processing technique based on shape and structure form (Sreedhar 2012). Morphological operations are important for filtering, segmentation and pattern recognition. The technique is a powerful non-linear tool for capturing image components; it is helpful in the representation and description of objects regions, such as boundaries, skeletons and convex hulls (Ciaccio 2011). Techniques such as erosion, dilation, opening and closing are used to perform morphological processing (Sreedhar 2012). Morphological operations are usually necessary to post-process the segmentation results of threshold methods, because the latter use only the intensity histogram and ignore the spatial information (Fasihi and Mikhael 2016).

A more comprehensive review of image quality enhancement techniques has been given by Gonzalez et al. (2010).

3.2.2 **Image Segmentation**

Image segmentation is a process of partitioning an image into semantically meaningful sub-regions. In medical imaging this can be achieved by identifying a surface for each tissue class individually, or by classifying every pixel in the image (Seerha 2013). Medical image segmentation is an essential stage to model different tissues and is an effective tool for clinical diagnosis, staging, screening, treatment planning and evaluating response to therapy.

In general, image segmentation can be divided into three types based on the degree of human interaction: manual segmentation, semi-automatic segmentation and fully automatic segmentation (Chitradevi and Sadasivam 2016; Fasihi and Mikhael 2016). Image objects can be divided into sub-regions manually, but it is time-consuming work, which has significant inter-observer and intra-observer variability; hence it is not reproducible.

In semi-automatic segmentation methods, minimum human involvement is required at some stages e.g. to initialise the method or to correct segmentation results manually (Fasihi and Mikhael 2016). In fully automatic methods, prior human knowledge such as location, size or shape is incorporated in the algorithms to produce robust techniques (Fasihi and Mikhael 2016).

Automatic segmentation methods are classified into three types based on the learning method that is used: supervised, semi-supervised and unsupervised segmentation. Supervised segmentation needs user interaction throughout the segmentation process whereas unsupervised methods generally require operator participation only after the segmentation stage is completed. In supervised segmentation, labelling the entire training data requires expert knowledge and is also time-consuming. In semi-supervised algorithms, a small amount of the data is labelled correctly and used for training, while

the rest is unlabelled (Fasihi and Mikhael 2016). Unsupervised methods do not depend on a labelled training set at all.

Alternatively, segmentation may be classified into classical and non-classical techniques (Bhattachayya et al. 2016; Chitradevi and Sadasivam 2016). Most of the classical segmentation approaches depend on filtering and statistical techniques. Classical methods engage thresholding techniques, edge detection or boundary-based techniques, region-based techniques, morphological techniques or k-means approaches. Non-classical techniques are represented in the fuzzy-neuro-genetic paradigm. Genetic algorithms are randomised search and optimization techniques guided by the principles of evolution and natural genetics; they work on the collection of probable solutions in parallel rather than the domain dependent knowledge. These techniques are ideal for those problems for which there is no knowledge about their domain theories or for problems that are difficult to formulate. Near optimal solutions with an objective or fitness function are provided by the genetic algorithms (Balla-Arabé et al. 2013; Bhattachayya and Chakraborty 2016). There are three properties that can influence the result of segmentation of an object in an image: object boundaries, object homogeneity and object shape. Image boundaries and homogeneity are dependent on the modality used. The most challenging problem is to segment images with insufficient object boundaries, such as edges, and/or lack of texture contrast between regions of interest (ROIs) and background (Huang et al. 2013).

There are various medical image segmentation methods; however, none of them works perfectly for all imaging modalities. Moreover, for each imaging modality, such as MRI or CT, there is no optimum method for segmentation. In addition, a suitable image segmentation method is dependent on the goal of the segmentation task.

In the rest of this Section, the most popular segmentation methods are discussed. These methods can be classified into four types: 1) Threshold based methods; 2) Region based

methods; 3) Model-based methods; 4) Clustering methods (Fasihi and Mikhael 2016; Chitradevi and Sadasivam 2016).

3.2.2.1 **Threshold based methods**

Thresholding is an important technique for image segmentation and pattern recognition and it is a basic component of many computer vision applications. The threshold is a value in a histogram that divides pixel intensities into two parts: the first part is the “foreground” that has pixels with intensity values greater than or equal to the threshold and the second part is the “background” that has pixel intensities less than the threshold (Norouzi 2014; Kaur and Goyal 2015).

Thresholding can be implemented locally or globally; for the global approach, the target objects are extracted from the background on the basis of the grey-level histogram by creating a binary partition of the image intensities (Seerha 2013; Norouzi 2014; Kaur and Goyal 2015). In global thresholding, a single threshold is chosen for the entire image. These methods cannot be used in some types of image such as those that do not have a constant background and have variation across the object (Kaur and Goyal 2015; Chitradevi and Sadasivam 2016).

The local thresholding method, that divides an image into sub images, can be used in order to define different threshold values for different parts of the image, after which multilevel thresholding can be applied (Norouzi 2014; Fasihi and Mikhael 2016; Chitradevi and Sadasivam 2016). Normally, local thresholding requires more time to segment an image compared to global thresholding.

In order to select the threshold value for each sub image, different statistical values are used such as mean, standard deviation, mean and standard deviation together, and mean of maximum and minimum (Norouzi 2014).

A popular method proposed by Otsu (1979) is one of the best thresholding methods for real-world images as regards uniformity and shape measures (Horng 2010). Otsu's thresholding method aims to find an optimal value for global thresholding (Norouzi 2014; Fasihi and Mikhael 2016). It is assumed that an image has two pixel classes or has a bi-modal histogram. Otsu's method has been widely used in medical imaging. For example, Wang et al. (2011) used Otsu threshold recognition combined with local entropy to extract prostatic lumina from histology images based on texture features of prostatic calculus. A multi-level Otsu method was proposed by Liao et al. (2001) to improve computation efficiency.

In general, threshold-based segmentation techniques are unable to segment most medical images properly. However, these techniques are used as the first step in a segmentation process (Kaur and Goyal 2015; Chitradevi and Sadasivam 2016).

At the same time, thresholding is a suitable choice for the segmentation of images that include objects with homogeneous regions, or when contrast between the objects and the background is high.

There are some drawbacks to global thresholding methods, such as their sensitivity to noise and intensity, the non-use of spatial information in the image and the fact that they cannot be applied in cases of irregular background illumination (Fasihi and Mikhael 2016).

These methods do not provide good results in MRI without pre-processing but are effective for CT images, which have less noise (Norouzi 2014). MRI images have many different parts, which make these methods non-applicable. Thus, loss of information from the image and misdiagnosis occur and this may mislead physicians in their clinical task.

3.2.2.2 **Region based methods**

Region based methods partition the image into regions with no overlap between two different regions. The most common region-based method is the region growing segmentation algorithm. This algorithm is the one of most popular region based techniques used for brain tumour segmentation (Fasihi and Mikhael 2016). A detailed summary of region based segmentation approaches is presented in a survey by Chitradevi and Sadasivam (2016).

Region growing is an approach to image segmentation based on the criterion of intensity information, in which neighbouring pixels are tested and connected to a region class if no edges are detected (Kaur and Goyal 2015; Fasihi and Mikhael 2016; Chitradevi and Sadasivam 2016). In region growing, the input image is adjusted into a set of homogeneous simple regions. Then, similar neighbouring regions are incorporated based on a certain decision rule and using an iterative merging process. At the beginning, the entire image is regarded as one rectangular region (Bhattachayya and Chakraborty 2016).

Region growing is used usually in combination with other segmentation techniques to delineate small, simple structures such as tumours and lesions in mammograms (Norouzi 2014). This method has been widely used in mammography in order to extract the potential lesion from its background (Norouzi 2014; Fasihi and Mikhael 2016).

The disadvantage of region growing is that it depends on the selection of a seed point, which requires human interaction (Nygren and Jasinski 2016). Seed point selection depends on human ability; thus, the extracted shape relies on the user to a considerable extent. Furthermore, the region may leak into nearby tissues with poor image contrast or when the boundary information is not good enough (Fasihi and Mikhael 2016).

3.2.2.3 **Model-based segmentation methods**

Normal human vision is able to detect objects in images whether or not they are completely represented. For this, specific knowledge about the geometric shape of the objects is required instead of local information. These types of model are also known as deformable models and they can be parametric or geometric (Fasihi and Mikhael 2016; Chitradevi and Sadasivam 2016). The regions need not be homogeneous because the methods use domain-specific prior information. Anatomic structures in an image can be segmented by building a connected and continuous model, which considers prior knowledge about the shape, location, size and direction of these structures.

Parametric deformable models, also known as snakes, include active contours (2D) and active surfaces (3D). The snake is a common initial active contour, used for supervised boundary identification in 2D images.

Active contour models or snakes have been widely applied in the segmentation of medical images to avoid the problem of intensity variation (Withey and Koles 2008). Snakes represent the objects as a set of landmarks. The initialisation step plays an important role in the use of snakes. In applications where the boundaries of objects are very close to each other, the starting position of the model should be placed close enough to the desired outlines to prevent converging to incorrect boundaries (Fasihi and Mikhael 2016).

Model based methods have become common because they are potentially able to solve problems associated with the complexity of structures and to improve results for noisy data. The most common approaches are active shape models (ASM) and active appearance models (AAM). These models have been proven in recent years to be very useful for medical image segmentation.

3.2.2.3.1 Shape Models

The statistical shape model was inspired by active contours. This model took existing shape information into account when deforming the contours used in the snakes algorithm (Cootes 2000). In medical image segmentation, prior shape information obtained from segmented objects is often used to build a shape model for various structures. The shape model is used to extract shapes from images. The ASM relies on the fact that searching is based on prior knowledge of the target object shape and position. This is an important characteristic of this algorithm as it allows the user to choose the images and to carefully place the “landmarks” for the creation of a model. Expert knowledge of the subject can be used in such tasks.

In medical imaging, shape models are mostly used for segmentation and recognition tasks (Melinska et al. 2015). A comprehensive review of shape models applied to medical image segmentation is presented in an article by Heimann and Meinzer (2009).

A step in the shape model searching process is the manual identification of landmark points in a set of training images, where the same landmark refers to the same anatomical location on each training image.

In respect of applying ASM for detection of the aorta or nearby vertebral structures, a number of studies have been published. Smyth et al. (1999) implemented an ASM to measure vertebral shape with high accuracy in VFA images obtained from DXA. The ASM was applied on full spine DXA scans obtained in a large cohort of patients with vertebral fractures. The authors reported the potential for the ASM to be used to detect fractures.

Bruijne et al. (2004) proposed semi-automatic delineation of abdominal aortic aneurysms using shape model fitting in sequential slices acquired by CT angiography. The shape model was used by Bruijne (2005) to segment the aorta on lateral spine images from radiographs. The abdominal aorta is made of soft tissue and cannot be detected from

background soft tissue on lateral spine images, and so segmentation was achieved by utilising its relative position to the four lumbar spine vertebrae (L1-L4).

A statistical model of shape also has been also used for cervical vertebra detection. The segmentation process included the following steps: (1) learning, in which landmarks were placed on the images to describe the vertebrae; (2) model design, in which a model was created after all the marked shapes had been aligned; (3) initialisation, in which the mean shape was placed on the area of interest and (4) segmentation, in which each point of the mean shape evolved in order to fit the edges of the vertebra (Benjelloun et al. 2011).

Shape models have also reported to be powerful algorithms for the prediction of bone strength in vivo, because they include information about bone shape and bone tissue material properties (Sarkalkan et al. 2014). Melinska et al. (2015) proposed a method to model the calcaneus by using the shape model.

The main advantage of ASMs is that the models are specific to the type of object under study, they describe only the variation observed in the training set and do not allow for other variation (Smyth et al. 1999).

In this study, the aim is to segment the calcified aorta in VFA images. The segmentation of the aorta is not easy because if the aorta is not calcified it cannot be seen of the x-ray images. Prior knowledge of the aorta location relative to the lumbar spine can be used to complete this task. Prior knowledge about the expected shape of the four lumbar vertebrae close to aorta can be very useful to detect the position of the calcified aorta in the VFA images.

3.2.2.3.2 Appearance Models

ASM has been extended to include local texture information when training and searching (Cootes and Taylor 2001; Cootes 2004). The active appearance model (AAM) algorithm attempts to find the model parameters, which generate a synthetic image similar to the target image. In each case, the best fitting model parameters can be used for another processing stage, such as measurement or classification (Cootes 2004). The advantage of AAM over ASM is that one can build a model with only a small number of landmarks (Cootes and Taylor 2001).

3.2.2.4 Clustering

Clustering is a process of arranging objects into groups based on the similarities of their features. The concept of similarity is used when dissimilar objects are grouped into different clusters. Similarity may be determined using measurement methods such as the Euclidean distance. Among the basic unsupervised clustering methods are Fuzzy C-means (FCM), k −means and Markov Random Fields (MRF) (Chitradevi and Sadasivam 2016). These algorithms are known as unsupervised learning methods because they do not need training data and they consume less time (Norouzi 2014; Chitradevi and Sadasivam 2016). Several FCM algorithms have been used for MRI segmentation.

In the k −means clustering algorithm, the data set is classified into a number k of clusters. Simplicity and low computation cost are the main advantages of k −means algorithm. However, their drawback is that the user has to select the k value in order to segment the image and different k values give different results (Norouzi 2014; Chitradevi and Sadasivam 2016).

In the Markov Random Fields (MRF) clustering method, spatial information is included in the clustering process and so clusters overlap and the effect of noise is reduced. MRF methods have been used to segment CT images of tumours (Fasihi and Mikhael 2016).

Supervised clustering methods include artificial neural network (ANN) and Bayes methods. ANN approaches are known as non-parametric techniques, because no parametric distribution (such as Gaussian) is proposed for the data (Chitradevi and Sadasivam 2016). The original concept behind the neural network was inspired by the mechanism of pattern recognition in the brain. A neural network can estimate the correlation between class labels and features, and it can also deal with multi-class data, where the data can be given more than two classes or categories.

An example of the application of neural networks in medical image processing is the Computer-Aided Diagnosis of solid breast nodules. Such an algorithm can identify breast nodule malignancy using multiple US features and an ANN classifier; the authors reported a sensitivity of 99% (Joo et al. 2004).

3.3 **Statistical Machine Learning**

Statistical machine learning involves data analysis and provides a framework, that allows researchers to decode what computed characteristics reveal, and how these characteristics could be used as a decision support tool. The characteristics are qualitative and they can be assigned to an ordinal scale. For example, AAC could be categorised as follows: (a) mild, (b) moderate and (c) severe. In this case, AAC severity could define a scale that would read: 1= mild, 2 = moderate, and 3 = severe.

Generally, data analysis starts by exploring the statistical characteristics of the data and producing plots to get the general impression of the spread of values. Initially, the probability densities of the features can be plotted. The simplest histograms are those that provide an overview of the distribution of values for each feature and for the response variable, which is equivalent to the diagnosis. The response variable can be divided into categories or classes. When the response variable takes any number of finite classes, the problem of predicting it is known as classification. When the response

variable takes any real value (any possible number from $-\infty\ to\ \infty$, the problem is known as regression (Tsanas et al. 2013).

In medical applications, classification is more frequently used, and hence this approach is used exclusively in this research.

3.3.1 **Classification**

Classification is important in data analysis, pattern recognition and machine learning. It provides intelligent decision making (Norouzi 2014). Classification is a process in which a model or classifier is constructed to predict class labels.

The aim of classification is to predict accurately the target class for each case in the data (Kesavaraj and Sukumaran 2013). The simplest type of classification problem is binary classification, where the target attribute has only two possible values. Multiclass targets have more than two values. To use the supervised classification technique, a subset of observations and measurements for which the target value is already known should be defined as training data. Training data include both the input and desired results.

Kotsiantis (2007) has described the classification process as a series of steps. The first step is collecting a dataset with known labels. In medical imaging, image labelling is often visually obvious and can be done through examination by an experienced operator (Pham et al. 1998). The second step is data preparation and data pre-processing. Pre-processing algorithms can be employed for issues such as missing values, discretisation and noise removal.

The third step is feature selection to identify and remove irrelevant and redundant features. This makes the algorithms work faster and also can improve their performance. The critical step is learning algorithm selection and how to choose a specific one. The final step is evaluation; in this context, there are at least three techniques that can be used to calculate classifier accuracy, all of them based on splitting the data into training

and testing sets (Mudry and Tjellström 2011). The difference between these techniques lies in how the division into these two data sets is made. Cross-validation and leave-one-out are examples of validation procedures.

The basic form of cross validation is K-fold cross-validation. Other forms are special cases of K-fold cross-validation or involve repeated cycles of K-fold cross-validation (Mudry and Tjellström 2011). In K-fold cross-validation, the data are divided into K equally or approximately equally sized folds. Subsequently, K iterations of training and validation are performed such that within each iteration a different fold of the data is set a side for validation, while the remaining $K - 1$ folds are used for learning.

Leave-one-out cross-validation (LOOCV) is a special case of K-fold cross-validation where K equals the number of observations in the data. Therefore, for each iteration all the data except for a single observation are used for training and the model is tested on that single observation. If the accuracy is unsatisfactory, the previous steps should be examined to establish the reasons. Poor accuracy can be due to using irrelevant features, imbalanced data, an inappropriate learning algorithm or high dimensionality.

Traditionally, supervised machine learning classifiers can be compared by performing a statistical examination of the accuracies of trained classifiers on specific data sets. The paired t-test is used to check the null hypothesis that the mean difference between the classifiers is zero (Kotsiantis 2007). Two types of errors can be created. Type I error is the probability that the test rejects the null hypothesis incorrectly. Type II error is the probability that the null hypothesis is not rejected, whereas actually there is a difference. Ideally, the test results should be independent of particular partitioning. This can be achieved by randomising the process of classification and repeating the process several times (Kotsiantis 2007).

Choosing the best algorithm for classification can be done by relying on previously published studies in particular medical imaging applications. In this context, some comparative classification studies are reviewed in this section.

Fernández-Delgado et al. (2014) evaluated 179 classifiers using 121 data sets over the UCI data base (Lichman 2013). The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms.

Random forest (RF) was reported to be the best classifier with a maximum accuracy 94.1%; the accuracy was greater than 90% in 84.3% of the data sets. The second best was support vector machine (SVM) with a Gaussian kernel implemented in C using LibSVM, which achieved 92.3% of the maximum accuracy. The authors also reported that there was no significant statistical difference between RF and SVM classifier results.

K-nearest neighbour (K-nn) is a popular parametric classifier method. The disadvantage of the SVM and K-nn methods is that they do not take spatial information into consideration. Singh (2016) has published a recent review of supervised machine learning algorithms in terms of the accuracy, speed of learning, complexity and risk of over fitting measures.

3.3.2 **Accuracy of Classification**

In general, the accuracy of a diagnostic test indicates the ability of the test to distinguish between patients with disease and those without (Leeflang et al. 2018).

Typically, the results of the test are categorised as positive or negative for the target condition. The terms positive and negative originate in early medical applications, where patients with some observed medical phenomenon (e.g. an illness) were denoted as positive, and the rest of the patients were described as negative. If a patient is positive for illness and is classified as positive by the test, this denotes a true positive (TP). If the

classifier prediction is negative, this is called a false negative (FN). Similarly, if the patient is negative for the illness and the classifier prediction is positive, this denotes a false positive (FP), while a negative prediction is called a true negative (TN).

In addition, sensitivity and specificity are also used to assess a medical diagnostic test. Sensitivity indicates the proportion of true positives correctly identified. In contrast, specificity refers to the proportion of true negatives correctly identified (Orrù et al. 2012; Harefa and Pratiwi 2016). Accuracy is the sum of TP and TN expressed a proportion of the total number of cases (Leeflang et al. 2018). Accuracy, sensitivity and specificity can be defined as follows (Majnik and Bosni 2013):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \qquad (3.1)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3.2)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (3.3)$$

3.4 Automatic Detection and Quantification of AAC

With regard to detecting and quantifying AAC automatically, only a few methods have been developed in the past. Bruijne (2005) proposed the first method for automatic detection and quantification of calcified plaques from radiographs of the lumbar aorta. In this study, medical experts outlined the calcification manually on a set of images of the part of the aorta adjacent to the lumbar vertebrae (L1-L4). The vertebrae were annotated

by placing 6 points on each one as is done routinely in vertebral morphology studies. The linear point distribution models (PDM) proposed by Cootes et al. (1995) were used to model object shape variations observed in the training set. To detect calcification, a pixel classifier was trained to differentiate between calcified pixels and the background. Pixels used for training were selected randomly from a region of interest including the aorta and its surroundings. Pixel classification into calcified and non-calcified was based on one threshold value. Nonetheless, the author reported an accuracy of 93.8%.

In another published work, the severity of atherosclerotic plaques on standard radiographs was estimated by Conrad-Hansen et al. (2007). Plaque severity was measured by subtracting the inpainting from the original image. The limitation of this work was the lack of accuracy in manual segmentation, which influenced the inpainting. The approach was based on comparing the observed image intensity to the image intensity that would be expected if the aorta were uncalcified. First, all calcified areas in the L1-L4 region were segmented. Images were annotated manually by radiologists. Thus, the non-calcified aorta appearance was reconstructed by interpolating the background image around the calcification using inpainting techniques. Plaque density was then estimated by subtracting this inpainting from the original x-ray.

Lillemark (2010) proposed the first method to describe the growth pattern of AAC over time, based on the assumption that growth pattern was a good estimator of the progression of atherosclerosis on radiographs. Follow-up of radiographic images were registered to the baseline images with thin plate spline (TPS) registration. TPS was chosen due to its ability for aligning the deformable aortas globally from a few landmarks. Landmarks on the intersections between the anterior and posterior aorta walls and the inter vertebral lines were used for the registration. Area, height, width, centre of mass position, movement of the centre of mass and upper and lower boundaries of matched AACs were measured on images obtained at two different time instances. AAC growth

was measured from individually matched AACs with area overlap used as the matching criterion.

A new morphological atherosclerotic calcification distribution (MACD) index was proposed by Barascuk et al. (2011). MACD was quantified from outlines of the calcified aorta on radiographs of the lumbar region. It was defined as the number of distinct calcified deposits (NCD) multiplied by the morphological atherosclerotic distribution (MAD) factor. The MAD factor describes the growth potential of the calcified plaques and was calculated by estimating the plaque area as visualised on the radiographs. The MACD factor was compared with scores derived using the AC-24 system. The study concluded that a refined index of aortic calcification, MACD, may provide additional and important information on risk related to CVD mortality compared to the traditional AC-24 index. This study was limited by a small sample size of 20 radiographs and the results were valid for a follow-up period of 8.5 years.

Recently, an analytical approach was developed for AAC quantification in VFA images (Grant et al. 2017). The study assessed a quantitative index to characterise AAC severity (QAAC). The technique relies on the relationship between plaque density and brightness on radiographs. AAC severity was quantified using an algorithm that measured pixel intensities in a defined ROI and that automatically exported all data points along its length as it moved horizontally across the observer specified ROI. The correlation between BMD and AAC measured by QAAC and a modified AC-24 score, which used only 12 points to score AAC, was determined in 110 patients. The study proved that there was a significant inverse relationship between bone density and aortic calcification measured by the QAAC technique. However, AAC was assessed only on segments of the aorta adjacent to L3-L4 and it was time-consuming to measure AAC severity in comparison with the Framingham technique.

3.5 **Summary**

This chapter has provided a review of image processing methods for segmentation and data analysis in medical applications. The chapter also surveyed available automatic methods for AAC measurements.

The main findings of the literature review were as follows:

1. There is a lack of studies describing automatic methods for AAC detection and quantification. Only one automatic method has been implemented in the past, but this was validated on radiographs. Furthermore, this method was not related to the accepted manual AAC measurement methods and thus it is unlikely to be adopted in the medical field. None of the proposed methods were tested in VFA images.

2. A new automatic method is required for AAC detection and quantification in VFA images. The new automatic quantification method needs to be related to the accepted manual assessment methods to ensure its possible future adoption by the medical community.

The next chapter describes dual energy x-ray absorptiometry, the modality used in this study to acquire VFA images.

Dual Energy X-ray Absorptiometry

4.1 Introduction

This chapter introduces the DXA modality utilised in this thesis for the detection of AAC. As outlined in Chapter 2, DXA is a standard diagnostic technique that is widely used to diagnose osteoporosis. The addition of VFA allows the detection of vertebral fractures. DXA is the gold standard for clinical trials designed to evaluate changes in patient BMD following therapeutic intervention. Osteoporosis is characterised by decreased BMD and poor bone quality resulting in an increased fracture risk. BMD is measured in g/cm^2 but commonly expressed as the T-score, the number of population SDs above or below the mean for healthy 30-year- old adults of the same sex and ethnicity as the patient (Oei et al. 2016). VFA has become a valuable and significantly used method to diagnose patients with osteoporosis by indicating the presence of vertebral fractures. The VFA method provides quick image acquisition at low cost and with low radiation exposure.

This chapter describes the physics of DXA and presents the characteristics of VFA images acquired by this modality. The properties of two models of Hologic DXA scanners (Discovery and Horizon) used in this study are presented.

4.2 Clinical Application of Dual Energy X-ray Absorptiometry

DXA is a two-dimensional non-invasive projection method that measures BMD in the lumbar spine, femur and forearm, which are the most common sites for osteoporotic fractures. Lumbar spine BMD measurements are usually made in the antero-posterior (AP) projection and may also be made in the lateral projection. The lateral spine image

allows BMD measurements in the vertebral bodies. These contain mainly trabecular bone without significant contribution from cortical bone, which is present in the posterior vertebral components (Finkelstein et al 1994).This technique is useful to diagnose osteoporosis and to evaluate the risk of fracture, so that therapeutic decisions can be taken and response to treatment assessed. In addition, whole body DXA imaging is a method for total and regional body composition measurement (Hologic 2003).

DXA was introduced in 1987 by Hologic and has become the gold standard for clinical BMD measurements (Wu et al. 2012). DXA is a popular technique for bone densitometry due to its high precision, low radiation dose, stable calibration and short scan times (Adams 1997). The World Health Organisation (WHO) has established DXA as a widely validated technique to measure BMD (World Health Organisation 2004). Two main manufacturers provide bone densitometers: Hologic and GE, previously Lunar.

For clinical interpretation, BMD is compared to the relevant reference population to calculate T and Z scores (Hricak 2013). The T-score compares the measured BMD with the mean BMD of a young (30-year-old) healthy population of the same gender. The WHO introduced the following classification system: if T-score at the spine, hip or forearm is less than -2.5, this indicates osteoporosis; if it is between -1.0 and -2.5, this indicates osteopenia and if it is greater than -1.0, this is regarded as normal. The Z-score compares measured BMD with that of an age and sex matched population and therefore accounts for the expected decline in BMD with age (World Health Organisation 2004).

The WHO has developed a Fracture Risk Assessment Tool (FRAX), which calculates the 10-year probability of a major osteoporotic fracture (Hricak 2013). This is an on-line tool that can be found at www.sheffield.ac.uk/FRAX. The risk factors incorporated in this software include age, sex, weight, height, use of glucocorticoid therapy, history of previous fracture, family history of hip fracture, smoking and alcohol consumption, and the presence of rheumatoid arthritis. Femoral neck and BMD may also be included if available.

DXA is based on two assumptions: first, the body is composed of only two components (soft tissue and bone mineral) and second, soft tissue overlying bone has the same composition as that adjacent to bone. Soft tissue includes muscle, fat, viscera, bone marrow, skin and the bone collagen matrix. Inhomogeneity of the fat distribution within the body will significantly affect DXA accuracy (Michael and Henderson 1998; El Maghraoui and Roux 2008). DXA scanners are usually calibrated assuming an abdominal thickness of 15 – 25 cm and errors may occur when scanning very thin patients (<10 cm) and the obese (>30 cm) (Cullum et al. 1989). Generally, the accuracy error of the most widely used bone densitometers is about 5 –10 % (Guglielmi et al. 2014).

Precision is a measure of the reproducibility of bone densitometry and is usually expressed as a coefficient of variation (CV). CV is the ratio of the standard deviation (SD) of a series of measurements to the mean value expressed as percentage. Ideally, DXA precision needs to be in the region of 1 %, and certainly better than 3 % in clinical use (Hricak 2013).

4.3 **Physics of Absorptiometry**

Three different photon interactions with tissue occur in the energy range that is relevant to bone densitometry: photoelectric effect, Raleigh scattering and Compton scattering (IAEA 2010).

The photoelectric effect converts the energy of a single photon into that of a single free electron and this type of interaction dominates at low energies. The photon is completely absorbed when it interacts with a bound electron and the electron emerges with kinetic energy equal to the photon energy $h\upsilon$ minus the electron binding energy. The probability of photoelectric interaction decreases rapidly as the photon energy increases. In general, the mass attenuation coefficient τ_m for photoelectric absorption varies roughly as $1/(h\nu)^3$ and with Z^3, where Z is the atomic number of the tissue (Hendee et al. 2003).

Rayleigh scattering occurs when the x-ray photon interacts with the whole atom. Photons are scattered in approximately the same direction as the incident photon with negligible loss of energy. This interaction occurs in media with high atomic number and is important in tissue only for low-energy photons. The importance of Rayleigh scattering is further reduced because little energy is deposited in the attenuating medium.

Compton scattering occurs when an incident x-ray photon ejects an electron from an atom and the photon is scattered with lower energy. The Compton mass attenuation coefficient varies directly with the electron density (electrons per kilogram) of the absorbing medium because Compton interactions occur primarily with loosely bound electrons.

The Compton mass attenuation coefficient is approximately independent of the atomic number of the attenuating medium (Z). For this reason, radiographs show very poor contrast when exposed to high-energy photons. When most of the photons in a beam of x-rays interact by Compton scattering, little selective attenuation occurs in materials with different atomic number. The image in a radiograph obtained by exposing a patient to high-energy photons is not the result of differences in atomic number between different regions of the patient, but reflects differences in physical density (kilograms per cubic meter) between the different regions (e.g., bone and soft tissue) (Hendee et al. 2003). The transmission of monoenergetic photons through the body is given by:

$$I = \ I_o \ \ exp\left(-\sum_{i=1}^{i=N} \mu_{\rho i} \ \ (\rho_i \, x_i)\right) \qquad (4.1)$$

where $I\,, I_o$ are the incident and transmitted intensities, $\mu_{\rho i}$ is the mass attenuation coefficient in $cm^2\ g^{-1}$ of the absorber material, ρ_i is the material density, x_i is the material thickness in cm, i represents an individual tissue and N is the number of different tissues.

The total attenuation depends on the thickness of the tissue and, in DXA, this will vary as bone mineral and soft tissue thickness varies. As a monoenergetic x-ray beam passes through bone that is surrounded by soft tissue, the transmitted intensity will be given by:

$$I = I_o \ exp - (\mu_s \rho_s x_s + \mu_b \rho_b x_b) \quad (4.2)$$

where b and s represent bone mineral and soft tissue respectively; x_b, x_s are the thickness of bone mineral and soft tissue in the x-ray path and ρ_b, ρ_s are the physical density of bone and soft tissue. The main chemical components of soft tissue are hydrogen, oxygen, nitrogen and carbon and these elements have lower physical density and lower atomic number than calcium and phosphorus that are found in bone mineral.

The volume of an attenuating element is given by:

$$V = \mathrm{A} \times x \quad (4.3)$$

where A is the projected area perpendicular to the x-ray beam. The physical density is given by:

$$\rho = \frac{m}{V} = \frac{m}{A \times x} \quad (4.4)$$

where m is the mass. Combining equations 4.3 and 4.4 gives the area density (M) *i.e.* mass per unit area as shown in equation 4.5.

$$M = \frac{m}{A} = \rho \, x \quad (4.5)$$

DXA is a two-dimensional projection technique and therefore BMD is an area density measurement not a volumetric density. Substituting for M in equation 4.2 gives:

$$I = \mathrm{I}_o exp -(\mu_s M_s + \mu_b M_b), \quad (4.6)$$

Taking the natural logarithm of equation 4.6 gives:

$$ln\ \left(\frac{I_0}{I}\right) = \mu_s M_s + \mu_b M_b \qquad (4.7)$$

The logarithm of the ratio of incident intensity to transmitted intensity (J), can be written as:

$$\mathrm{J} = \mu_s M_s + \mu_b M_b \qquad (4.8)$$

In dual energy mode, transmission measurements are made at two different photon energies (low and high) are given by:

Low energy:

$$J' = \mu_s' M_s + \mu_b' M_b \qquad (4.9)$$

High energy:

$$J = \mu_s M_s + \mu_b M_b \qquad (4.10)$$

These simultaneous equations are solved for M_b and M_s :

$$M_b = \frac{J' - \left(\mu_s'/\mu_s\right) J}{\mu_b' - \left(\mu_s'/\mu_s\right)\mu_b} \qquad (4.11)$$

$$M_s = \frac{\left(\mu_b'/\mu_b\right) J - J'}{\left(\mu_b'/\mu_b\right)\mu_s - \mu_s'} \qquad (4.12)$$

The ratio of the soft tissue attenuation coefficients at the two energies is defined as R_s and this depends on the soft tissue composition.

$$R_s = \frac{\mu_s'}{\mu_s} \qquad (4.13)$$

$$M_b = \frac{J' - R_s J}{\mu_b' - R_s \mu_b} \quad (4.14)$$

In a bone-free region, $M_b = 0$ and so

$$J' = \mu_s' M_s \quad (4.15)$$

$$J = \mu_s M_s \quad (4.16)$$

$$R_s = \frac{\mu_s'}{\mu_s} = \frac{J'}{J} \quad (4.17)$$

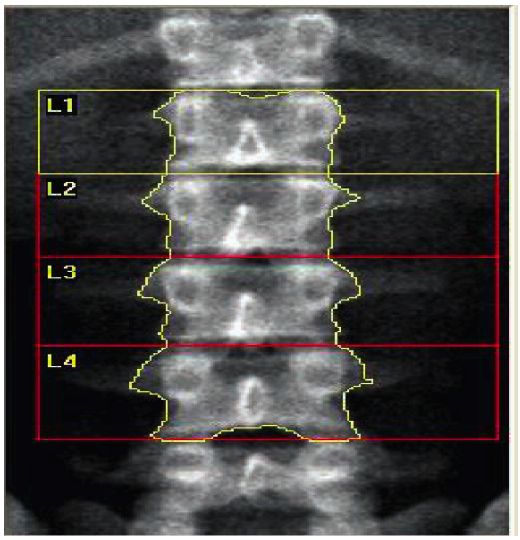
R_s is a measure of the percent fat of the soft tissue (IAEA 2010). It varies with the lean-to fat ratio within soft tissue with an increase in fat reflected by a decrease in R_s. BMD is calculated by assuming that the thickness and composition of the soft tissue in the bone-free region is identical to tissue anterior and posterior to bone and within bone so that they have identical $R_s$values.

The measurement of M_b is described in equation 4.14 for a single x-ray path through bone and soft tissue, whereas in practice measurements over a volume of bone are needed. To achieve this, a linear series of measurements across the part of the patient that contains bone must be made. The region of tissue adjacent to bone but within the analysis region is known as the soft tissue baseline. The result is an attenuation profile for each of the x-ray beams. The high-energy absorption profile is multiplied by R_s and then subtracted from the low energy profile to leave bone mineral.

BMD values are presented as a digital image with each pixel corresponding to a measurement point through the patient. In order to cover a larger area of bone, numerous attenuation profiles are acquired along the length of the bone to form an

image. The R_s value is estimated line-by-line over the image and averaged over soft tissue on either side of bone within the DXA algorithm. The projected area of bone (BA) in cm^2 is the sum of the pixels within the bone edge. Within this area, the BMD of individual pixels is averaged and multiplied by the BA to calculate bone mineral content (BMC) in g.

A lumbar spine DXA image showing analysis regions and an example of BMD, BMC and BA results is shown in Figure 4-1 (IAEA 2010). Lumbar spine BMD measurements are usually made in the antero-posterior (AP) projection although it is also possible to measure BMD in the lateral projection.



Region	Area (cm²)	BMC (g)	BMD (g/cm²)	T-Score
L1	13.21	13.59	1.029	0.2
L2	14.23	14.85	1.044	-0.5
L3	16.04	16.86	1.051	-0.5
L4	17.93	18.90	1.054	-0.8
TOTAL	61.41	64.20	1.045	-0.4

Figure 4-1: Lumbar spine DXA image showing analysis regions and an example of BMD, BMC, and BA results

4.4 Dual Energy X-ray Absorptiometry Technology

Typically, a DXA system consists of an x-ray tube emitting photons, which are filtered and collimated to pass through the patient's body as a beam and then enter a detector. The photon source, collimator and detector are aligned and connected mechanically on a scanning arm (Shepherd 2009). Figure 4-2 shows a schematic diagram of a DXA system with a fan- shaped x-ray beam in the anterior-posterior (AP) direction. In some DXA scanners, the arm can rotate through 90^0 from the AP position so that the beam is in to the lateral direction. Figure 4-3 shows a Hologic (Discovery) DXA scanner with the arm in the in lateral position.

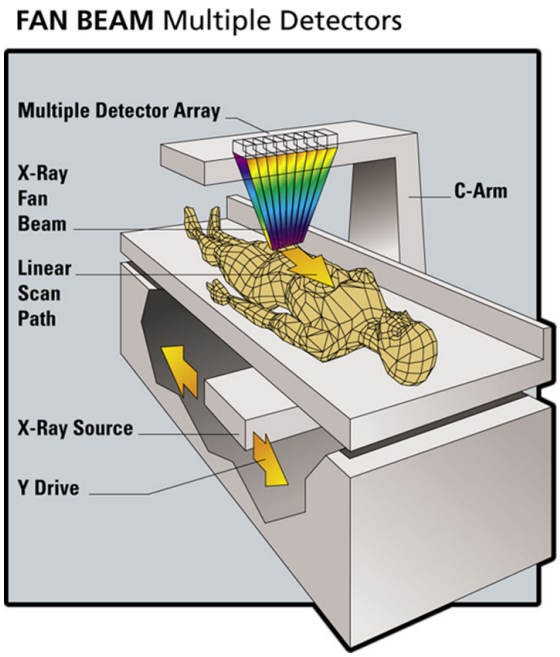


Figure 4-2: DXA scanner components AP scan. (https://ufhealthjax.org/womens-imaging/images/dxa-procedure.jpg)

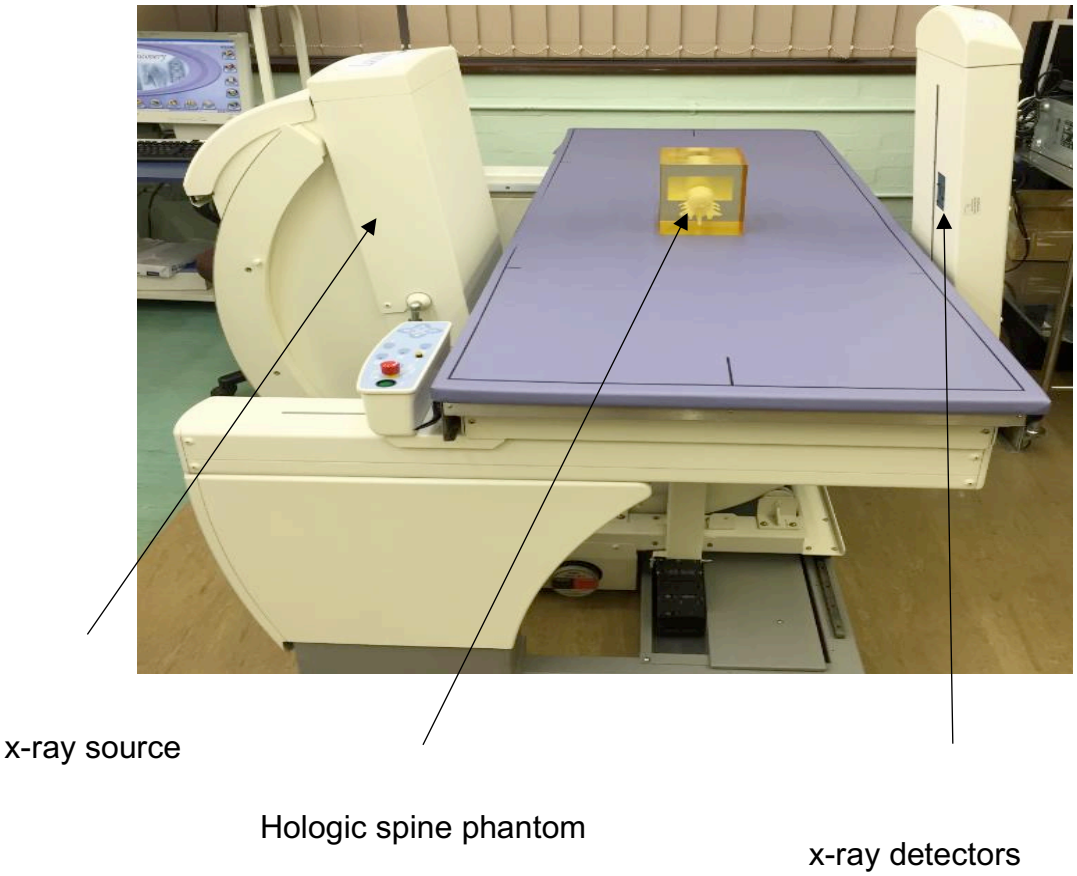


Figure 4-3: Hologic Discovery DXA scanner in position for a lateral spine scan

4.4.1 **Production of Dual Energy X-ray beam**

DXA produces a spectrum with two x-ray energy peaks. The low and high energies are selected to optimise the separation of bone mineral and soft tissue in the scanned sites. x-ray tubes produce radiation with a continuous spectrum of photon energies. In single energy (SE) mode, only one effective energy is produced by the x-ray tube. In dual energy (DE) mode, two x-ray energies are produced either by continuously switching the voltage of the x-ray tube between high and low values, known as kV switching, or by K-edge filtration (Blake and Fogelman 1997). The K-edge is the binding energy of the K shell electron of an atom. A sudden increase in the attenuation coefficient of photons

occurs at a photon energy just above the binding energy of the K shell electrons of the atoms interacting with the photons.

Hologic uses a voltage switching system. In the original QDR1000 DXA scanner, the x-ray tube voltage was switched between 70 kV (as low energy) and 140 kV (as high energy) during half-cycles of the electrical power supply (Shepherd 2009). In later models, including the Discovery and Horizon scanners used in this research, the voltages were 100 kV and 140 kV (Shepherd et al. 2002).

The higher x-ray energy beam undergoes hardening to remove the low energy part of the x- ray spectrum by using copper or brass filters during the corresponding half-cycle. The effective photon energy alternates between 40 keV and 70 keV (Lorente-ramos et al. 2011).

