Computer-aided Landmark based Phenotyping to Identify the Genetic Basis of Skeletal Abnormalities from Mouse X-ray Images

By

Omar Munthir Al Okashi

Applied Computing Department

The University of Buckingham

United Kingdom

A THESIS

Submitted for the degree of DOCTOR OF PHILOSOPHY

in Computer Science to the School of Science in the

University of Buckingham

January 2018

Abstract

Human genetics has attracted a great deal of interest in recent years due to the increasing amount of evidence of its impact on our health and behaviour. A number of exciting research projects have been launched to investigate mammalian genes in terms of sequencing, functions and their links to different diseases or abnormalities. The Human Genome Project (HGP) launched in 1990 achieved the main goal of determining the location of each of the 20,000 genes in the genetic map by 2003. The Mouse Genome Project (MGP) afterwards had more promising objectives in discovering the functionality of all the genes and their role in human diseases. The mouse genome was chosen as a mammalian model due to the sufficient similarity between human and mouse genes in addition to the feasibility of applying effective manipulation on mouse's genes. To observe the effect of the systematic knockout of individual genes, X-ray imaging was used as a key step in the phenotyping pipeline. This process has resulted in a large volume of mouse X-ray images that need to be annotated. Arguably, the manual annotation of hundreds of thousands of such images is not only time and resource consuming but also affected by subjective decisions and human errors due to various levels of skills and tiredness.

This thesis is primarily concerned with the development of computerised tools to segment, label and measure accurately different bones and organs of the mouse's body from X-ray images generated by MGP. The annotation requires multi-level reliability measures to increase the accuracy of the final measurements. The thesis then explores the effectiveness of the automated measurements (i.e. brain size and spine curvature) in identifying a list of genes associated with specific skeletal abnormalities. The main contributions of the thesis are summarised as follows:

1- Proposing a method that combines image processing approaches with supervised learning based capabilities to estimate the size of the mouse's brain from X-ray images which leads to identify a list of candidate genes responsible for brain size abnormalities.
2- Developing a method for automatic detection and segmentation of the spine from full body mouse X-ray images followed by automatic estimation of spine curvatures using different metrics to identify a list of candidate genes responsible for spine abnormality such as Scoliosis.

3- Developing an automated method for rib detection and counting. The method relies on a sequence of heuristic approaches to locate initial rib masks. This is then followed by a two-level classification scheme to detect ribs based on their geometrical and texture features.
4- Developing a consistent image quality control scheme and embedding it in all the proposed methods in order to achieve acceptable level of reliability of the automatic annotations in a high throughput situation.

In addition to the afore-mentioned completed work, the thesis presents works in progress in automatic phenotyping on other parts of the mouse's body such as pelvis shape and extra measurements in the head of the mouse.

The thesis finally argues that the significance of the presented work is not only related to automating a manual image annotation process exercised in the MGP, but also it scales up to biomedical image processing and pattern recognition in general that may translate to practical medical applications.

Acknowledgments

All praise and thanks are due to the Almighty Allah who has helped me and given me the power to complete this thesis.

My Sponsor

Firstly, I would like to thank my sponsor in Iraq, **The Higher Committee for Education Development in Iraq (HCED)** for offering me this scholarship and for their supporting to achieve my degree. I would also like to thank the **Ministry of Higher Education and Scientific Research** in Iraq for their supporting to achieve my degree.

My Supervisors

I would really like to express all my gratitude and thankful to my supervisors **Dr. Hisham Al-Assam** and **Mr. Hongbo Du** for their unlimited support and guidance. **Dr. Assam** and **Mr. Du** generously offered their time, advices and support to me to pass and overcome all the encountered difficulties throughout my study.

I sincerely appreciate all the support and help offered by the Head of Department, **Dr. Harin Sellahewa** and all academic staff in the Applied Computing Department/University of Buckingham.

Collaborators

I would like to thank the Wellcome Trust Sanger Institute Cambridge and Mr Chris Lelliott and Mr Simon Maguire in particular for providing the test images for the research and useful discussions.

My Friends and Colleagues

I should also thank all my friends who keeping supported me during my study.

This thesis is dedicated to Soul of My Father, My great Mother, My best Brothers, My wonderful Wife and my lovely kids Yousof and Mayar.