In a K-edge filter system, such as that used by GE (Lunar), the x-ray tube is operated at a constant voltage and filtered into low and high energy bands (IAEA 2010). The K-edge filter separates the x-ray spectrum into two separate components of ‘‘high-energy’’ and ‘‘low- energy’’ photons: using either a cerium filter; (70 and 40 keV) or using a samarium filter (45 and 80 KeV) (Hricak 2013).

4.4.2 **DXA Detectors and Collimators**

DXA scanning geometries vary; the original DXA scanners used a pencil x-ray beam. Scanning time was approximately10–15 min per single site, and up to 30–40 min for whole-body imaging in a large patient. Modern DXA scanners have a fan shaped x-ray beam and a bank of detectors, which allows faster scanning and improved image quality and spatial resolution. The spatial resolution of modern scanners is between 0.5 and 0.35 mm (Hricak 2013). Cone beam scanners are also available, in particular for small-animal applications.

In pencil beam DXA systems, the source collimator is a small hole that illuminates a single detector. Images are formed by linearly translating the x-ray gantry across an entire image row and then shifting down one pixel and scanning the next row in the opposite direction (i.e. raster-scanning).

Fan beam DXA systems have slit collimators and a linear array of 64 or more solid-state detectors. Images are formed by scanning linearly in the direction perpendicular to the length of the detector array, with the entire array being exposed to radiation. This makes fan beam systems faster compared with pencil beam systems in proportion to the number of elements, because a larger portion of the x-ray tube flux is used (Wu et al. 2012; Speller et al. 2006). In addition, the detectors in fan beam system are smaller than those in pencil beam systems, resulting in smaller pixel sizes (Shepherd 2009).

Cone beam systems have a rectangular source aperture (collimator) that illuminates an entire region of interest for each exposure. In these systems, the effect of scattered radiation can be significant.

The function of the detector is to determine the tissue attenuation at a certain point in the patient by comparing the x-ray intensity measured at that point to the x-ray intensity measured with no patient present. The detector may be a scintillation crystal coupled to a photo multiplier tube (PMT) or, in more modern systems, a photodiode. When an x-ray interacts with a scintillating material, the energy of the x-ray is converted into a scintillation consisting of many photons of light. A PMT or photodiode converts this scintillation into an electrical voltage pulse. Photodiode detectors consist of either a light-sensitive photodiode coupled to a scintillator or an x-ray-sensitive photodiode. Their advantage is that they can be manufactured in small sizes of 1 mm^2 to 4 mm^2, compared to approximately 10 mm^2 or larger for PMT (Shepherd 2009). Therefore, photodiodes are used in high spatial resolution DXA systems that use fan or cone beam geometry. However, their disadvantage is that they do not contain any internal gain; thus, noise is higher and for comparable image quality, a higher x-ray flux is required.

4.5 **Hologic DXA**

In this study, two types Hologic DXA scanner were used: Discovery and Horizon. The Discovery was the first system to integrate the two factors for determining osteoporotic fracture risk: BMD measurements and VFA. The Discovery uses a high-resolution detector array of 64 to 216 elements in addition to a fan beam technology. This provides a technical foundation for high image quality and improves scanning time. The system captures the hip and spine with 10-second scanning time and dose of less than 10μSv (Ram 2012).

The Hologic Horizon is the latest version of the Hologic QDR series; like the Discovery, it integrates BMD measurement and VFA (Wu et al. 2012). Ultra-fast, high output ceramic detectors, a high frequency pulsing x-ray generator and an internal reference system are the key technology innovations of the Horizon scanner. The Horizon scanner uses a high-resolution multi-element detector array with gadolinium sulfoxylate (GADOX) scintillator technology (64 to 216 detectors, model dependent) (Wilson and Kelly 2013). Pulsing x-ray generators of high frequency are used in fluoroscopy to reduce radiation dose while maintaining image quality.

In this research three DXA scanners were used: Discovery A (S/N: 70902) upgraded from Delphi in 2014, Discovery A (S/N: 86224) and Horizon A (S/N200253). The x-ray source alternate between high (140 kV) and low (100 kV).

4.6 **Quality Control**

Each Hologic DXA scanner is provided with an anthropomorphic spine phantom for quality control (QC) purposes.

Hologic recommends that users should scan the QC phantom at the beginning of each day on which patient BMD is measured and at least three times a week. Spine phantom images are analysed using the standard Hologic BMD software.

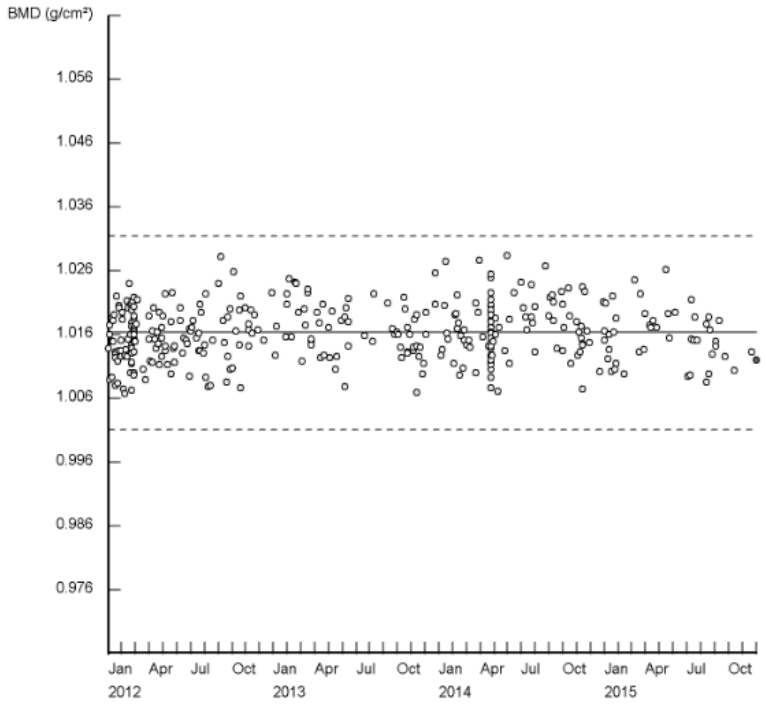
QC scans were retrieved from the DXA scanner archives at the University Hospital of Wales, Cardiff to cover the period over which VFA scans were acquired in this work. Table 4-1 lists the spine phantom L1 to L4 BMD QC results for three Hologic DXA scanners over the time period of the first part of this research. Table 4-2 lists the corresponding QC results over the period of the phantom study in this research. Figure 4-4 shows a QC BMD plot for the Discovery A scanner (S/N70902) from 2012 to 2015. All CV% values were less than 0.5% and there was no significant change of BMD with time, indicating good scanner stability.

Table 4-1: QC results for three DXA scanners over the acquisition period of the clinical VFA scans

DXA scanner	Type	N	Period	Mean BMD (g/cm^2)	SD (g/cm^2)	CV%
S/N: 70902	Discovery A	793	Jan 2010- Dec 2011	1.013	0.005	0.457
S/N: 86224	Discovery A	1477	Jan 2012- Dec 2016	0.998	0.003	0.264
S/N: 200253	Horizon A	564	Jan 2016- Dec-2017	0.921	0.002	0.225

Table 4-2: QC results of three DXA scanners used experimental phantom scan

DXA scanner	Type	N	Period	Mean BMD (g/cm^2)	SD (g/cm^2)	CV%
S/N: 70902	Discovery A	1159	Jan 2010- Dec 2015	1.014	0.005	0.464
S/N: 86224	Discovery A	316	Dec 2017-Jan 2018	0.996	0.003	0.270
S/N: 200253	Horizon A	383	Aug 2016- Dec 2017	0.921	0.0019	0.209



Setup	Reference Values	Plot Statistics
a Lumbar Spine phantom #10203 System S/N: 70902	Limits: ±1.5% of mean Mean: 1.016 (g/cm²) SD: 0.005 (g/cm²)	Number of Points: 366 Mean: 1.016 (g/cm²) SD: 0.004 (g/cm²) CV: 0.429 %

Figure 4-4: Quality control plot for L1 to L4 BMD of the Hologic spine phantom BMD for one DXA scanner between 2012 and 2015

4.7 **Vertebral Fracture Assessment (VFA)**

The most important clinical outcome in osteoporosis is the occurrence of fracture. Vertebral fractures usually occur in postmenopausal women and older men, with an estimated prevalence of 10% - 26% in both men and women aged 50 years and older (Drampalos 2015). Highly effective drugs are available for osteoporosis treatment. The most efficient is the bisphosphonate zoledronic acid, which can reduce the risk of vertebral fractures by 76% and of non-vertebral fractures by 24% (Drampalos 2015; Oei

et al. 2016). The presence of one or more prevalent vertebral fractures on lateral spine radiographs is a strong predictor of future incident vertebral fractures and a moderate predictor of non-vertebral fractures, independent of BMD. Vertebral Fracture Assessment (VFA) is a relatively new DXA scanner technology that permits imaging of the thoracic and lumbar spine for the identification of vertebral fractures.

Although the standard method for the detection of vertebral fractures (VFs) is radiography of the spine, VFA has several advantages. These include substantially lower radiation dose, lower cost, higher patient convenience, and less operator-dependent variance (Hospers et al. 2009). Another disadvantage of radiography is the problem of geometric distortion, because radiographs are acquired using cone-beam geometry. This causes an increase in size of the image compared to the object due to divergent path of the photons producing the image.

VFs can be detected on other modalities such as CT or MRI as well. Fractures can be categorised according to skeletal site and shape and measured according to the amount of height loss and the number of fractured vertebrae.

VFA images allow evaluation for prevalent and incident VFs using qualitative, semi-quantitative or quantitative morphometric methods (Schousboe et al. 2010). Several methods are available to determine vertebral deformities and distinguish them from vertebrae with a normal shape. A frequently used method is that based on semi quantitative morphometry (QM) to evaluate vertebral height, as proposed by Genant et al. (1993). The method involves measuring a reduction of anterior and middle height relative to the posterior height within a vertebra and/or reductions of these height relative to the corresponding heights of adjacent vertebrae. This method uses fixed threshold values of loss of height ratio (0.60, 0.75 and 0.80) (Drampalos 2015).

Another approach is based on the visual identification of VFs and is known as the algorithm-based qualitative method (ABQ). This technique focuses on the appearance

of the centre vertebral endplates to identify VFs without a minimum threshold of vertebral height reduction. In this methods endplate integrity is identified regardless of vertebral height reduction (Drampalos 2015).

The Hologic Delphi scanner was the first DXA system to employ VFA, for which the Hologic name is instant vertebral imaging (IVA). This enabled experts to identify fractures in the spine with a rapid, low-dose single-energy image (Wu et al. 2012). VFA images have been shown to be a good alternative to standard radiography for the detection of vertebral fracture (Schousboe et al. 2006).

There is sufficient soft tissue anterior to the lumbar spine in VFA images to allow the detection of calcification in the abdominal aorta. Thus, two highly prevalent public health problems can be assessed through the same diagnostic test: osteoporosis and CVD (Schousboe et al 2006; Schousboe et al. 2007). Figure 4-5 shows a Hologic Discovery scanner set up for lateral spine imaging. VFA images can be acquired in two energy modes: SE mode and DE mode. VFA images obtained in SE mode include prominent soft tissue features. In contrast, images acquired with DE use a second x-ray beam energy to account for the x-ray absorption by soft tissues. DE images are noisier but have higher contrast for bone and greatly diminished soft tissue shadows. Advantages of SE imaging are that the images can be obtained faster, and the vertebral endplates and cortices are slightly sharper than on DE images. However, these images are disadvantaged in that shadows created by soft tissues can obscure visualization of the vertebrae, especially in areas where the contrast between adjacent soft tissues is considerable, such as the diaphragm.

Typical lateral VFA images acquired with the Discovery scanner in two energy modes are shown in Figure 4-6 (Hologic 2007). Many studies have been employed VFA images for AAC detection and measurement; this was discussed in Section 2.3 in detail.

Figure 4-5: Hologic Discovery DXA system in position for lateral spine scan. (http://www.hologic.com/products/imaging/skeletal-health/discovery-dxa-system

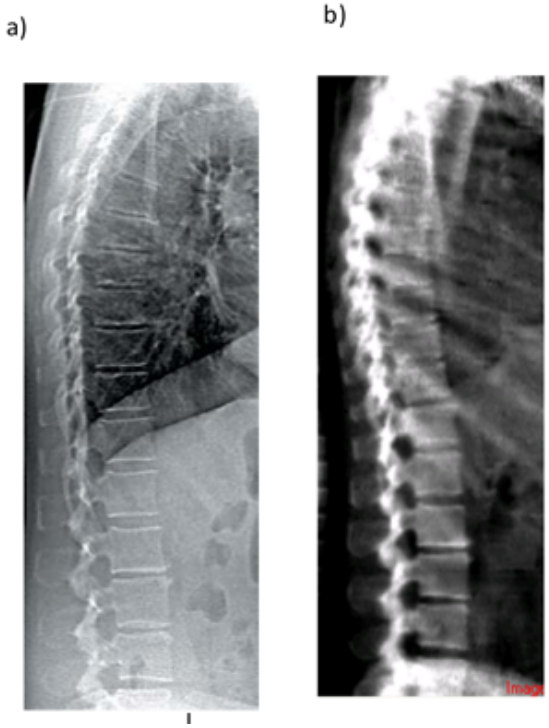


Figure 4-6: Lateral VFA images acquired with a Hologic Discovery DXA scanner, (a) in single energy mode, b) in dual energy mode

4.8 **Summary**

This chapter has discussed the medical imaging modality that is employed in this research. The physical principles of DXA have been briefly introduced. Two different Hologic DXA scanner models used in this research have been described.

The next chapter introduces a new method for the detection and quantification of AAC in VFA images produced by both models of Hologic DXA scanner.

Chapter: 5

Automatic Detection and Quantification of AAC

5.1 Introduction

As outlined in Chapter 2, VFA images can show AAC with good sensitivity and specificity compared to radiography (Lewis et al. 2016; Grant et al. 2017). The main limitations of VFA images are their relatively limited spatial resolution of these images compared to the radiographs and low signal to noise ratio (SNR) (Guglielmi et al. 2008). VFA image noise can be structural or statistical. Structural noise occurs due to the presence of bones, such as ribs and different soft tissues in the abdomen region. Statistical noise occurs due to the variation in number of detected x-ray photons per pixel.

In general, there are several factors affecting the accuracy of the DXA technique: (a) variation in soft tissue composition within the x-ray beam, (b) patient positioning and scan analysis, (c) artefacts due to metal or clothing, (d) random errors due to photon and electronic noise (e) drifts in scanner calibration. Problems with patient positioning for lateral spine imaging are reduced in DXA scanners in which C-arm can rotate to 90^0 from the anterior-posterior position to the lateral position (Guglielmi et al. 2008).

AAC appears in VFA images as small and elongated bright structures on the anterior and posterior walls of the aorta, but these also can extend to its centre. The assessment of AAC adjacent fractured vertebrae is difficult and this becomes even more challenging if the posterior vertebral wall was collapsed. At this time, there is no consensus on how to score AAC adjacent to a crushed vertebra. In some cases, distinguishing calcified aortic walls from other calcifications and artefacts, such as bowel wall adjacent to a gas bubble, is not easy. For example, contrast between air and soft tissue can create

curvilinear borders mimicking calcification. Calcified kidney stones, the anterior cortices of the vertebrae themselves, other calcified tissues and the edges of the lower ribs extending down into the abdominal region sometimes need to be carefully distinguished from true AAC (Schousboe et al. 2017).

A variety of studies have been conducted previously to score AAC manually on radiographs and VFA images using the AC-24 points system. However, at the time of this research, only few automatic approaches for AAC detection and quantification on radiographs had been described and they were not related to the clinically accepted manual methods. Furthermore, none of these methods were tested on VFA images characterised by lower SNR compared to standard radiographs.

The main contribution of this chapter is the development of a new method for automatic detection and quantification of AAC related to existing manual scoring techniques. The method employs shape and appearance models for automatic segmentation of the aorta and uses an automatic classification method for the purpose of quantification. The new automatic AAC quantification method is based directly on the AC-24 manual method. Evaluation of the automatic method in relation to the manual AC-24 results is also presented.

5.2 **Data set**

This research used data provided by the bone densitometry service at the University Hospital of Wales, Cardiff. A set of 73 anonymised VFA images of patients referred for the investigation of osteoporosis in the period between 2010 and 2015 were collected. These images were selected visually because of obvious aortic calcification. The VFA images were acquired by two Hologic Discovery A DXA scanners operating in SE mode with a current of 5.0 mA and a voltage of 140 kV, giving an average x-ray energy of about 70 keV. The pixel size was 0.35945 mm x 0.361840 mm, the line spacing was 0.0362 cm and the point resolution was 0.1258 cm for both scanners. Scanner S/N70902

was used in the period between 2010 and 2012 and scanner S/N86224 in the period between 2012 and 2015. Approval for the study was obtained from Cardiff University School of Engineering Ethics Committee.

The selected patient images had three different degrees of aortic calcification, namely mild, moderate and severe. This classification was made visually based on the degree of calcification that was seen. Images in the anterior-posterior (AP) and lateral views were extracted in digital imaging and communications in medicine (DICOM) format. AP images were used to identify the four lumbar spine vertebrae (L1-L4) on the lateral images.

Figure 5-1 shows examples of VFA images with various degrees of AAC. Three labelled VFA images cropped to a lumbar region of interest (ROI) and indicating severe, moderate and mild AAC are shown in Figure 5-2. Image constant has been enhanced by applying a widowing technique for better visualisation of the calcification.

Unless otherwise specified, image display and processing was done using Matlab (Mathworks, MA, USA) software, While statistical analysis was done with SPSS (IBM, New York USA).

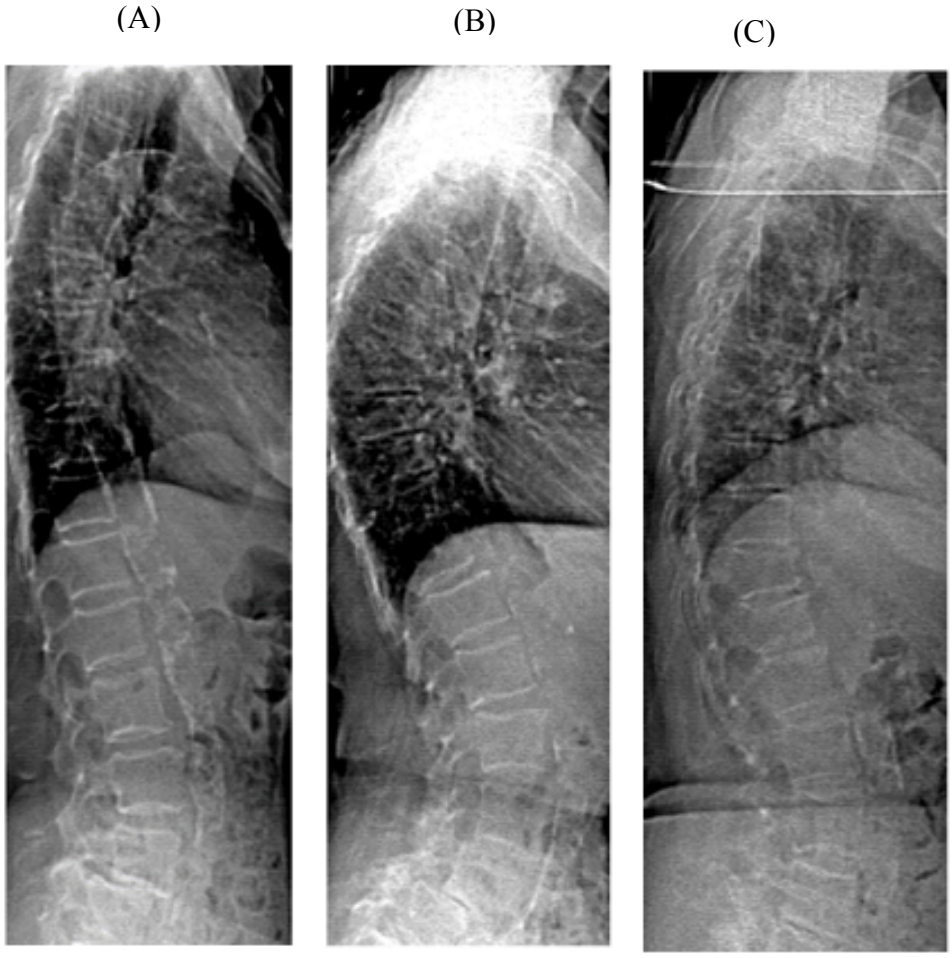


Figure 5-1: Examples of VFA images retrieved from the bone densitometry service archive at the University Hospital of Wales, Cardiff. (A) 91- year-old female, (B) 83-year-old female, (C) 79-year-old female. The images have been windowed to enhanced contrast

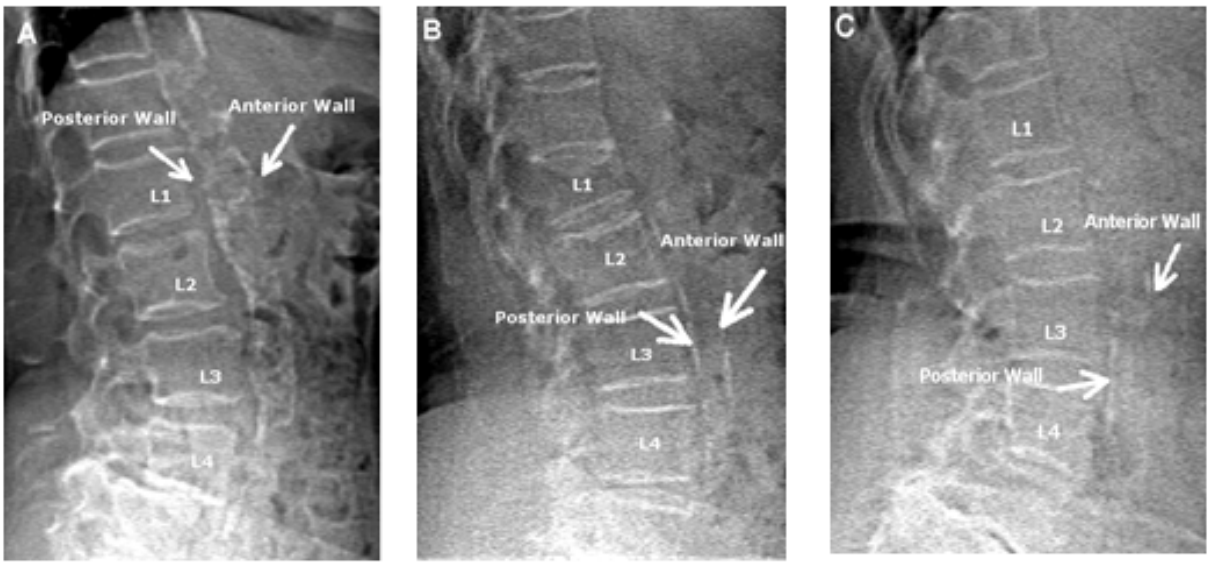


Figure 5-2: Cropped and labelled VFA images of female patients with (A) severe AAC, (B) moderate AAC, (C) mild AAC

5.3 **AC-24 Manual Scoring System**

AAC was scored visually on the set of 73 VFA images using the same method (AC-24) as the Framingham Heart Study (Kauppila et al. 1997; Walsh et al. 2002; Schousboe et al. 2006; Barascuk et al. 2011; Ganz et al. 2012). One reader who was trained to interpret VFA scores scored all 73 images.

AC-24 scoring was done by visual inspection of AAC to in the aortic walls. The abdominal aorta was divided into eight sections (four anterior wall and four posterior wall) immediately adjacent to each vertebra L1-L4 in every VFA image (Figure 5-3). The calcification in each segment was graded as follows: 0 for no AAC, 1 for AAC occupying one-third or less of the aortic wall within the segment, 2 for AACs occupying more than one-third, but less than two-third of the length, or 3 for more than two thirds occupation of the wall length. The summation of scores gives a composite AC-24 score ranging from 0 to 24. Table 5-1 shows the AC-24 scoring for the example shown in Figure 5-3.

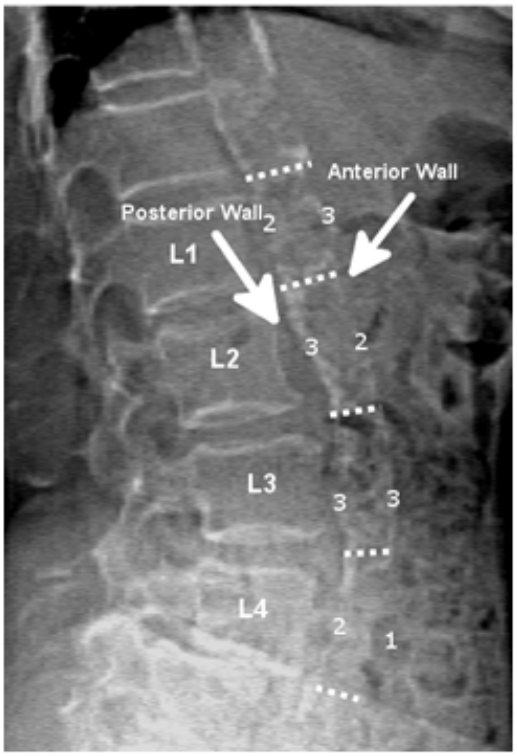


Figure 5-3: Labelled VFA image with AC-24 scores in 8 aortic segments. The dotted lines are drawn at the position and orientation of the intravertebral spaces

Table 5-1: AC24 scoring for the example VFA scan of the aorta shown in Figure 5-3

Vertebral	AC24 scores		
	Anterior wall	Posterior wall	Anterior + Posterior
	(0-3)	(0-3)	(0-6)
L1	2	3	5
L2	3	2	5
L3	3	3	6
L4	2	1	3
Sum (L1-L4)	10	9	19
Maximum	12	12	24

5.4 AAC Categorisation using the manual AC-24

The total AC-24 score was used to categorise AAC into three classes following the method of Mohammad et al. (2014): mild AAC (score 1-4), moderate AAC (score 5-12) and severe AAC (score >12). A similar approach has been used in other previous studies. It was not feasible to categorise AAC in to 24 separate classes since the average of images per class would have been about three (with a total of 73 images). This would have been insufficient for further analysis.

5.5 Segmentation of Lumbar Spine and Aorta

The main problem in developing an automatic calcification detection technique for VFA images is the presence of other structures, particularly bones. Bones have similar appearance to calcification due to the fact that the chemical composition of calcification is identical to hydroxyapatite, which is a main component of bone.

A general review of segmentation methods was presented in Section 3.2. The segmentation of the aorta is very difficult when it is not calcified, as soft tissue cannot be seen on x-ray images. However, the location of the aorta is constrained by the position of the spine, which can be detected easily in the image. Thus, prior knowledge of the

shape and position of the lumbar spine can be used to guess the position of the abdominal aorta even if it is not calcified. A suitable image segmentation approach for this task is the application of model-based algorithms, which use prior knowledge of what is expected in the image. Typically, the best fit of the model to the data is output as a new image. Statistical models of shape and appearance are powerful tools for interpreting medical and are well-studied to dealing with VFA images, which are inherently and noisy (Roberts et al. 2003; Bruijne et al. 2004; Benjelloun et al. 2011; Sarkalkan et al. 2014).

5.5.1 **Statistical Shape Model**

The aim of this step in the work package is to derive models that allow the analysis of both new shapes and the synthesis of shapes similar to those in a training set. Typically, the training set is developed by manual annotation of a number of images. The shape model can be built by analysing variations in shape over the training set.

5.5.1.1 **Landmarks**

Building a statistical shape model consists of representing the mean shape and a number of modes of variation using a set of training instances. This method strongly depends on the chosen shape representation (Cootes 2004). The shape of an object is represented by a set of n points, which may be in two or three dimensions. The process starts by choosing suitable landmarks, which must be consistently located from one image to another. The simplest method for generating a training set is by annotation of a series of images with a set of corresponding points by a human expert. Two dimensional images can be annotated using 2D points, which are usually placed at clear corners of object boundaries, 'T' junctions between boundaries or easily located biological landmarks. Object shape, which is described by n points in d dimensions, can be represented by a nd element vector produced by concatenating the elements of the

individual point position vectors. A single 2-D image can be represented by n landmark points,$\{(x_i, y_i)\}$, as the $2n$ element vector, $\mathbf{x}$, where

$$\mathbf{x} = \left(x_1, x_2, x_3 \ \ldots\ldots x_{\mathrm{n}}, y_1, y_2, y_{3,\ldots.} y_{\mathrm{n}}\right)^T \tag{5.1}$$

Here T, represents the transpose of $\mathbf{x}$. For s training instances, s such vectors $\mathbf{x}_j$ can be generated.

5.5.1.2 **Aligning the Training Set**

In order to obtain a description of the object's shape independent of its position, orientation and scale, the next step in this modelling process is the representation of all shapes aligned in the same coordinate system. The most common approach is Procrustes Analysis, which minimises the sum D of the distances of each shape from the mean shape $\bar{x}$ (Goodall et al.2014) :

$$D = \sum |x_i - \bar{x}|^2 \tag{5.2}$$

5.5.1.3 **Modelling Shape Variation**

The next step is dimensionality reduction of the data set from the initial nd values. For this, principal component analysis (PCA) is applied to the aligned shapes. The data form a cloud of points in the nd space. PCA computes the main axes of this cloud of points and this allows the approximation of any of the original points using a model with fewer than nd parameters. PCA is applied using the following steps:

Calculate the mean of the data:

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{x}_i \tag{5.3}$$

Calculate the covariance matrix of the data:

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^{s} (\mathbf{x_i} - \bar{\mathbf{x}}) \ (\mathbf{x_i} - \bar{\mathbf{x}})^T \qquad (5.4)$$

Calculate the eigenvectors $\mathbf{\Phi}_i$ and eigenvalues λ_i of $\boldsymbol{S}$.

If $\mathbf{\Phi}$ contains the t eigenvectors corresponding to the largest eigenvalues, then the training $\mathbf{x}$ can be approximated using

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{\Phi b} \qquad (5.5)$$

where $\mathbf{\Phi} = (\phi_1|\phi_2|\ldots|\phi_t)$ and **b** is a t dimensional vector given by

$$\mathbf{b} = \mathbf{\Phi}^{\boldsymbol{T}} (\mathbf{x} - \bar{\mathbf{x}}) \qquad (5.6)$$

The vector $\boldsymbol{b}$ defines a set of parameters for a deformable model. As the elements of $\boldsymbol{b}$ may vary, the shape can be varied, using Equation 5.5. The variance of the i^{th} parameter b_i across the training set is given by λ_i. By applying limits of $\pm 3\sqrt{\lambda_i}$ to the parameter b_i, the generated shape will be similar to shapes in the original training set. The number t of eigenvectors to retained can be chosen, so that the model represents some proportion (e.g. 98%) of the total variance of the data or the residual terms can be considered noise.

5.5.2 **Applying statistical shape model to VFA images**

To build a shape model, a set of 20 VFA images with clear evidence of aortic calcification was selected as training data. A statistical SM was created using the method described above. For each image in the training set, landmarks were positioned manually at the corners and at the centres of the end-plates of each of the first four lumbar vertebrae (L1-L4). A further 32 landmarks were placed along the anterior and posterior walls of the calcified aorta adjacent to L1-L4, creating 56 landmarks in total. Figure 5-4 shows an example of VFA image enhanced for better visualisation (A) and a cropped image with

landmarks (B). Figure 5-5 shows the first three modes of shape variation. Here, the first three modes of variation explain ≈83% of shape variation in the training data set.

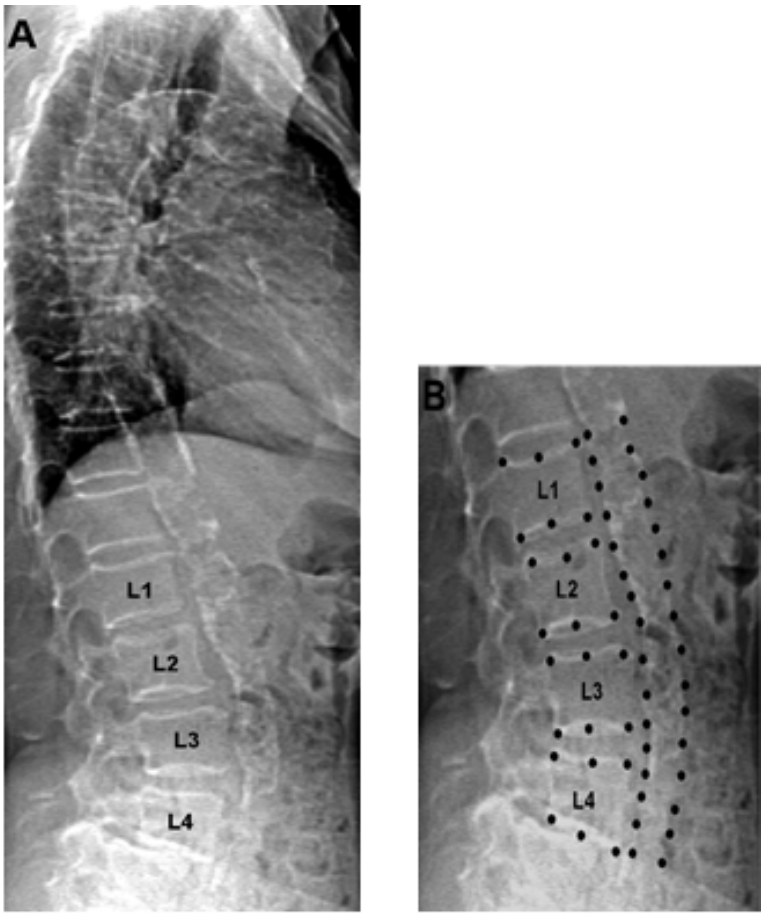
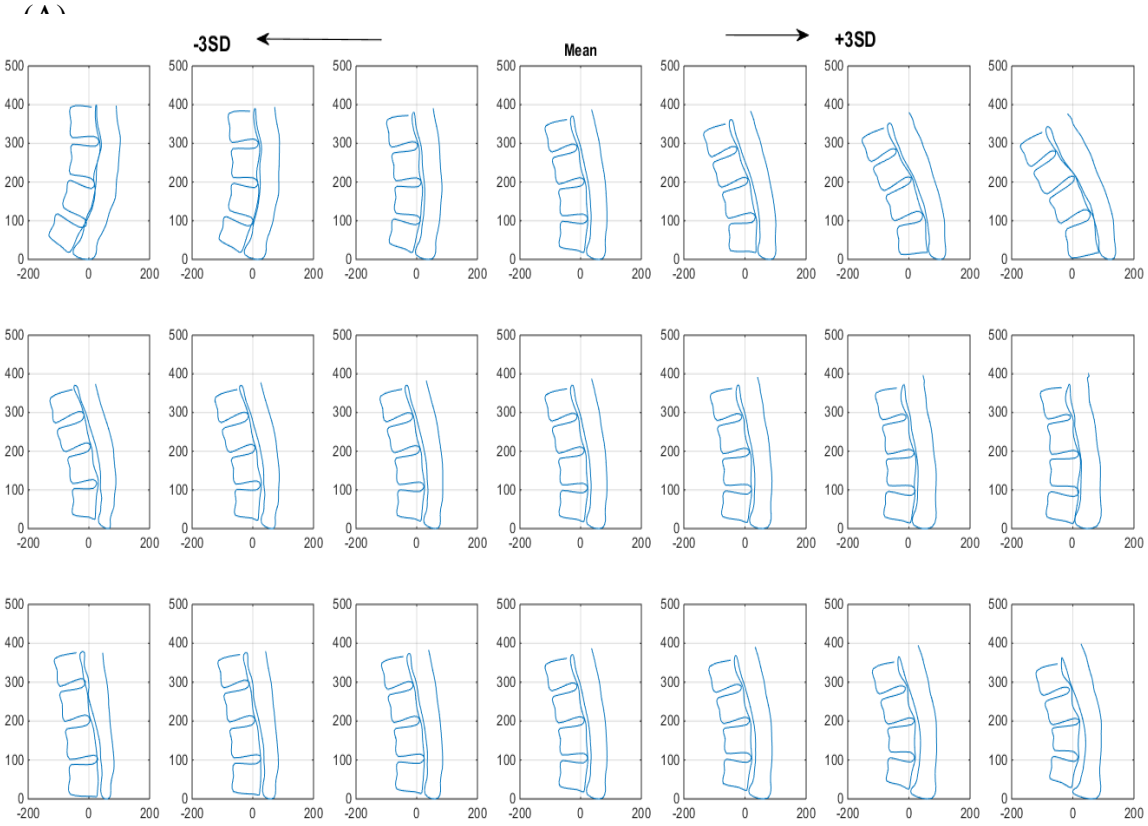


Figure 5-4: Example of an annotated VFA image with 56 landmarks (shown as black dots): (A) original image; (B) cropped image with landmarks



(B)

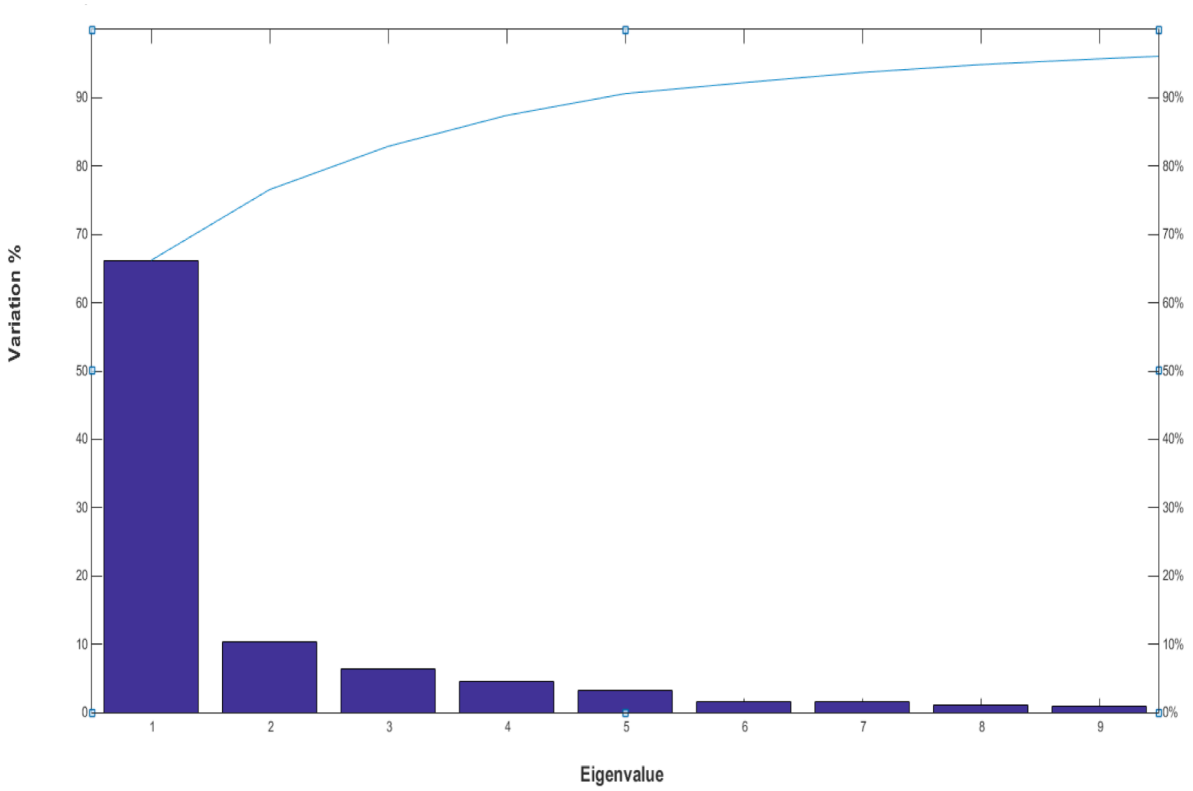


Figure 5-5: (A) First three modes of shape variation $\pm 3SD$ in turn, (B) Shape variation with number of eigenvalues

5.5.3 **Statistical model of texture**

In order to synthesise a complete image of an object, a shape and texture model is used. This section describes how a statistical model can be built to represent both shape variation and texture variation.

A statistical model of the grey-level texture can be built by warping each example image so that the control points match the mean shape. This can then be done using a triangulation algorithm (Cootes 2004). The resulting samples can be then normalised to minimise the global lighting variation by applying a scaling factor α and an offset β on the texture vector g_{im}:

$$\mathbf{g} = (\mathbf{g}_{im} + \beta\mathbf{1})/\alpha \qquad (5.6)$$

The values of α and β are chosen to best match the vector to the normalised mean. Assume that $\bar{\mathbf{g}}$ is the mean of the normalised data, scaled and offset so that the sum of elements is zero and the variance of elements is unity. The values of α and β required to normalise g_{im} are given by:

$$\alpha = \mathbf{g}_{im} \cdot \bar{\mathbf{g}} \quad , \qquad \beta = (\mathbf{g}_{im} \cdot \mathbf{1})/\mathrm{n} \qquad (5.7)$$

where n is the number of elements in the vectors.

Thus, for every training set, the profile $\mathbf{g}$ can be represented by the mean profile $\bar{\mathbf{g}}$ and variation $\boldsymbol{P}_g$ by applying PCA:

$$\mathbf{g} = \bar{\mathbf{g}} + \boldsymbol{P}_g \, \boldsymbol{b}_g \qquad (5.8)$$

where $\bar{\mathbf{g}}$ is the mean normalised grey-level vector, $\boldsymbol{P}_g$ is a set of orthogonal modes of intensity variation and $\boldsymbol{b}_g$ is a set of grey-level parameters.

5.5.4 **Combined Appearance Model**

The shape and texture of the training set can be represented by the parameter vectors $\boldsymbol{b}_s$ and $\boldsymbol{b}_g$. A further PCA can be applied to the data after concatenating the shape and grey-level variations in one vector to produce a model of the form:

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{W}_s \, \boldsymbol{b}_s \\ \boldsymbol{b}_g \end{pmatrix} = \begin{pmatrix} \boldsymbol{W}_s \, \boldsymbol{P}_s^T \, (\mathbf{x} - \bar{\mathbf{x}}) \\ \boldsymbol{P}_g^T \, (\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} \tag{5.9}$$

where $\boldsymbol{W}_s$ is the diagonal matrix of weights for each shape parameter that gives appropriate balance between the shape and the texture models. By applying the PCA on this vector, a further new model can be generated:

$$\mathbf{b} = \mathbf{P}_\mathbf{c} \, \mathbf{c} \tag{5.10}$$

where $\boldsymbol{P}c$ are the eigenvectors and $\mathbf{c}$ is a vector of appearance parameters controlling both the shape and grey-levels of the model.

5.5.4.1 **Applying statistical shape and texture models to VFA images**

Figure 5-6 shows the mean and the first three eigenvectors created from the texture model. The original texture, the texture described by the combined model and the difference between them are shown in Figure 5-7.

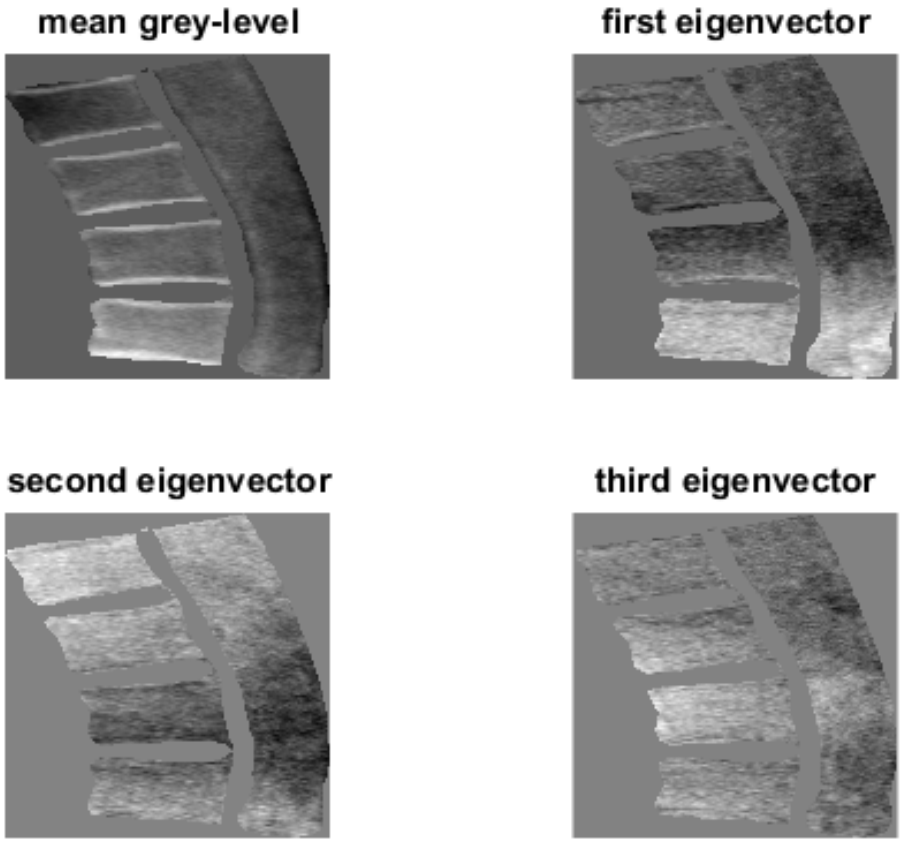


Figure 5-6: Mean appearance model and first three eigenvectors

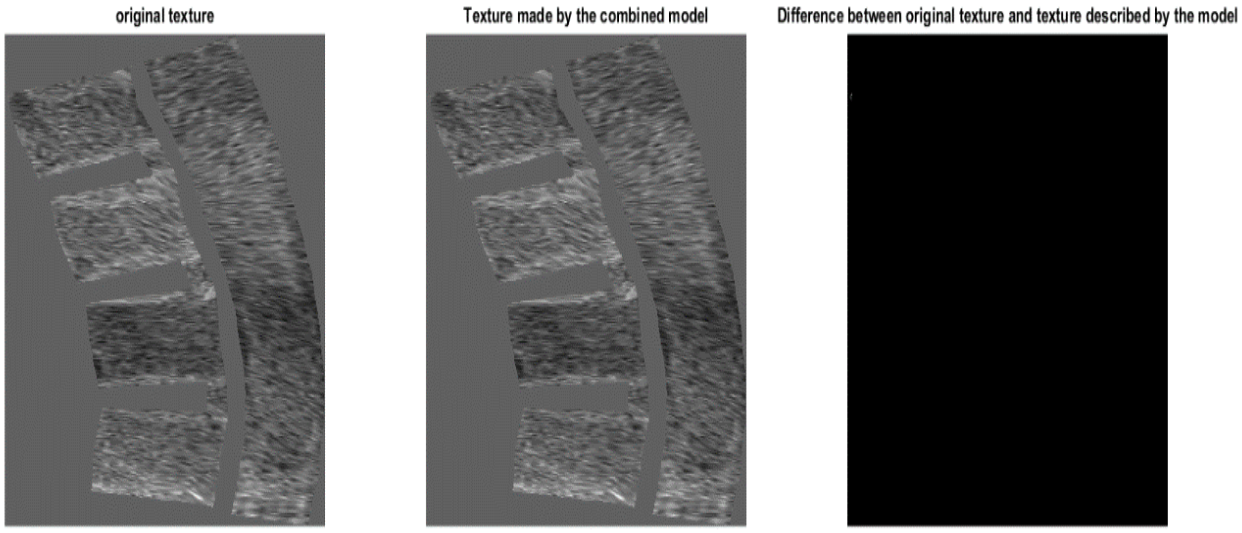


Figure 5-7: The original texture, the texture described by combined model and the difference between them

5.6 **Active appearance model (AAM) searching algorithm**

In order to detect the lumbar spine and the aorta in VFA images, AAM searching algorithm developed by Cootes et al. (2001) was used. The mean shape produced was placed manually in an approximate correct position in the VFA images. The algorithm was then used to fit spine and aorta edges for all images in the data set.

The process produced a new set of images of segmented lumbar spine (L1-L4) and the calcified aorta and its surrounding area. The algorithm was applied successfully to the entire set of 73 images and it took 23 seconds for 100 iterations to capture the desired two objects for each image. Figure 5-8 shows the searching process using the AAM.

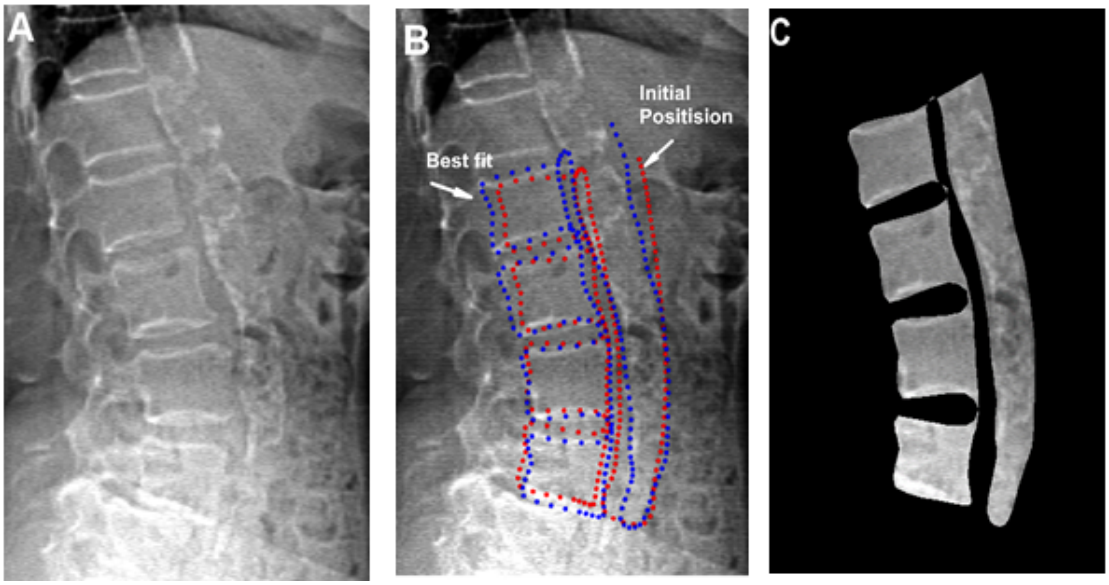


Figure 5-8: (A) Input image; (B) finding the best fit, initial position (red points) the best fit (blue points); (C) segmented spine and aorta

5.7 **Extracting the aorta from the segmented image**

In order to extract the aorta from the segmented image produced by the AAM searching algorithm, the grey-scale image was converted to a binary image by thresholding. A connected-component labelling algorithm based on pixel connectivity was then implemented. The algorithm works by scanning the image pixel-by-pixel (from top to

bottom and left to right) in order to identify connected pixel regions (Burger and Burge 2013). Each binary image was labelled into 8 connected objects. The pixels labelled 0 were the background. The pixels labelled 1 made up the first object; the pixels labelled 2 made up the second object; and so on. The area of each labelled object was computed. This was done for all 73 VFA images in the data set. The aorta was extracted as the labelled object that had the greatest area in the entire image; all other objects related to the background and the spine were removed. Figure 5-9 shows the extracted aorta from the example used in Section 5.3. Further examples of calcified aortas extracted from various segmented images produced by the AAM are presented in Figure 5-10.

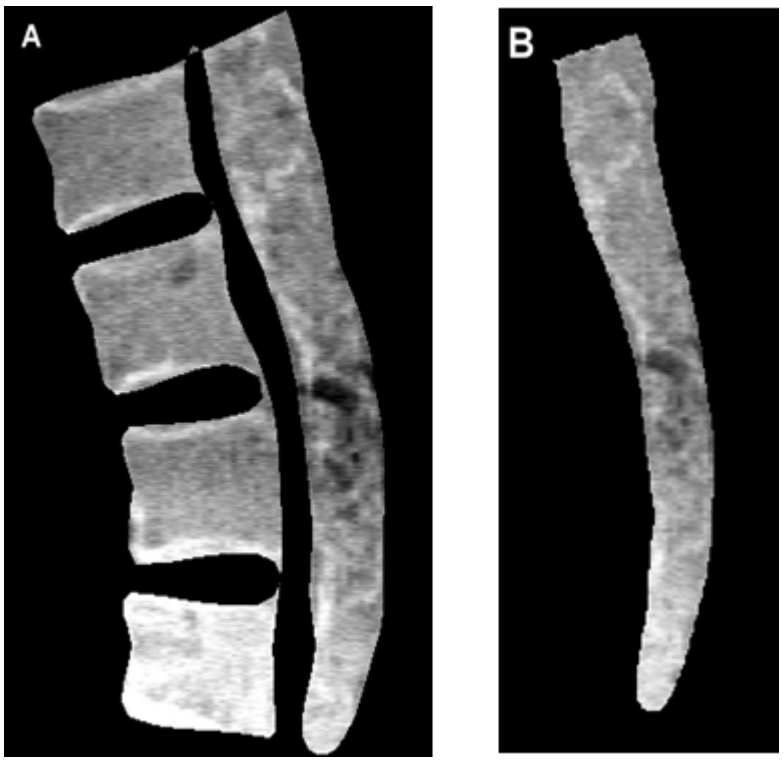


Figure 5-9: (A) Segmented calcified aorta and spine produced by AAM; (B) largest region area (aorta)

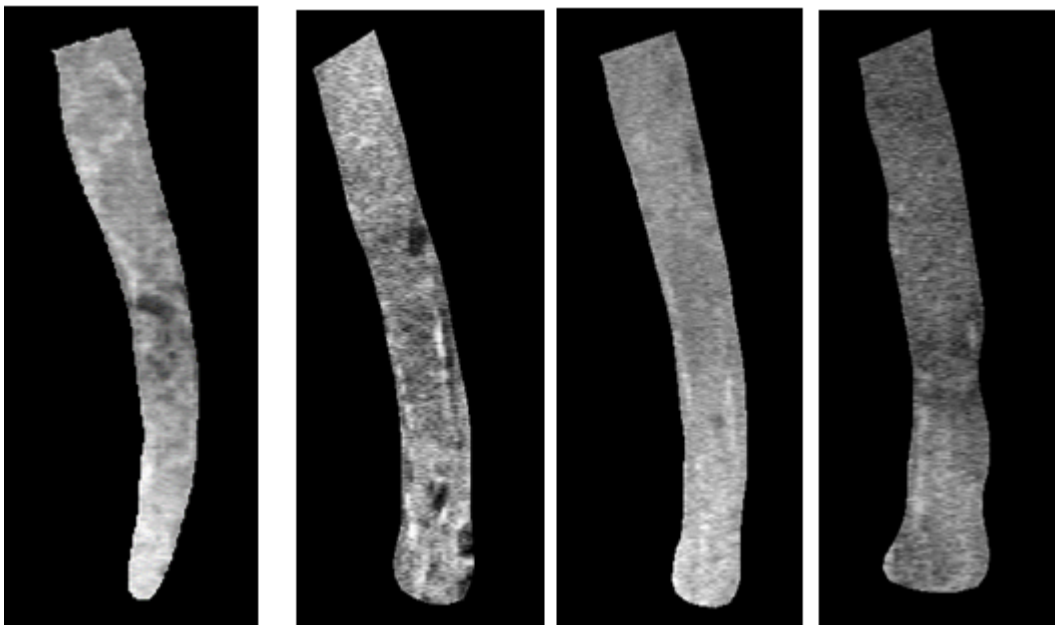


Figure 5-10: Examples of segmented aortas with different degrees of calcification and different brightness

5.8 **Calcification Detection**

Although the segmented aorta obtained from the AAM clearly masked out all the calcified pixels, sometimes it also detected parts of the aorta that had high brightness due to changes in soft tissue composition. This variation in brightness and the influence of obesity made the use of a single threshold not applicable in the next stage. Multilevel thresholding is a technique that segments a grey level image into several distinct regions (Horng 2010). The technique defines more than one threshold for the specified image and it segments the image into regions of certain brightness that identify the background and several objects. In the application of multilevel thresholding, every segmented aorta was processed using specified quantisation levels and output values.

The popular performance indicator, peak signal to noise ratio (PSNR), was used to compare the segmentation results obtained by the multilevel image threshold techniques (Horng 2010):

$$PSNR = 20log_{10}\left(\frac{255}{RMSE}\right) \tag{5.9}$$

where RMSE is the root mean square error, defined as:

$$RMSE = \frac{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j)-\hat{I}(i,j))^2}}{MN} \tag{5.10}$$

where I , $\hat{I}$ are the original and segmented images respectively, each of size MxN.

Figure 5-11 shows that for three different examples of aorta with low, moderate and high-calcification the $PSNR$ reaches saturation after seven levels of threshold levels. Images were processed using the maximum threshold values. This technique is called image quantisation. As the number of the thresholds increased, the thresholded image tended towards the original image and this was evaluated visually. An example of a quantised image with different threshold levels is presented in Figure 5-12. It is clear that after 14 levels of thresholding, the image (F) tended towards the original (A).

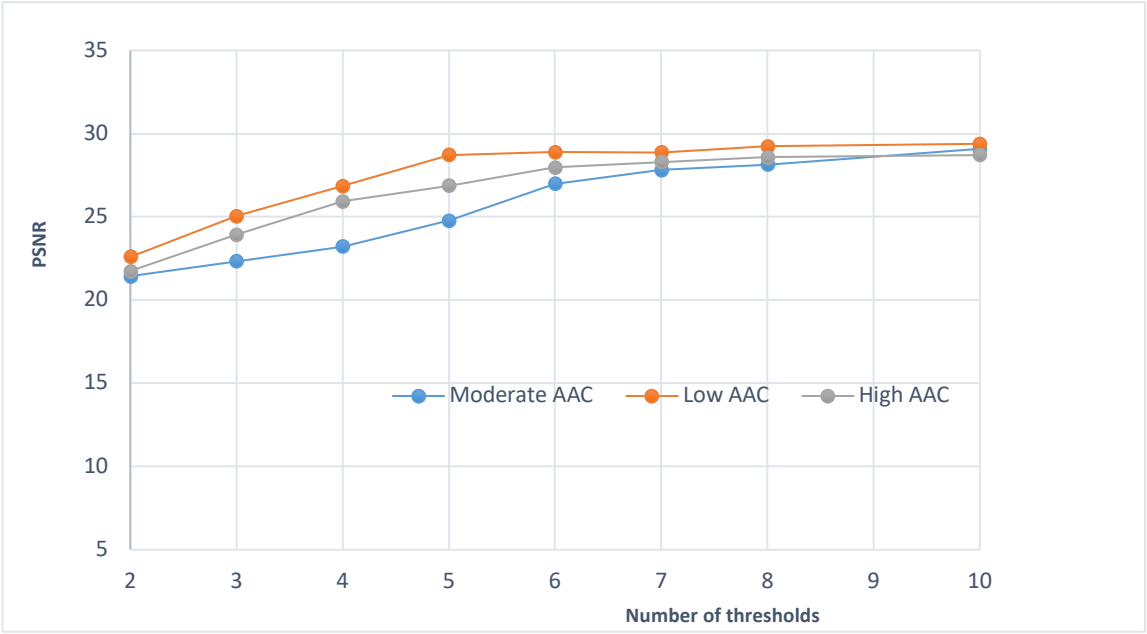


Figure 5-11: Plots of PSNR vs. number of thresholds for three different calcified aortas: moderate AAC; low AAC; high AAC

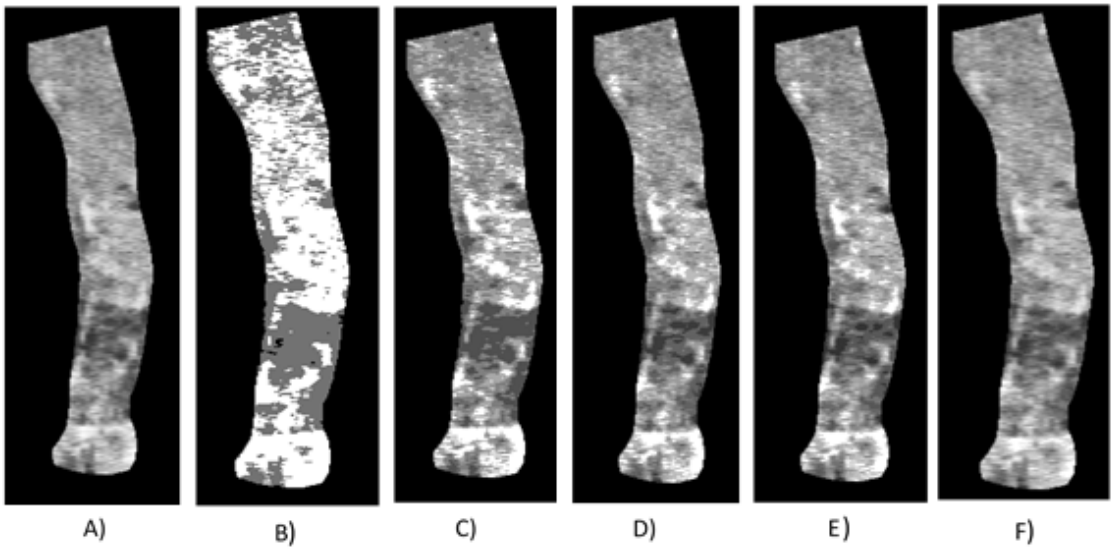


Figure 5-12: (A) original aorta image; (B) 2-level thresholding; (C) 4-level; (D) 6-level; (E) 7-level; (F) 14-level

5.9 **Quantification of AAC**

5.9.1 **Feature Extraction**

First-order texture analysis was implemented to extract the features of every segmented aorta. The main advantage of this approach is the simplicity of using standard descriptors (e.g. mean and variance). Regions that did not represent calcification in the aorta were removed. This was done by examining connected regions in the binary image were that had been automatically labelled. Any region consisting of fewer than 50 pixels was removed as it was most likely represented noise not calcification. After that, quantitative features related to the degree of the calcification were extracted from the processed aorta images.

The first feature extracted from each segmented aorta was the fraction of calcified pixels to the total number of pixels. Seven further features were obtained by computing the grey-level co-occurrence matrix (GLCM)) $P(i,j)$ (Honeycutt and Plotnick 2008; Preethi and Sornagopal 2014; Harefa and Pratiwi 2016). GLCM is a popular texture feature extraction technique due to its simplicity and efficiency in comparison to other methods such as wavelet transform. This method is used widely in many texture analysis applications (Zhuocai et al. 2011; Harefa and Pratiwi 2016). The additional features were: mean, variance, energy, entropy, homogeneity, contrast and correlation. The features were computed using the following equations (Jain et al.1995):

$$fraction\ of\ calcified\ pixels = \frac{number\ of\ calcified\ pixels}{total\ number\ of\ nonzero\ pixels} \tag{5.11}$$

$$Energy = \sum_{i,j} P(i,j)^2 \tag{5.12}$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j} \tag{5.13}$$

$$\text{Contrast} = |i-j|^2\, P(i,j) \tag{5.11}$$

$$\text{Homogenity} = \sum_{i,j} \frac{P(i,j)}{1+|i-j|} \tag{5.15}$$

$$\text{Entropy} = -\sum_{i,j} P(i,j)\log P(i,j) \tag{5.16}$$

The mean μ and variance σ are given by:

$$\mu_i = \sum_{i=1}^{n} i \sum_{i,j}^{n} P(i,j) \quad , \quad \mu_j = \sum_{j=1}^{n} j \sum_{i,j}^{n} P(i,j) \qquad (5.17,18)$$

$$\sigma_i = \sum_{i=1} (i - \mu_i)^2 \sum_{i,j}^{n} P(i,j) \quad , \quad \sigma_j = \sum_{j=1} (j - \mu_j)^2 \sum_{i,j}^{n} P(i,j) \qquad (5.19,20)$$

Figure 5-13 shows an example of a calcified aorta in the original VFA image and after segmentation, quantisation and thresholding to reveal pixels containing calcification. On the binary image, pixels that had value of 1 were labelled as calcified. The binary image shown in Figure 5-13 (C) was used to calculate the feature represented by the fraction of calcified pixels to the total number of pixels. Another example of the calcified aorta is shown in Figure 5-14.

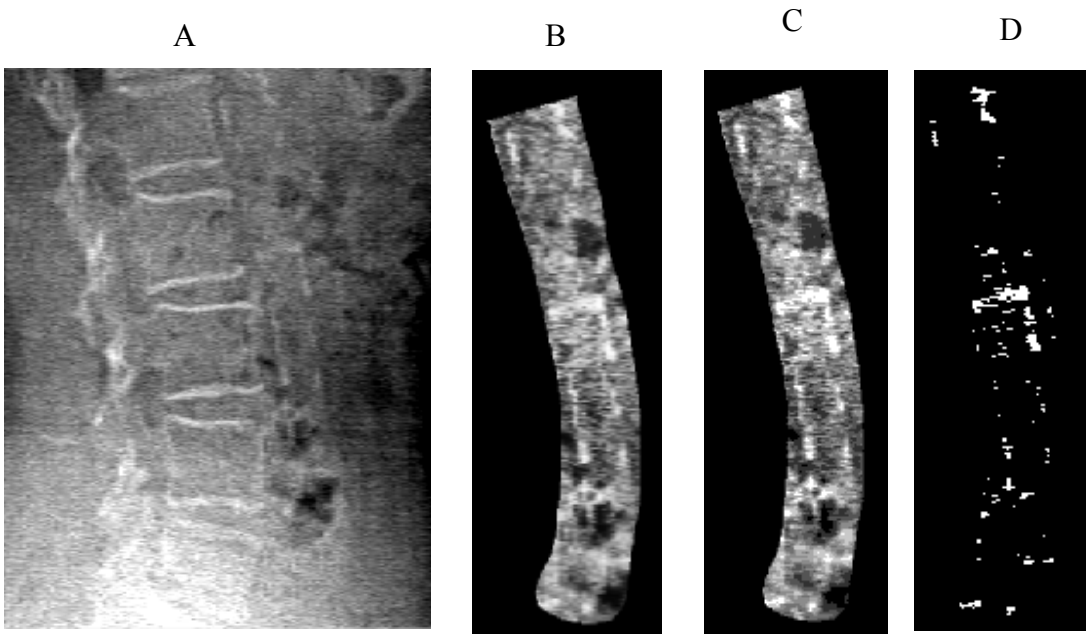


Figure 5-13: (A) Original image; (B) segmented aorta, (C) segmented aorta quantised using 7 threshold levels; (D) binary image after thresholding

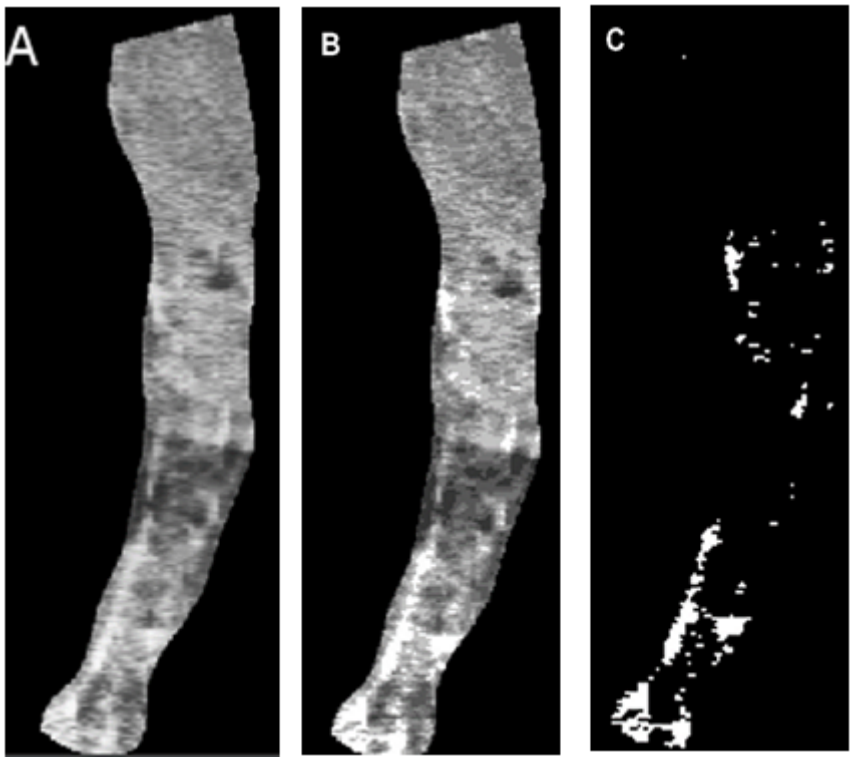


Figure 5-14: (A) Segmented aorta; (B) segmented aorta quantised using 7 threshold levels, (C) binary image after thresholding

5.10 **Automatic Calcification Classification**

The prediction stage is built on two classifiers trained on the training data: Support vector machine (SVM) using linear kernel from LIBSVM and K- nearest neighbour (K-nn). K-nn is a regular non-parametric and widely used classification method. This method is as called non-parametric because it does not require information about the statistical properties of pixels (Singh 2016) It is suitable for multi-modal classes and in the applications where an object can have many labels.

SVMs are powerful data classification algorithms, which are able to provide high accuracy (Chih-Wei Hsu et al. 2008; Singh 2016). SVMs belong to the general category of kernel methods. A kernel method is an algorithm that depends on the data through dot-products (Kotsiantis 2007; Orrù et al. 2012; Harefa and Pratiwi 2016).

A linear kernel has been reported to provide the best performance in many applications and requires only one parameter to be tuned (Walker 2010). In essence, SVMs are two-class classifiers but multiclass problems can be solved with multiclass extensions. A standard method is the so-called one-versus-the-rest approach, where for every class a classifier is trained to compare that class against the rest of the classes.

The above classifiers were used to evaluate the automatic method for AAC quantification against visual scoring. The classification stage was done using the Orange software package (Demšar et al. 2013). A K-folds cross validation method was implemented with 5 folds to partition the data set into two parts: a set to train the classification model and a set to validate the model. The output of this implementation was the assignment of each aorta image into one of three classes based on the values of the extracted features. Again, the three classes used were as follows: class 1 for mild AAC, class 2 for moderate AAC and class 3 for severe AAC.

5.10.1 Feature selection

In general, the performance of any automatic classifier is not optimised when all features are used. In order to assess the relative importance of the 8 extracted features, the s-fold cross-validation popular method was used for estimating the correct classification rates (CCR). In this method, the available data set is divided into a sub-set used for classifier design (i.e., the training set) and a sub-set used for testing the classifier (i.e., the test set) (Ververidis and Kotropoulos 2008). Figure 5-15 shows the CCR for the individual features: 1 ratio of calcified area/total area, 2 entropy, 3 variance, 4 energy, 5 correlation, 6 mean, 7 contrast and 8 homogeneity. The first 5 features were selected for the classification stage as they gave the best accuracies were achieved by using these features.

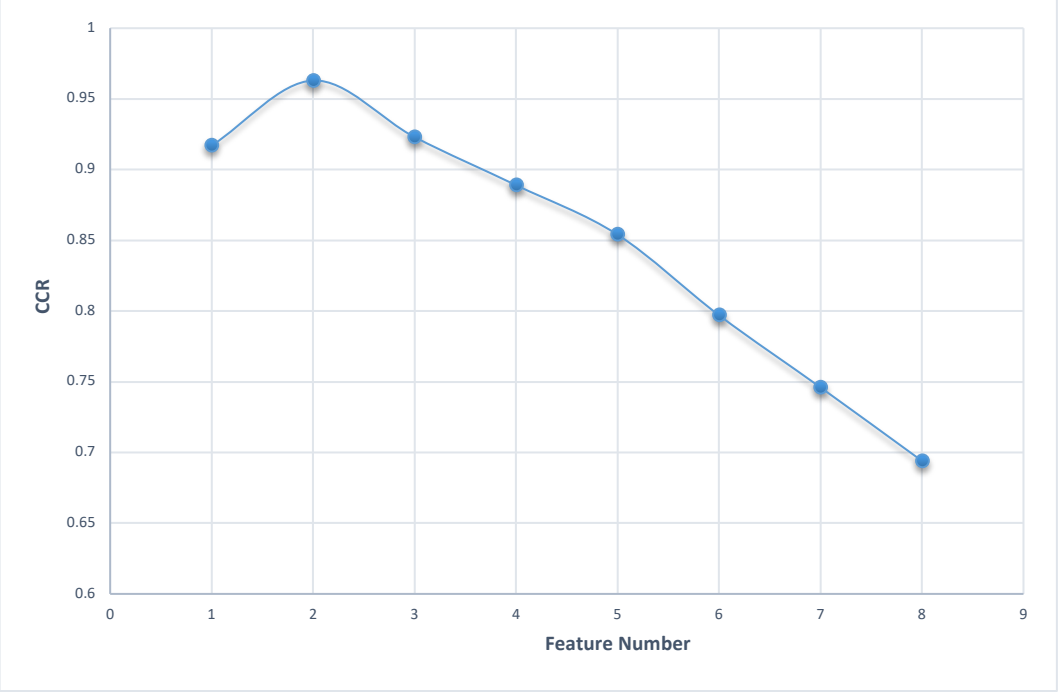
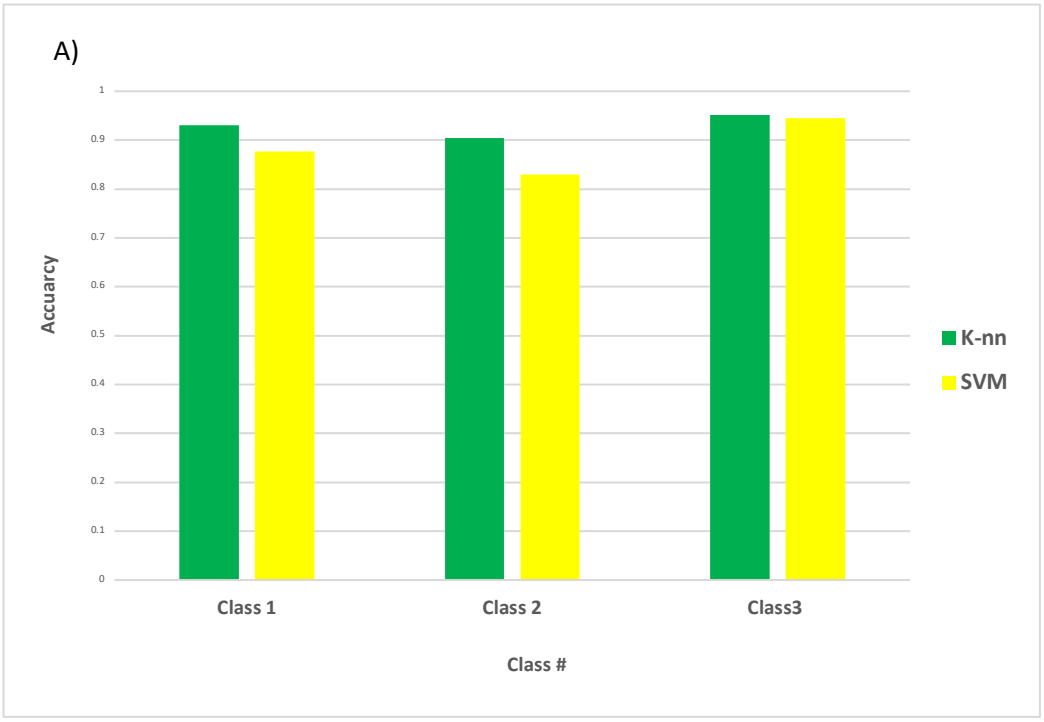


Figure 5-15: Variation of CCR with feature number

5.11 Comparison of Automatic and Manual Classification

The results of automatic AAC classification compared with manual visual classification are shown in Figure 5-16. The accuracy of K-nn (93.1%, 90.4% and 95.2% for classes 1, 2 and 3 respectively) was better than that of SVM (87.7%, 83.3% and 94.5%) as shown in Figure 5-16(A). The performance of the two automatic classifiers was also assessed by receiver operating characteristic (ROC) analysis. Figure 5-16(B) shows the area under the curve (AUC) of sensitivity vs. (1 – specificity) for each classifier. Similar performance was found for all classes.



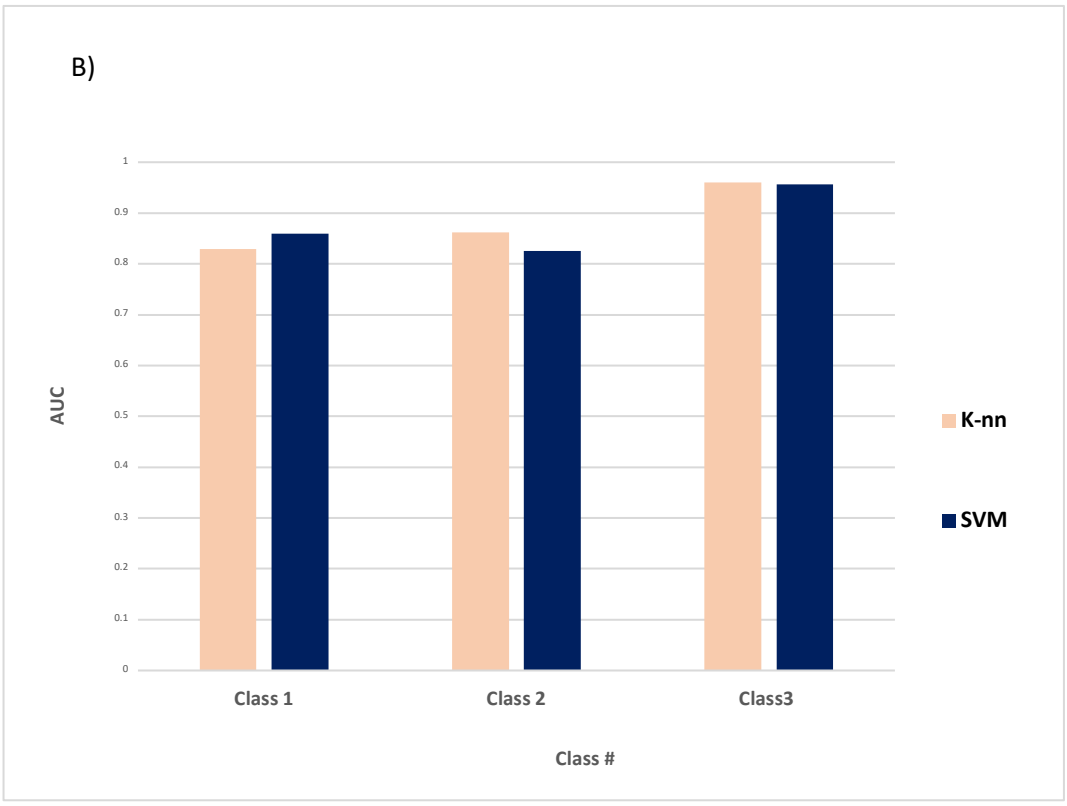


Figure 5-16: (A) accuracy of K-nn and SVM classifiers for AAC classes 1, 2 and 3; (B) AUC for each classifier and AAC class

5.12 **Summary**

This chapter has presented a method for measuring the severity of abdominal aortic calcification automatically in VFA images acquired by DXA in SE mode.

The automatic method consists of two stages. In the first stage an active appearance model has been employed for automatic segmentation. The model was trained on 20 VFA images and tested on another 53 unseen VFA images. AAM was able to extract two relevant objects (the aorta and the spine). The method was robust, i.e. it worked for all test images, and it took only 23 seconds to complete the searching process with 100 iterations

To quantify AAC automatically, a new method based on popular classification techniques was developed and validated against an established manual technique (AC-24). For the selected data set of 73 images, the automatic method achieved high accuracy for all classes of AAC severity.

In the next chapter, the work is extended to include VFA images from patients who have little or no aortic calcification and validated using a large dataset consisting of 390 VFA images.

Evaluation of Automatic Method for Detection and Quantification of AAC in VFA Images

6.1 Introduction

The development of an automated system for the detection and quantification of AAC based on the analysis of VFA images requires a large dataset for the training and validation of the proposed system.

This chapter evaluates the automatic approach developed in Chapter 5 for AAC detection and quantification in VFA images acquired by Hologic DXA scanners in single energy mode. The preliminary study proposed an algorithm, which was tested on 73 VFA images with clear evidence of calcification. The selected images included three calcification degrees: mild, moderate and severe.

Considering that this algorithm allowed automatic segmentation and quantification of AAC in images with three degrees of AAC, it was hypothesised that the same approach would enable assessment of AAC in a large number of VFA images including those not showing any calcification.

This chapter presents a validation of the automatic system for all possible AAC categories on a large dataset consisting of several hundred images.

6.2 Study Population

A set of VFA images from 390 female patients was selected; these included 69 females from the first data set. The additional images were retrieved sequentially by patient name

in alphabetical order from the DXA scanner archives. Those images with clear artefacts, such as metal inside the body, were excluded. Only the first image for each patient was taken; no follow up images were included. Each image was anonymised and given a study number. The Cardiff University School of Engineering Ethics Committee approved this retrospective study.

The VFA images were of patients referred for the investigation of osteoporosis between 2010 and 2017 at the University Hospital of Wales, Cardiff. Images from only females were used because the large majority of patients referred for DXA are female. The images were acquired by three Hologic DXA scanners: Discovery A (S/N70902) in the period between 2010 and 2012, Discovery A (S/N86224) in the period between 2012 and 2015 and Horizon A (S/N200253) in the period between 2016 and 2017. Images in this set had different degrees of aortic calcification: no-calcification, mild, moderate and severe calcification. Examples that have been cropped for the abdominal region and whose contrast has been enhanced by windowing are shown in Figure 6-1. A total of 15 images with anatomical or technical artefacts (e.g patient movement) or inadequate space anterior to the aorta for AAC assessment were excluded. Figure 6-2 provides some examples of the excluded images. Table 6-1 presents the characteristics of the remaining data set of 375 images.