Abbreviations

ADHE	Adaptive Histogram Equalization
ARL	Angle of Regression Line
ASM	Active Shape Model
BA	Bland Altman
BHPF	Butterworth High-pass filter
CLAHE	Contrast Limited Adaptive Histogram Equalization
CPM	Charged Particle Model
CSC	Central Spinal Curve
CT	Computed Tomography
DNN	Deep Neural Network
FHT	Fuzzy Hough Transform
GHT	Generalized Hough Transform
GLCM	Grey Level Co-occurrence Matrix
GUI	Graphical User Interface
HE	Histogram Equalization
HGP	Human Genome Project
HOG	Histogram of Oriented Gradient
HT	Hough Transform
IMPC	International Mouse Phenotyping Consortium
kNN	k-Nearest-Neighbour
LB	Lower Bound
LBP	Local Binary Pattern
LoA	Limit of Agreement
MAE	Mean Absolute Error

MGP	Mouse Genetics Project
MRI	Magnetic Resonance Imaging
OOI	Object of Interest
PCA	Principle Component Analysis
PCNN	Pulse Coupled Neural Network
RF	Random Forest
RMSE	Root Mean Square Error
ROI	Region of Interest
RR	Reference Range
SRM	Spinal Reference Model
SRS	Scoliotic Research Society
SSM	Statistical Shape Model
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UB	Upper Bound
WT	Wild Type
WTSI	Wellcome Trust Sanger Institute

Table of Contents

Chapter 1: Introduction 1

1.1 Research Background 1

1.2 Diseases and Abnormalities related to Mouse Skeleton 3

1.3 Research Aim and Objectives 5

1.4 Research Challenges 6

1.5 Contributions of the Thesis 10

1.6 Structure of the Thesis 11

Chapter 2: Biological and Computational Background 12

2.1 Biological Backgrounds 12

2.1.1 The Human Genome Project 13

2.1.2 The Mouse Genome Project 13

2.2 Computational Background 15

2.2.1 Image Enhancements 15

2.2.2 Image Segmentation 26

2.2.3 Machine Learning based Objects Detection and Classification 32

2.2.4 Texture-based Features 34

2.2.5 Summary 38

Chapter 3: Research Framework and Methodology 39

3.1 Overview of the Research Framework 39

3.2 Data set 41

3.3 Ground Truth 42

3.4 Experimental Protocols and Performance Evaluation Metrics 43

3.4.1 Identifying Genetic Markers using Reference Range (RR) 43

3.4.2 Comparing Manual versus Automatic Measurements 44

3.4.3 Statistical Measures of Classification Performance 47

3.4.4 Image Classification into either Normal or Abnormal 48

3.5 Summary 49

Chapter 4: Automatic Detection of Landmark Points to Estimate the Brain's Dimensions of the Mouse in X-ray Images 50

4.1 Problem Statement 51

4.2 Literature Review 54

4.3 Reliable Automatic Estimation for the Mouse Brains' Dimensions 57

4.3.1 Automatic Alignment 59

4.3.2 Skull Modelling 60

4.3.3 Reliability Check on Skull Model and Skull Model Refinement 61

4.3.4 Initial Detection of Width and Length Landmarks 66

4.3.5 Refining the Locations of the Brain Width Landmarks 69

4.3.6 Refining Length Landmarks and Estimating Brain Length 77

4.4 Experiment Results and Analysis 81

4.4.1 Manual vs. Automatic Measurements 81

4.4.2 Genes Identification 86

4.5 Discussion 92

4.5.1 Results Discussion 92

4.5.2 Parameter Sensitivity of the Reliability Checks 94

4.5.3 The impact of Patch Size around the landmark on the Accuracy of its Location 96

4.5.4 Effectiveness of Proposed Method when Applied on Different Data set 97

4.5.5 MATLAB Tool for Brain Size Estimation 99

4.6 Summary 100

Chapter 5: Automatic Spine Segmentation and Curvature Quantification 102

5.1 Problem Statement 103

5.2 Literature Review 105

5.3 The Proposed Method 109

5.3.1 Alignment for Full Body Dorsoventral Input Image 111

5.3.2 Spine ROI Segmentation 113
5.3.3 Spine ROI Enhancement 115
5.3.4 First Level Spine Segmentation 116
5.3.5 Reliability Check and Refinements of the Segmented Spine 119
5.3.6 Second Level Spine Segmentation 124
5.3.7 Spine Curvature Quantification 129
5.4 Experiment Results and Analysis 132
5.4.1 Manual versus Automatic Measurements 133
5.4.2 Automatic Identification of Scoliosis Cases 136
5.5 Discussions 140
5.5.1 Investigation of the Vertebrae based Solutions 141
5.5.2 Importance of Alignment of the Mouse Body 143
5.5.3 Global Landmarks Detection Using Machine Learning 144
5.5.4 Sources of Inaccurate Normal/Abnormal Spine Classification 146
5.5.5 MATLAB tool for Spine Curvature Estimation 147
5.6 Summary 148
Chapter 6: Ribs Detection and Counting from Mouse X-ray Images..... 150
6.1 Problem Statement 151
6.2 Literature Review 152
6.3 The Proposed Method 156
6.3.1 Segmentation and Enhancements for Rib ROI 157
6.3.2 Reliability Check of Rib ROI 160
6.3.3 Segmentation of Rib Objects 161
6.3.4 Enhancements and Smoothing of Rib Objects 166
6.3.5 Two-level Classification Scheme to Remove non-Rib Objects 172
6.3.6 Reliable Rib Count 176
6.4 Experiment Results and Analysis 179
6.5 Discussions 183