Table 6-1: Characteristics of the final study data set

Study Population n= 375		
	Estimated Mean	Standard deviation(SD)
Age, y	74.17	13.32
Height, cm	154.94	7.18
Weight, Kg	61.07	11.45

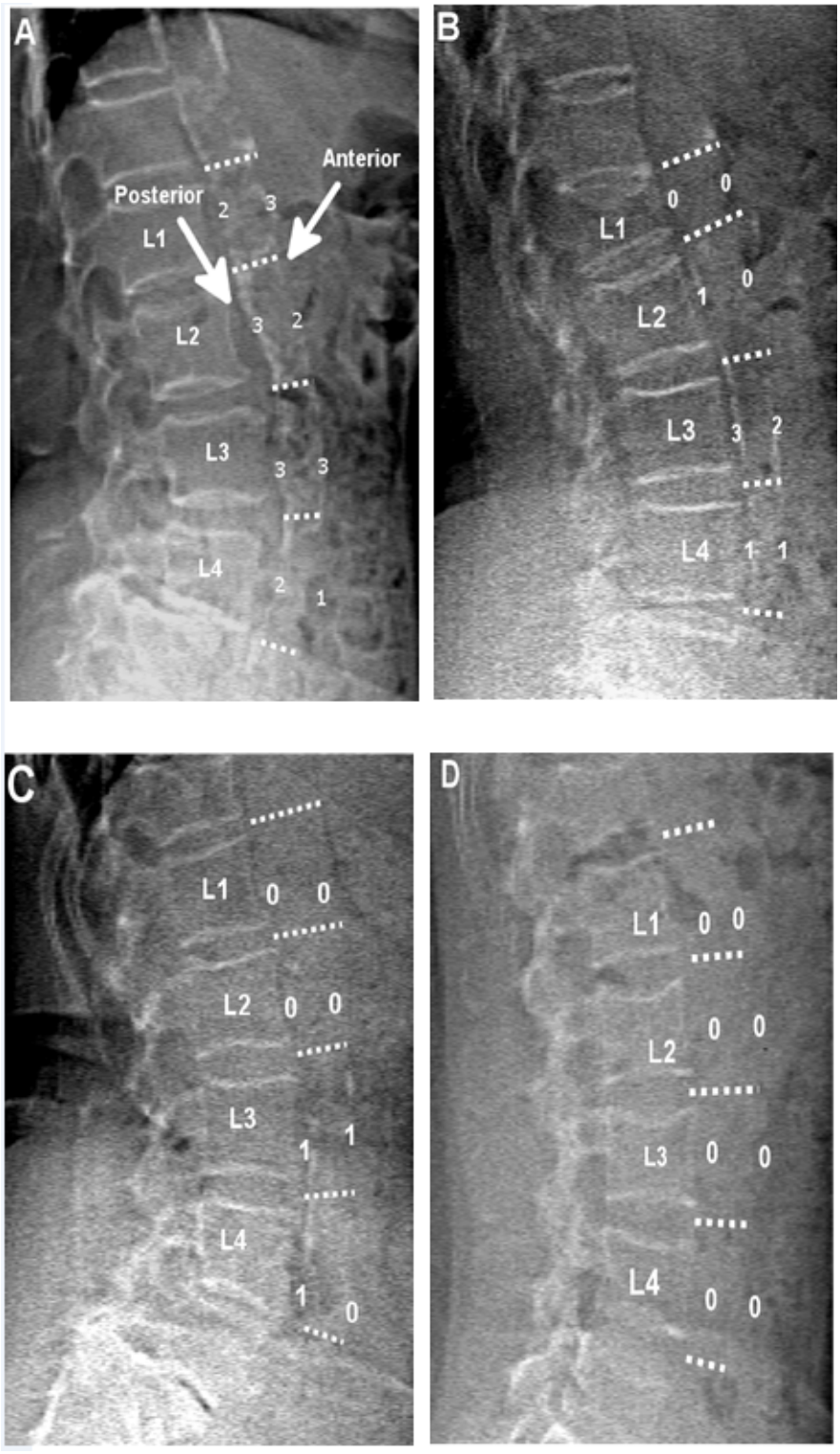


Figure 6-1:.Examples of VFA images obtained by DXA: (A) a 91- year-old patient with highly calcified aorta (AC-24 =19), (B) a 74-year-old-patient with moderate calcified aortic walls (AC-24= 8), (C) an 82 year-old- patient with mild calcified aorta (AC-24= 3), (D) a 38 year-old- patient with no aortic calcification (AC-24=0)

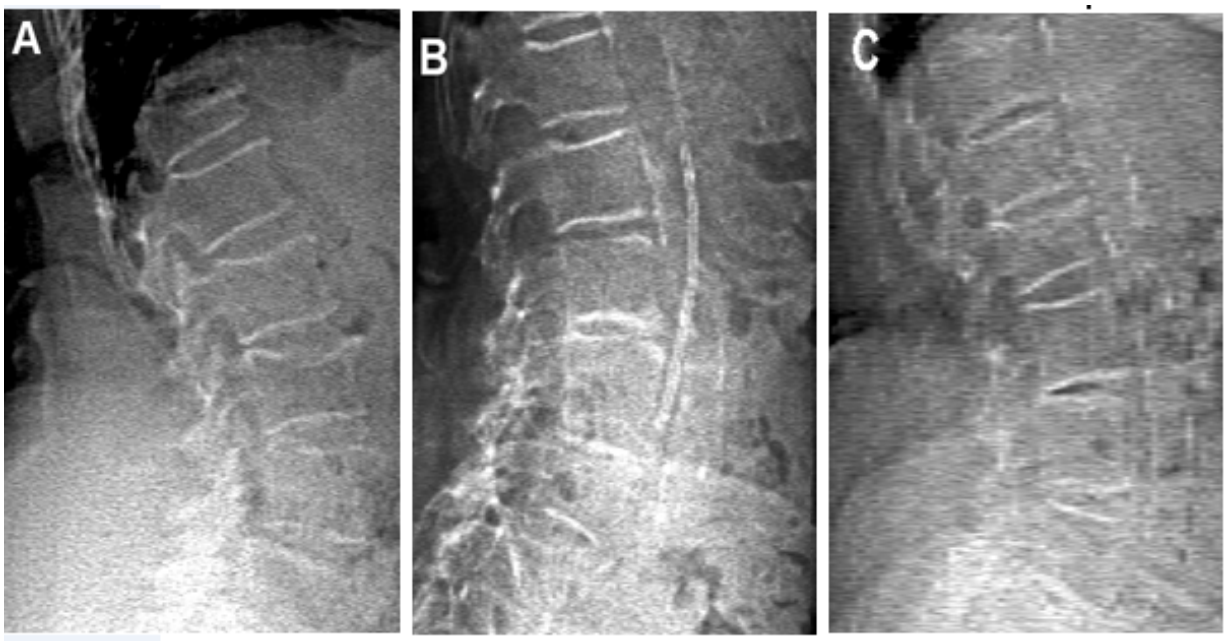


Figure 6-2: Examples of VFA images excluded from the dataset due to: (A) inadequate space anterior to the aorta; (B) anatomical artefacts, (C) artefacts due to patient movement

6.3 **Variability of AAC Measurement in VFA Images**

The manual scoring system AC-24 was implemented as described in Chapter 2 and AAC in each VFA image was categorised as explained in Section 5.4. In addition to the previous three categories (mild, moderate and severe AAC), a new category corresponding to an AC-24 score of 0 was added. This was done to examine whether the algorithm is suitable for patients with no calcification, since this is true in many cases.

Two tests of the variability of AC-24 scores, with respect to data observers were made. These were the variability across multiple observers (inter-observer variability), and variability of a single observer (intra-observer variability) (Mchugh 2012).

The presence and severity of AAC were assessed by two blinded readers. Reader A recorded AC-24 scores twice for the whole set of 375 VFA images. Reader B recorded AC-24 scores twice on a sub-set of 100 VFA images selected randomly. Images were viewed with Matlab (Mathworks, MA, USA) and the scores were recorded on Microsoft Excel spread sheet (2013).

Intra-class correlation coefficients (ICC) between the two sets of AC-24 readings and AAC categories were calculated for each reader individually. The ICC is an index of repeated measures. It is widely accepted and has been used in many studies to examine the intra-observer variability of AAC scores in radiographs and VFA images (Lewis et al. 2016; Grant et al. 2017; Schousboe et al. 2017).

The inter-observer agreement of AAC scores between reader A and B was measured as a simple percent agreement and inter-class correlation coefficient ICC. The Pearson's correlation coefficient is a statistical test that measures the statistical relationship, or association, between two continuous variables. It is regarded as the best method of measuring the association between variables of interest because it is based on the method of covariance. The pearson's correlation coefficient was used to assess the correlation between the two readers AC-24 scores.

In addition, inter-class variability for AAC categories was assessed using cohen's Kappa (k) cross tabulations algorithm statistics (Nelitz et al. 1999; Davies and Fleiss 1982). The kappa algorithm was designed particularly as a measure of agreement between two readers for categorical rating; it incorporates correction for chance agreement (Banerjee 1999; Mchugh 2012). Kappa has been applied by other researches to evaluate the inter-observer variability in the measurement of the degree of vertebral fracture (Hospers et al. 2009; Drampalos 2015).

Generally, a kappa coefficient of 0–0.20 is considered poor agreement, 0.21–0.4 is fair, 0.41–0.6 is moderate, 0.61–0.8 is strong, and > 0.8 is considered near-complete

agreement (Mchugh 2012; Ridge et al. 2016). Calculation of Cohen's kappa may be performed according to the following formula:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \qquad (6.1)$$

where Pr(a) represents the actual observed agreement and Pr(e) represents chance agreement. The confidence interval (CI) was calculated to assess the precession of k estimation and the probability value $(p\ value)$ was used to test the significance of this estimation. A p value of < 0.05 was considered statistically significant, as most public health professionals use this value as a standard.

The above statistical analysis was computed using SPSS (IBM, New York USA), with 95% confidence intervals CIs.

6.3.1 Intra-observer variability of calcification scores

Table 6-2 summarises the intra-class correlation for the two readers. Reader A showed high ICC > 0.98 for both AC-24 scores and AAC categories over the entire dataset. The ICC of Reader B on 100 images also was > 0.95 for both AC-24 scores and AAC categories.

Table 6-2: Intra- observer variability for two readers for AC-24 scores and AAC categories (ICC with 95% CI)

	Number of VFA (n)	**Intra-observer variability**		***P* value**
		AC-24 score	4 AAC categories	
Reader A	369	0.988 (0.985 to 0.991)	0.986 (0.982 to 0.989)	<0.001
Reader B	100	0.960 (0.940 to 0.973)	0.953 (0.929 to 0.968)	<0.001

6.3.2 **Inter-observer Variability of Calcification Scores**

There was a strong and significant linear correlation between the AC-24 scores given by the two readers with a Pearson's correlation coefficient of ($r^2 = 0.956, p < 0.001$) Figure (6-3).

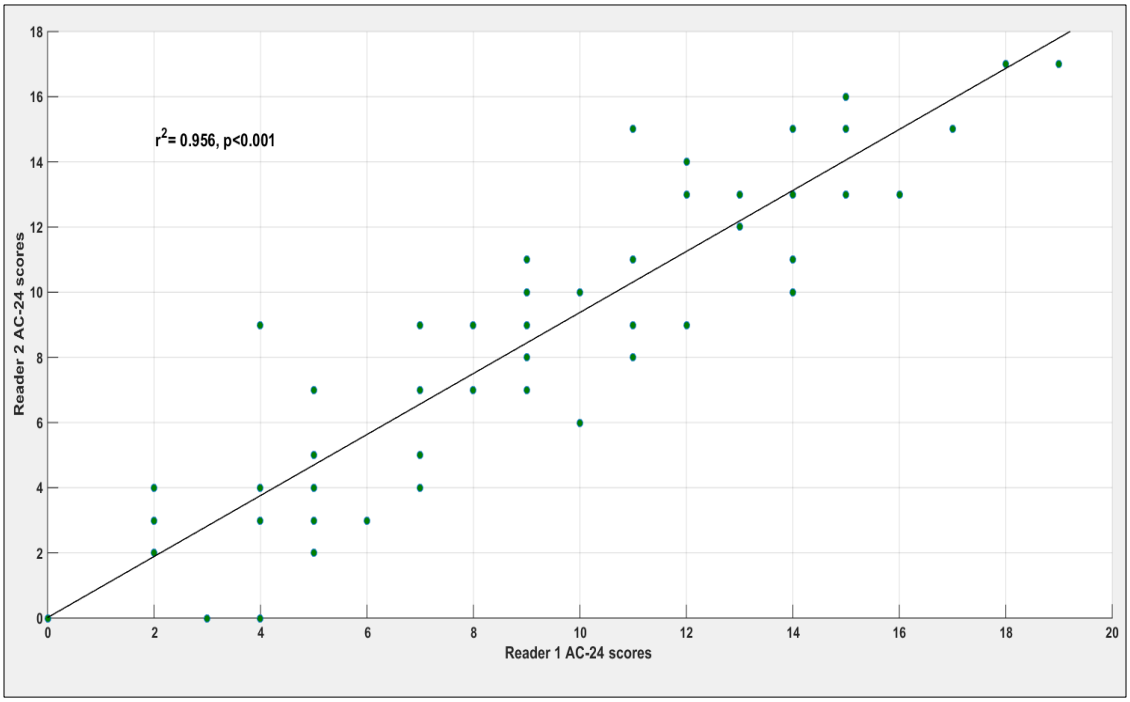


Figure 6-3: The correlation of AC-24 scores between two readers

The other results of the inter-observer variability tests (ICC, kappa correlation and percent agreement) as shown in Table 6-3.

Table 6-3: Inter-observer agreement measured for two Readers for AC-24 and AAC categories with 95% CI

Method	**Reader A vs Reader B n =100**				
	AC-24 scores	**4 AAC categories**			
ICC inter class correlation	0.953 (0.931 to 0.969)	0.934 (0.903 to 0.956)			
Kappa correlation (κ)	-	0.634 (0.505 to 0.743)			
Percentage of agreement	-	Class 0	Class 1	Class 2	Class 3
		88.9%	73.1%	64.5%	63.0%

There was good agreement between the two readers with ICC >0.9. The intra-observer reliability between the two observers was > 0.9 for both AC-24 scores and AAC categories.

In addition, the percent inter-observer agreement was high in both cases. This agreement was 89% on images with no AAC, while it was 63% on images of high AAC. Overall, there was strong agreement between the two readers with $k = 0.634$ as regards AAC categories.

6.4 **Relationship between Patient Age and AAC**

As outlined in Section 2.3, there is a strong association between patient age and the presence and extent of AAC (Okuno et al. 2007; Schousboe et al. 2007; Allison et al. 2008; Honkanen et al. 2008; Grant et al. 2017; Schousboe et al. 2017).

The association of AAC severity score with age was analysed using linear regression analysis for reader A. Reader A had scored the entire data set with ICC >0.98 for AC-24 score and AAC category.

Figure 6-4 shows that, overall, AAC score increased with age and that there was a strong and significant relationship between them. No calcification was observed in patients aged ≤ 40 years; conversely, the majority of patients aged > 40 years were given scores indicating some degree of calcification in this dataset. The Pearson's correlation coefficient and 95% CIs between AC-24 scores were, $0.602\ (0.543\ to\ 0.661, p\ < 0.001)$. Similar results have been obtained in many other studies (Honkanen et al., 2008; Kim et al. 2013; Leckstroem et al. 2014; Grant et al. 2017). Furthermore, about 56% of patients had AC-24 score of >1 at level L4; this percentage was slightly less at levels L3 to L1: 48%, 41% and 31% respectively. Overall, the number of posterior aortic wall segments that were given AC-24 scores from 1-3 was greater than that for the anterior aortic wall.

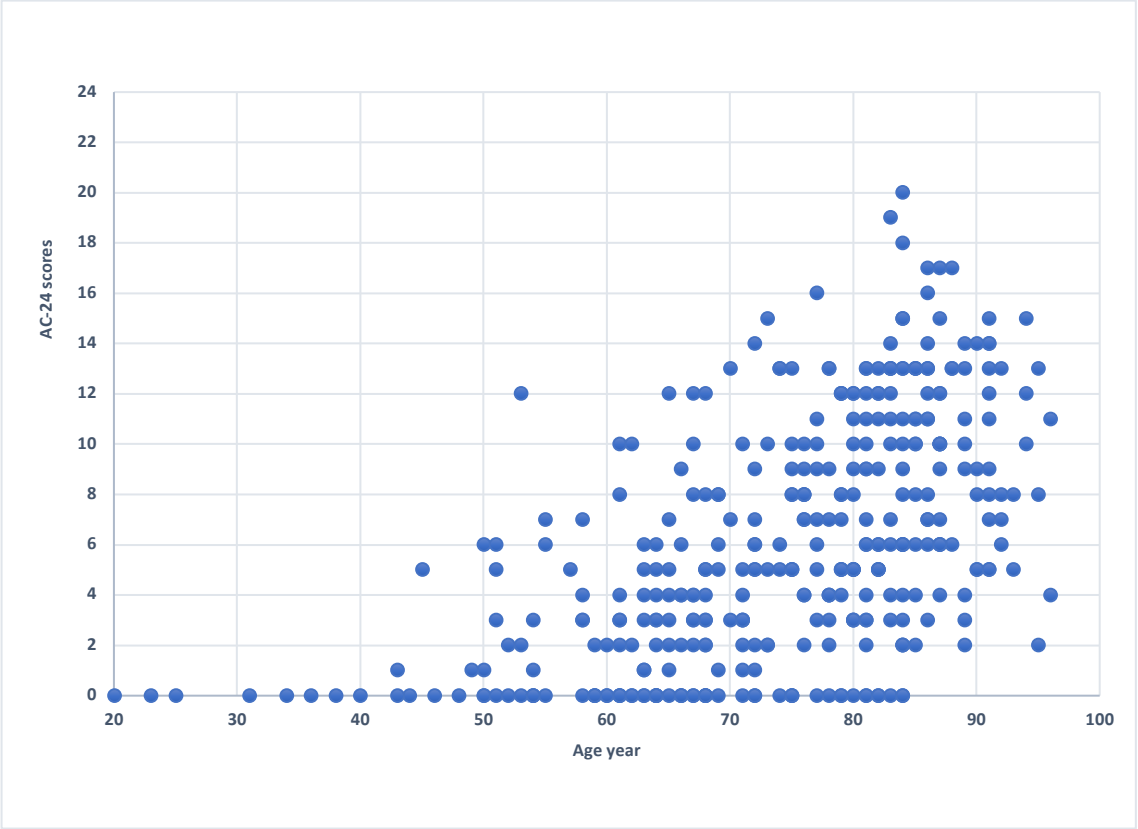


Figure 6-4: Scatter plot between AC-24 score and age

6.5 Evaluation of the Automatic method for AAC detection and Quantification

6.5.1 Manual Categorisation of AAC Severity

The total number of aortic regions in VFA images that were successfully segmented by the automatic method was 369 out of 375. Further analysis was restricted to this set of 375.

The distribution of AAC severity as assessed by reader A using the AC-24 scoring system is shown in Figure 6-5. The largest group (127) comprised those images with no AAC while the smallest (64) comprised those with severe AAC. There were 93 and 85 instances of mild and moderate AAC respectively.

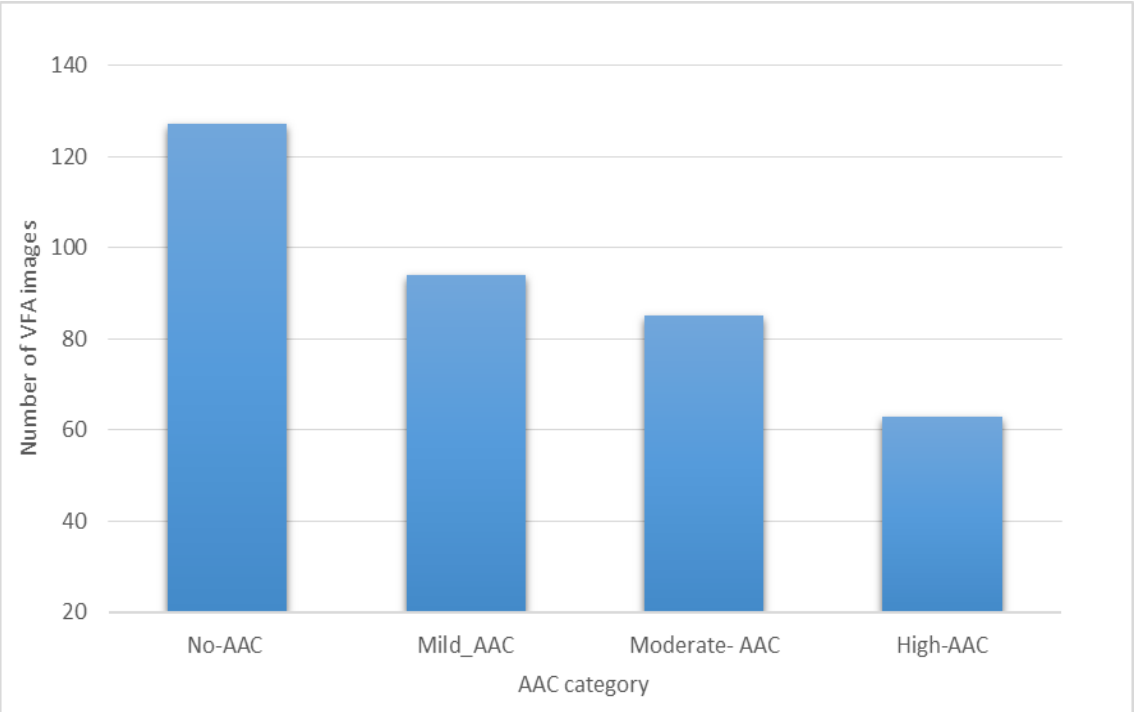


Figure 6-5: Distribution of AAC severity by reader A based on the AC-24 system: no AAC (0), mild AAC (1-4), moderate AAC (5-12) and high AAC (>12)

6.5.2 Classification Accuracy and Feature Selection

6.5.2.1 Classification into 4 AAC Categories

The performance of three classifiers, namely support vector machine (SVM), K-nearest neighbour (K-nn) and random forest (RF), was compared at this stage. The data set was partitioned using 5 folds cross validation into a training group to train the classifiers and a test group to validate the algorithm.

The classification stage was done using the Orange software package (Demšar et al. 2013). In this analysis, a step to rank the features was applied to evaluate their significance. Features were extracted as described in section 5.9.1 and ordered after applying the ranking.

The classification accuracy obtained with these features was tested by adding one feature at a time in the rank order. The achieved accuracy was recorded and each calculation was repeated with different 5-folds 10 times. Figure 6-6 shows the weighted average classification accuracy of the three classifiers as a function of the number of features used.

The accuracy of SVM increased with the number of features to reach 89% using the first 6 features. RF accuracy increased until 7 features had been used but decreased slightly when the eighth was added. For K-nn, accuracy received its maximum value with 4 features and remained constant until the addition of the eighth feature where it decreased by about 3%. Figure 6-7 shows the effect of increasing the value of K on average weighted accuracy for the K-nn classifier; the highest accuracy was achieved with K=5.

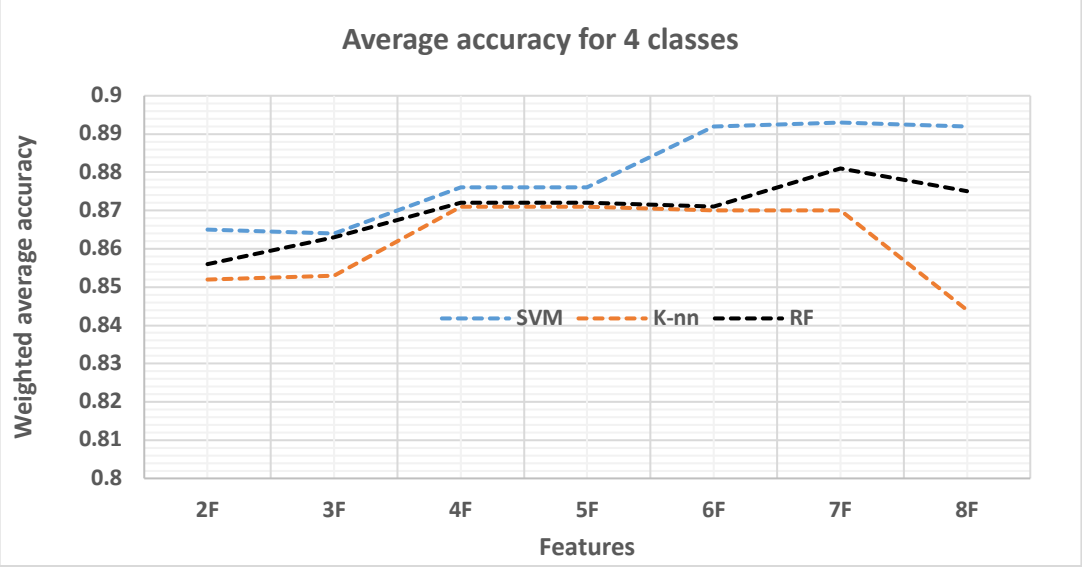


Figure 6-6: Variation of weighted average classification accuracy with the number of features for the three classifiers based on four categories of AAC severity

The optimal weighted average accuracy, sensitivity and specificity of each AAC category obtained by the three classifiers are listed in Table 6-4. Overall, the best accuracy, sensitivity and specificity were achieved with the SVM classifier.

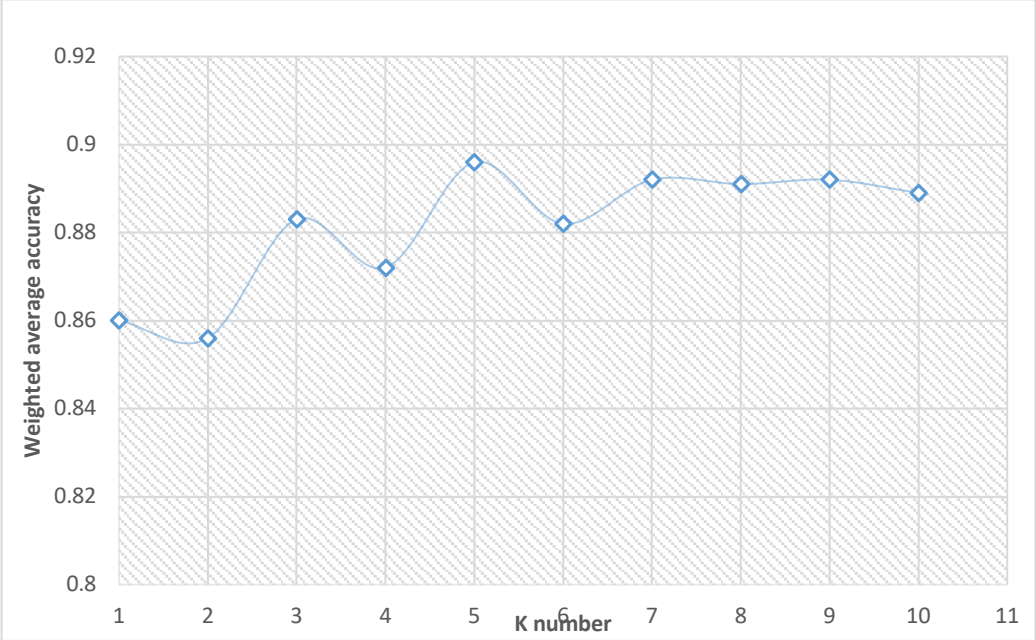


Figure 6-7: Weighted average accuracy obtained by K-nn for different K values

An unpaired t-test revealed a significant difference between the average accuracy achieved by SVM and K-nn: $0.022\,(\mathrm{CI},\ 0.0169 - 0.027, \mathrm{p} < 0.001)$, SVM and RF: $0.0116\,(\mathrm{CI}, 0.0044 - 0.0189, \mathrm{p} = 0.0033)$ and also RF and K-nn, $0.0103\,(\mathrm{CI}, 0.0038 - 0.0168, \mathrm{p} = 0.0043)$.

Table 6-4: Accuracy, sensitivity, specificity and SD of three classifiers for 4 AAC categories

Classifier	AAC Category	Accuracy	Sensitivity	Specificity
SVM	Class 0	0.910± 0.008	0.855± 0.012	0.940± 0.002
	Class1	0.832± 0.011	0.675± 0.017	0.877± 0.006
	Class 2	0.891± 0.008	0.743± 0.023	0.933± 0.004
	Class 3	0.950± 0.003	0.864± 0.018	0.946± 0.006
	Weighted Average	0.892± 0.048	0.785± 0.140	0.923± 0.074
K-nn	Class 0	0.916± 0.004	0.854± 0.010	0.954± 0.003
	Class1	0.810± 0.009	0.628± 0.016	0.867± 0.008
	Class 2	0.830± 0.007	0.678± 0.084	0.868± 0.004
	Class 3	0.912± 0.007	0.761± 0.017	0.917± 0.007
	Weighted Average	0.872± 0.056	0.740± 0.113	0.906 ±0.012
RF	Class 0	0.911± 0.026	0.852± 0.007	0.941± 0.011
	Class1	0.821± 0.009	0.627± 0.044	0.866± 0.009
	Class 2	0.861± 0.007	0.644± 0.045	0.901± 0.013
	Class 3	0.933± 0.005	0.797± 0.042	0.92± 0.0140
	Weighted Average	0.881± 0.052	0.740± 0.123	0.910± 0.067

The SVM classifier correctly predicted 112 cases out of 127 for class 0 and 57 cases out of 64 for class 3 with sensitivity and specificity > 85%. However, there was misclassification for classes 1 and 2. Tables 6-5 - 6-7 show the confusion matrices for the SVM, K-nn and RF classifiers respectively.

Table 6-5: Confusion matrix obtained using the SVM classifier for 4 AAC categories

Actual	Predicted					
		0	1	2	3	Total
	0	112	14	1	0	127
	1	19	61	11	2	93
	2	0	10	64	11	85
	3	0	2	5	57	64
	Total	131	87	81	70	369

Table 6-6: Confusion matrix obtained using the K-nn classifier for 4 AAC categories

Actual	Predicted					
		0	1	2	3	Total
	0	117	10	0	0	127
	1	21	52	18	2	93
	2	0	21	50	14	85
	3	0	1	9	54	64
	Total	138	84	77	70	369

Table 6-7: Confusion matrix obtained using the RF classifier for 4 AAC categories

Actual	Predicted					
		0	1	2	3	Total
	0	113	12	2	0	127
	1	20	55	14	4	93
	2	0	18	55	12	85
	3	0	2	10	52	64
	Total	133	87	81	68	369

6.5.2.2 **Classification into 3- AAC Categories**

To improve their sensitivity, groups mild and moderate (class 1 and class 2) were merged. The new group comprised VFA images with AC-24 scores in the range 1-12. Class 0 (no AAC) and class 3 (severe AAC) (1-12) remained unchanged as the goal was to validate the algorithm that had previously been developed.

Figure 6-8 shows variation of the weighted average classification accuracy for the 3 AAC categories with the number of features. Features were ranked and ordered as before.

The best weighted average accuracy, sensitivity and specificity for each AAC category obtained with the three classifiers are shown in Table 6-8.

Application of the unpaired t-test between three classifiers revealed non-significant differences in accuracy between three used classifiers: SVM and K-nn, $0.0043\ (\mathrm{CI}, -0.0014 - 0.010, p = 0.128)$; SVM and RF, $0.0028\ (\mathrm{CI}, -0.0046 - 0102\ p = 0.4403)$; K-nn and RF $0.0015(\mathrm{CI}, -0.005 - \ 0.0081, p = 0.63)$.

The overall accuracy obtained by the three classifiers was about 88%. However, both sensitivity and specificity were improved by merging the mild and moderate AAC classes. The confusion matrices for the SVM, K-nn and RF classifiers are shown in Tables 6-9 - 6-11.

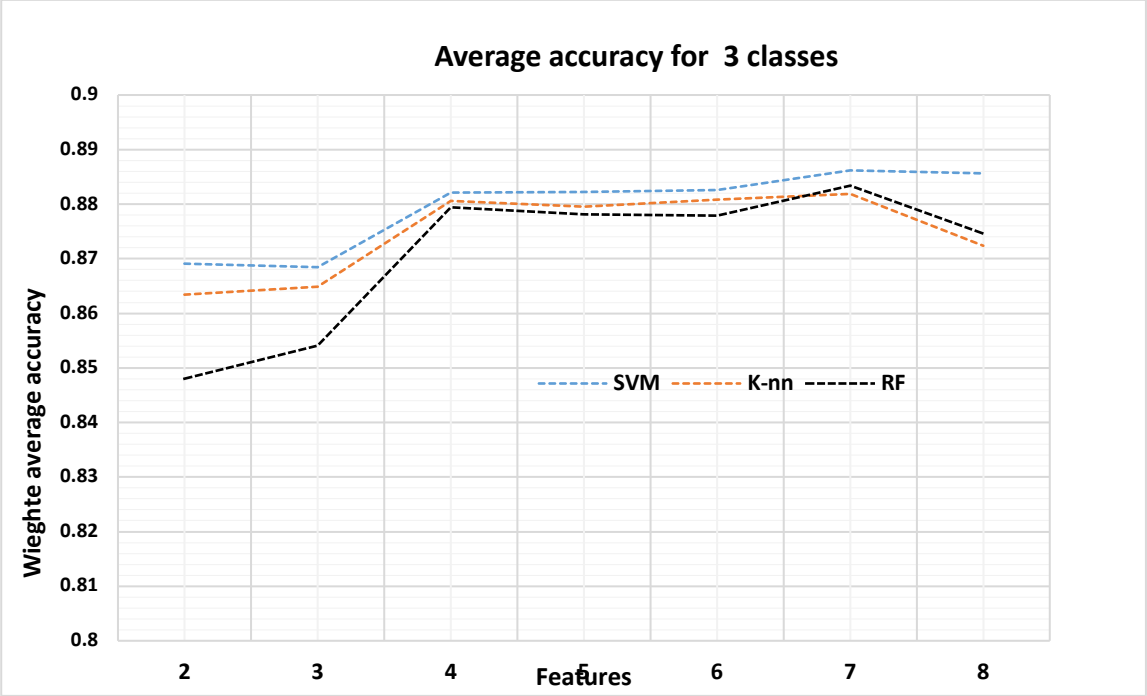


Figure 6-8: Variation of weighted average classification accuracy with the number of features for the three classifiers based on three categories of AAC severity

Table 6-8: Accuracy, sensitivity, specificity and SD of three classifiers for 3 AAC categories

Classifier	AAC Category	Accuracy	Sensitivity	Specificity
SVM	Class 0	0.914± 0.005	0.859± 0.009	0.945± 0.005
	Class 1&2	0.860± 0.007	0.858± 0.009	0.861± 0.008
	Class 3	0.944± 0.005	0.862± 0.015	0.943± 0.005
	Weighted average accuracy	0.886± 0.050	0.860± 0.003	0.904± 0.060
K-nn	Class 0	0.922± 0.002	0.863± 0.006	0.955± 0.003
	Class 1&2	0.840± 0.007	0.848± 0.001	0.834± 0.007
	Class 3	0.918± 0.006	0.771± 0.015	0.887± 0.006
	Weighted average accuracy	0.882±0.057	0.840± 0.045	0.885± 0.076
RF	Class 0	0.911± 0.008	0.843± 0.012	0.933± 0.009
	Class 1&2	0.845± 0.030	0.823± 0.016	0.857± 0.013
	Class 3	0.917± 0.032	0.840± 0.033	0.927± 0.009
	Weighted average accuracy	0.880±0.0450	0.833± 0.014	0.896±0.052

Table 6-9: Confusion matrix obtained using the SVM classifier for 3 AAC categories

	Predicted				
Actual		0	1 and 2	3	Total
	0	116	11	0	127
	1 and 2	17	153	8	178
	3	0	8	56	64
	Total	133	172	64	369

Table 6-10: Confusion matrix obtained using the K-nn classifier for 3 AAC categories

	Predicted				
Actual		0	1 and 2	3	Total
	0	113	14	0	127
	1 and 2	17	150	11	178
	3	0	10	54	64
	Total	130	174	65	369

Table 6-11: Confusion matrix obtained using the RF classifier for 3 AAC categories

	Predicted				
Actual		0	1 and 2	3	Total
	0	116	11	0	127
	1 and 2	16	149	13	178
	3	0	11	53	64
	Total	132	171	66	369

6.6 **Summary**

This chapter has presented a validation of the automatic technique that was developed for AAC detection and quantification as discussed in Chapter 5.

AAC scores obtained by the AC-24 system were strongly related to patient age. The pattern of AAC distribution was similar to the findings of many previous studies with the highest calcification score being at L4 level. The developed algorithm achieved high accuracy, sensitivity and specificity.

The present approach may be helpful in identifying patients with atherosclerosis before the development of symptoms of cardiovascular disease. It provides a fast (less than 1 minute) and low-cost assessment of the location and severity of AAC.

The next chapter presents the design and construction of a phantom to investigate the performance of VFA images acquired by DXA in the detection of AAC.

Phantom Design and Construction

7.1 Introduction

In the first part of this thesis, the ability of DXA to provide useful information on abdominal aortic calcification has been demonstrated. By analysing VFA images acquired using this modality, a new method to quantify AAC in the lumbar region has been developed and validated on a large data set with different degrees of calcification.

The aim of the second part of this study is to challenge the VFA technique under various conditions to investigate the capability of the scanner to detect calcification in its early stages when its thickness is small. Apart from calcification thickness, detectability may be affected by factors such as VFA energy mode and patient body width. A Perspex phantom was used to mimic abdominal soft tissue with a strip of aluminium inserted at its centre to mimic calcification.

The aim of this work was to examine the capability of VFA to detect a small uniform thickness of material that is equivalent to calcification. To the author's knowledge, such a systematic study has not been done previously.

In diagnostic radiology, phantoms are often made of Perspex and aluminium, where the Perspex is used to simulate soft tissue and the aluminium is used as a bone equivalent material. There is no commercial phantom that is suitable for this study. The Hologic spine phantom, which is used for QC, was not used for this work as it consists of a model of vertebrae L1 to L4 moulded from calcium hydroxyapatite mixed in epoxy resin and embedded in a homogeneous epoxy resin block.

In order to model both abdomen soft tissue and calcification of the aortic region, a new phantom of appropriate dimensions and made of appropriate materials was designed and constructed.

This chapter presents the phantom materials used in this study. The chapter also discusses the initial experiments that were made to estimate suitable phantom dimensions. These preliminary experiments were conducted using the Hologic Discovery A DXA scanner (S/N70902) at the University Hospital of Wales (UHW) and slabs of Perspex. Imaging VFA was done in single energy mode.

7.2 **Modelling of Abdomen Soft Tissue**

In diagnostic radiology, some measurements must be done with phantoms because it is impractical to make such measurements on human subjects. For example, test phantoms with tissue substitutes are used to estimate the radiation dose delivered to human organs and to evaluate image quality. Various types of phantom model have been made to represent specific human anatomy and phantom design has evolved to simulate the real patient (Wood et al. 2017). In practice, the radiological properties of phantom materials should be as similar as possible to those of tissue.

In recent years, many studies have addressed the design of experimental phantoms as a useful resource in x-ray modelling. Several radiation applications have been tested by employing phantoms of inexpensive design and material.

The transmission of x-radiation through a region of the body depends on the properties of tissue, such as effective atomic number Z_{eff} and physical density ρ. The effective atomic number Z_{eff}, of a material is the atomic number of a hypothetical element that attenuates photons to the same extent as the material. In Hologic DXA, the x-ray tube voltage is switched between 100 kV (as low energy) and 140 kV (as high energy) during half-cycles of the power supply (Shepherd 2009; Lorente-Ramos et al. 2011). In the

corresponding energy range, the dominant interactions are the photoelectric effect and Compton scattering. Generally, the mass attenuation coefficient τ_m for photoelectric absorption is proportional to Z_{eff}^3 and $1/E^3$, where E is the photon energy. Consequently, at low energy the photoelectric effect is the dominant interaction in a relatively high atomic number materials such as bone. On the other hand, Compton scattering is the dominant interaction in soft tissue, especially at high energy. The mass attenuation coefficient σ_e is proportional to electron density and is nearly independent of the atomic number of the attenuating medium.

7.2.1 **Phantom Material**

Water and plastic materials are frequently used for dosimetry phantoms as recommended by major dosimetry protocols (International Atomic Energy Agency (IAEA) 2000; International Committee on Radiation Units 2010). Water is used because human soft tissue contains about 70% of water.

In a study by Dae-Cheol et al. (2010), acrylic and water were used as tissue equivalent material to examine the effect of the thickness of soft tissue on bone density measured using DXA. The most popular acrylic plastic is polymethyl methacrylate (PMMA), which is sold under the brand names of Plexiglas, Lucite, Perspex and Crystallite. The elemental composition for PMMA, water and soft tissue are summarised in Table 7-1.

Table 7-1: Elemental composition of water, Perspex and soft tissue (Mihailescu and Borcia 2006; Singh et al. 2014)

Element Properties			Elemental mass fraction		
Symbol	Atomic number (Z)	Atomic weight	Water	Perspex	Soft Tissue[1]
H	1	1.008	0.112	0.081	0.1012
C	6	12.011	-	0.600	0.1110
O	8	15.999	0.888	0.320	0.7620
N	7	14.007	-	-	0.0260

[1] https://physics.nist.gov/cgi-bin/Star/compos.pl?matno=262

Perspex possesses similar x-ray attenuation properties to soft tissue and has been used in many studies (Anderson et al. 2000; Caldas et al. 2010; İrem et al. 2016). Perspex is used due to its availability and homogeneity and the fact that it is easy to machine. However, it is not exactly tissue equivalent as its physical density is greater than that of water and its effective atomic number is slightly less. Moreover, the relative electron density of Perspex to water is 1.16 (Standard Imaging 2013). The mass attenuation coefficients of water and Perspex at 70 keV are 0.1945 and 0.1836 cm^2 /g respectively (Lewis et al. 2012).

However, studies have shown that Perspex is suitable for simulating soft tissue (Kelly et al. 1998; Anderson et al. 2000; Palm et al. 2002; Losasso 2005; Desai et al. 2010; Caldas et al. 2010; İrem et al. 2016). The effective atomic number Z_{eff} and other properties of soft tissue, Perspex and water are listed in Table 7-2. Z_{eff} is expressed by Hendee et al. (2003) as follows:

$$Z_{eff} = \left(a_1 z_1 + a_2 z_2 + a_3 z_3 + \cdots + a_n z_n \right)^{\frac{1}{2.94}} \quad (7\text{-}1)$$

where, $Z_1, Z_2, \ldots. Z_n$ are the atomic numbers of the elements in the material and $a_1, a_2 \ldots\ldots a_n$ are the fractional contribution of every element to the total number of electrons in the mixture. The chemical composition of Perspex is $(C_5O_2H_8)_n$ with a density of 1.19 g/cm^3 and Z_{eff} of 6.48 (Lewis et al. 2012; Ali et al. 2015).

Table 7-2: Properties of soft tissue, Perspex and water (Lewis et al. 2012)

Element	Effective atomic number Z_{eff}	Density (g/cm^3)	Mass attenuation (cm^2/gm)	
			40 keV	70 keV
Soft Tissue	7.47	1.00	0.2688	0.193
Perspex	6.48	1. 19	0.2430	0.1836
Water	6.60	1.00	0.2683	0.1945

7.2.2 **Phantom Dimensions**

The estimation of phantom width was based on human body waist and hip circumference, while the determination of the length and height of this phantom was based on the characteristics of the DXA scanner.

7.2.2.1 **Phantom Width**

In order to estimate the width of the abdominal region, the results of a study by Henderson et al. (2002) were used. This published work is a survey of a national sample of adults aged between 19 to 64 years; the survey aimed to gather information about the nutritional status of the British population. The research provided anthropometric data on a representative sample of 2,251 participants. The mean of two recorded measurements was taken. Agreement between these two measurements was checked and data were included only if the percentage difference was less than 15%.

According to Henderson et al. (2002), the waist is defined as the mid-point between the iliac crest and the lower rib. The hip circumference is defined as the maximum circumference over the buttocks and below the iliac crest. Waist and hip measurements

were taken from a total of 1,782 subjects, 808 men and 974 women. Subsequently, waist to hip ratios were calculated (Henderson et al. 2002).

In order to find the width W of the Perspex phantom that models soft tissue in the abdomen, this region was assumed to be elliptic. The circumference C of this ellipse (Figure 7-1A and 7-1B) was expressed using the common Ramanujan's Approximation Theorem II (Villarino 2005).

$$C = \pi\,(a + b)\left(1 + \frac{3h}{10\sqrt{4 - 3h}}\right), \tag{7-2}$$

where

$$h = \frac{(a - b)^2}{(a + b)^2} \tag{7-3}$$

and a and b are the semi-major axis and semi-minor axis of the ellipse respectively.

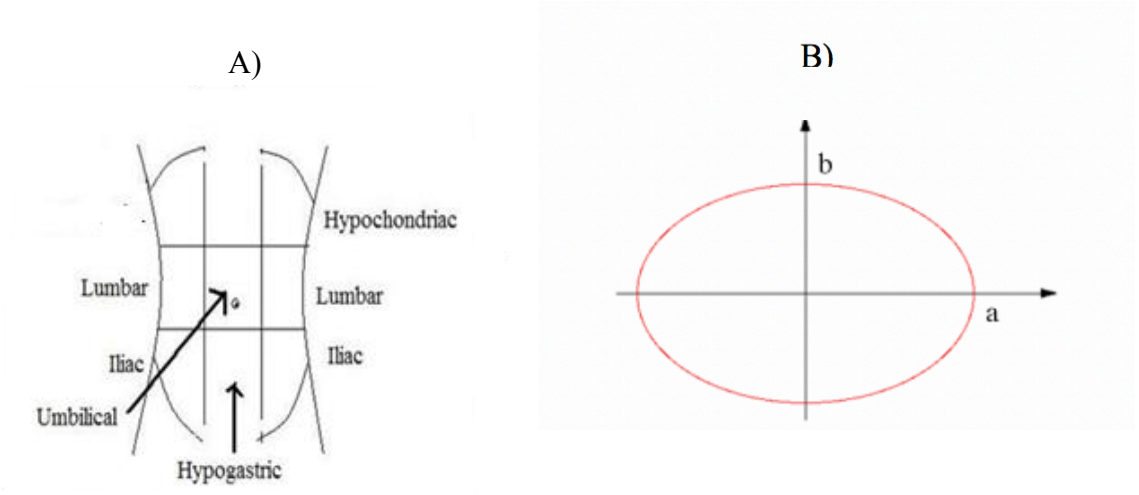


Figure 7-1: (A) abdominal region (B) ellipse shape.

The mean ± SD waist and hip circumference of men were 100 ±11.7 cm and 105 ±8.3 cm respectively. For women, the waist and hip circumference were 86 ±12.1 cm and 106 ±10.8 cm respectively. The range of body width was calculated from the average circumference of waist and hip for ± 1SD, ± 2SD and ± 3SD.

Body width at the lumbar spine region was assumed to be 2a. Taking into account the diversity of the shape of human the human abdomen in both genders, three assumptions were made to estimate body width: a=b (the minimum), a =1.5b and a=2b (the maximum). Table 7-3 and 7-4 show the results of this calculation for men and women respectively.

Table 7-5 summarises the minimum and maximum body width for both women and men. The calculation suggested that the phantom width should range between 20 and 50 cm.

Table 7-3: Estimated body width for men

	Waist Circumference (cm)	Hip Circumference (cm)	Average Body Circumference	Body Width a=b	Body Width a=1.5b	Body Width a=2b
Mean +3Sd	135.1	129.9	132.5	42.197	50.134	54.732
Mean+ 2Sd	123.4	121.6	122.5	39.013	46.351	50.601
Mean +Sd	111.7	113.3	112.5	35.828	42.567	46.471
Mean	100	105	102.5	32.643	38.783	42.34
Mean-1Sd	88.3	96.7	92.5	29.459	34.999	38.209
Mean-2Sd	76.6	88.4	82.5	26.274	31.216	34.078
Mean-3Sd	64.9	80.1	72.5	23.089	27.432	29.948

Table 7-4: Estimated body width for women

	Waist Circumference (cm)	Hip Circumference (cm)	Average Body Circumference (cm)	Body Width a=b	Body Width a=1.5b	Body Width a=2b
Mean +3Sd	122.3	138.4	130.35	41.513	49.321	53.844
Mean+ 2Sd	110.2	127.6	118.9	37.866	44.988	49.114
Mean +Sd	98.1	116.8	107.45	34.22	40.656	44.385
Mean	86	106	96	30.573	36.324	39.655
Mean-1Sd	73.9	95.2	84.55	26.927	31.991	34.925
Mean-2Sd	61.8	84.4	73.1	23.28	27.659	30.196
Mean-3Sd	49.7	73.6	61.65	19.634	23.327	25.466

Table 7-5: Maximum and minimum estimated body width estimated for men and women

Gender	Waist Circumference (cm) ±SD	Hip Circumference (cm) ± SD	Body Width (cm)	
			Minimum	Maximum
Men	100 ± 11.7	105 ± 8.3	23	55
Women	86 ± 12.1	106 ±10.8	20	54

After calculating the desired width, the Perspex phantom was constructed in several configurations to cover the range between 20 cm and 50 cm, taking into account the fact that an aluminium strip would need to be sandwiched between two halves of the phantom. The phantom was designed using slabs of Perspex of thickness 1 cm, 2.5 cm, 5 cm and 10 cm. Figure 7-2 shows the Perspex phantom and aluminium strip placed on the DXA scanner table. Figure 7-3 shows a schematic diagram of the Perspex phantom and its dimensions from three views: superior-inferior (S-I), anterior-posterior (A-P) and lateral.

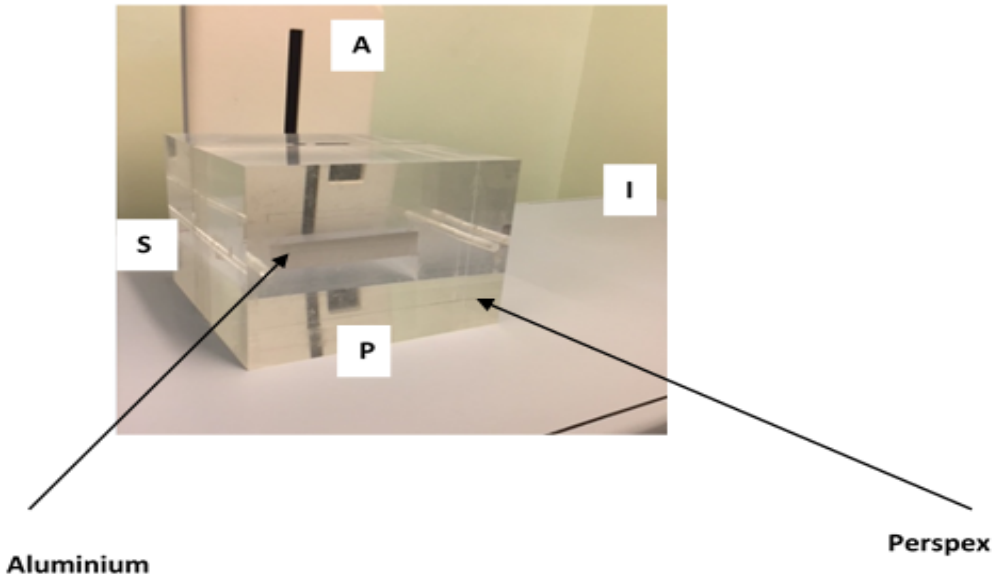


Figure 7-2: Perspex phantom with (aluminium strip) on the DXA scanner table: S=superior, I= inferior, A= anterior, P= posterior

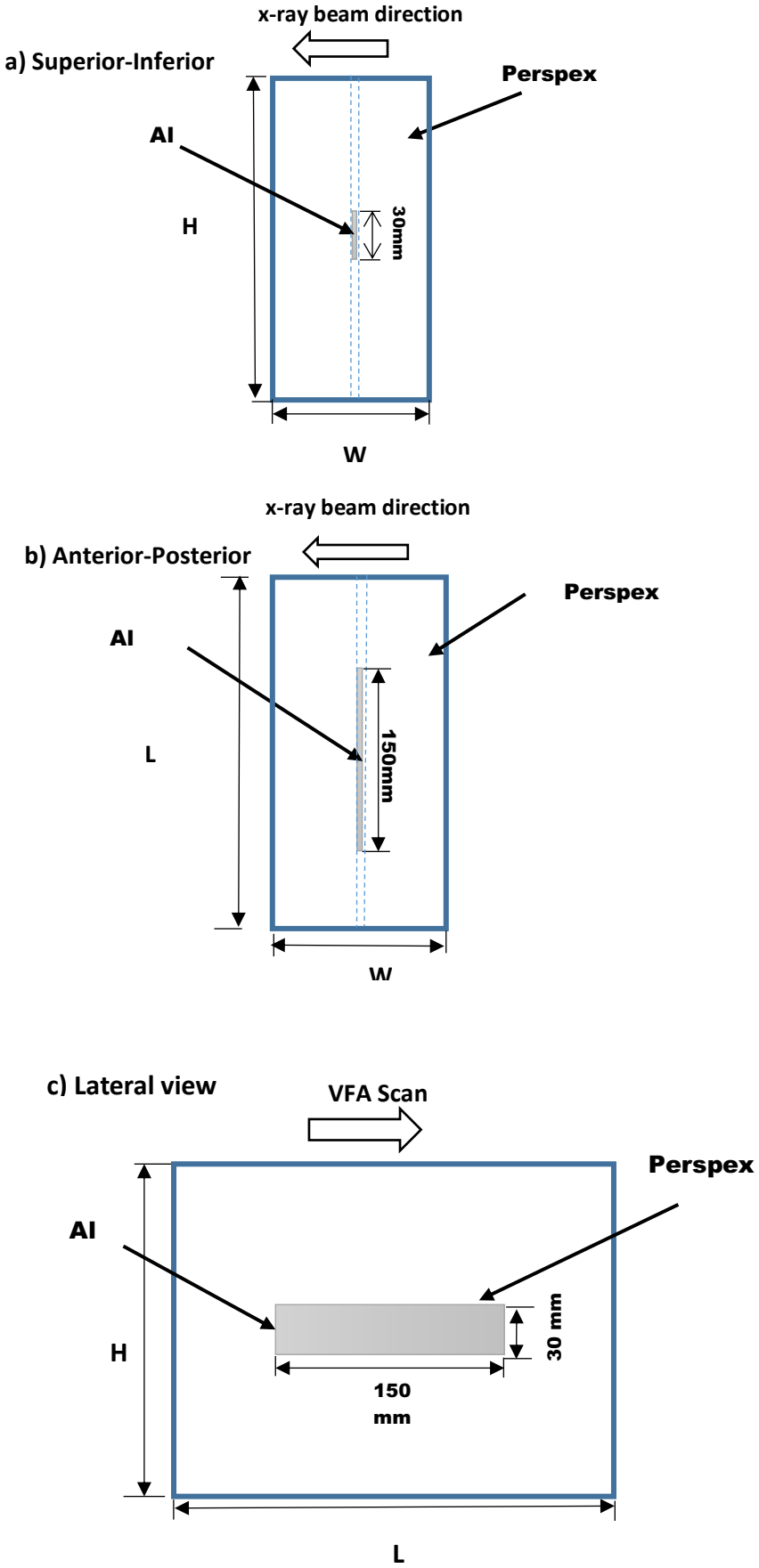


Figure 7-3: Schematic views of the Perspex phantom and its dimensions: H= height, W= width, L= length

7.3 **Phantom Height and Length**

Experiments were done in order to determine an appropriate length (L) and height (H) (Figure 7-3) for the phantom.

a) Phantom Height

For fan-beam Hologic DXA systems, the scan width is fixed by the width of the collimator and the operator has no control over this factor. During the time period over which the VFA scans used in this phantom study were acquired, the scan width was 14.5 cm. The height of the Perspex phantom should be greater than the scan width to avoid undesired air anterior the phantom. To determine the phantom height, a series of Perspex slabs of 1 cm thickness were used. An initial total height of 9 cm was scanned, and this value was increased in steps of 1 cm. The acquired VFA images were visually inspected using the Hologic software. Air was clearly present anterior to the phantom (viewed in the lateral direction) on VFA images with Perspex height in the range of 9 cm to 16 cm, while at 17cm, the Perspex covered the entire width of x-ray beam.

The outcome of this experiment was that the height of the Perspex phantom should be ≥17 cm to avoid the presence of air in the image. A height of 20 cm was chosen.

b) Phantom Length

To complete the Perspex phantom design, the length of the phantom was determined. In this experiment, a Hologic QDR -1 anthropomorphic spine phantom (Serial no. 10203) of length of 20 cm was used. Two VFA images were acquired with different scan length: 17 cm and 30 cm. Figure 7-4 shows the experimental set up. Images obtained in single energy mode are shown in Figure 7-5a and 7-5b.

These images were analysed using Hologic software to investigate whether there was a difference in pixel values at corresponding locations on the phantom. To do this, a series of landmarks were identified; these included points within the vertebrae, at the edges of the vertebrae and outside the vertebrae. The results indicated that there were differences in pixel value at corresponding locations on the image acquired with a 30 cm scan length. Since this was not the case with the image acquired with a 17 cm scan length, it was concluded that the anomalies were due to the presence of air.

The outcome of this experiment can be summarised as follows: the scan length should be less than the Perspex phantom length to avoid air superior and inferior to the phantom. For convenience, the length of the phantom was chosen to be 25 cm.

Figure 7-4: Experimental set up for lateral VFA scan of the Hologic spine phantom

a b)

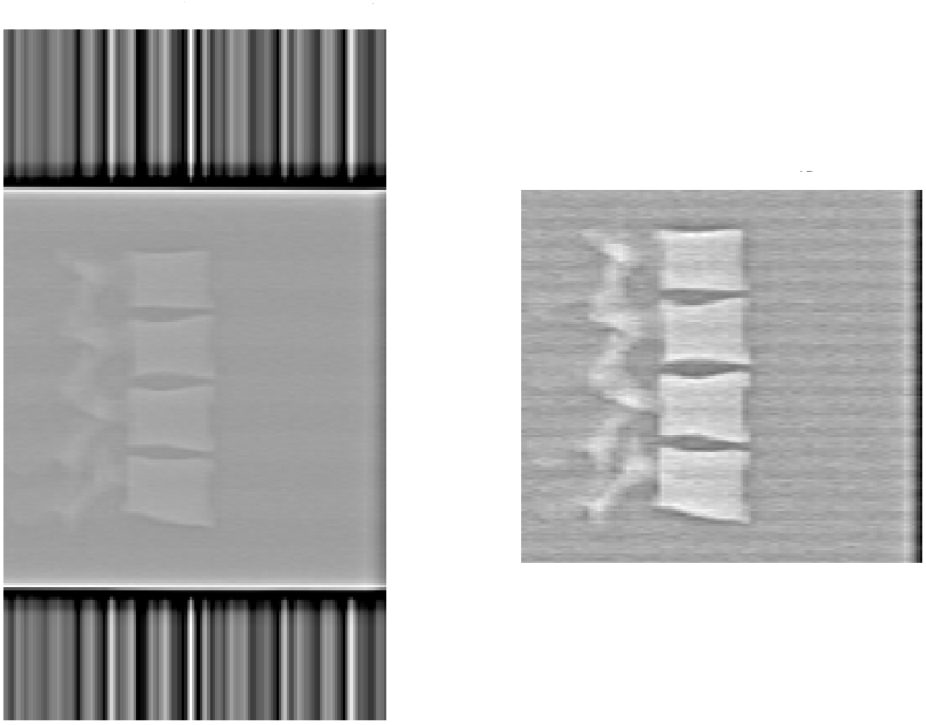


Figure 7-5: Lateral VFA images of Hologic spine phantom with: a) 17 cm scan length, b) 30 cm scan length

7.4 **Modelling of Calcification**

7.4.1 **Phantom Material**

The composition of calcification is identical to that of calcium hydroxyapatite (section 2.4) and aluminium has been used in many studies to simulate calcification (Dmitrovsky and Boskey 1985; Higgins et al. 2005; Warren et al. 2013). Table 7-6 lists some physical properties of hydroxyapatite and aluminium (Jonson 1993; Manjunatha and Rudraswamy 2012). However, there is a difference between these two materials in regards their physical properties.

Table 7-6: Properties of calcium hydroxyapatite and aluminium (Jonson 1993)

Material	Atomic effective number Z_{eff}	Density g/cm^3	Mass attenuation (cm^2/gm)	
			40 keV	70 keV
Hydroxyapatite	15.86	3.15	0.988	0.313
Aluminium	13.26	2.70	0.568	0.242

Aortic calcification takes the form of deposits that are embedded into the soft tissue. Ideally, the phantom should be identical to the real case to ensure an accurate simulation result. However, in this study, the intention was not to simulate the real pattern of calcification within the aorta, but to investigate the performance of VFA to detect uniform thicknesses of calcification represented by an equivalent material. This was the simplest approach to simulate different thicknesses of a calcification.

7.4.2 Aluminium Strips Thickness Measurements

In this phantom study, the calcification was modelled using aluminium strips (99% purity) and six different thicknesses: 0.05, 0.1, 0.2, 0.3, 0.5 and 1 mm of dimensions 150 mm x 30 mm. The tolerance was 15% for a thickness of 0.05 mm and 10% for thicknesses > 0.05 mm.

The aluminium strips were given identification numbers and their nominal thickness, as given by the manufacturer was checked using a Digimatic screw gage micrometre (MDC-1" PXT) with 0.00005" resolution (Figure 7-6).

Thickness was measured at fifteen points on each strip. Sets of five measurements were made along each edge of the strip and along the central line. Table 7-7 shows an

example of fifteen points measured on a strip of aluminium with a nominal thickness of 0.5 mm ± 0.05 mm. The nominal thickness and measured mean (SD) thickness of a representative set of strips are shown in Table 7-8.

For the experimental work, different combination of the strips were used to give a range of aluminium thickness between 0.05 mm and 2 mm.

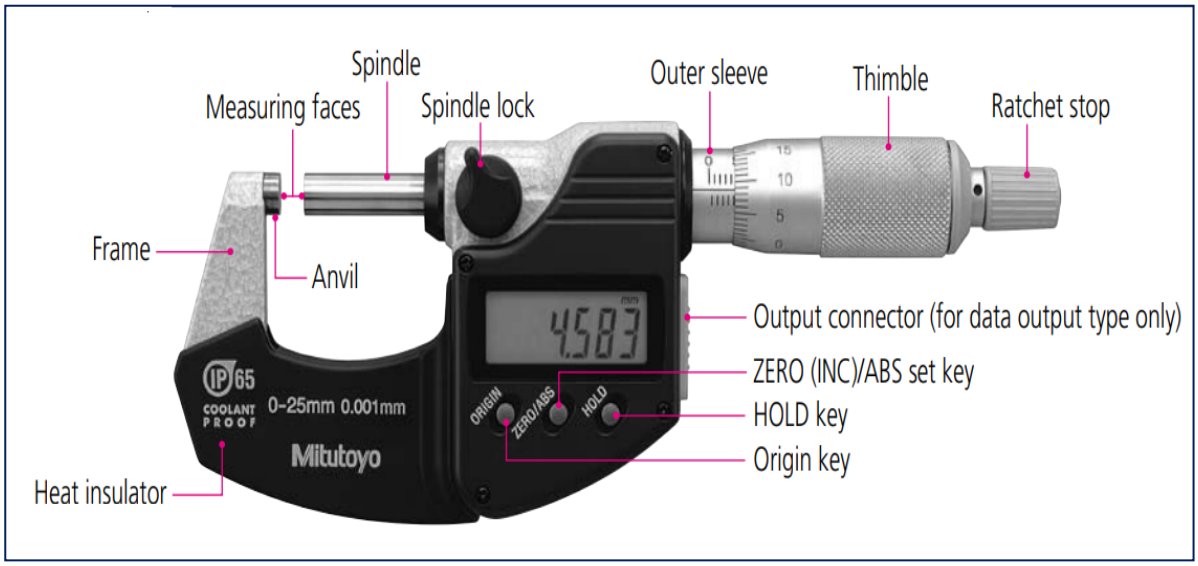


Figure 7-6: Digimatic micrometre screw gage

Table 7-7: Thickness measurements for one aluminium strip with nominal thickness 0.5 mm

Point #	Centre	Left edge	Right edge
1	0.472	0.471	0.469
2	0.475	0.474	0.469
3	0.469	0.470	0.475
4	0.473	0.474	0.467
5	0.473	0.472	0.475

Table 7-8: Thickness measurements for six aluminium strips with nominal thickness of 0.05,0.1,0.2,0.3,0.5 and 1 mm

Nominal thickness mm	Measured thickness (SD) mm	Difference %
0.05	0.0482(0.001)	-3.60%
0.1	0.102(0.0013)	+2.30%
0.2	0.202(0.0023)	+1.10%
0.3	0.288(0.0010)	-3.97%
0.5	0.472(0.0026)	-5.60%
1	0.901(0.0021)	-9.88%

7.5 **Summary**

This chapter has presented the phantom designed and constructed to simulate abdomen soft tissue and uniform aortic calcification.

The main features of the phantom can be summarised as follows:

- The abdomen was simulated with Perspex of width in the range 20 to 50 cm.
- The phantom height and length were set at 20 cm and 25 cm respectively,
- In order to mimic the calcification, aluminium strips of thickness ranging between 0.05 mm and 2 mm were used.

The next chapter presents a phantom study conducted in this research to investigate the performance of DXA in the detection of small calcification thickness in the aorta through VFA images acquired in single and dual energy and with different phantom configurations.

Chapter: 8

Performance Evaluation of DXA for AAC Detection

8.1 Introduction

This chapter presents and discusses the phantom study conducted in the second part of this research. The purpose of performing these experiments was to examine the capability of VFA images acquired by DXA to detect calcification in the abdominal aorta. The minimum thickness of calcification equivalent material that can be detected in VFA images was investigated.

Two types of DXA scanner (Hologic Horizon and Hologic Discovery) were used in this study and scans were acquired in two energy modes: single energy (SE) and dual energy (DE). The aim of this study was to examine how well VFA images acquired by the DXA modality are able detect calcification and to define the limitations of these images. The study included investigating:

- The effect of energy mode on images of different phantom configurations;
- The effect of the width of the abdomen region, represented by a soft tissue equivalent material (Perspex), on VFA image quality;
- The minimum detectable thickness of uniform calcification, represented by aluminium;
- The repeatability and reproducibility of image contrast and contrast to noise ratio.

8.2 **VFA Image Acquisition Protocol**

The following steps were developed as the protocol used to acquire VFA images with the two Hologic DXA scanners for all experiments in this study. The manufacturer's instructions were followed, and the daily quality control check applied.

Laser Positioning

On the control panel, the "Table" switch was pressed to move the table and C-arm to the centre position. The position of laser indicator was aligned to be at the centre of the table.

Phantom Setup

The Perspex phantom was positioned at the centre of the table, so that the longitudinal the laser indicator line (superior-inferior direction) coincided with the midline of the phantom. The transverse laser indicator (lateral direction) crossed the phantom at a distance of 12.5 cm from both ends of the phantom along its length.

Scan Acquisition

VFA scans were performed with the default option; this meant that only lateral scans were made and that anterior-posterior views were not obtained.

Instant vertebral assessment (IVA), as VFA is called in the Hologic system, was highlighted on the scan type selection window. Scan length was chosen to be less than the phantom length (25 cm). The laser indicator was placed centrally 3.3 cm from the superior edge of the phantom and the scan length was set to 18.2 cm. The C-arm was positioned using the arm controls on the control panel to acquire the VFA images.

This protocol was repeated for all phantom configurations, i.e. each configuration of Perspex width and aluminium thickness. Aluminium strips of nominal thickness 0.05, 0.15, 0.2, 0.25, 0.3, 0.5, 0.8, 1.0, 1.3, 1.5, 1.8 and 2 mm were inserted at the centre of

the phantom between two Perspex blocks of total width 15, 20, 25, 30, 35 and 40 cm. Figure 8-1 shows the experiment setup.

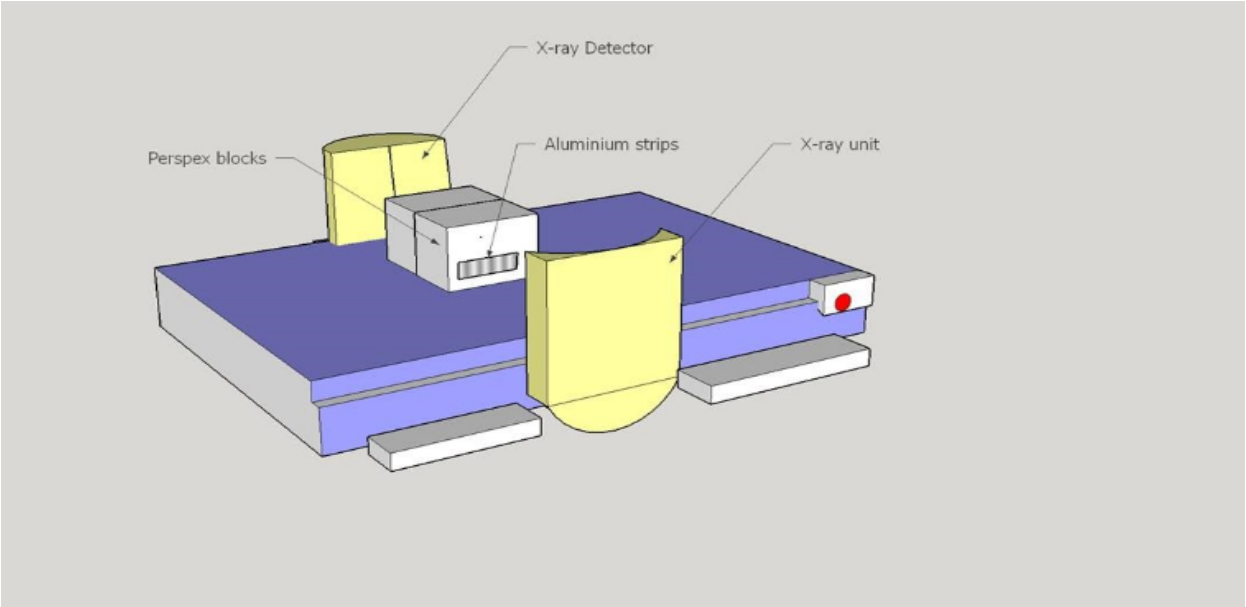


Figure 8-1: Experimental setup for VFA imaging of the phantom

8.3 Preparation of Images for Analysis

VFA images were exported using Hologic software in DICOM format. Image size was $502x402$ pixels (12 bits) in SE mode and $502x399$ pixels (8 bits) in DE mode for both scanners. The raw images were analysed using Matlab 2014 software with no pre-processing and the visibility of the Al strips was enhanced by applying image windowing. Artefacts appeared at the edges of images acquired by both scanners in both modes. Since these artefacts affected pixel values if they were included within regions of interest, they were removed by cropping the images manually. Figures 8-2 and 8-3 depict these artefacts on SE and DE images of a phantom configuration comprising15 cm Perspex combined with 0.5 mm of aluminium, with suitable windowing for better visualisation.

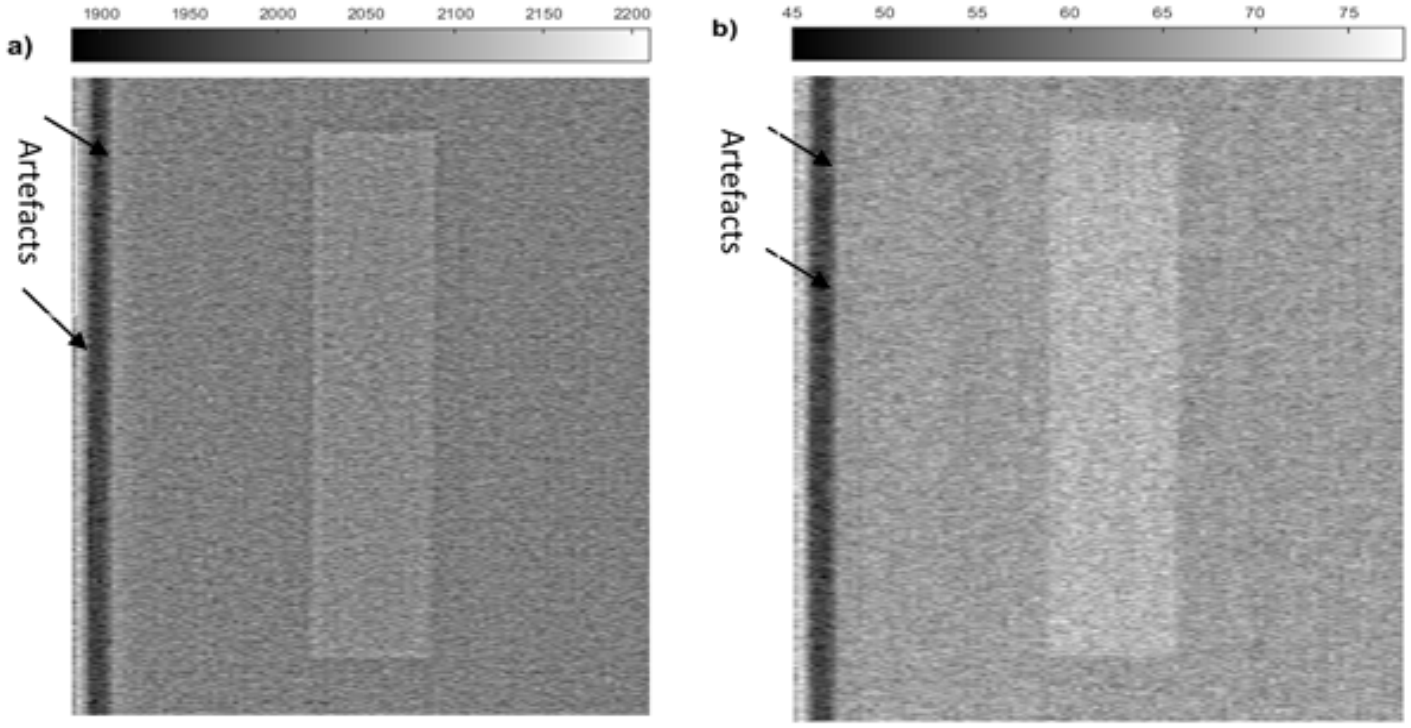


Figure 8-2: Original phantom VFA images acquired by the Hologic Horizon scanner for 15 cm Perspex and 0.5 mm Al: (a) single energy mode, (b) dual energy mode

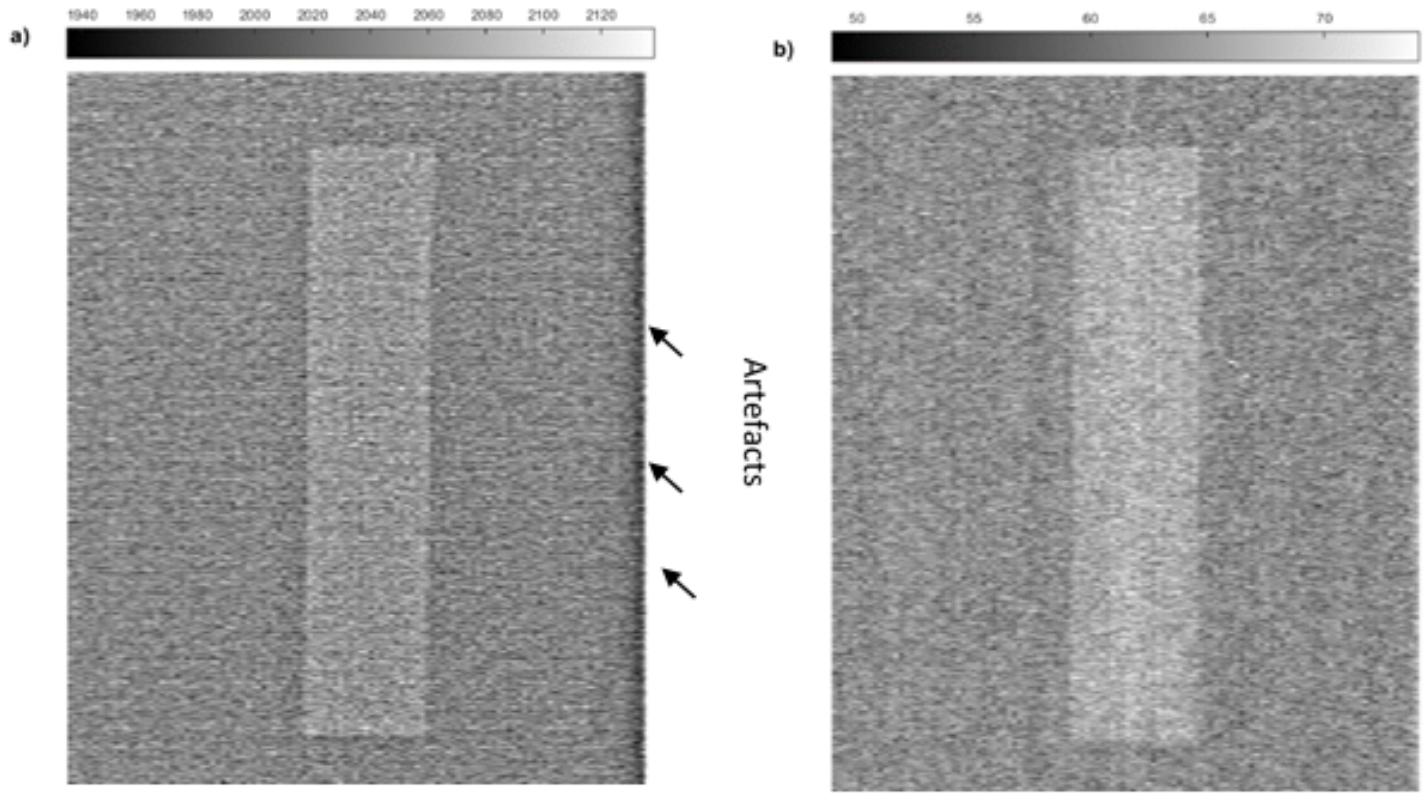


Figure 8-3: Original phantom VFA images acquired by the Hologic Discovery scanner for 15 cm Perspex and 0.5 mm Al: (a) single energy mode, (b) dual energy mode

For some of the analyses, the entire cropped VFA image was selected as a region of interest (ROI). However, for another type of analysis, three rectangular ROIs were selected on each image: a central ROI within the aluminium strip and two lateral ROIs within Perspex on each side. Figures 8-4 and 8-5 show three ROIs placed on a phantom image comprising 15 cm Perspex combined with 0.5 mm of Al in SE and DE modes for the two scanners.

To determine the position and size of the three ROIs, profiles were created of pixel values along the entire phantom height for a configuration of 15 cm Perspex combined with greatest Al thickness (2 mm). For each point along the profile, the pixel value was the mean of 450 rows along the phantom length. This was done so that the central ROI could be positioned optimally within the image of the aluminium strip with its edges well away from the edges of the strip. Figures 8-6 and 8-7 show examples of the effect of the aluminium edges on mean pixel values.

On Horizon scanner images, the central ROI was defined on column number 135 with a width of 50 pixels; this was done to avoid the effect of aluminium edges. The two other ROIs were positioned at a distance of 70 pixels horizontally from the right and left sides of centre ROI. The right and left ROIs started at column number 15 and 255 respectively since they had the same width as the central ROI.

For the Discovery scanner, the central, right and left ROIs were drawn on columns 30, 150 and 255 with a width of 50 pixels. Again the distance between the central ROI and left/right ROIs was 70 pixels on both sides.

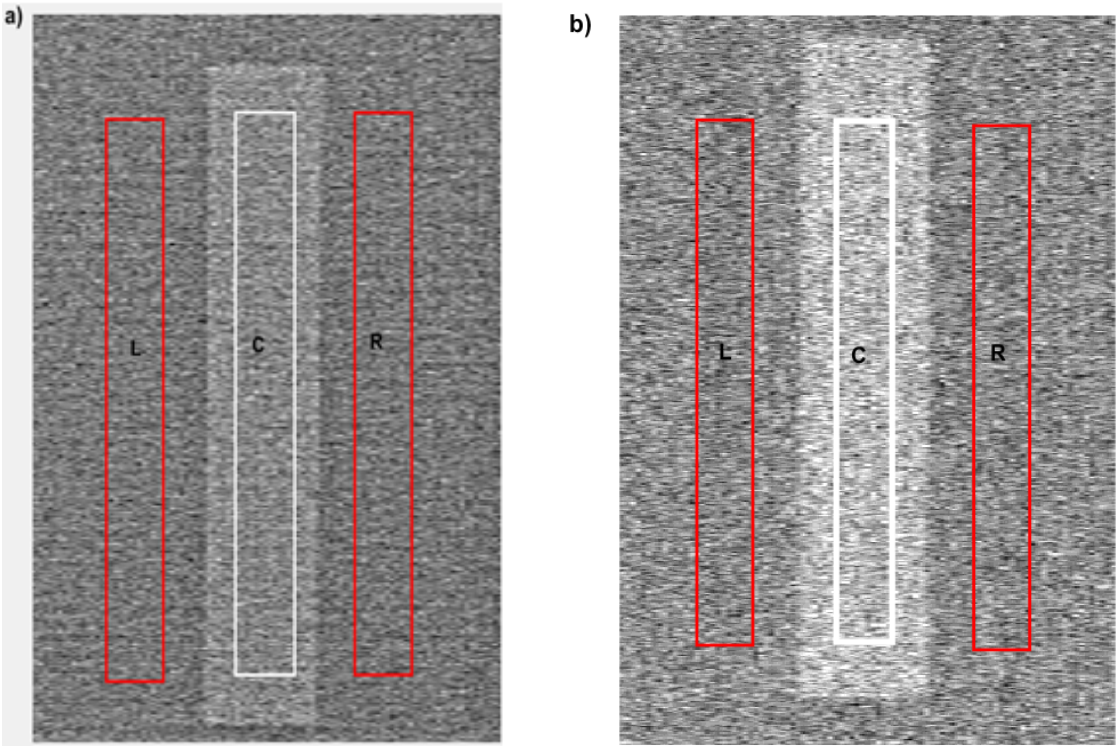


Figure 8-4: ROIs on a phantom images of 15 cm Perspex (left, L and right, R) and 0.5 mm aluminium (centre, C) obtained with a Horizon scanner: (a) single energy mode, (b) dual energy mode

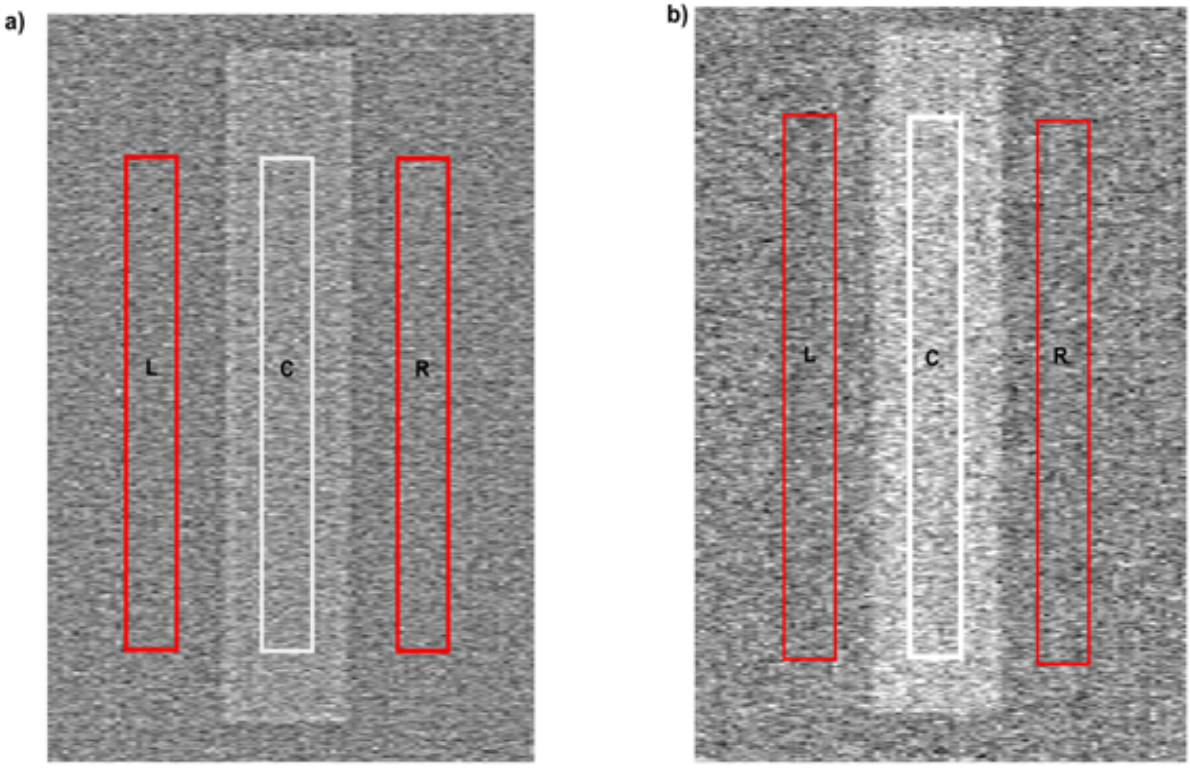
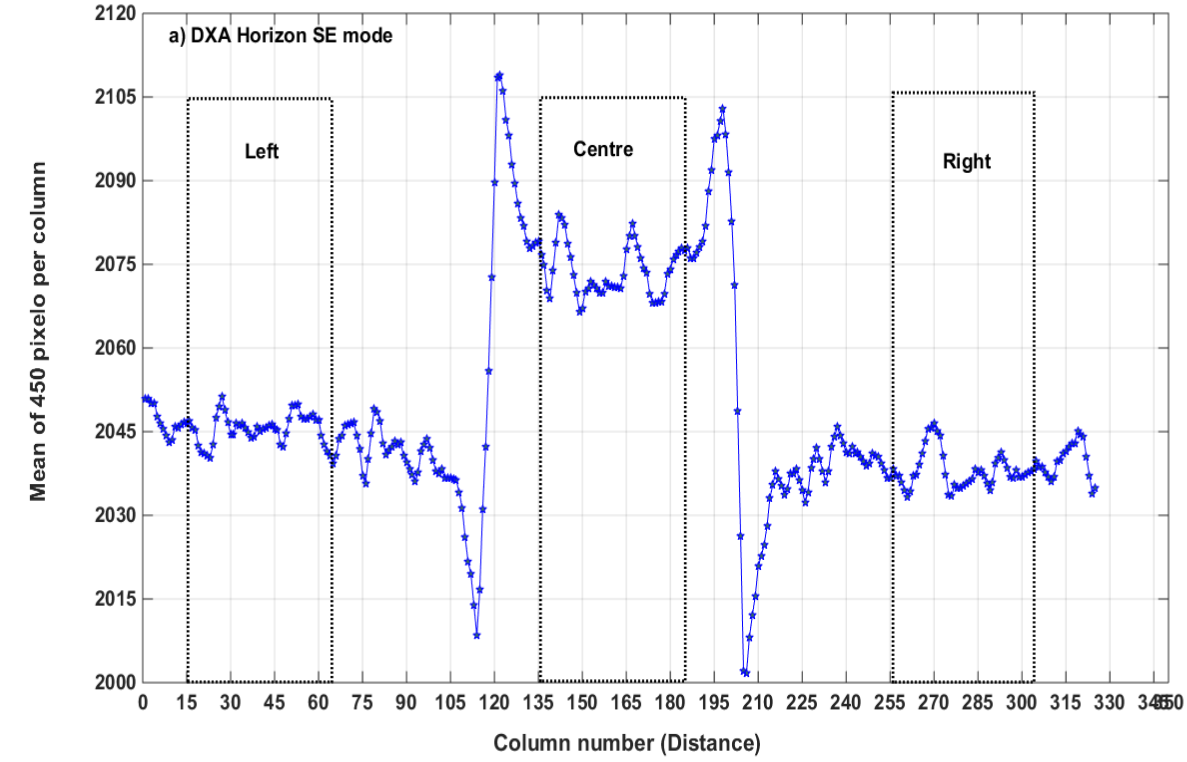