6.5.1 Main Reasons of Inaccuracies 183

6.5.2 Parameters Sensitivity 184

6.5.3 Using Different View to Count the Ribs 188

6.6 Summary 189

Chapter 7: Work in Progress: Other Automated Measurements of Mouse Body from X-ray Images 190

7.1 The Proposed Methods for the Additional Measurements of the Mouse Head 190

7.1.1 Brain Case Perimeter Estimation 190

7.1.2 Lambdoid Sutures Measure Estimation 191

7.1.3 Maxillary Width Estimation 192

7.2 Pelvis Segmentation 192

7.3 Initial Results and Analysis 194

Chapter 8: Conclusion and Future works 196

8.1 The Main Contributions of the Thesis 197

8.2 Impacts of the Research and its Applicability to the Human Images and other Mouse Data 199

8.3 Future Work 201

References 205

Table of Figures

Figure 1.1 The manual count of caudal vertebrae from mouse X-ray image. The normal count of caudal vertebrae should be from 27 to 30 (image is provided by Sanger Institute). 2
Figure 1.2 Normal and abnormal brain size of mouse X-ray images (images are provided by Sanger Institute). 4
Figure 1.3 Normal and abnormal spine curvature of mouse X-ray images (images are provided by Sanger Institute). 4
Figure 1.4 Normal and abnormal rib count from mouse X-ray images. 5
Figure 1.5 Low contrast of rib area in the mouse X-ray image (image is provided by Sanger Institute). Image shows low contrast of ribs 12 and 13 in the right and left sides of mouse rib area. 7
Figure 1.6 Good and poor quality of mouse skull X-ray images (images are provided by Sanger Institute). 8
Figure 1.7 Inconsistent head placement in mouse X-ray images 9
Figure 1.8 Different ROIs in full body mouse X-ray image. Spine ROI (green rectangle), rib area ROI (red rectangle) and pelvis ROI (blue rectangle). 10
Figure 2.1 Examples of different postures of mouse X-ray images (images are provided by Sanger Institute). 14
Figure 2.2 Increasing the contrast of the spine X-ray image of mouse using Gamma transformation. 16
Figure 2.3 Applying histogram equalization techniques on rib area of mouse X-ray image. . 18
Figure 2.4 Results of applying bilateral filter on rib area of mouse X-ray image. 19
Figure 2.5 Enhancing the rib ROI of mouse X-ray image using minimum order filter. 20
Figure 2.6 Highlighting the rib from rib area of mouse X-ray image using BHPF. 22
Figure 2.7 Increasing the contrast of the limbs in X-ray image of mouse body using BHPF and HE- BHPF. 23
Figure 2.8 Result of applying binary closing morphology operation to close the spine mask. 24
Figure 2.9 Combination of grey-scale morphology operations to enhance the pelvis area. 26
Figure 2.10 Region growing segmentation of limbs and skull mask from full body mouse X-ray image. 27
Figure 2.11 Applying maximum entropy and Otsu methods to segment various parts of from mouse body X-ray images. 29

Figure 2.12 Circle representation in geometric and accumulator spaces for Hough transform. 30
Figure 2.13 Circle detection using HT. The red circles represent the circles in Hough space which votes to select the required circle on the image (green circle) with an intersection point as a centre for the new circle (yellow point). 31
Figure 2.14 Schematic representation of RF (Breiman, 1999) and SVM (Cortes and Vapnik, 1995). 34
Figure 2.15 LBP code generation. Adapted from (Nixon and Aguado, 2008). 35
Figure 2.16 Comparision of the mean squared error of LBP histograms of rib vs rib and rib vs non-rib objects from mouse X-ray image. 36
Figure 2.17 Main steps of HOG feature extraction. In the first step, the image is divided into equal-size cells. After that, the gradient histogram is calculated for each cell. This is followed by histogram normalization for group of cells (block) and then collecting the final HOG features from all the generated histograms. 37
Figure 2.18 HOG features comparisons for positive versus positive and positive versus negative patch samples were sampled from mouse skull X-ray image. 38
Figure 3.1 Research framework. The framework pipeline starts by automatically align the body and the head of the mouse X-ray images to the vertical position to enable modelling of the skull and ROI segmentation for rib and spine. After that the enhancements and initial segmentation operations are applied to extract the initial mask for object of interest. Next, the reliability check is done to either correct the model/mask or reject them. After that the final refinement is applied to enable measurements estimation and then classify the images to normal and abnormal and finding candidate genes. 40
Figure 3.2 Two mouse X-ray images views selected for different measurements extraction for the thesis. 42
Figure 3.3 X-ray image of mouse head with manual annotation for width and length of brain. 43
Figure 3.4 Normal distribution of width of brain measurements estimation for WT male mice. The lower and upper bounds of the RR are shown as black lines. 44
Figure 3.5 Scatter plot for actual versus predctions measurements. The regression line of the plot is shown as blue line. 45
Figure 3.6 Bland Altman plot to estimate the agreement between manual and automatic measurements. The dashed lines represent the upper and lower limit of agreement. The solid

line represents the bias of the data (mean of differences). The figure demonstrate a good agreement between the manual and automatic measurements. 46
Figure 3.7 Confusion matrix to evaluate classification result.....47
Figure 3.8 Spine image classification to either normal or abnormal using kNN classifier and cross validation. The process starts by selecting one normal and abnormal samples from random normal and abnormal sets. The kNN then classify those samples before repeating the selection of new samples. The entire process is repeated 50 times to produce the average sensitivity and specificity.....49
Figure 4.1 Manual Annotation of mouse brain dimension landmark points (images are provided by Sanger Institute). 52
Figure 4.2 Examples of good and poor ROI of the four width and length key landmarks.....53
Figure 4.3 The main stages of the proposed method for mouse's brain dimensions estimation from mouse head X-ray image. The method starts by automatically aligning the mouse head to the vertical position. After that the skull is modelled using a circle. The reliability of the circle is assessed to correct or reject it. Next the circle position is used to estimate the four initial width and length key landmarks. Finally the initial locations of the key landmarks are refined to the final position. 58
Figure 4.4 Inconsistent initial position of mouse head X-ray images. The approximation of the angle of the head with x-axis is ranging from -35° to -55°.....59
Figure 4.5 Head alignment process.....60
Figure 4.6 Skull modelling using Hough transform. 61
Figure 4.7 Main steps of skull model reliability check and refinement process. The process starts by checking the size and the position of the circle. If the circle is high or in abnormal size, the refinement process is done to correct it. After that, the circle's position is checked and refined if it is too low.....62
Figure 4.8 Inaccurate circle position detection (too-high circle) from mouse X-ray image and its refinement. 63
Figure 4.9 Inaccurate circle position detection using HT (too-low circle) from mouse X-ray image and its refinement.....65
Figure 4.10 Inaccurate initial circle detection (big green circle) because of inaccurate estimation of the radius (red points) and inaccurate corrected circle (blue circle) because of inaccurate y-position of the red points which leads to low circle.....66

Figure 4.11 Initial W_1 and W_2 landmarks (blue points) located on the circumference of the circle using an angle of 45° for right point and angle of 135° for left point. The right and left most points of the skull mask (red points). 67
Figure 4.12 Detection of the initial position for lower length landmark (L_1). Lower point of the circle (yellow point) and the detected upper edge of the occipital bone (green point)..... 68
Figure 4.13 Detection of the initial position of the upper length landmark (L_2). 69
Figure 4.14 Different examples of initial width landmarks positions. In all cases, the initial detected landmarks (blue points) are not located on the desired position to estimate the width of the brain. 70
Figure 4.15 Coronoid process detection. 71
Figure 4.16 Width landmarks detection. Initial W_1 landmark position (blue point) located using circle model and final W_1 landmark position (green point) estimation using intensity based approach. 72
Figure 4.17 Samples of different sizes segmented coronoid processes. The accepted size for the coronoid process should range from 500 to 3750 pixels. The size of each coronoid process in the figure is as follow..... 73
Figure 4.18 Selected training pixels from W_1 ROI in training stage and W_1 landmark point location in prediction stage. 74
Figure 4.19 OOB classification error rate for RF-LBP (blue) RF-HOG (red), RF-Gabor (yellow) and RF-GLCM (purple) models for W_1 landmark. In general, the classification error rate of the RF model decreases when the number of trees increases until 100 trees where it starts to stabilize. The classification error rates for RF-LBP and RF-HOG models are smaller than other models (less than 0.005). 75
Figure 4.20 An example of prediction final W_1 landmark. 76
Figure 4.21 Final brain width estimation from mouse X-ray image..... 77
Figure 4.22 Lower length landmark (L_1) detection from enhanced ROI using intensity-based approach. Line setting (green line) by using the position of the initial L_1 landmark, migration of the points of the line (blue points) and the final L_1 position (yellow points) which located as the median of y-coordinates of the distributed points of the line. 78
Figure 4.23 Upper length landmark (L_2) detection using intensity-based approach. 79
Figure 4.24 Different windows in mouse head X-ray image in the training and prediction stages for L_1 and L_2 landmarks. 80
Figure 4.25 Final brain length estimation from mouse X-ray image..... 81