Figure 8-5: ROIs on a phantom images of 15 cm Perspex (left, L and right, R) and 0.5 mm aluminium (centre, C) obtained with a Discovery scanner: (a) single energy mode, (b) dual energy mode



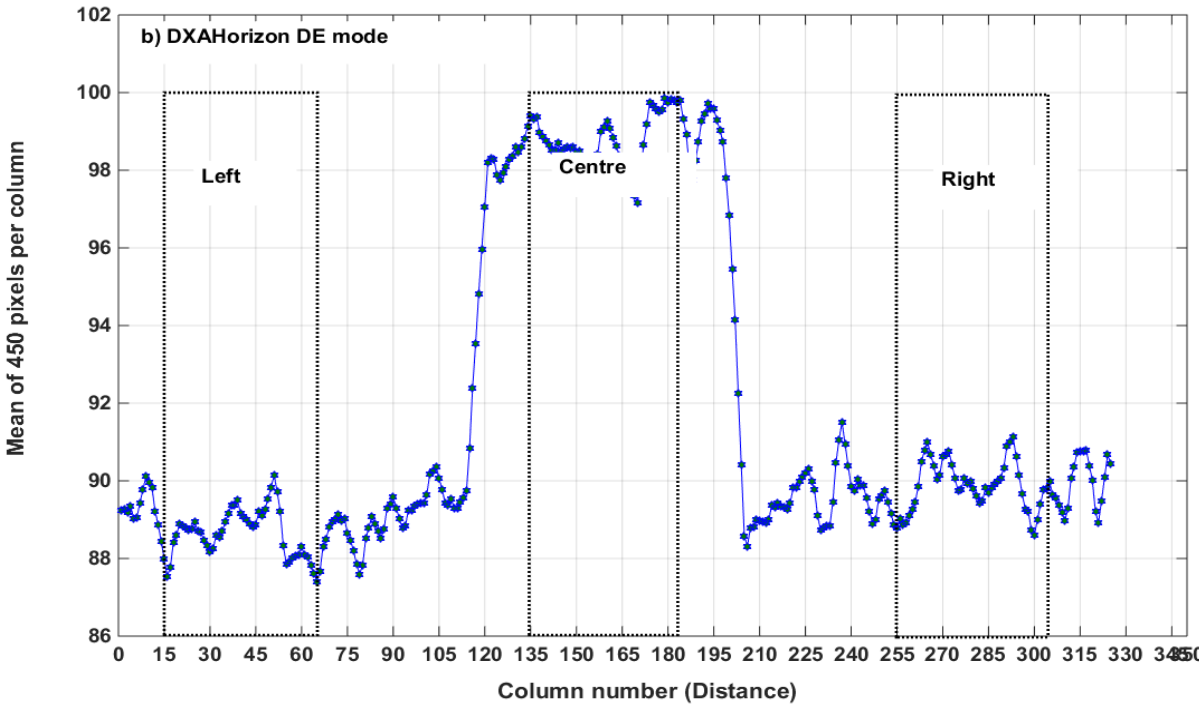


Figure 8-6: VFA image profile showing the aluminium strip edge effect for 15 cm Perspex combined with 2 mm aluminium obtained with the Horizon scanner in (a) single energy mode, (b) dual energy mode

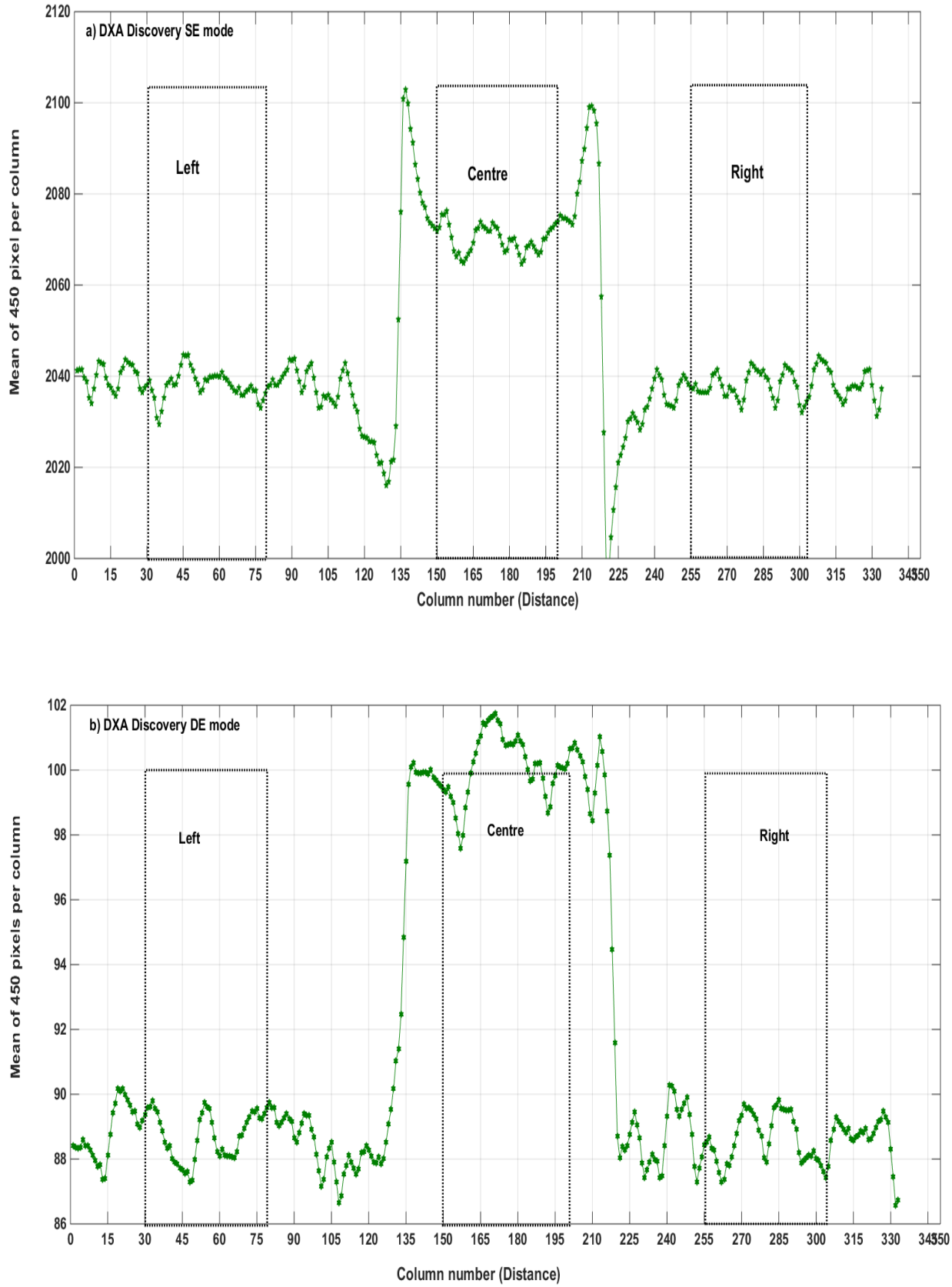


Figure 8-7: VFA image profile showing the aluminium strip edge effect for 15 cm Perspex combined with 2 mm aluminium obtained with the Discovery scanner in (a) single energy mode, (b) dual energy mode

8.4 **Uniformity and Noise**

In this study, the influence of Perspex width on VFA images acquired in two energy modes was investigated with the two DXA scanners. Images were obtained using the protocol discussed in section 8.2 and prepared for analysis as described in the previous section. Image quality was assessed in terms of uniformity and noise.

8.4.1 **Method**

VFA images of the Perspex phantom were acquired with no aluminium insert and using a range Perspex width: 15, 20, 25, 30, 35, 40 cm. Profiles were drawn of the mean of 450 pixels in 325 columns in the x direction (phantom length) and also for the mean of 325 pixels in 450 rows in the y direction (phantom length). The profiles were analysed by linear regression and correlation.

The variations of pixel values in the entire VFA images was also examined. This was done by computing the mean and standard deviation for the whole image after cropping the artefacts. The percentage coefficient of variation (%CV) was computed:

$$\%CV = \frac{SD\ of\ the\ mean\ pixel\ values\ on\ the\ entire\ image}{mean\ of\ the\ pixels\ of\ the\ entire\ image} \times 100 \qquad (8\text{-}1)$$

8.4.2 **Results**

Representative image profiles for Perspex widths of 15 cm, 25 cm and 40 cm in SE and DE modes for two scanners are shown in Figures 8-8 to 8-11.

Figures 8-12a and 8-12b, show the variation of the mean pixel values with Perspex width for the Horizon scanner.

Corresponding results are shown for the Discovery scanner in Figures 8-13a and 8-13b. It is clear that as the width of Perspex increased, noise, represented by the standard deviation, also increased. This reflects the fact that the number of photons received by the detectors decreased as Perspex width increased.

Tables 8-1 and 8-2 show the mean, SD and %CV of the pixel values in the entire VFA images acquired with the Horizon scanner in SE and DE modes while Tables 8-3 and 8-4 show corresponding data for the Discovery scanner. The %CV increased with increasing Perspex width as the noise in these images increased.

The two scanners revealed the same trend. The mean pixel value was approximately constant at 2048 and 2045 for the Horizon and Discovery respectively in SE mode for Perspex widths between 15 cm and 30 cm. It then increased to reach 2053 for the Horizon and 2047 for the Discovery for a Perspex width at 40 cm. On the other hands, in DE mode, the mean value increased consistently as the Perspex width increased to 27 cm and decreased thereafter.

Calculated %CV was consistent between the two scanners. In SE mode, it varied between 0.6 % and 16% for the Horizon scanner and 0.7% and 14% for the Discovery scanner. In DE mode, %CV increased from 2% to 16% for the Horizon and from 2% to 17% for the Discovery.

There was no linear correlation between mean pixel value and distance parallel to the scan direction (y direction); $r^2 < 0.1$ for all Perspex widths. However, in some instances there was a correlation between pixel values and distance in the x direction. For example, with a Perspex width of 15 cm $r^2 = 0.366, (p < 0.001)$ for the Horizon scanner and $r^2 = 0.182, (p < 0.001)$ for the Discovery in SE made. In DE mode there was

significant and strong correlation between pixel value and distance; for example, $r^2 = 0.693, (p < 0.001$ on images of 15 cm of Perspex and $r^2 = 0.403, (p < 0.001)$ with 30 cm Perspex for the Horizon scanner. There was no corresponding linear correlation $r^2 < 0.1$ in DE images for the Discovery scanner. A complete comparison between the two DXA scanners is presented in Appendix A.

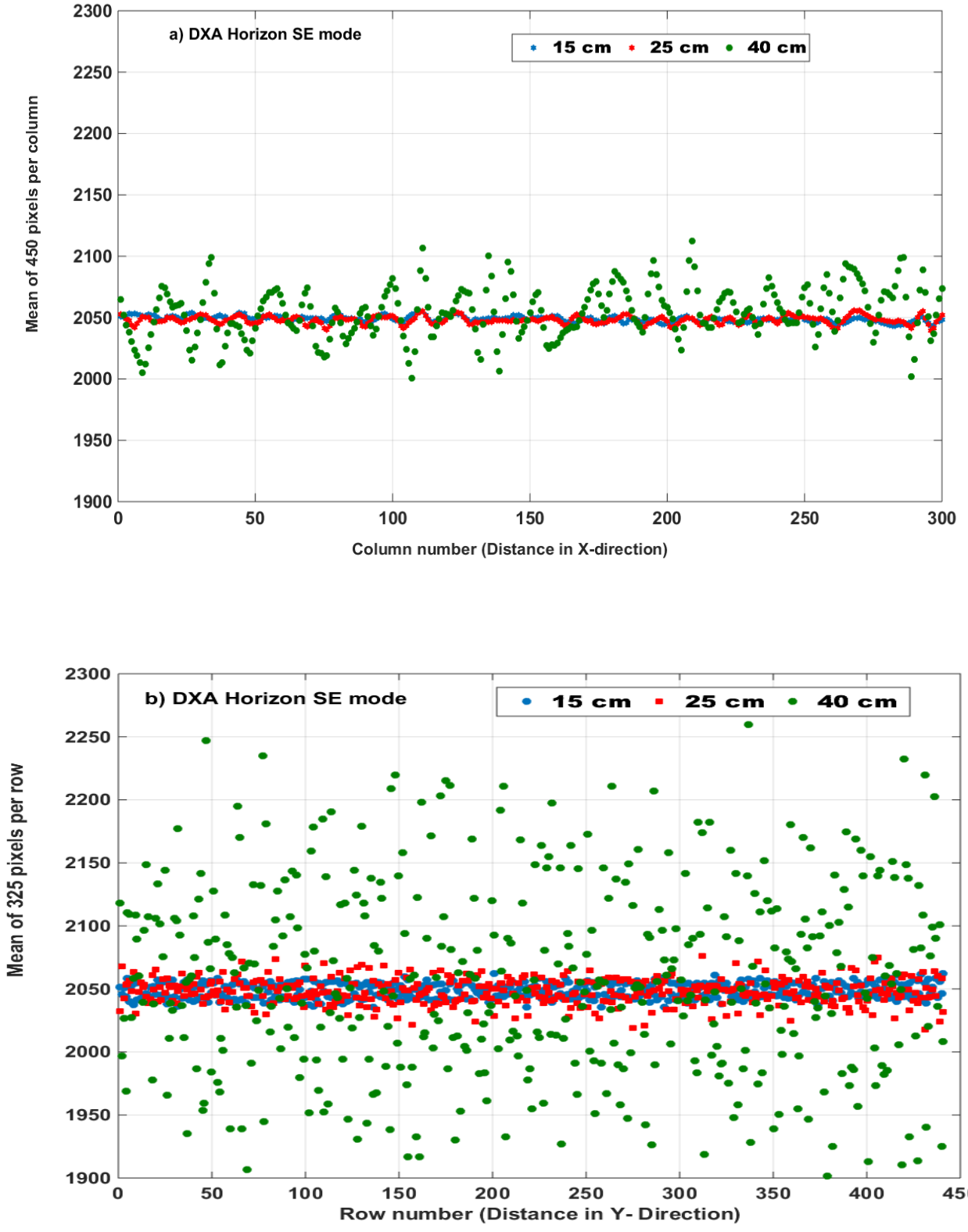


Figure 8-8: Image profiles with Perspex widths of 15 cm, 25 cm and 40 cm in single energy mode for the Horizon scanner: (a) in the X direction (left), (b) the Y direction (right)

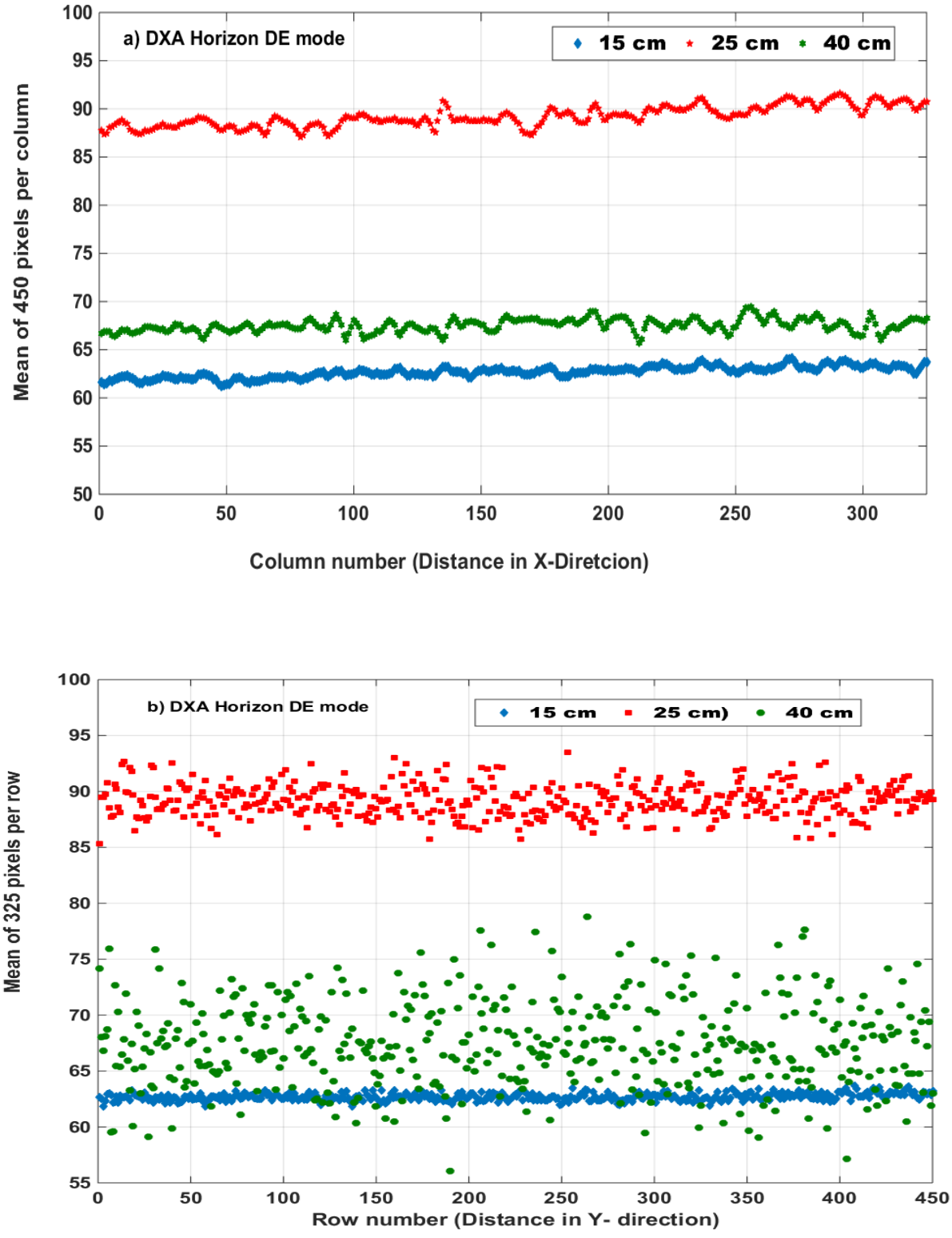
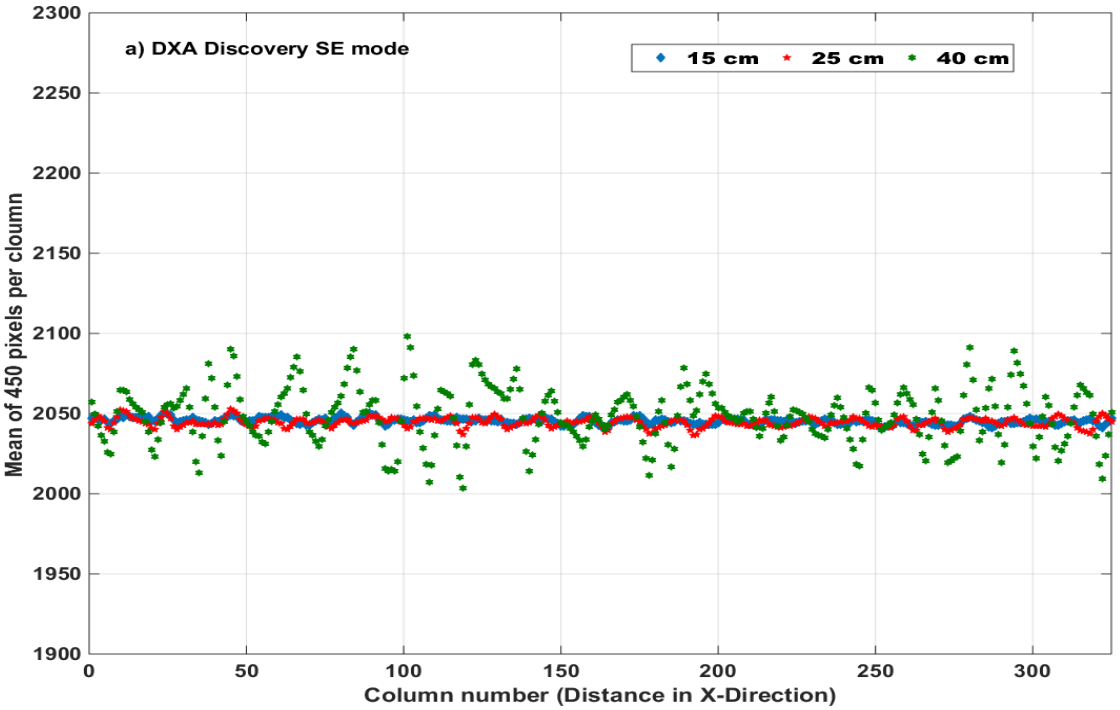


Figure 8-9: Image profiles with the Perspex widths of 15 cm, 25 cm and 40 cm in dual energy mode for Horizon scanner: (a) in the X direction (left), (b) in the Y direction (right)



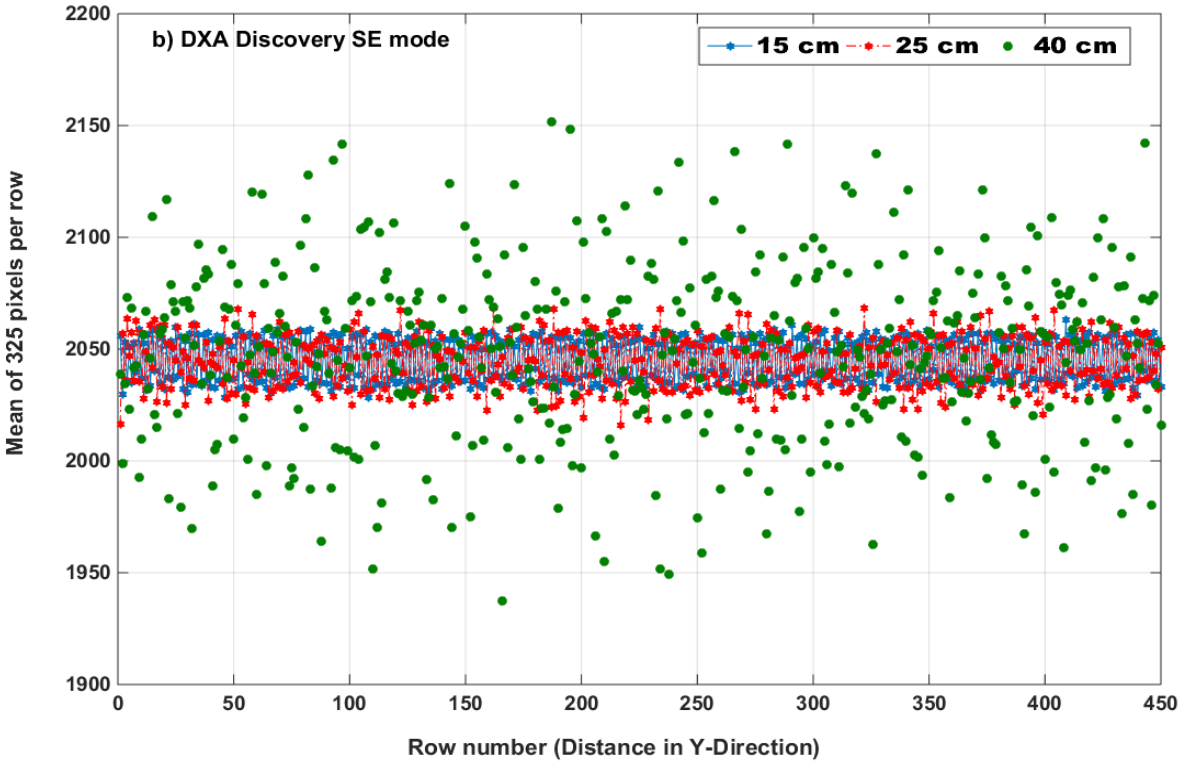
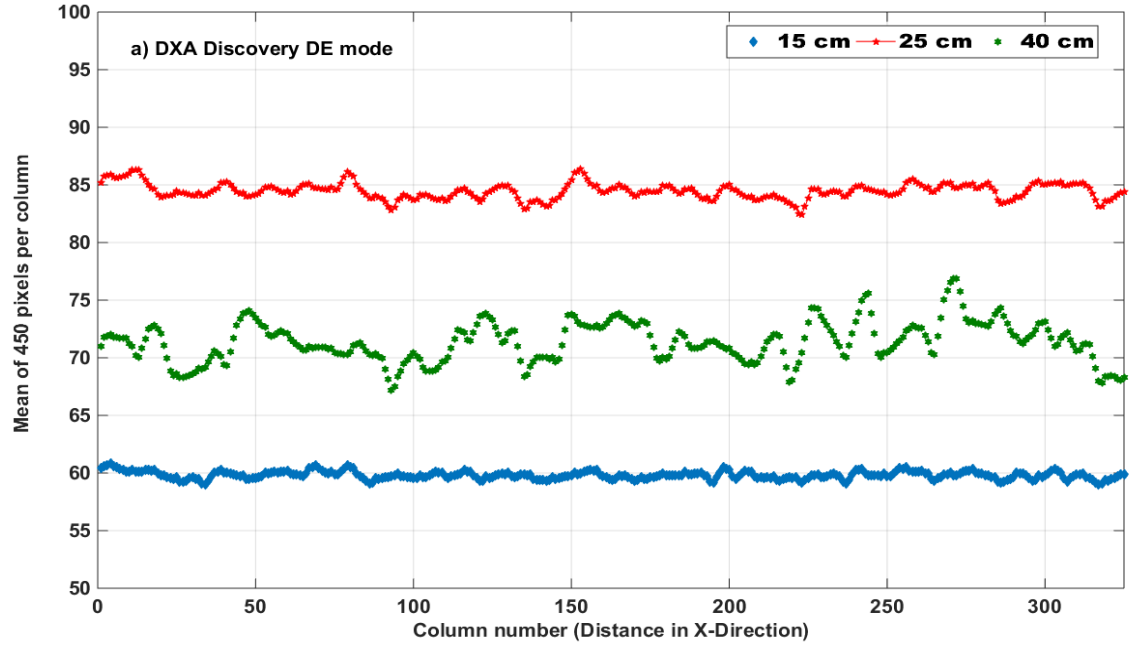


Figure 8-10: Image profiles with Perspex widths of 15 cm, 25 cm and 40 cm in single energy mode for Discovery scanner: (a) in the X direction, (b) in the Y direction



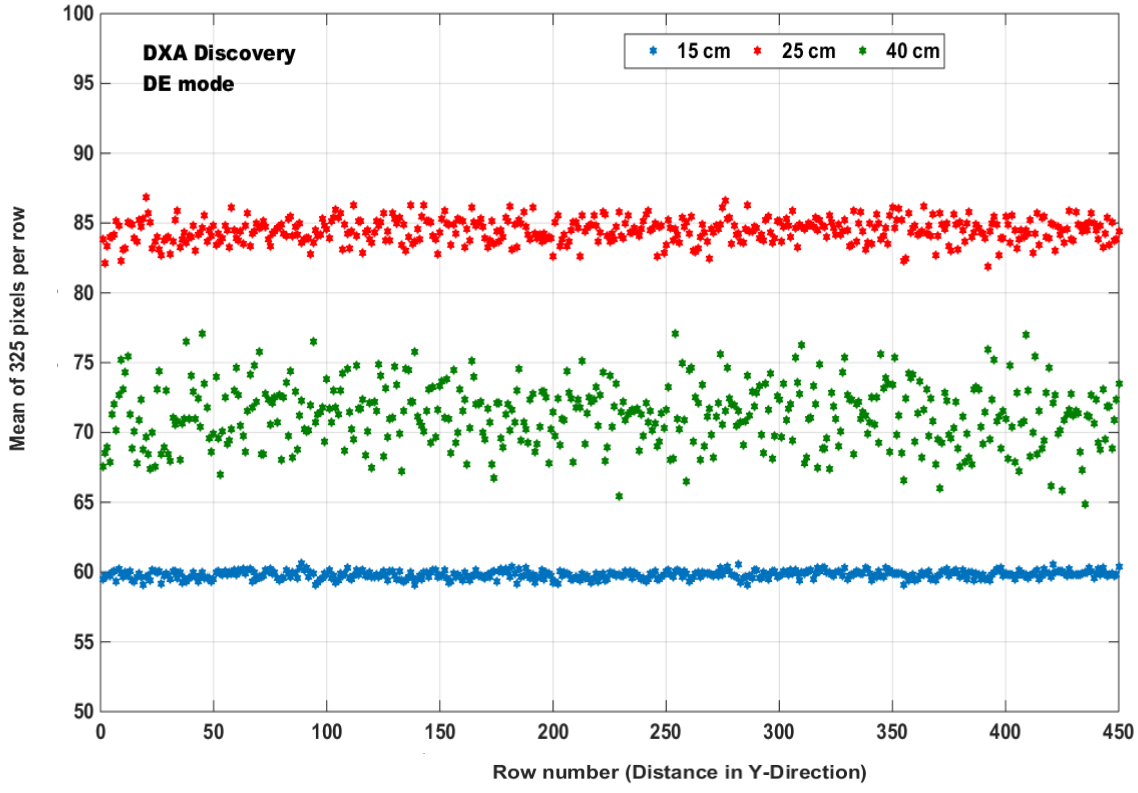


Figure 8-11: Image profile with Perspex widths of 15 cm, 25 cm and 40 cm in dual energy mode for the Discovery scanner: (a) in the X direction, (b) in the Y direction

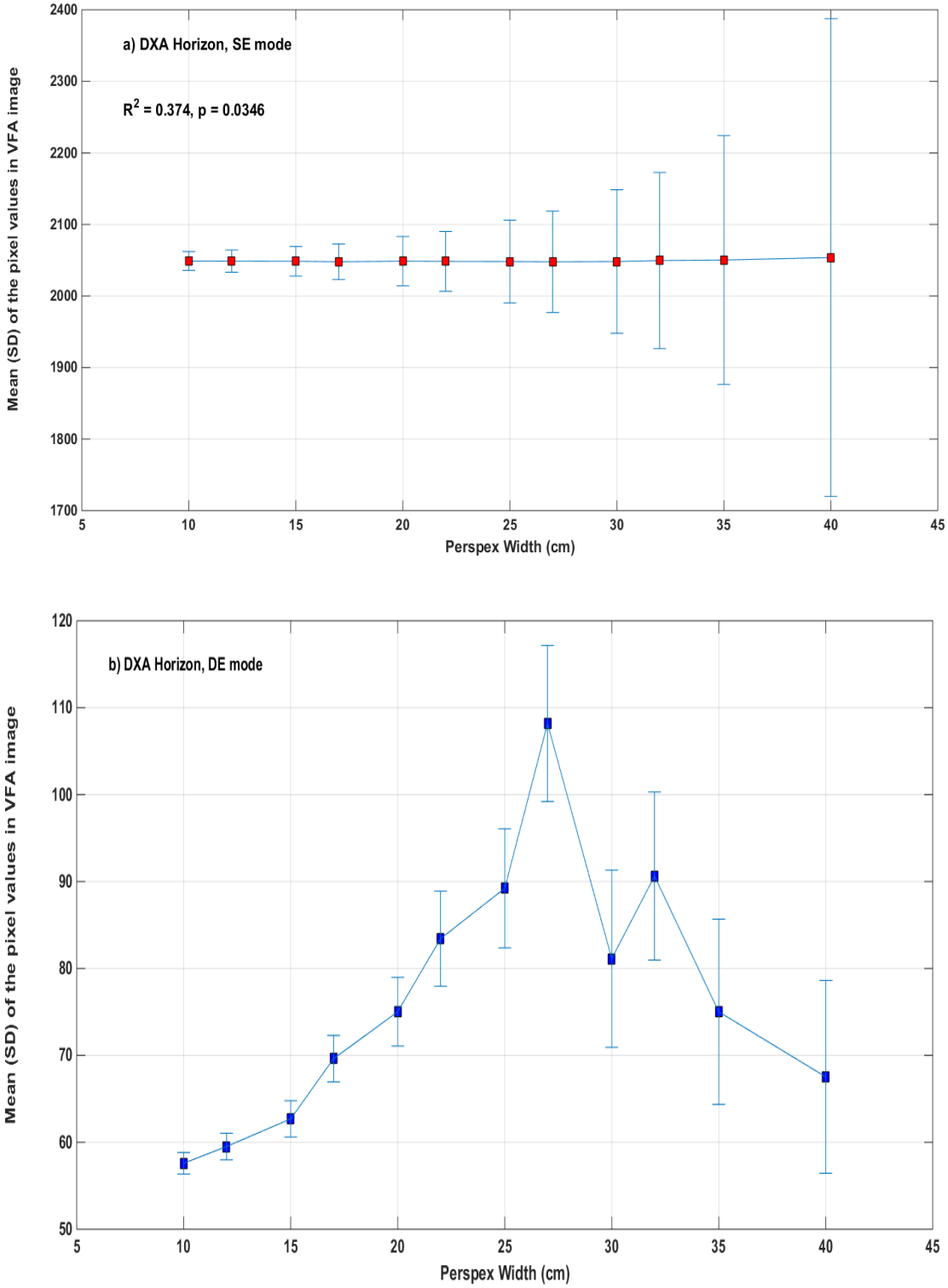
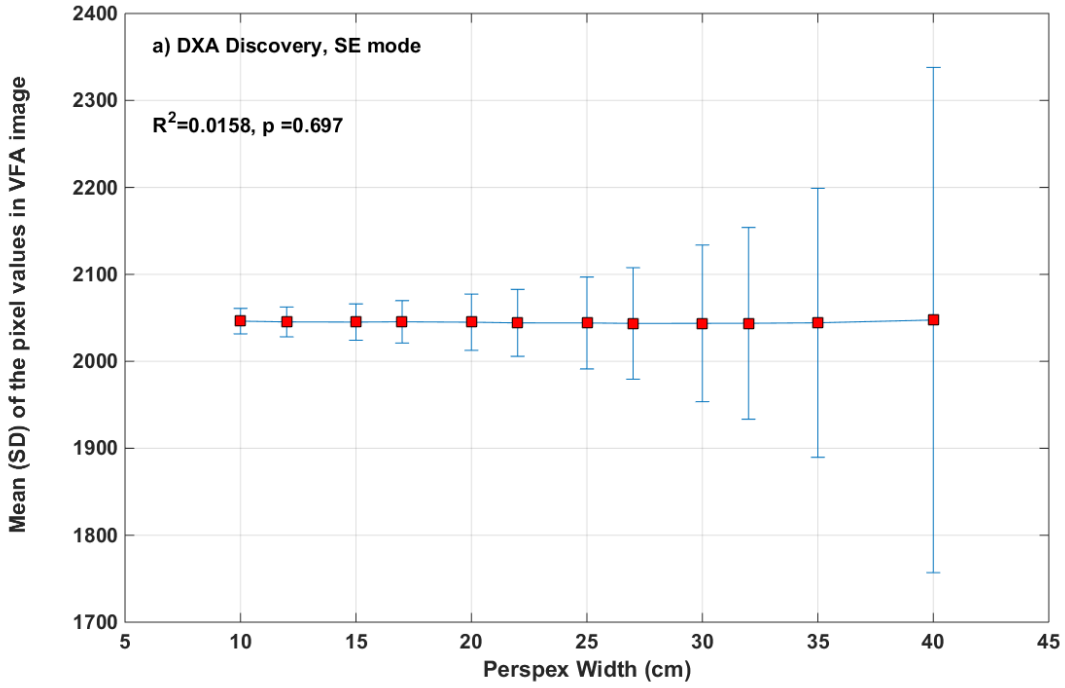


Figure 8-12: Variation of the mean (SD) of pixel values with Perspex width for VFA images obtained the Horizon scanner in: (a) single energy mode (b) dual energy mode



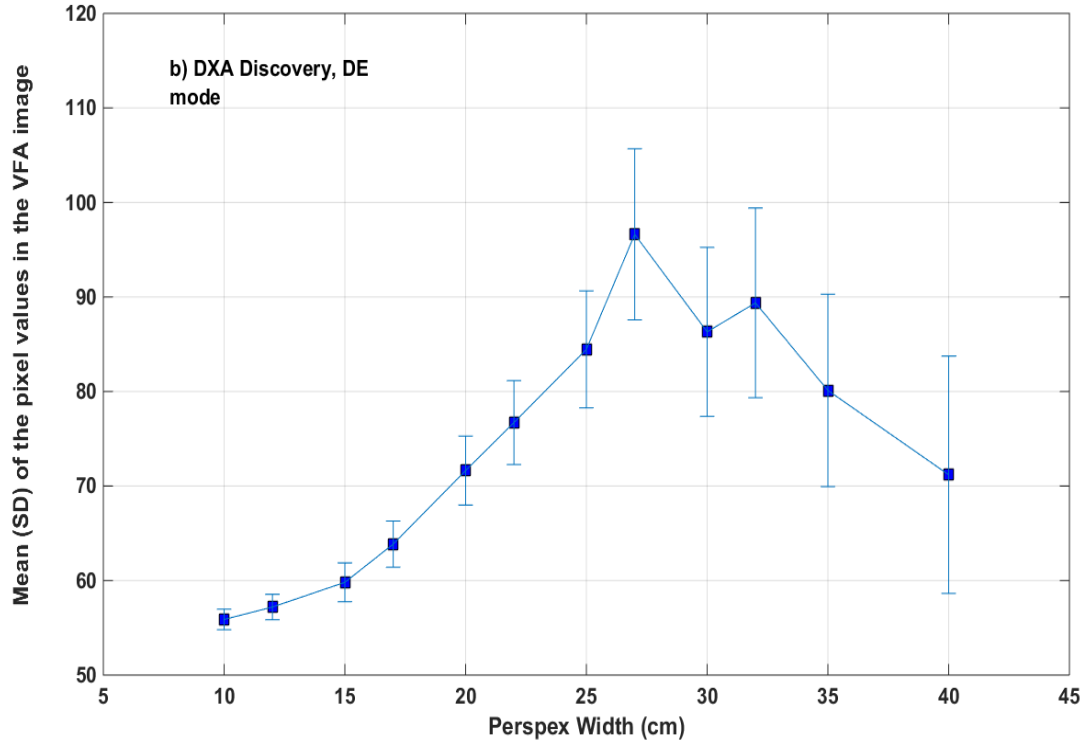


Figure 8-13: Variation of the mean (SD) of pixel values with Perspex width for VFA images obtained the Discovery scanner in: (a) single energy mode (b) dual energy mode

Table 8-1: Mean, SD and %CV of pixel values for VFA images acquired with the Horizon scanner in single energy mode.

Perspex width	Mean for the whole image	SD	% CV
10 cm	2048.9455	13.1774	0.6431
12 cm	2048.7597	15.4851	0.7558
15 cm	2048.5679	20.6625	1.0086
17cm	2047.8691	24.8719	1.2145
20 cm	2048.7065	34.4343	1.6808
22 cm	2048.3918	41.8541	2.0433
25 cm	2048.1942	57.8862	2.8262
27 cm	2047.8236	70.8389	3.4592
30 cm	2048.191	100.3036	4.8972
32 cm	2049.5182	123.1397	6.0082
35 cm	2050.2657	173.8707	8.4804
40 cm	2053.8053	333.8863	16.257

Table 8-2: Mean, SD and %CV of pixel values for VFA images obtained with the Horizon scanner in dual energy mode.

Perspex width	Mean for the whole image	SD	%CV
10 cm	57.593	1.2338	2.1423
12 cm	59.5157	1.5324	2.5747
15 cm	62.6967	2.0875	3.3296
17cm	69.6189	2.6741	3.8411
20 cm	75.0257	3.9531	5.269
22 cm	83.4262	5.4674	6.5535
25 cm	89.2148	6.8467	7.6744
27 cm	108.1815	8.9764	8.2975
30 cm	81.1224	10.1927	12.5646
32 cm	90.6329	9.6674	10.6666
35 cm	75.0146	10.6595	14.2099
40 cm	67.5282	11.0892	16.4215

Table 8-3: Mean, SD and %CV of pixel values for VFA on images obtained with Discovery scanner in single energy mode

Perspex width	Mean for whole image	SD	%CV
10 cm	2046.1828	14.6425	0.7156
12 cm	2045.3142	17.1163	0.8369
15 cm	2045.1101	20.9119	1.0225
17cm	2045.4316	24.368	1.1913
20 cm	2044.9602	32.3009	1.5795
22 cm	2044.183	38.5146	1.8841
25 cm	2044.1134	52.8592	2.5859
27 cm	2043.5641	64.1312	3.1382
30 cm	2043.6415	90.0949	4.4085
32 cm	2043.7055	110.254	5.3948
35 cm	2044.2767	154.6461	7.5648
40 cm	2047.4683	290.476	14.1871

Table 8-4: Mean, SD and %CV of pixel values for VFA images obtained with the DXA Discovery scanner in dual energy mode

Perspex width	Mean for whole image	SD	%CV
10 cm	55.8848	1.083	1.938
12 cm	57.1971	1.3484	2.3575
15 cm	59.8112	2.0522	3.4311
17cm	63.8523	2.4389	3.8196
20 cm	71.6345	3.6502	5.0956
22 cm	76.7139	4.4367	5.7834
25 cm	84.4554	6.1892	7.3283
27 cm	96.6264	9.0547	9.3708
30 cm	86.3086	8.9345	10.3518
32 cm	89.3758	10.0348	11.2277
35 cm	80.1153	10.1765	12.7023
40 cm	71.1926	12.5564	17.6373

8.5 **Detectability assessment**

As a part of this research, the minimum thickness of calcification equivalent material that can be detected in VFA images was investigated. Several approaches were applied to assess detectability of a uniform thickness of aluminium embedded in Perspex. These comprised human vision and the measurement of contrast (C) and (CNR). This section discusses these approaches and describes the methods used. Image analysis was done using the three ROIs described in section 8.3.

8.5.1 **Contrast and Contrast to Noise Ratio**

The contrast C that indicates the brightness difference in between the aluminium strip and background (Perspex) was computed using the following equation (Dickerscheid et al. 2013; Jakubiak et al. 2013):

$$\boldsymbol{C} = \frac{(N_{Al} - N_P)}{N_P} \tag{8-2}$$

Here, N_{Al} is the mean pixel value in the centre ROI (aluminium) and N_P is the mean pixel values of within the left and right ROIs (Perspex):

$$N_P = \frac{\left(N_{P\,(Left)} + N_{P(Right)}\right)}{2} \tag{8-3}$$

where $N_{P\,(Left)}$ and $N_{P\,(Right)}$ are the mean pixel values in the left and right ROIs respectively. The detection limit of an imaging system depends on the lesion to background contrast and the noise in the background (Bao and Chatziioannou 2010). Image noise was expressed as follows:

$$Noise = \frac{\sigma_P}{N_P} \tag{8-4}$$

where σ_P is the standard deviation of the pixel values in two ROIs within Perspex. This was calculated using equation (8-5) (Baldelli et al. 2009; Taylor 2016);

$$\sigma_P = \sqrt{\frac{{\sigma_{P\ Left}}^2 + {\sigma_{P\ Right}}^2}{2}} \tag{8-5}$$

where $\sigma_{\mathrm{P\ Left}}$ and $\sigma_{\mathrm{P\ Right}}$ are the pixel values in standard deviations of the pixel values within the left and right Perspex ROIs respectively.