Figure 4.26 Examples of identifiable and non-identifiable coronoid processes from contrast enhanced mouse head X-ray images. 83
Figure 4.27 Correlation of manual versus hybrid automatic measurements estimation for the width and length of the mouse brain. All the measurements in mm. 84
Figure 4.28 BA analysis for manual versus automatic brain width and length estimation. (All measurements are in mm). The y-axis is the difference between the manual and automatic measurements while the x-axis is the mean between them for each point. 85
Figure 4.29 Normal distribution of mouse brain width measurement for WT population with LB and UB obtained using the RR method. 86
Figure 4.30 Normal distribution of width estimation for mutant population with RR bound. 87
Figure 4.31 Scatter plot of the brain width-length measurements for male WT and mutant mice. 89
Figure 4.32 Width estimation for "Cdk5rap2" gene male mouse instances. All the measurements are less than the lower bound of the normal range (hit gene). 90
Figure 4.33 Width estimation for 7 WT male instances. All the measurements are in the normal range (i.e. normal width). 91
Figure 4.34 Non-identifiable coronoid processes from mouse X-ray image. 93
Figure 4.35 Unreliable L1 and L2 landmarks detection using intensity based approach. 94
Figure 4.36 Accurate position detected circles imposed on a mouse X-ray image. The vertical distances for each circle are as follow. 95
Figure 4.37 Inaccurate detected circles before refinement (green) and corrected circle (blue) imposed on mouse X-ray images. The caption below shows the vertical distances for the circles before refinement (green circles). 95
Figure 4.38 Patches of Several different sizes are sampled around manually annotated landmark (red points) from W1 ROI of mouse head X-ray image. The sizes for each patch are as follows. 96
Figure 4.39 OOB classification error for three different RF-LBP models were trained by features extracted from patches of different sizes ((17 × 17), (33 × 33) and (65 × 65) pixels). The lowest error rates are associated with the RF trained with the features extracted from the patches with (33 × 33) and (65 × 65) pixels size. 97
Figure 4.40 Circles and landmarks detection for MRC and Sanger mouse head X-ray images. 98
Figure 4.41 Width landmarks detection from MRC mouse head X-ray images using RF model. 99

Figure 4.42 Snapshot of the brain size estimation MATLAB tool. The GUI enable the user to select the image using "Image Browser" button. In addition, it offers determination of the gender and whether the mouse is mutant or WT. The final results are expressed in terms of the measurements in mm and whether the measurements are in normal or abnormal range. 100
Figure 5.1 Spinal curvature measurements, adapted from (Vrtovec, Pernuš and Likar, 2009) 104
Figure 5.2 Comparison between spine X-ray images of mouse and human where the red boxes zoomed into the upper parts of the chest area. 105
Figure 5.3 Main stages of spine segmentation and curvature quantification from mouse X-ray image method. The method starts by aligning the image vertically and then cropping the spine ROI. This is followed by enhancement and segmentation of spine. At this stage, the reliability check is done to either accept, refine or reject the segmented spine. Finally, the curvature quantification of the spine are estimated as a six angle and index based metrics. 110
Figure 5.4 Full body mouse X-ray image alignment process. 112
Figure 5.5 Global landmarks annotation on the aligned mouse body in X-ray image using mask based approach. 114
Figure 5.6 Spine ROI segmentation using global landmark points. 115
Figure 5.7 Enhancing the spine ROI cropped from mouse X-ray image 116
Figure 5.8 Sensitivity of length of structure element to close the spine mask 117
Figure 5.9 Initial segmentation of spine from mouse X-ray image. 117
Figure 5.10 Some geometrical properties of unwanted objects near the spine mask. The red line represents the length of the unwanted object while the yellow line represents the distance between the unwanted object and the nearest border of the image as follow. 118
Figure 5.11 Unreliable spine masks with irregular thickness problem. 120
Figure 5.12 Process of discovering and refining of irregular thickness of mouse spine mask. The process starts by measuring the distances between every pair of opposite points on the right and left borders. If any irregular distance is discovered, a relocation of the out-line point is done. After that the reliability of the refined borders is done to decide whether it is accepted or not. 121
Figure 5.13 Borders of the spine before and after refinement process. 122
Figure 5.14 Examples of correcting irregular thickness of spine mask. The upper image in the column represents the spine borders before refinement and the lower one is the result of refinements. The right and left maximum differences for the borders after refinements are as follows 123

Figure 5.15 Intensity based refinements of the spine borders. 125
Figure 5.16 Texture based refining of spine borders. 126
Figure 5.17 Classification error rate for RF-LBP (blue line) and RF-HOG (red line) models. The error rate is calculated across different number of trees in the RF models. 127
Figure 5.18 BA analysis for intensity and texture based landmarks detection versus manually annotated landmarks. 128
Figure 5.19 Locating CSC, SRMs and Landmark Points. 130
Figure 5.20 Angle based and distance-based index measurements. 132
Figure 5.21 Scatter plot for average of manual measurements of A_1 versus automatic measurement obtained by using automatically obtained measurements $Angle_1$ and $Angle_2$ 135
Figure 5.22 Scatter plot for BA analysis for manual measure versus $Angle_1$ and $Angle_2$..... 136
Figure 5.23 Confusion matrix for classification results of $Angle_2$ measurements of mouse spine. The matrix shows the calculation for the sensitivity, specificity and the total accuracy..... 137
Figure 5.24 Examples of normal and abnormal spine curvature of mouse X-ray images..... 140
Figure 5.25 Top and bottom vertebrae detection using RF based approach. Several patches automatically set near the upper and lower vertebrae based on U2 and L2 (red points) to predict the centre of the vertebrae (yellow point). 141
Figure 5.26 Lumbar vertebrae detection using intensity profile approach. 142
Figure 5.27 Importance of alignment of the full body mouse from X-ray image. 143
Figure 5.28 Global landmarks annotation using RF approach. 144
Figure 5.29 Applying machine learning and mask based approaches on image from MRC Harwell (image is provided by MRC Harwell)..... 146
Figure 5.30 Misclassification of mouse spine from X-ray image by automatic proposed method. 147
Figure 5.31 Screenshot of spine curvature estimation MATLAB tool. Image browser button can be used to browse the image file. To estimate the metrics, "Estimate Metrics" button is used. After that, the "Classify Image" button shows whether the spine is normal or abnormal. 148
Figure 6.1 Manual counting and labelling of ribs in mouse X-ray image. The image shows the overlapping rib shapes that results from projecting 3D rib object to a 2D image space. 151
Figure 6.2 Main steps of rib detection and counting method. Method starts by cropping and enhancing of the ribs ROI from the full body mouse X-ray image. This is followed by a reliability check for the ribs ROI. After that, segmentation and post processing operations are

done to obtain the rib masks. Finally, the two-level classification scheme is applied to remove irrelevant objects and then counting the ribs. 156
Figure 6.3 Segmentation of rib ROI from full body mouse X-ray image using 3 different horizontal shifting values. 157
Figure 6.4 Estimation of the accurate upper cut point for rib ROI. 158
Figure 6.5 Different length for estimation of accurate upper cut point for rib ROI. The red, green and yellow lines show the lower edge of the upper part of rib area segmented with length 80, 100 and 120 pixels respectively. All the three distances lead to include the upper first rib in the cropped part 159
Figure 6.6 Different samples of the mask of the upper part of rib area and the x-coordinate difference values that enable detection of the first rib location. All the difference values for the right and left ribs are greater than 10 pixels as follows. 159
Figure 6.7 Rib ROI enhancements. 160
Figure 6.8 Effect of spine mask on the rib area reliability 161
Figure 6.9 Thresholding the enhanced rib area of figure 6.7(c) with different threshold values. Relative small values lead to segmenting most of the objects and tissues and so no clear separation between rib and non-rib objects. Larger values lead to diminishing the rib objects completely. The used threshold values as follows 162
Figure 6.10 Isolating of rib masks from mouse rib area. 163
Figure 6.11 Exclusion of spine object from the rib area using spine mask. 164
Figure 6.12 Removing the edge of the ribcage. 165
Figure 6.13 Smoothing rib objects in mouse rib area. 167
Figure 6.14 Divided rib mask merging using adaptive closing operation. 167
Figure 6.15 Several examples of minor axis length of ellipse around single and connected rib masks 169
Figure 6.16 Disconnecting jointed ribs mask. 170
Figure 6.17 Disconnecting upper jointed ribs mask. 171
Figure 6.18 Enhancing rib mask segmentation. 172
Figure 6.19 Enhancements and smoothing of rib area 172
Figure 6.20 Uniform LBP features extraction from rib object. 174
Figure 6.21 Results of applying the two-level classifier scheme to remove non-rib objects from rib area. 176
Figure 6.22 Several examples of detected and labelled rib mask superimposed on original rib area images 176