The CNR may be used to identify the detectability of aluminium against the background (Perspex) and may be defied as:

$$CNR = \frac{(N_{AL} - N_P)}{\sigma_P} \tag{8-6}$$

Linear regression and correlation were used to investigate the relationship between Al thickness and both C and CNR for different width of Perspex.

8.5.1.1 **Results**

In general, C and CNR increased with increasing thickness of aluminium for all Perspex widths as expected. Figures 8-14 to 8-21 show the results for images acquired in SE and DE mode with two scanners for different Perspex widths combined with Al thickness between 0.05 mm and 2 mm. For a given Al thickness, C decreased with Perspex width; this was even more obvious for CNR since noise also increased with Perspex width.

Generally, Images obtained in DE mode showed a better contrast and CNR than in SE mode for all phantom configurations. DE contrast was about 10 times greater than SE mode contrast for both scanners. In addition, DE CNR was approximately 3 times greater than SE CNR.

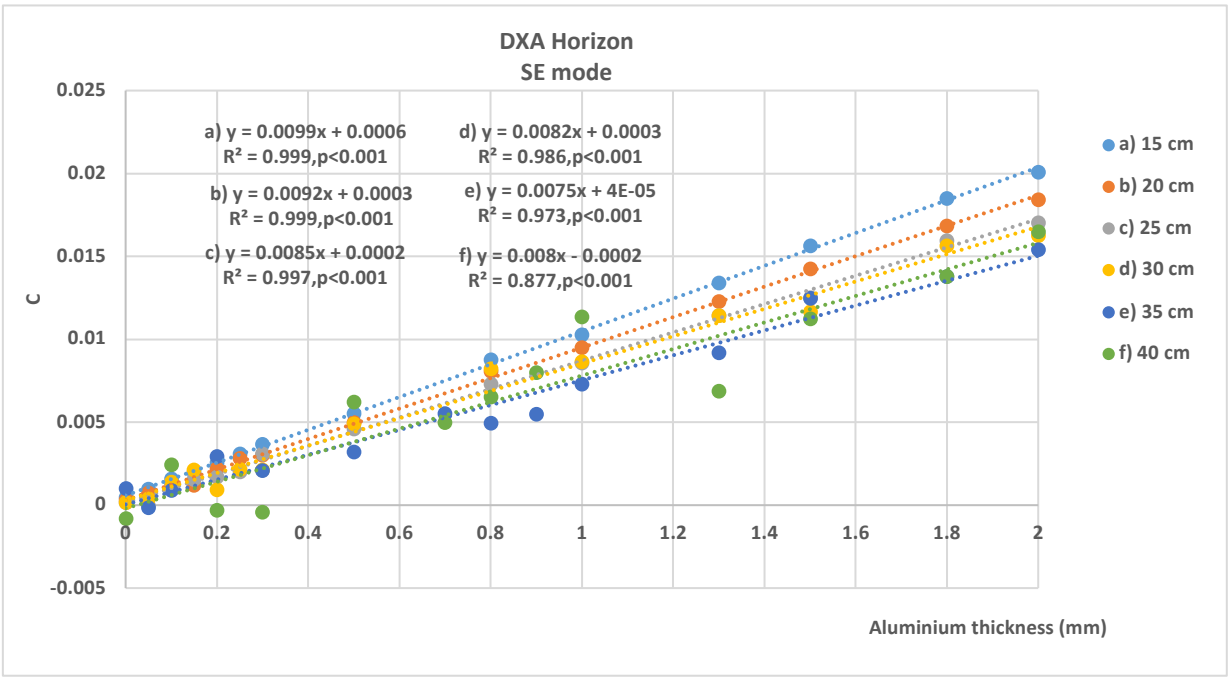


Figure 8-14: Contrast as a function of Al thickness for different Perspex widths in SE mode images obtained with the Horizon scanner

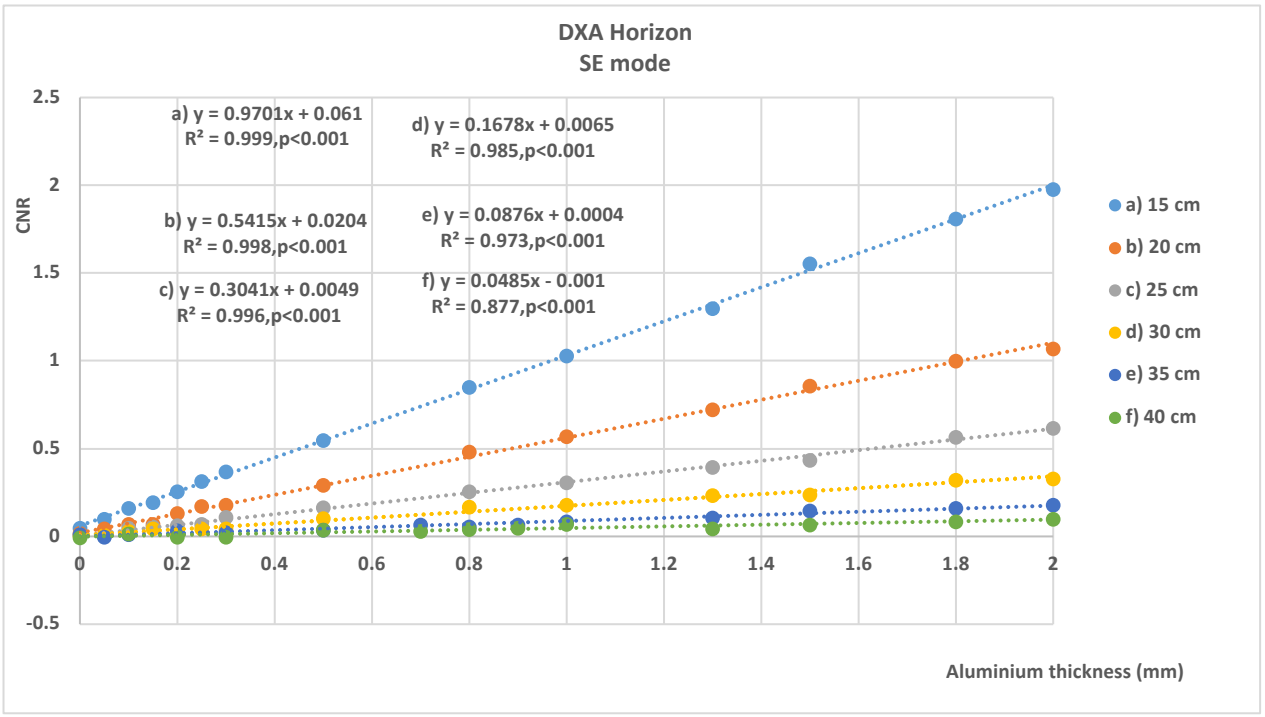


Figure 8-15: CNR as a function of Al thickness for different Perspex widths in SE mode images obtained with the Horizon scanner

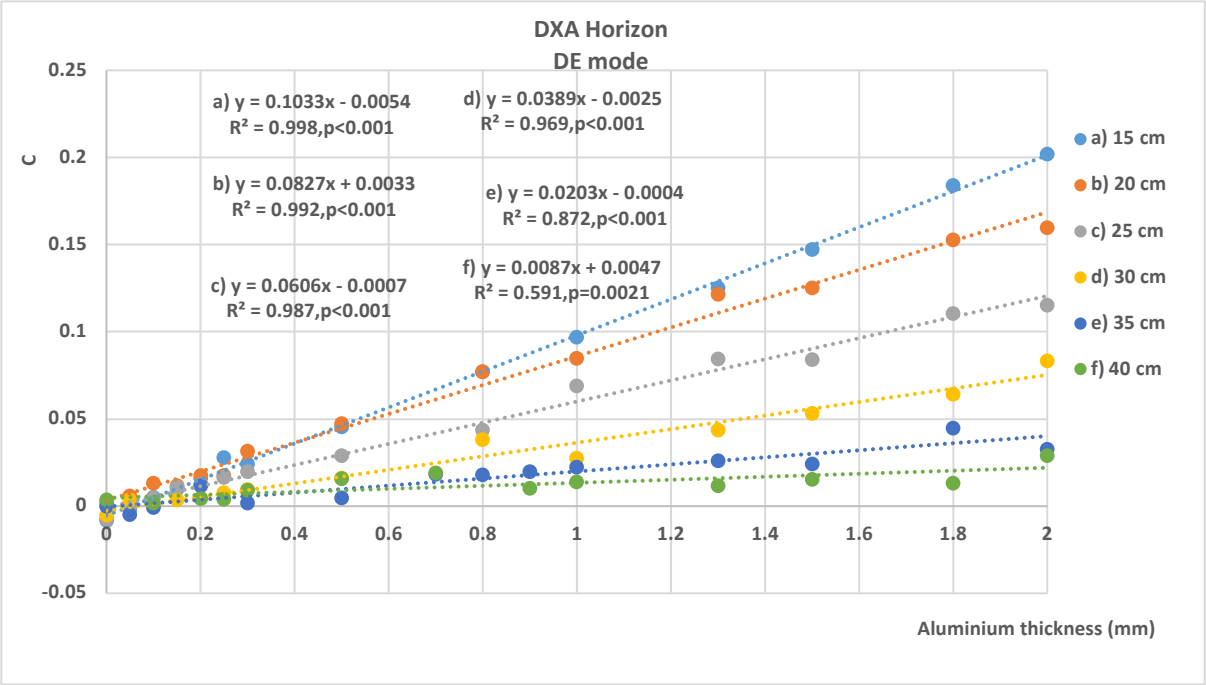


Figure 8-16: Contrast as a function of Al thickness for different Perspex widths in DE mode images obtained with the Horizon scanner

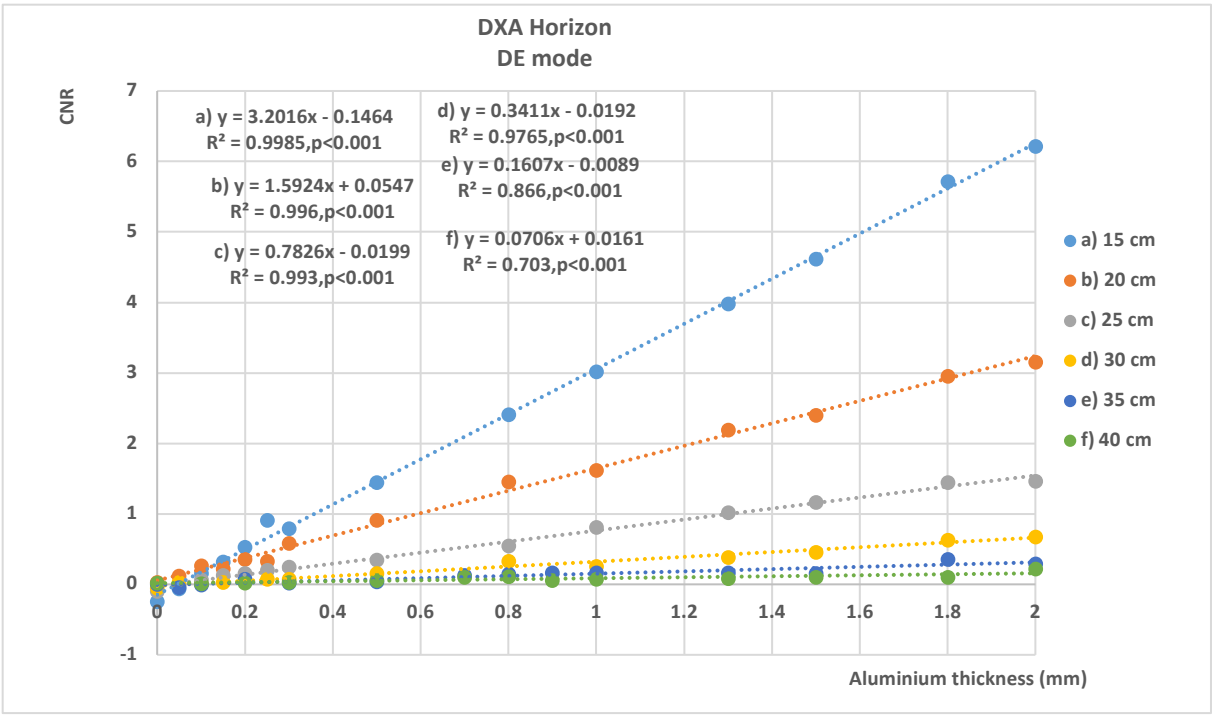


Figure 8-17: CNR as a function of Al thickness for different Perspex widths in DE mode images obtained with the Horizon scanner

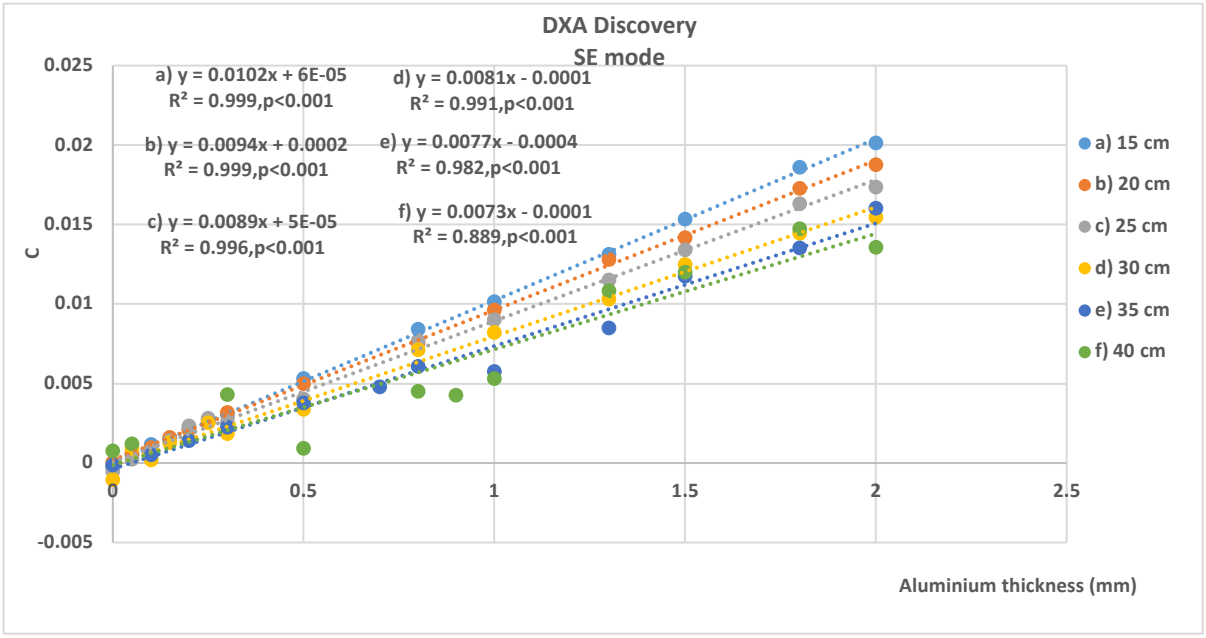


Figure 8-18: Contrast as a function of Al thickness for different Perspex widths in SE mode images obtained with the Discovery scanner

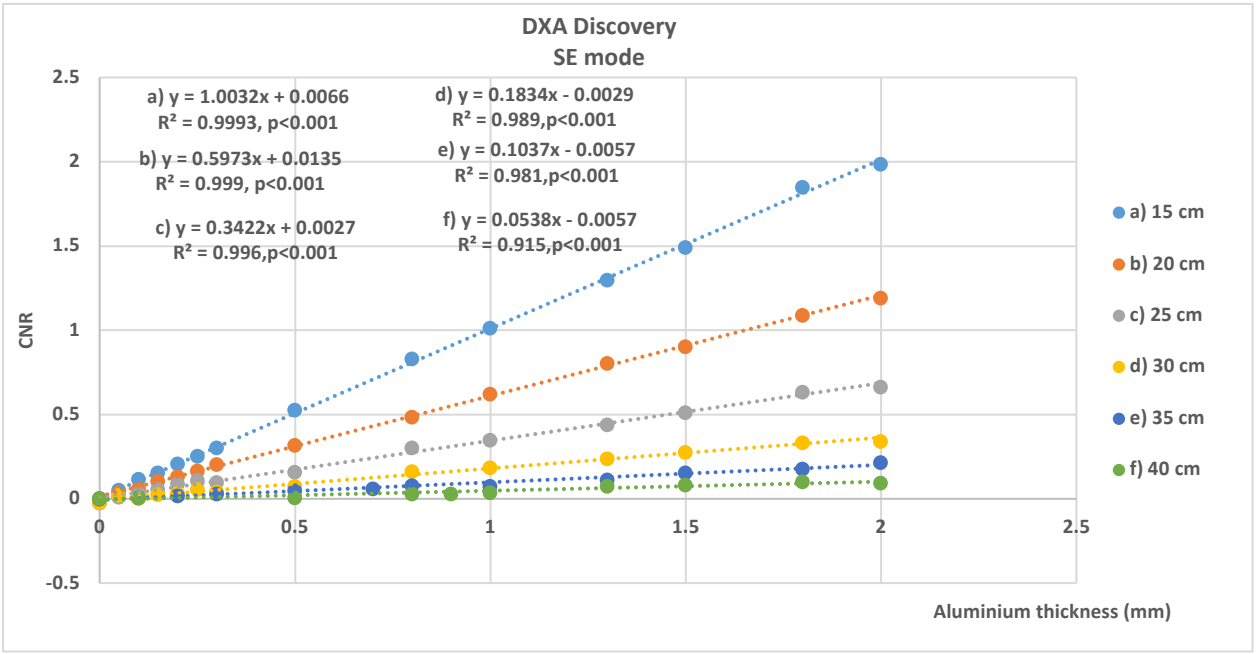


Figure 8-19: CNR as a function of Al thickness for different Perspex widths In SE mode images obtained with the Discovery scanner

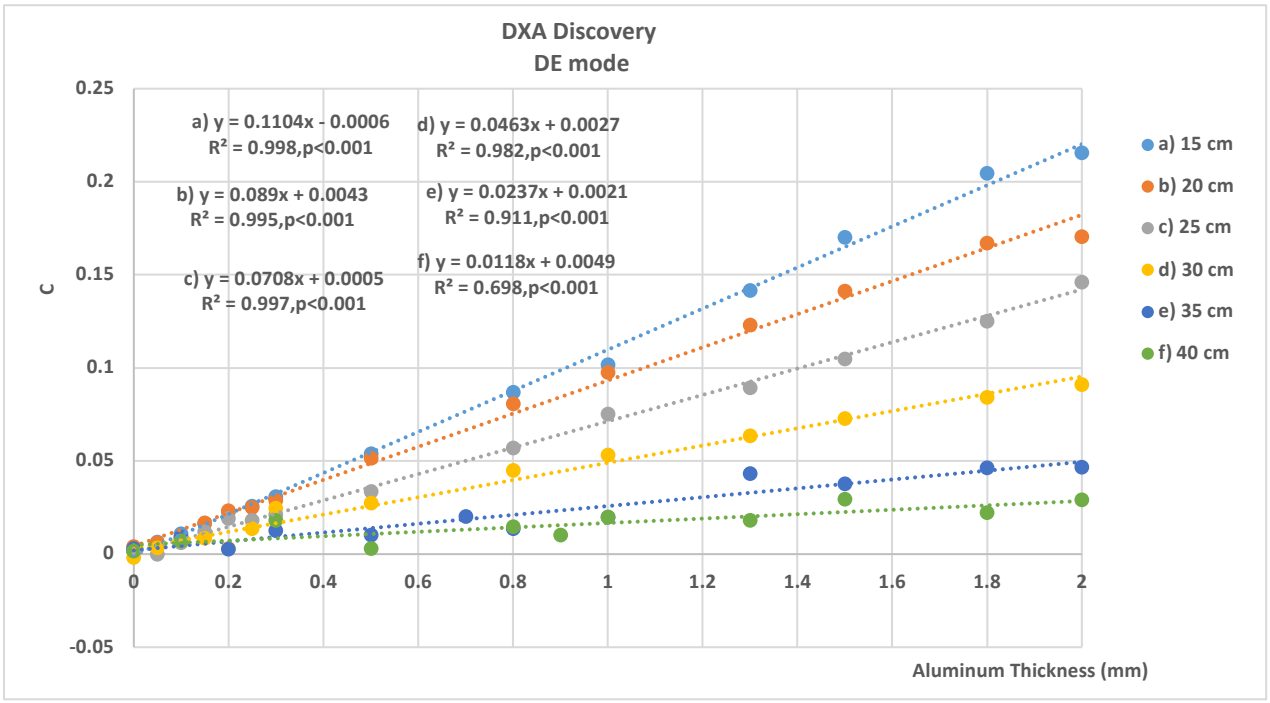


Figure 8-20: Contrast as a function of Al thickness for different Perspex widths in DE mode images obtained with the Discovery scanner

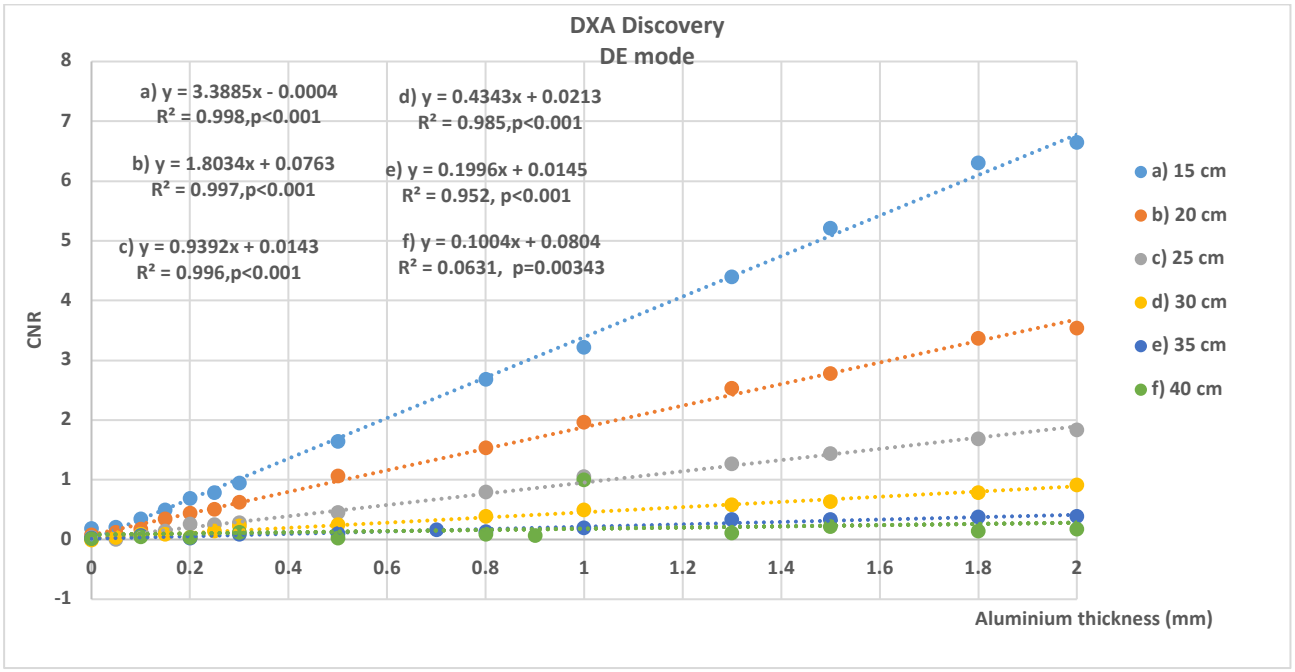


Figure 8-21: CNR as a function of the thickness of Al for different Perspex width in DE mode images performed with the Discovery

8.5.2 **Visual Assessment**

In addition to the objective method described in the previous section, a subjective approach also was used to investigate the minimum thickness of Al that could be seen on VFA images of the phantom.

The presence of aluminium the within Perspex was assessed visually on raw VFA images. Images were viewed using both Hologic and Matlab 2014 software windowing was applied for better visualisation.

Figures 8-22 and 8-23 show representative examples of images acquired with phantom configurations comprising 25 cm of Perspex combined with 0.15 mm, 0.20 mm and 0.25 mm of Al in both energy modes for both scanners.

The Al thickness that could be visualised varied from 0.05 mm at 15 cm Perspex width to 1.8 mm at 40 cm Perspex width on the Horizon and from 0.1 mm to 1.8 mm on the Discovery. Table 8-5 shows the minimum detectable aluminium thickness that could be seen on the images in both SE and DE modes together with the corresponding CNR values. These CNR values are thresholds for visibility of the strips.

An Al thickness of 0.05 mm was visible within 15 cm of Perspex on SE images obtained with the Horizon scanner and the corresponding CNR was 0.099. The same Al thickness was not clearly detectable with the same Perspex width for the Discovery scanner, for which the smallest thickness was 0.1 mm with a CNR of 0.117. This suggests a SE CNR threshold value of about 0.1 strip detectability with 15 cm Perspex.

The CNR visibility threshold for the Horizon scanner was less than for the Discovery scanner for all Perspex widths in both modes, indicating better performance.

For both scanners, the CNR visibility threshold was less in SE mode than DE mode, indicating that improved contrast with DE is offset by greater noise.

Table 8-5: The minimum detectable thickness of Al assessed visually on VFA images and the corresponding CNR values

Horizon scanner				Discovery scanner	
Perspex thickness	Energy Mode	Al Thickness (mm)	Corresponding CNR	Al Thickness (mm)	Corresponding CNR
15 cm	SE	0.05	0.0992	0.1	0.1172
	DE	0.1	0.159	0.1	0.3456
20cm	SE	0.15	0.0703	0.15	0.1050
	DE	0.15	0.2247	0.15	0.3405
25 cm	SE	0.2	0.061	0.2	0.0915
	DE	0.25	0.2004	0.2	0.2544
30cm	SE	0.5	0.101	0.5	0.0766
	DE	0.5	0.1613	0.5	0.2498
35cm	SE	0.8	0.0569	0.7	0.0618
	DE	0.9	0.1683	0.7	0.1653
40cm	SE	1.8	0.0833	1.8	0.1012
	DE	1.8	0.1003	1.8	0.1424
Mean	SE	0.079		0.092	
	DE	0.169		0.250	

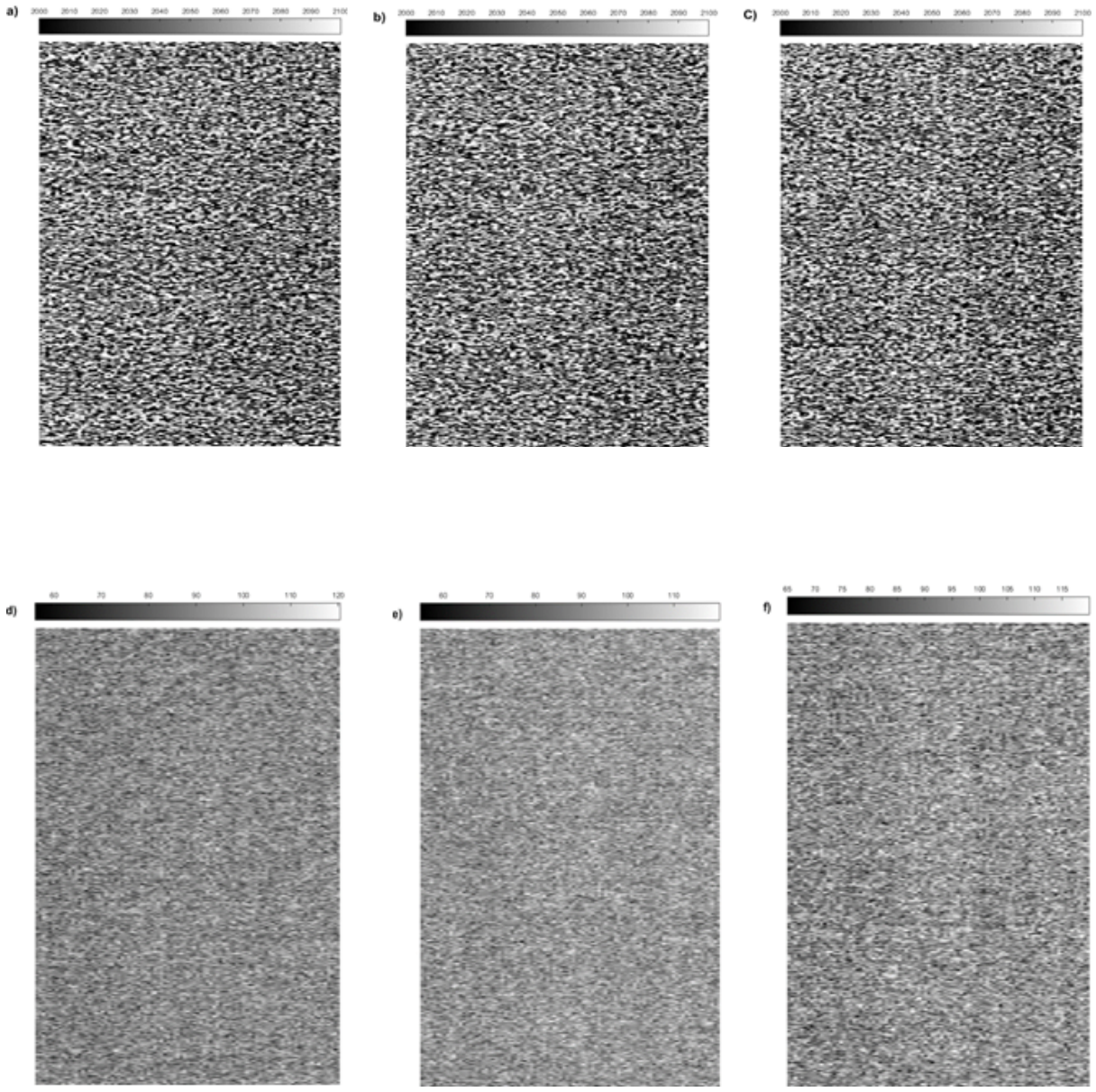


Figure 8-22: Horizon scanner VFA images of 25 cm Perspex combined with 0.15 mm Al (a) SE (not visible) and (d) DE (not visible); 0.2 mm Al (b) SE (not visible) and (e) DE (not visible) ; 0.25 mm Al (c) SE (visible) and (f) DE (visible)

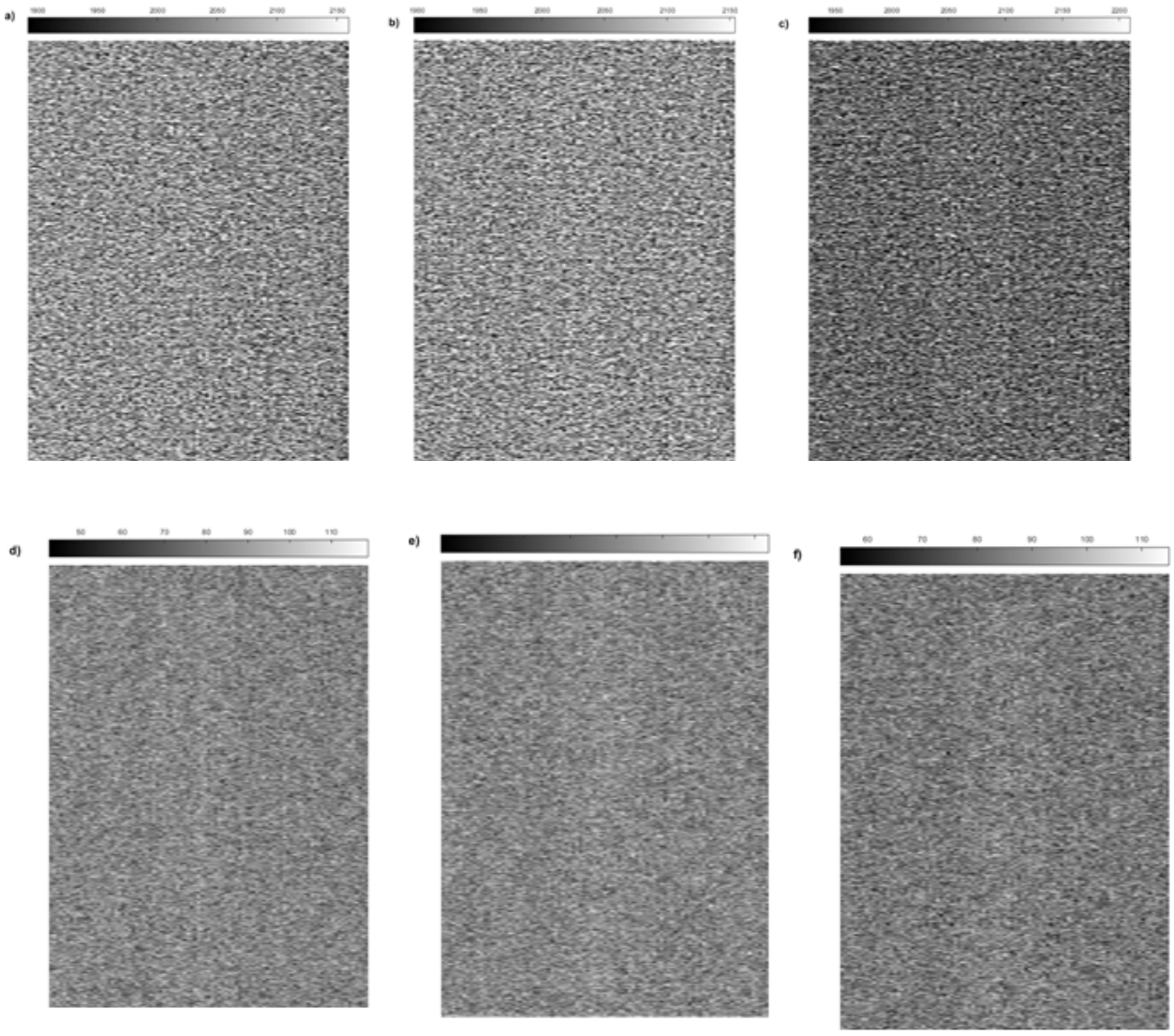


Figure 8-23: Discovery scanner VFA images of 25 cm Perspex combined with 0.15 mm Al (a) SE (not visible) and (d) DE (not visible); 0.2 mm Al (b) SE (not visible) and (e) DE (not visible); 0.25 mm Al (c) SE (visible) and (f) DE (visible)

8.6 **Repeatability and Reproducibility of C and CNR**

To evaluate the repeatability of contrast and CNR, images of a phantom combination of 15 cm, 25 cm and 40 cm Perspex blocks with Al strips of thickness 0.1 mm, 0.5 mm and 1 mm acquired five times in both SE and DE modes. All five scans were performed on the same day and under the same conditions with no repositioning. The number of repeated scans was chosen to be consistent with the availability of the clinical DXA scanners for this research and also to match previous studies (Johnson 1991).

In addition, the reproducibility of the same two quantities was measured for each phantom configuration with the phantom repositioned five times. %CV was used to measure both the repeatability and reproducibility. The %CV of C and CNR was calculated as follows:

$$\%CV\ of\ C = \frac{\sigma_C}{C} \times 100 \tag{8-7}$$

$$\%CV\ of\ CNR = \frac{\sigma_{CNR}}{CNR} \times 100 \tag{8-8}$$

where C and CNR are the mean of 5 values and σ_C and σ_{CNR} are the corresponding standard deviations.

8.6.1 **Results**

Tables 8-6 to 8-9 show the repeatability and reproducibility (%CV) of C and CNR measured with the Horizon scanner. The %CV of the pixel mean value in the Perspex ROIs also shown. Corresponding data for the Discovery scanner is presented in Appendix B.

In general, the mean Perspex pixel value was showed good repeatability and reproducibility: $\leq$ 0.2% in SE mode and $\leq$ 10 % in DE mode for Perspex widths of $\leq$ 25 cm.

With the Horizon scanner, the repeatability of the mean Perspex pixel value for Perspex widths of 15 and 25 cm was about 0.02 - 0.04% in SE mode and varied from 1% up to 9% in DE mode. For a Perspex width of 40 cm this value was about 0.1 - 0.2% in SE mode and up to 25% in DE mode. With the Discovery scanner, the SE mode results were similar for Perspex width of 15 and 25 cm but better (maximum 8%) for a 40 cm width.

For Perspex widths of 15 and 25 cm, reproducibility of the mean pixel value was similar to repeatability. In SE mode, the same was true for a Perspex width of 40 cm but in DE mode, repeatability was worse than reproducibility in general. The pattern was similar for the Discovery scanner. Again, DE reproducibility was better in the Discovery scanner than the Horizon scanner.

In all cases, the repeatability and reproducibility of C and CNR were very similar. They became worse as the width of Perspex increased (for a given thickness of aluminium) and improved as the thickness of aluminium increased (for a given Perspex width). With few exceptions, repeatability was better than reproducibility. These results are much as expected.

With the Horizon scanner, the repeatability of C and CNR of 15 cm Perspex combined with 0.1, 0.5 and 1.0 mm of aluminium was ≤ 6% in SE mode and < 9 % in DE mode. For 25 cm Perspex in SE mode, the repeatability was about 25% with 0.1 mm Al, 3% with 0.5 mm Al and <1 % with 1.0 mm Al. It was worse in DE mode, varying between about 5% with 1.0 mm Al and about 50% with 0.1 mm Al. For 40 cm of Perspex, the reproducibility was always >20%.

The Discovery scanner revealed a different trend. A Perspex width of 15 cm combined with 0.1, 0.5 and 1.0 mm of Al gave a repeatability of < 10 % in SE and DE modes. A Perspex width of 25 cm gave a repeatability about 20% with 0.1 mm Al, and < 5 % with 0.5 mm and 1.0 mm Al in SE mode, whereas corresponding values in DE mode were 18% and <10%. The repeatability with Perspex was worse overall with values between 12% and 70% in SE and DE modes.

With the Horizon scanner and 15 cm Perspex width, reproducibility was < 13% in SE and DE modes. For 25 cm Perspex with 0.1 mm Al, it was about 60% in SE mode and 70% in DE but decreased to < 5% in SE mode and $\leq$ 40% in DE mode with 0.5 mm Al and to <15% in both SE and DE modes with 1.0 mm Al. For a Perspex width of 40 cm, reproducibility was worse being >30% in both SE and DE modes.

The Discovery scanner showed a similar trend. The reproducibility was worse than the repeatability. On images with 15 cm Perspex, the reproducibility was <10% for all aluminium thicknesses in both SE and DE modes. On images with 25 cm Perspex, it was worse in SE mode; the minimum was <1% with 1 mm Al while the maximum was 30% with 0.1 mm Al.

In general, contrast and CNR was repeatable and reproduceable with a Perspex width of $\leq$ 25 cm in combination with aluminium thickness $\geq$ 0.5 mm. Both repeatability and reproducibility were poor with Perspex width of 40 cm.

Table 8-6: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 15 cm with 0.1, 0.5 and 1.0 mm of Al measured with Horizon scanner

15 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	SE	Mean of Perspex pixel value	0.0183	0.0186
		C	5.3901	5.7744
		CNR	5.4232	6.3331
	DE	Mean of Perspex pixel value	1.8507	1.1339
		C	8.3052	12.9255
		CNR	8.8743	11.8961
Al 0.5 mm	SE	Mean of Perspex pixel value	0.0262	0.0154
		C	1.9874	2.9785
		CNR	1.3162	2.9835
	DE	Mean of Perspex pixel value	1.1656	2.2599
		C	2.8351	3.2269
		CNR	2.0843	2.0934
Al 1.0 mm	SE	Mean of Perspex pixel value	0.0283	0.0344
		C	0.8810	1.8423
		CNR	0.8695	1.9267
	DE	Mean of Perspex pixel value	1.3034	1.4658
		C	2.5026	3.0112
		CNR	1.2175	1.8986

Table 8-7: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 25 cm with 0.1, 0.5 and 1.0 mm of Al measured with Horizon scanner

25 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	SE	Mean of Perspex pixel value	0.0292	0.0889
		C	24.5579	57.2288
		CNR	23.9756	57.467
	DE	Mean of Perspex pixel value	7.3071	10.034
		C	50.9945	67.427
		CNR	52.9731	69.2533
Al 0.5 mm	SE	Mean of Perspex pixel value	0.0366	0.0569
		C	2.8936	5.3308
		CNR	2.7564	4.8791
	DE	Mean of Perspex pixel value	3.775	5.6422
		C	12.2467	39.8495
		CNR	8.2234	33.147
Al 1.0 mm	SE	Mean of Perspex pixel value	0.0237	0.0216
		C	0.823	3.1526
		CNR	0.8279	3.1576
	DE	Mean of Perspex pixel value	8.6093	7.0138
		C	6.2123	14.7526
		CNR	4.4597	7.1786

Table 8-8: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 40 cm with 0.1, 0.5 and 1.0 mm of Al measured with Horizon scanner

40 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	SE	Mean of Perspex pixel value	0.1314	0.1332
		C	44.6171	61.7815
		CNR	45.0061	61.8532
	DE	Mean of Perspex pixel value	8.0813	11.2486
		C	67.4902	74.6334
		CNR	68.5326	74.3851
Al 0.5 mm	SE	Mean of Perspex pixel value	0.1524	0.1084
		C	30.165	46.1281
		CNR	29.6026	46.1779
	DE	Mean of Perspex pixel value	8.6984	14.3077
		C	45.4079	55.7257
		CNR	45.1173	48.9551
Al 1.0 mm	SE	Mean of Perspex pixel value	0.1734	0.1865
		C	26.7242	31.7483
		CNR	28.1028	33.8105
	DE	Mean of Perspex pixel value	25.8578	25.813
		C	62.6387	75.708
		CNR	50.2352	62.972

8.7 **Conclusion**

DXA is considered as the most accurate modality for the clinical diagnosis of osteoporosis and VFA imaging has emerged as a useful method for AAC detection and quantification in recent studies (Goldstein et al. 2010; Lewis et al. 2016; Elmasri et al. 2016; Grant et al. 2017). This is the first phantom study concerning the capability of DXA to detect AAC through VFA images acquired in SE and DE modes.

Two Hologic scanners were in this study: Horizon, which is the latest version, and Discovery. An experimental protocol was developed for this research and was used to obtain VFA images from both DXA scanners.