Figure 6.23 Rib count from connected ribs mask. 177
Figure 6.24 Different examples of unreliable rib count. 179
Figure 6.25 Confusion matrix for the reliable rib count category. 180
Figure 6.26: Example of uncertainty in the ground truth. 181
Figure 6.27: Confusion matrix for the unsure rib count category. 182
Figure 6.28 Using images from unsure rib count in manual rib count. 183
Figure 6.29 Examples of inaccurate rib count because of inaccurate upper cut. 183
Figure 6.30 Results of applying BHPF with a various cut off frequency values on the same rib area image 185
Figure 6.31 Accuracy figures of rib count resulting from using different cut off frequency values (25, 35, 40, 45, 50, 55, and 60) for BHPF. The Figures shows that the best accuracy obtained when the value of cut off frequency for BHPF is 50. 186
Figure 6.32 Results of applying BHPF with the selected cut off frequency value (50) on four different mouse rib area images. Each column repressents the mouse rib area image before and after applying BHPF. 186
Figure 6.33 Results of applying CLAHE enhancements on same rib area image using five various values for number of tiles. All the result images show better contrast than the original image. 187
Figure 6.34 Accuracy figures of rib count resulting using different number of tile values for CLAHE ([7 × 7], [8 × 7], [8 × 8], [8 × 9], [9 × 9], [10 × 10]). The Figures shows that the best accuracy obtained when the number of tiles for CLAHE is [8 × 8]. 188
Figure 6.35 Assessing mouse X-ray image of full body lateral view mouse X-ray image for rib count. 188
Figure 7.1: Main stages of the automatic brain case perimeter estimation. The process starts by trimming the upper and lower area of the skull mask. After that the edge of the prepared mask is obtained and superimposed on the input skull image (yellow border). 191
Figure 7.2 Lambdoid suture length measure. 191
Figure 7.3 Automatic maxillary width estimation. The green and red points represent the edge of the nasal bone. The maxillary bone landmark is determined as the first sudden change of the x-coordinates of the nasal bone points. The maxillary width is the distance between the two maxillary bone landmarks (blue line). 192
Figure 7.4 Pelvis segmentation process from mouse X-ray image. The process starts by preparing pelvis mask by applying grey scale morphology enhancements and Gabor filter on

the pelvis ROI image. The pelvis mask is then used to mask the pelvis object from an enhanced pelvis image. 193
Figure 7.5 Measure the distance between the ischium of the pelvis girdles. 194
Figure 7.6 Accuracy of extra measurements in head area. 195
Figure 7.7 Different examples of pelvis segmentation from mouse X-ray image. 195
Figure 8.1 The automatic measurements and annotations obtained by our proposed solutions superimposed on a full body mouse X-ray image. Those annotations include different measurements and landmarks annotation in the head of the mouse. Also, the spine and pelvis segmentation. Finally, it includes the rib detection and count. 198
Figure 8.2 Human spine curvature estimation from X-ray image. 200
Figure 8.3 Initial rib segmentation from human X-ray image. 201
Figure 8.4 Samples of annotated head X-ray images with different measurements (images are provided by MRC Harwell). 201

Table of Tables

Table 4.1 Results of all landmark points in terms of mean and standard deviation of the Euclidean distances between the automated landmark points and the ground truth..... 82
Table 4.2 Manual versus automatic width and length measurements evaluation. Table shows the correlation between the manual and automatic measurements R^2. It also shows the angle of regression line (ARL), mean of absolute error (MAE) and the standard deviation of the absolute error (σ). 84
Table 4.3 BA analysis for the width and length of mouse brain estimation. (All measurements are in mm). 85
Table 4.4 Possible gene hits which defined using automatically estimated width and length measurements of mouse brain. The Hit ratio determine whether the gene is a hit (hit ratio >= 60%) or not (hit ratio < 60%)..... 88
Table 5.1: Intra-Observer and Inter-Observer variations in the manual measurements of spine curvature Cobb angles using R^2 and ARL 133
Table 5.2: Intra-Observer and Inter-Observer variations in the manual measurements of spine curvature Cobb angles using BA analysis. 133
Table 5.3 Automatic versus manual angle-based measurements of mouse spine curvature using R^2 and ARL. 134
Table 5.4 BA analysis for automatic versus manual angle-based measurements of mouse spine curvature. 134
Table 5.5 Classification results using automatic spine curvature measures of mouse spine X-ray image..... 137
Table 5.6 Classification results using majority voting decision-level fusion among all the six measurements..... 138
Table 5.7 Correlation between $Angle_1$ and Index-based measures. The results show high correlation for whole data set and mutant data set and low correlation for the WT data set. 138
Table 6.1 Different categories of rib count (reliable count, unsure count and rejected images). 179
Table 6.2 Manual and automatic phenotyping of 8 mutant lines for mouse rib count. 180
Table 6.3 Accuracy of rib detection. The proposed method succeeded to detect most of the ribs in the tested images. 182
Table 7.1 Percentage of accuracy for proposed methods to estimate different measurements in the head area and the pelvis segmentation. 194

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas, work or words have been included, I have adequately cited and referenced the original sources.

Omar M. Al Okashi

Chapter 1: Introduction

1.1 Research Background

The Human Genome Project (HGP) was initiated in 1990 with the aim of providing complete and accurate sequence of all human genes. By 2003, the National Human Genome Research Institute announced the completion of HGP successfully (Collins, Morgan and Patrinos, 2003). Despite the success in identification of over 20000 protein-coding genes, the functionalities of many genes are still unknown and thus the genetic basis of many diseases remains obscure (White et al., 2013).