Different phantom configurations were investigated. Perspex blocks of width 15-40 cm were used as tissue equivalent material. These were combined with aluminium strips of thickness ranging between 0.05 mm and 2 mm to simulate calcification in the form of calcium hydroxyapatite.

The phantom study concentrated only on the smallest thickness of aluminium could be seen using this technique, and not on the whole range of calcification thickness. Thicknesses of up to 5 mm or more have been reported in heavily calcified arteries.

The study focused on different aspects of VFA performance related to the detection of calcification. These included studying the uniformity and noise content of VFA images acquired under different conditions. To assess uniformity profiles of mean pixel value were generates two directions: perpendicular and parallel to the scan direction. In addition, mean (SD) pixel value was plotted against Perspex thickness.

The detectability limits of calcification were also was investigated. Several approaches were applied: these included human vision and the measurement of contrast (C) and CNR. For these measurements' ROIs were selected carefully to avoid edge effects due to the aluminium strips.

With the Horizon scanner in SE mode, the mean pixel value was about 2048 for Perspex widths of 15 cm to 32 cm. This value increased to 2050 and 2054 for Perspex widths of 35 cm and 40 cm respectively. With the Discovery scanner in SE mode, the mean pixel value was 2045 for Perspex widths between 15 and 32 cm; it then slightly decreased to 2044 for 35 cm and increased to 2047 for 40 cm.

DE images revealed the same trend in the two scanners; the mean pixels value increased for Perspex width between 15 cm and 27 cm and then decreased for Perspex widths $\geq$ 30 cm.

In all cases, noise (expressed as the %CV of pixel values) increased as Perspex width increased.

Visual assessment showed that a thickness of aluminium as small as 0.05 mm and 0.1 mm can be seen on VFA images with 15 cm of Perspex with Horizon and Discovery scanners respectively. This thickness increased with Perspex width reach 1.8 mm with 40 cm of Perspex due to the increase in noise. The CNR threshold for detectability was about 0.1 in SE mode and about 0.2 - 0.3 in DE mode.

For each Perspex width, contrast and CNR increased linearly with aluminium thickness and there was a strong and significant correlation in both SE and DE modes for images acquired with both scanners. For a given thickness of aluminium, these two decreased as Perspex width increased. This was especially true for CNR indices because noise increased with |Perspex width.

8.8 **Summary**

This chapter has described a phantom study to evaluate the capability of VFA imaging with DXA to detect AAC in the early stages of its formation, where the thickness of calcification is small. The study was done in the simplest way by mimicking calcified areas of uniform thickness within the abdominal aorta.

The results of the study suggest that under idealised imaging conditions, SE and DE VFA is capable of detecting small thicknesses of calcification with good linearity and acceptable repeatability and reproducibility for body widths in the range 15 and 30.

Conclusions and Future Work

9.1 Introduction

The goal of the research presented in this thesis was to assess the suitability of DXA modality for the detection of AAC using VFA images, with the purpose of finding a low-cost method of CVD diagnosis.

The main focus of the thesis was on the automatic quantification of AAC in VFA images. The research was divided into two main parts. In the first part, an automatic method for AAC detection and quantification in VFA images was developed and evaluated in a large number of images.

Although it was possible to see calcification in the abdominal aorta on VFA images, it was difficult to measure its extent using a simple thresholding technique. There was no clear difference between calcified and normal pixels. This led to curiosity about VFA images and a desire to learn more about them and to assess their suitability for AAC detection.

In the second part of this thesis, the limitations of VFA for the detection of small calcification thicknesses were investigated. To this end, a physical phantom study was conducted; a Perspex and aluminium phantom simulating soft tissue in the abdomen and calcification in the aorta was designed and built. VFA images of different phantom configurations were acquired in single energy (SE) and dual energy (DE) modes. The minimum detectable aluminium thickness was assessed visually and related CNR.

9.2 Contributions to Knowledge

The main contributions to knowledge presented in this thesis are as follows:

1. Development of a new automatic method for AAC detection and quantification, which might allow clinicians to identify AAC as well as vertebral fractures on VFA images.

2. Validation of the developed automatic system for a wide range of AAC severity on a large dataset consisting of several hundred images;

3. Design and construction of a physical phantom to simulate both soft tissue in the abdomen and aortic calcification within this region with the capability to represent a range of abdominal width and calcification thickness;

4. Examination of the sensitivity of VFA imaging with respect to AAC detection including:

    a. Investigation of the effect of energy mode on images of different phantom configurations;

    b. Investigation of the minimum detectable thickness of calcification represented by the phantom material.

9.3 Conclusions

This research has provided further evidence of the capability of VFA to evaluate AAC. The presence of AAC on VFA images is currently considered an incidental finding, but the extent and morphology of the calcification may be considered as an additional risk factor for future CVD. Although the manual AC-24 scoring system is reliable for AAC measurements, it is limited by time and cost.

In the first part of this thesis a new approach was developed for AAC detection and quantification in VFA images. Initially the method was applied on a small set of selected VFA images with only three degrees of AAC severity. Afterwards, the same method was extended to include 4 AAC categories, including no AAC, in a large patient image data set.

The main conclusion of this part is that the new method may assist in the identification of patients with AAC and classify its severity. This may be exploited in the diagnosis of atherosclerosis before symptoms of cardiovascular disease develop. When severe AAC is detected, this should be considered in the context of patient disease characteristics and history such as advanced age, renal function, chronic inflammatory conditions, CVD, diabetes and medication (especially drugs such as statins). Including the degree of AAC alongside a BMD report may assist in patient follow-up as regards the progression of atherosclerosis and the development of CVD.

In the second part of this thesis, a physical phantom was used to investigate the limitations of VFA for the detection of AAC. The principal conclusion of the second part is that under idealised imaging conditions, VFA has the potential to be used for detecting small thicknesses of calcification with good linearity and acceptable repeatability and reproducibility in SE and DE modes for patients with body width < 30 cm.

9.4 **Limitations**

The main limitations of the research are discussed here. First of all, it is necessary to mention that the manual scoring of AAC was done by a reader (A), who, although trained, was not an expert in the interpretation of VFA images. Another limitation is that the proposed algorithm is not applicable to VFA images that contain anatomical and technical artefacts.

With regards to the phantom study, it does not reflect the real distribution of calcification within the aorta. It is an initial pilot study, which may aid in confirming the ability of VFA images to detect calcification in the aorta.

9.5 **Future Work**

For further validation of the automatic approach, interaction with relevant experts needed should be persuaded. This could include land-marking and AAC scoring of VFA images. This could be followed by clinical studies to relate AAC severity to indicies of CVD.

Furthermore, the development and use of a phantom that mimics the real distribution of calcification in the aortic walls would be more effective to assess the performance of VFA for the detection of AAC.

Reference

Adams, J.E. 1998. Single-and dual-energy: X-ray absorptiometry. In Bone densitometry and osteoporosis. Springer, Berlin, Heidelberg. pp. 305-334.

Alexopoulos, D. et al. 2003. Prognostic significance of coronary artery calcium in asymptomatic subjects with usual cardiovascular risk. American heart journal, 145(3), pp.542-548.

Ali, U. et al. 2015. A review of the properties and applications of poly (methyl methacrylate) (PMMA). Polymer Reviews, 55(4), pp.678-705.

Allison, M. A. et al. 2008. The epidemiology of abdominal aortic diameter', Journal of Vascular Surgery, 48(1), pp. 121–127.

Anderson, J. A. et al. 2000. Choice of Phantom Material and Test Protocols to Determine Radiation Exposure Rates for Fluoroscopy. Radiographics: a review publication of the Radiological Society of North America, Inc, 20(4), p.1033.

Baldelli, P. et al. 2009. A novel method for contrast-to-noise ratio (CNR) evaluation of digital mammography detectors. European radiology, 19(9), pp.2275-2285.

Balla-Arabé, S. et al. 2013. A fast and robust level set method for image segmentation using fuzzy clustering and lattice Boltzmann method. IEEE transactions on cybernetics, 43(3), pp.910-920.

Banerjee, M. et al. 1999. Beyond kappa: A review of interrater agreement measures. Canadian journal of statistics, 27(1), pp.3-23.

Bao, Q. and Chatziioannou, A.F. 2010. Estimation of the minimum detectable activity of preclinical PET imaging systems with an analytical method. Medical physics, 37(11), pp.6070-6083.

Barascuk, N. et al. 2011. Abdominal aortic calcification quantified by the Morphological Atherosclerotic Calcification Distribution (MACD) index is associated with features of the metabolic syndrome. BMC cardiovascular disorders, 11(1), p.75.

Bastos, F. et al. 2012. Calcification of the abdominal aorta as an independent predictor of cardiovascular events: a meta-analysis., Heart (British Cardiac Society), 98(13), pp. 988–94.

Benjelloun, M.et al. 2011. A framework of vertebra segmentation using the active shape model-based approach. Journal of Biomedical Imaging, 2011, p.9.

Bentzon, J.F.et al. 2014. Mechanisms of plaque formation and rupture. Circulation research, 114(12), pp.1852-1866.

Benvenuti, L.A. et al. 2005. Different patterns of atherosclerotic remodeling in the thoracic and abdominal aorta. Clinics, 60(5), pp.355-360.

Bhattachayya. 2016. Image Segmentation: A Review, springer, (978-3-319-47523–3), pp. 29–41.

Blake, G.M. and Fogelman, I. 1997. Technical principles of dual energy x-ray absorptiometry. In Seminars in nuclear medicine (Vol. 27, No. 3, pp. 210-228). Elsevier.

Bolland, M.J.et al. 2010. Abdominal aortic calcification on vertebral morphometry images predicts incident myocardial infarction. Journal of Bone and Mineral Research, 25(3), pp.505-512.

Bots, M.L. et al.1993. Carotid intima-media wall thickness in elderly women with and without atherosclerosis of the abdominal aorta. Atherosclerosis, 102(1), pp.99-105.

Braillon, P. et al.1998. Precision of body composition measurements by dual energy X-ray absorptiometry. Applied radiation and isotopes, 5(49), pp.501-502.

Bruijne, M.et al. 2004. Interactive segmentation of abdominal aortic aneurysms in CTA images. Medical Image Analysis, 8(2), pp.127-138.

Bruijne, M. 2005. A pattern classification approach to aorta calcium scoring in radiographs. In International Workshop on Computer Vision for Biomedical Image Applications (pp. 170-177). Springer, Berlin, Heidelberg.

Bukowska, J.et al. Optical spectroscopy and computational methods in biology and medicine (Vol. 14). Springer Science & Business Media.

Burger, W., et al. 2009. Principles of digital image processing (p. 221). London: Springer.

Caldas, M.D.P. et al. 2010. Comparative evaluation among different materials to replace soft tissue in oral radiology studies. Journal of Applied Oral Science, 18(3), pp.264-267.

Cecelja, M. et al. 2013. Abdominal aortic calcification detection using dual-energy X-ray absorptiometry: validation study in healthy women compared to computed tomography. Calcified tissue international, 92(6), pp.495-500.

Chitradevi, A. and Sadasivam, V. 2016. Various approaches for medical image segmentation: A survey, Current Medical Imaging Reviews, 12(2), pp. 77–94.

Chow, J.T. et al. 2008. Abdominal Aortic Calcification, BMD, and Bone Microstructure: A Population-Based Study. Journal of Bone and Mineral Research, 23(10), pp.1601-1612.

Ciaccio, E.J. 2011. Information Applied Medical Image Processing: A Basic Course, by Wolfgang Birkfellner, Associate Professor, Department of Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria. 10(1), p. 101.

Conrad-Hansen, L.A. et al. 2007. October. Quantifying calcification in the lumbar aorta on X-ray images. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg. pp. 352-359

Cootes, T.et al. 2000. An introduction to active shape models. Image processing and analysis, pp.223-248.

Cootes, T.F. et al.1995. Active shape models-their training and application. Computer vision and image understanding, 61(1), pp.38-59.

Cootes, T.F.et al. 2001. Active appearance models. IEEE Transactions on Pattern Analysis & Machine Intelligence, (6), pp.681-685.

Cootes, T.F. and Taylor, C.J. 2001. Statistical models of appearance for medical image analysis and computer vision. In Medical Imaging 2001: Image Processing. International Society for Optics and Photonics. Vol. 4322, pp. 236-249

Criqui, M.H.et al. 2014. Abdominal aortic calcium, coronary artery calcium, and cardiovascular morbidity and mortality in the Multi-Ethnic Study of Atherosclerosis. Arteriosclerosis, thrombosis, and vascular biology, 34(7), pp.1574-1579.

Cullum, I.D.et al. 1989. X-ray dual-photon absorptiometry: a new method for the measurement of bone density. The British journal of radiology, 62(739), pp.587-592.

Dae-Cheol, K. et al. 2010. Assessment of the effect of bone density and soft tissue thickness on phantom measurements. Journal of the Korean Physical Society, 57(5), pp.1263-1269.

Maroules, C.D.et al. 2013. Abdominal aortic atherosclerosis at MR imaging is associated with cardiovascular events: the Dallas heart study. Radiology, 269(1), pp.84-91.

Davies, M. and Fleiss, J.L., 1982. Measuring agreement for multinomial data. Biometrics, pp.1047-1051.

Demšar, J.et al. 2013. Orange: data mining toolbox in Python. The Journal of Machine Learning Research, 14(1), pp.2349-2353.

Desai, N. et al. 2010. Practical evaluation of image quality in computed radiographic (CR) imaging systems. In Medical Imaging 2010: Physics of Medical Imaging. International Society for Optics and Photonics. Vol. 7622, p. 76224Q.

Dey, N. et al. 2012. Detection and measurement of arc of lumen calcification from intravascular ultrasound using harris corner detection. In Computing and Communication Systems (NCCCS), 2012 National Conference on (pp. 1-6). IEEE.

Dickerscheid, D.et al. 2013. Contrast-noise-ratio (CNR) analysis and optimisation of breast-specific gamma imaging (BSGI) acquisition protocols. EJNMMI research, 3(1), p.21.

Dmitrovsky, E. and Boskey, A.L., 1985. Calcium-acidic phospholipid-phosphate complexes in human atherosclerotic aortas. Calcified tissue international, 37(2), pp.121-125.

Doherty, T.M. et al. 2004. Genetic determinants of arterial calcification associated with atherosclerosis. In Mayo Clinic Proceedings, Vol. 79, No. 2, pp. 197-210. Elsevier.

Drampalos, E.et al. 2015. Vertebral fracture assessment: Current research status and application in patients with kyphoplasty. World journal of orthopedics, 6(9), p.680.

Dunmore-Buyze, P.J.et al. 2002. In vitro investigation of calcium distribution and tissue thickness in the human thoracic aorta. Physiological measurement, 23(3), p.555.

Elmasri, K. et al. 2016. Automatic detection and quantification of abdominal aortic calcification in dual energy X-ray absorptiometry. Procedia Computer Science, 96, pp.1011-1021.

Erbel, R. and Eggebrecht, H. 2006. Aortic dimensions and the risk of dissection. Heart, 92(1), pp.137-142.

İrem, E.R.K. et al. 2016. Determination of Photon Mass Attenuation Coefficient for Some Phantom Materials using GATE Code and Comparison with Experimental and XCOM

Data. International Journal of Nuclear and Radiation Science and Technology (IJNURASAT), 1(2), pp.11-13.

Fasihi, M.S. and Mikhael, W.B. 2016. Overview of current Biomedical Image segmentation methods. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 803-808). IEEE.

Fayad, Z.A. and Fuster, V. 2001. Clinical imaging of the high-risk or vulnerable atherosclerotic plaque. Circulation research, 89(4), pp.305-316.

Fernández-Delgado, M. et al. 2014. Do we need hundreds of classifiers to solve real world classification problems?. The Journal of Machine Learning Research, 15(1), pp.3133-3181.

Finkelstein, J.S.et al.1994. A comparison of lateral versus anterior-posterior spine dual energy x-ray absorptiometry for the diagnosis of osteopenia. The Journal of Clinical Endocrinology & Metabolism, 78(3), pp.724-730.

Fishbein, M.C. 2010. The vulnerable and unstable atherosclerotic plaque. Cardiovascular Pathology, 19(1), pp.6-11.

Fitzpatrick, L.A., et al. 1994. Diffuse calcification in human coronary arteries. Association of osteopontin with atherosclerosis. The Journal of clinical investigation, 94(4), pp.1597-1604.

Chuang, M. L. et al. 2014. Distribution of abdominal aortic calcium by computed tomography: impact of analysis method on quantitative calcium score. Academic radiology 20, no. 11 (2013): 1422-1428.

Ganz, M. et al. 2012. Distribution, size, and shape of abdominal aortic calcified deposits and their relationship to mortality in postmenopausal women. Journal of Biomedical Imaging, 2012, p.8.

Genant, H. K. et al.1993. Vertebral fracture assessment using a semiquantitative technique, Journal of Bone and Mineral Research, 8(9), pp. 1137–1148.

Golestani, R.et al. 2010. Abdominal aortic calcification detected by dual X-ray absorptiometry: a strong predictor for cardiovascular events. Annals of medicine, 42(7), pp.539-545.

Gonzalez, R.C. et al. 2004. Digital image processing using MATLAB (Vol. 624). Upper Saddle River, New Jersey: Pearson-Prentice-Hall.

Goodall, C., 1991. Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society. Series B (Methodological), pp.285-339.

Goyal, A. et al. 2012. A comprehensive review of image smoothing techniques. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(4), pp.pp-315.

Grant, M. et al. 2017. A novel quantitative approach to the measurement of abdominal aortic calcification as applied to the Canadian Multicenter Osteoporosis Study (CaMOS). Bone, 97, pp.201-208.

Guglielmi, G. et al. 2008. Vertebral morphometry: current methods and recent advances. European radiology, 18(7), pp.1484-1496.

Guglielmi, G., et al. 2014. Quality assurance in bone densitometry. Current Radiology Reports, 2(2), p.33.

Haka, A.S. et al. 2002. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. Cancer research, 62(18), pp.5375-5380.

Harefa, J. et al. 2017. Comparison classifier: support vector machine (SVM) and K-nearest neighbor (K-NN) in digital mammogram images. Jurnal Informatika dan Sistem Informasi, 2(2), pp.35-40.

Hartshorne, T.C.et al. 2011. Ultrasound measurement of aortic diameter in a national screening programme. European Journal of Vascular and Endovascular Surgery, 42(2), pp.195-199.

Heimann, T. and Meinzer, H.P. 2009. Statistical shape models for 3D medical image segmentation: a review. Medical image analysis, 13(4), pp.543-563.

Hendee, W.R., 2003. Medical imaging physics. John Wiley & Sons.

Henderson, L. 2004. The National Diet and Nutrition Survey: adults aged 19 to 64 years. Energy, protein, carbohydrate, fat and alcohol intake, 2.

Higgins, C. L.et al. 2005. Quantification of calcification in atherosclerotic lesions, Arteriosclerosis, Thrombosis, and Vascular Biology, 25, pp. 1567–1576.

Hollander, M. et al. 2003. Comparison between measures of atherosclerosis and risk of stroke: The Rotterdam study, Stroke, 34(10), pp. 2367–2372.

Hologic 2003. Advanced Point-of-Care Bone Health Assessment. available at: http://www.brownsmedicalimaging.com/wp-content/uploads/HOLOGIC-DXA-Discovery-Specs.

Hologic 2007. Discovery Better Patient Care Rests on Bone Health Testing. Available at: http://www.brownsmedicalimaging.com/wp-content/uploads/HOLOGIC-DXA-Discovery-Brochure.

Honeycutt, C.E. and Plotnick, R., 2008. Image analysis techniques and gray-level co-occurrence matrices (GLCM) for calculating bioturbation indices and characterizing biogenic sedimentary structures. Computers & Geosciences, 34(11), pp.1461-1472.

Hong, C., Bae, K.T. and Pilgram, T.K., 2003. Coronary artery calcium: accuracy and reproducibility of measurements with multi–detector row CT—assessment of effects of different thresholds and quantification methods. Radiology, 227(3), pp.795-801.

Honkanen, E.et al. 2008. Abdominal aortic calcification in dialysis patients: results of the CORD study. Nephrology Dialysis Transplantation, 23(12), pp.4009-4015.

Horng, M.H., 2010. Multilevel minimum cross entropy threshold selection based on the honey bee mating optimization. Expert Systems with Applications, 37(6), pp.4580-4592.

Hospers, I.C.et al. 2009. Vertebral fracture assessment in supine position: comparison by using conventional semiquantitative radiography and visual radiography. Radiology, 251(3), pp.822-828.

Hricak, H. 2013. Osteoporosis and bone densitometry measurements. New York, pp.123-132.

Huang at al. 2013. Medical Image Segmentation, Advanced Materials Research, (i), pp. 1–35.

IAEA 2010. Dual energy x ray absorbitiometry for bone mineral density and body composition assessment, IAEA HUMAN HEALTH SERIES PUBLICATIONS.

Ichii, M. et al. 2013. Quantitative analysis of abdominal aortic calcification in CKD patients without dialysis therapy by use of the Agatston score., Kidney & blood pressure research, 38(2–3), pp. 196–204.

Imaging, S. 2013. Densities and Characteristics of Materials used in Standard Imaging Products, (608), p. 53562. Available at: https://www.standardimaging.com/uploads/manuals/Lucy_MNL_80542-10.pdf

Jain, R., Kasturi, R. and Schunck, B. G. (1995) Machine Vision Chapter 7 Texture, pp. 234–248.
Jakubiak, R. R. et al. (2013). Image quality, threshold contrast and mean glandular dose in CR mammography., Physics in medicine and biology, 58(18), pp. 6565–83.

Jang, S. et al. 2012. Original article Relation of aortic calcification, wall thickness, and distensibility with severity of coronary artery disease : evaluation with coronary CT angiography, pp. 839–844.

Jayalath, R.W.et al. 1991. Precision and stability of dual-energy X-ray absorptiometry measurements, 49(3), pp. 174–178.

Jonson, R. 1993. Mass attenuation coefficients, quantities and units for use in bone mineral determinations., Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA, 3, pp. 103–106.

K.Sreedhar 2012. Enhancement of Images Using Morphological Transformations, International Journal of Advanced Research in Computer Science and softwae engineering, 4(1), pp. 33–50.

Kauppila et al. 1997. New indices to classify location, severity and progression of calcific lesions in the abdominal aorta: a 25-year follow-up study., Atherosclerosis, 132(2), pp. 245–50.

Kaur, M. and Goyal, P. 2015. A Review on Region Based Segmentation, International Journal of Science and Research (IJSR), 4(4), pp. 3194–3197.

Kaur, S. and Singh, E. R. 2015. Image De-noising Techniques : A review paper, 2(8), pp. 1649–1653.

Kelly, T. L.et al. 1998. DXA body composition: theory and practice., Applied radiation and isotopes : including data, instrumentation and methods for use in agriculture, industry and medicine, 49(5–6), pp. 511–3.

Kim, E.-D. et al. 2013. Association of abdominal aortic calcification with lifestyle and risk factors of cardiovascular disease., Korean journal of family medicine, 34(3), pp. 213–20.

Kotsiantis, S. B. 2007. Supervised machine learning: A review of classification techniques, Informatica, 31, pp. 249–268.

Kumar, M. and Rana, A. 2016. Image Enhancement using Contrast Limited Adaptive Histogram Equalization and Wiener filter, 5(16977), pp. 16977–16979.

Laroche, M. and Delmotte, A. 2005. Clinical Investigations Increased Arterial Calcification in Paget Õ s Disease of Bone, pp. 129–133.

Laughlin, G. a et al. 2011. Abdominal aortic diameter and vascular atherosclerosis: the Multi-Ethnic Study of Atherosclerosis., European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery. Elsevier Ltd, 41(4), pp. 481–7.

Leckstroem, D. C. T. et al. 2014. Prevalence and predictors of abdominal aortic calcification in healthy living kidney donors., International urology and nephrology, 46(1), pp. 63–70.

Lee, J. S. et al. 2012. Detection of hydroxyapatite in calcified cardiovascular tissues, Atherosclerosis, 224(2), pp. 340–347.

Lee, S.M. et al. 2017. Abdominal aortic calcification score among several vascular calcification scores of plain radiograph is the most reliable predictor of severe coronary artery calcification in dialysis patients. *Renal failure*, *39*(1), pp.729-735.

Leeflang, M. M. G. et al. 2018. Annals of Internal Medicine Academia and Clinic, Academia and Clinic, 149(12), pp. 889–898.

Lewis, J. R. et al. 2016. Abdominal Aortic Calcification Identified on Lateral Spine Images from Bone Densitometers Are a Marker of Generalized Atherosclerosis in Elderly Women, Arteriosclerosis, Thrombosis, and Vascular Biology, 36(1), pp. 166–173.

Lewis, M. et al. 2012. Comparative CTDI measurements in Perspex and water equivalent dosimetry phantoms.13(7). pp.279-280

Liao, P. S. at al. 2001. A fast algorithm for multilevel thresholding, Journal of Information Science and Engineering, 17(5), pp. 713–727.

Liao, S. et al. 2009. Fabrication of nano-hydroxyapatite/collagen/osteonectin composites for bone graft applications., Biomedical materials (Bristol, England), 4(2), p. 025019.

Lichman, M. 2013. UCI Machine Learning Repository, School of Information and Computer Science, pp. 21-24.

Lillemark, et al. 2010. Growth patterns of abdominal atherosclerotic calcified deposits from lumbar lateral X-rays., The international journal of cardiovascular imaging, 26(7), pp. 751–61.

Lorente-Ramos, R. et al. 2011. Dual-energy x-ray absorptiometry in the diagnosis of osteoporosis: A practical guide, American Journal of Roentgenology, 196(4), pp. 897–904.

Lorkowski, S. and Cullen, P. 2007. Atherosclerosis: pathogenesis, clinical features and treatment, eLS, pp. 1–11.

El Maghraoui, A. and Roux, C. 2008. DXA scanning in clinical practice., QJM : monthly journal of the Association of Physicians, 101(8), pp. 605–17.

Majnik, M. and Bosni, Z. 2013. ROC analysis of classifiers in machine learning: A survey, Intelligent Data Analysis, 17(3), pp. 531–558.

Makandar, A. 2015. Image Enhancement Techniques using Highpass and Lowpass Filters, 109(14), pp. 12–15.

Manjunatha, H. C. and Rudraswamy, B. 2012. Energy absorption and exposure build-up factors in hydroxyapatite, Radiation Measurements. Elsevier Ltd, 47(5), pp. 364–370.

Mark Grant et al. 2017. A novel quantitative approach to the measurement of abdominal aortic calcification as applied to the Canadian Multicenter Osteoporosis Study (CaMOS), Bone. Elsevier Inc., 97, pp. 201–208.

Mchugh, M. L. 2012. Lessons in biostatistics Interrater reliability: the kappa statistic, pp. 276–282.

Melinska, A. U. et al. 2015. Statistical, Morphometric , Anatomical Shape Model (Atlas) of Calcaneus, pp. 1–15.

Michael, G. J. and Henderson, C. J. 1998. Monte Carlo modelling of an extended DXA technique., Physics in medicine and biology, 43, pp. 2583–2596.

Mihailescu, D. and Borcia, C. 2006. Water equivalency of some plastic materials used in electron dosimetry: A Monte Carlo investigation, Romanian Reports in Physics, 58(4), pp. 415–425.

Mohammad, A. et al. 2014. Vertebral fracture assessment-detected abdominal aortic calcification and cardiovascular disease in rheumatoid arthritis., Seminars in arthritis and rheumatism. Elsevier, 43(5), pp. 632–7.

Mudry, A. and Tjellström, A. 2011. Historical background of bone conduction hearing devices and bone conduction hearing aids, Advances in Oto-Rhino-Laryngology, 71, pp. 1–9.

Nelitz, M. et al. 1999. Reliability of radiological measurements in the assessment of hip dysplasia in adults, British Journal of Radiology, 72(APR.), pp. 331–334.

Netto, O. S. et al. 2013. Detection of abdominal aortic calcification by densitometry, Radiol Bras, 46(1), pp. 35–38.

Nick Townsend at al. 2015. Cardiovascular Disease Statistic 2015.

Nielsen, M. et al. 2010. Distribution, size, shape, growth potential and extent of abdominal aortic calcified deposits predict mortality in postmenopausal women, BMC Cardiovascular Disorders, 10(1), p. 56.

Nissen, S. E. and Yock, P. 2001. Clinical Cardiology : New Frontiers Novel Pathophysiological Insights and Current Clinical Applications Rationale for Ultrasound Imaging, pp. 604–617.

Norouzi, A. 2014. Medical Image Segmentation Methods , Algorithms , and Applications, IETE TECHNICAL REVIEW, (June).

Nygren, P. and Jasinski, M. 2016. A Comparative Study of Segmentation and Classification Methods for 3D Point Clouds.

ODonnell, C. J. et al. 2002. Evidence for heritability of abdominal aortic calcific deposits in the Framingham Heart Study, Circulation, 106(3), pp. 337–341.

Oei, H. H. S. et al. 2002. The association between coronary calcification assessed by electron beam computed tomography and measures of extracoronary atherosclerosis: The Rotterdam Coronary Calcification Study, Journal of the American College of Cardiology. Elsevier Masson SAS, 39(11), pp. 1745–1751.

Oei, L. et al. 2016. Quantitative imaging methods in osteoporosis, Quantitative Imaging in Medicine and Surgery, 6(6), pp. 680–698.

Okuno, S. et al. 2007. Presence of Abdominal Aortic Calcification Is Significantly Associated With All-Cause and Cardiovascular Mortality in Maintenance Hemodialysis Patients, American Journal of Kidney Diseases, 49(3), pp. 417–425.

Organization, W. H. 2004. WHO SCIENTIFIC GROUP ON THE ASSESSMENT OF OSTEOPOROSIS AT PRIMARY HEALTH Care Level, World Health Organization, May(May 2004), pp. 5–7.

Orrù, G. et al. 2012. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review, Neuroscience and Biobehavioral Reviews, 36(4), pp. 1140–1152.

Ortlepp, J. R. et al. 2004. The amount of calcium-deficient hexagonal hydroxyapatite in aortic valves is influenced by gender and associated with genetic polymorphisms in patients with severe calcific aortic stenosis., European heart journal, 25(6), pp. 514–22.

Palm, A. and LoSasso, T. 2005. Influence of phantom material and phantom size on radiographic film response in therapy photon beams., Medical physics, 32(8), pp. 2434–2442.

Pariente-rodrigo, E.et al. 2016. Reliability of radiologic evaluation of abdominal aortic calcification using the 24-point scale ☪, 58(1), pp. 46–54.

Palm, D.L. et al.1998. A survey of current methods in medical image segmentation. Technical report, Johns Hopkins University, Baltimore.

Preethi, G. and Sornagopal, V. 2014. MRI Image Classification Using GLCM Texture Features, 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), pp. 1–6.

Prieto, R. M. et al. 2011. Study on the structure and composition of aortic valve calcific deposits : Etiological aspects, 2(1), pp. 19–25.

Rafael C. Gonzalez, R. E. W. 2002. Digital Image Processing. 2nd edn. Prentice Hall.

Ram, G. B. (2012). Discovery DXA System. Available at: http://www.brownsmedicalimaging.com/wp-content/uploads/HOLOGIC-DXA-Discovery-Brochure.

Ricotta, J. J. et al. 2008. Cardiovascular disease management: the need for better diagnostics., Medical & biological engineering & computing, 46, pp. 1059–1068.

Ridge, C. A. et al. 2016. Differentiating between Subsolid at CT : Inter- and Intraobserver Thoracic Radiologists 1, Radiology, 278(3), pp. 1–9.

Roberts, M. G.et al. 2003. Linking Sequences of Active Appearance Sub-Models via Constraints: An Application in Automated Vertebral Morphometry, Procedings of the British Machine Vision Conference 2003, p. 38.1-38.10.

Sanz, J. and Fayad, Z. a 2008. Imaging of atherosclerotic cardiovascular disease., Nature, 451(February), pp. 953–957.

Sarkalkan, N. et al. 2014. Statistical shape and appearance models for fast and automated estimation of proximal femur fracture load using 2D finite element models., Journal of biomechanics. Elsevier, 47(12), pp. 3107–14.

Sasi, N. M. and Jayasree, V. K. 2013. Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images, Engineering, 5(October), pp. 326–331.

Schmitz, B. L. et al. 2005. Noninvasive Coronary Angiography With Multislice Computed Tomography, 293(20).

Schousboe, J.T. et al. 2010. Densitometric vertebral fracture assessment (VFA), International Osteoporosis Foundation, 12, 1–24.

Schousboe, J. T. et al. 2008. Stroke in Older Women, 23(3).

Schousboe, J. T. et al. 2017. Abdominal aortic calci fi cation on dual-energy X-ray absorptiometry : Methods of assessment and clinical signi fi cance , 104, pp. 91–100.

Schousboe, J. T.et al. 2007. Detection of aortic calcification during vertebral fracture assessment (VFA) compared to digital radiography., PloS one, 2(8).

Schousboe, J. T. et al. 2006. Detection of abdominal aortic calcification with lateral spine imaging using DXA., Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry, 9(3), pp. 302–8.

Seerha, G. 2013. Review on Recent Image Segmentation Techniques, 5(02), pp. 109–112.

Shakeri, A. et al. 2013. Correlation between Aortic Wall Thickness and Coronary Artery Disease by 64 Slice Multidetector Computed Tomography., Journal of cardiovascular and thoracic research, 5(3), pp. 91–5.

Shepherd, John A. et al. 2002. Measurement of breast density with dual X-ray absorptiometry: feasibility." Radiology 223, no. 2 (2002): 554-557.

Shepherd, W. et al. 2009. X-ray Absorptiometry, Journal of the ICRU, 9(1), pp. 37–58.

Singh, A. 2016. A Review of Supervised Machine Learning Algorithms, pp. 1310–1315.

Singh, V. P. et al. 2014. Determination of Effective Atomic Numbers Using Different Methods for Some Low- Z Materials, Journal of Nuclear Chemistry, 2014, pp. 1–7.

Smyth, P. P.,et al. 1999. Vertebral Shape : Automatic Measurement with Active, Radiology.

Songur, A. et al. 2010. Abdominal aorta and its branches: morphometry—variations in autopsy cases, Eur J Gen Med, 7, pp. 321–325.

Speller, R. D. et al. 2006. Photon absorptiometry , bone densitometry and the challenge of osteoporosis.

Szulc, P. et al. 2013. Severity of aortic calcification is positively associated with vertebral fracture in older men--a densitometry study in the STRAMBO cohort., Osteoporosis international : a j, 24(4), pp. 1177–84.

Tatami, Y. et al. 2015. Impact of abdominal aortic calci fi cation on long-term cardiovascular outcomes in patients with chronic kidney disease, Atherosclerosis. Elsevier Ltd, 243(2), pp. 349–355.

Taylor, C. 2016. Evaluation of the effects of positioning and configuration on contrast-to-noise ratio in the quality control of a 3D Accuitomo 170 dental CBCT system, Dentomaxillofacial

Radiology, 45(5).

Toussaint, N. D. et al. 2009. Determination and validation of aortic calcification measurement from lateral bone densitometry in dialysis patients, Clinical Journal of the American Society of Nephrology, 4, pp. 119–127.

Tsanas, A., Little, M.A. and McSharry, P.E., 2013. A methodology for the analysis of medical data. In Handbook of Systems and Complexity in Health (pp. 113-125). Springer, New York, NY.

Tuttolomondo, A. et.al. 2012. Atherosclerosis as an inflammatory disease. Current pharmaceutical design, 18(28), pp.4266-4288.

Van Metter, R.L. 2000a. Handbook of medical imaging, Volume 1. Physics and psychophysics. SPIE Press, 0819436216, p. 949.

Van Metter, R.L. 2000b. Handbook of Medical Imaging Volume 1, SPIE Press, 0819436216, p. 2000.

Vasuki, P. and Devi, M. B. 2017. A Survey on Image Preprocessing Techniques for Diverse Fields of Medical Imagery, IEEE, pp. 1–6.

Ververidis, D. and Kotropoulos, C. 2008. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition, Signal Processing, 88(12), pp. 2956–2970.

Villarino, M. 2005. Ramanujans Perimeter of an Ellipse, arXiv preprint math/0506384, 3(July 2005), pp. 1–12. Available at: http://arxiv.org/abs/math/0506384.

Walsh, C. R. et al. 2002. Abdominal aortic calcific deposits are associated with increased risk for congestive heart failure: The Framingham Heart Study, American Heart Journal, 144(4), pp. 733–739.

Wang, J. and Tan, Y. 2011. Morphological Image Enhancement Procedure Design By Using Genetic Programming Categories and Subject Descriptors, 2011, pp. 1435–1442.

Warren, L. M. et al. 2013. Comparison of the x-ray attenuation properties of breast calcifications, aluminium, hydroxyapatite and calcium oxalate., Physics in medicine and biology, 58(7), pp. N103-13.

Wilson, K.E., 2006. Detection of abdominal aortic calcification with IVA. Bedford MA, USA: Hologic Inc.

Wilson, K.E. and Kelly, T.L., 2013. Horizon™ DXA System: Technical and Clinical Advantages.

Wilson, P. W. et al. 2001. Abdominal aortic calcific deposits are an important predictor of vascular morbidity and mortality., Circulation, 103(CVD), pp. 1529–1534.

Withey, D. J. and Koles, Z. J. 2008. A Review of Medical Image Segmentation : Methods and Available Software, Methods, 10(3), pp. 125–148.

Wong, N.D.et al. 2011. Abdominal aortic calcium and multi-site atherosclerosis: the Multiethnic Study of Atherosclerosis. *Atherosclerosis*, *214*(2), pp.436-441.

Wu, T. et al. (2012) Hologic Bone Densitometry and the Evolution of DXA, pp. 1–4. Available at: http://www.hologic.ca/sites/default/files/white-papers/WP-00068%20HistoryDXA.pdf

Yamamoto, D.et al. 2016. Predictors of abdominal aortic calcification progression in patients with chronic kidney disease without hemodialysis. *Atherosclerosis*, *253*, pp.15-21.

Zhuocai, W.et al., 2011. Extraction of prostatic lumina and automated recognition for prostatic calculus image using PCA-SVM. Computational and mathematical methods in medicine, 2011.

Appendix A: Linear regression and correlation results

Table A1: Linear regression and correlation of mean pixel value with distance in the X-direction for a range of Perspex widths

Perspex width	Horizon scanner								Discovery scanner							
	Slope	Intercept	r^2	p-value	Slope	Intercept	r^2	p-value	Slope	Intercept	r^2	p-value	Slope	Intercept	r^2	p-value
	SE				DE				SE				DE			
15 cm	-0.015	2051.1	0.366	<0.001	0.0048	61.941	0.693	<0.001	-0.009	2046.8	0.182	<0.001	-0.001	59.93	0.03	0.00306
20 cm	-0.005	2049.5	0.056	<0.001	0.0066	73.972	0.696	<0.001	-0.002	2045.7	0.008	0.112	-0.001	71.83	0.07	<0.001
25 cm	0.0007	2048.0	0.0006	0.652	0.0080	87.855	0.631	<0.001	-0.004	2045.1	0.015	0.035	-0.001	84.66	0.018	0.0199
30 cm	0.0032	2047.8	0.005	0.201	0.0062	80.017	0.403	<0.001	-0.002	2044.8	0.001	0.585	0.001	86.35	0.0025	0.386
35 cm	0.01043	2048.0	0.011	0.050	0.0042	74.253	0.274	<0.001	0.005	2044.5	0.0018	0.46	0.003	79.74	0.053	<0.001
40 cm	0.0463	2045.8	0.047	<0.001	0.0015	7.048	0.0429	<0.001	-0.014	2051.4	0.0045	0.228	0.007	70.36	0.140	<0.001

Appendix B: Repeatability and Reproducibility

Table B1: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 15 cm with 0.1, 0.5 and 1.0 mm of Al measured with Discovery scanner

15 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	**SE**	Mean of Perspex pixel value	0.0119	0.0241
		C	8.4815	9.505
		CNR	8.8361	9.81
	DE	Mean of Perspex pixel value	0.9399	0.9109
		C	5.5126	6.9489
		CNR	6.0092	6.6894
Al 0.5 mm	**SE**	Mean of Perspex pixel value	0.0179	0.0307
		C	0.4931	0.7434
		CNR	2.2844	2.3801
	DE	Mean of Perspex pixel value	2.0487	2.1392
		C	3.7019	4.9793
		CNR	2.3256	3.3942
Al 1.0 mm	**SE**	Mean of Perspex pixel value	0.0223	0.0379
		C	1.2773	3.7591
		CNR	1.0897	4.5508
	DE	Mean of Perspex pixel value	1.4544	0.7388
		C	0.6367	1.9863
		CNR	2.1952	2.2024

Table B2: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 25 cm with 0.1, 0.5 and 1.0 mm of Al measured with Discovery scanner

25 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	**SE**	Mean of Perspex pixel value	0.0364	0.0505
		C	21.9436	28.2489
		CNR	21.5114	29.099
	DE	Mean of Perspex pixel value	4.9545	4.996
		C	18.0404	22.5493
		CNR	16.5005	20.3084
Al 0.5 mm	**SE**	Mean of Perspex pixel value	0.0335	0.0347
		C	6.2942	7.8785
		CNR	6.3736	7.3293
	DE	Mean of Perspex pixel value	2.2472	2.5022
		C	1.6763	11.432
		CNR	2.6346	10.832
Al 1.0 mm	**SE**	Mean of Perspex pixel value	0.0236	0.0182
		C	2.2053	0.5538
		CNR	3.1145	0.7956
	DE	Mean of Perspex pixel value	4.1274	5.4389
		C	6.2986	10.1902
		CNR	3.3258	7.9559

Table B3: Repeatability and reproducibility (%CV) of the mean Perspex pixel value, C and CNR for phantom configuration of 40 cm with 0.1, 0.5 and 1.0 mm of Al measured with Discovery scanner

40 cm	Perspex		Repeatability	Reproducibility
Al 0.1 mm	SE	Mean of Perspex pixel value	0.1804	0.2265
		C	61.9452	74.7114
		CNR	62.1688	73.9251
	DE	Mean of Perspex pixel value	4.7267	7.8225
		C	67.8801	108.5103
		CNR	71.2194	113.8021
Al 0.5 mm	SE	Mean of Perspex pixel value	0.2159	0.1253
		C	72.6538	109.7533
		CNR	72.0311	108.8998
	DE	Mean of Perspex pixel value	6.1221	6.6927
		C	38.4167	38.5028
		CNR	38.9035	40.6274
Al 1.0 mm	SE	Mean of Perspex pixel value	0.148	0.0836
		C	12.9894	19.949
		CNR	12.08	19.7891
	DE	Mean of Perspex pixel value	8.0297	13.0142
		C	18.3559	37.104
		CNR	18.8747	32.2055