Over many years of research, the mouse has been considered and accepted as a key model to assess mammalian gene functions (Simon et al., 2013). Different genetic techniques like "knock-out"- replacement of an inactive gene for the normal one- was used to produce custom-made mouse models for a wide array of specific diseases(Griffiths et al., 2005). Developed mouse models can help the scientists to understand many types of human disease such as Down syndrome (Reeves et al., 1995), cancer (Roby et al., 2000), epilepsy (Löscher, 2011), etc.

In 2006, the Wellcome Trust Sanger Institute (WTSI) initiated the Mouse Genome Project (MGP) that would provide a step toward the aim of knocking out all genes and screening each mutant line for a broad range of traits (White et al., 2013). These different gene knockouts lead to changes in the expression of a particular trait (i.e. phenotypes) including skeleton of the mouse body. The Sanger Institute monitors and assesses alterations in mouse model using X-ray images. To find any specific phenotype resulted by a single gene knock-out, several mouse models (7 males and 7 females) are considered. The used mouse model belongs to the C57BL/6N strain, which is one of the most widely used strains due to their easy breeding (Simon et al., 2013). For each mouse model, different organs, bones and limbs need to be examined against any resulting changes which leads to five different mouse postures (see Section 2.1.2 for more details). Considering the number of phenotyping pipelines, the number of different postures per mouse model, number of mouse model for each pipeline, a huge amount of X-ray images has been produced (20,000 mutant lines (one line for each gene) × 14 mouse models × 5 images). All these biomedical images are currently examined and annotated manually. The manual examination/annotation, counting and measuring various parts of skeletal structures to identify the underlying genetic basis of any abnormality in the phenotype is associated with specific diseases and abnormalities, for instances scoliosis and change in

brain size. This is typically done by comparing the resulting information extracted from the images of mutant mice against the ground truth information extracted from the annotated Wild Type (WT) mouse images - the normal, non-mutated version of a gene common in nature (Griffiths et al., 2005) – to establish if there are any possible hits - big changes from control baseline (White et al., 2013). Figure 1.1 shows an example of manual annotation that annotates and counts the caudal vertebrae from mouse X-ray image in Sanger Institute (Sanger, 2013). The MGP has around 43 different count/annotations in various parts of the mouse body to assess any changes in the skeletal of the mouse. The counting is required to check if different bones or limbs are in normal count or not, for instance ribs, vertebrae and digits. The annotation involves investigation and examination whether specific organs, limbs or bones (i.e. spine, vertebrae, pelvis, skull, ulna and teeth) has normal shape and measurements or not.

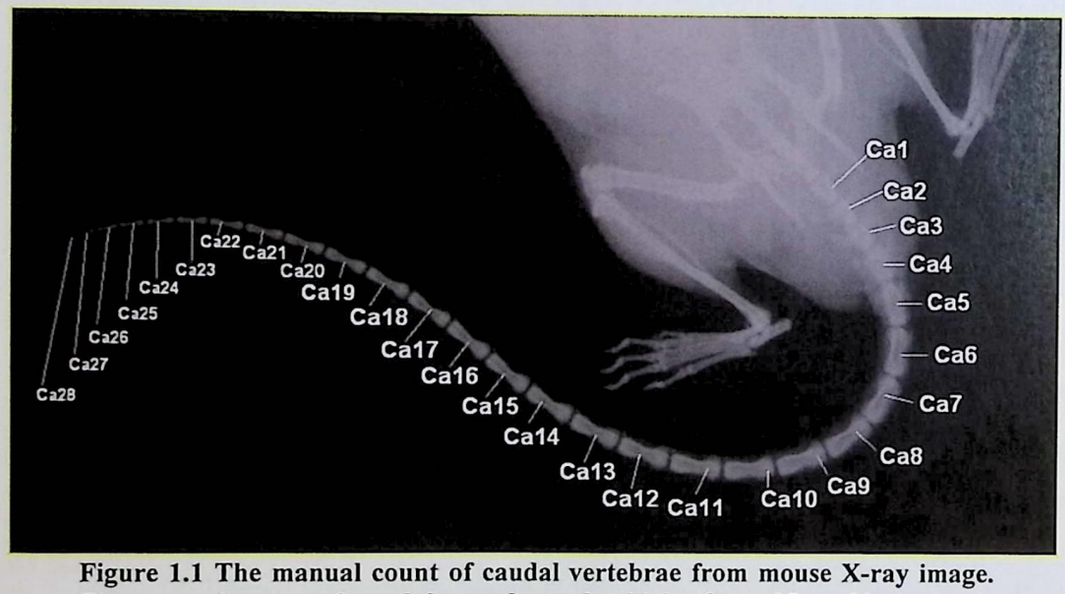


Figure 1.1 The manual count of caudal vertebrae from mouse X-ray image. The normal count of caudal vertebrae should be from 27 to 30 (image is provided by Sanger Institute).

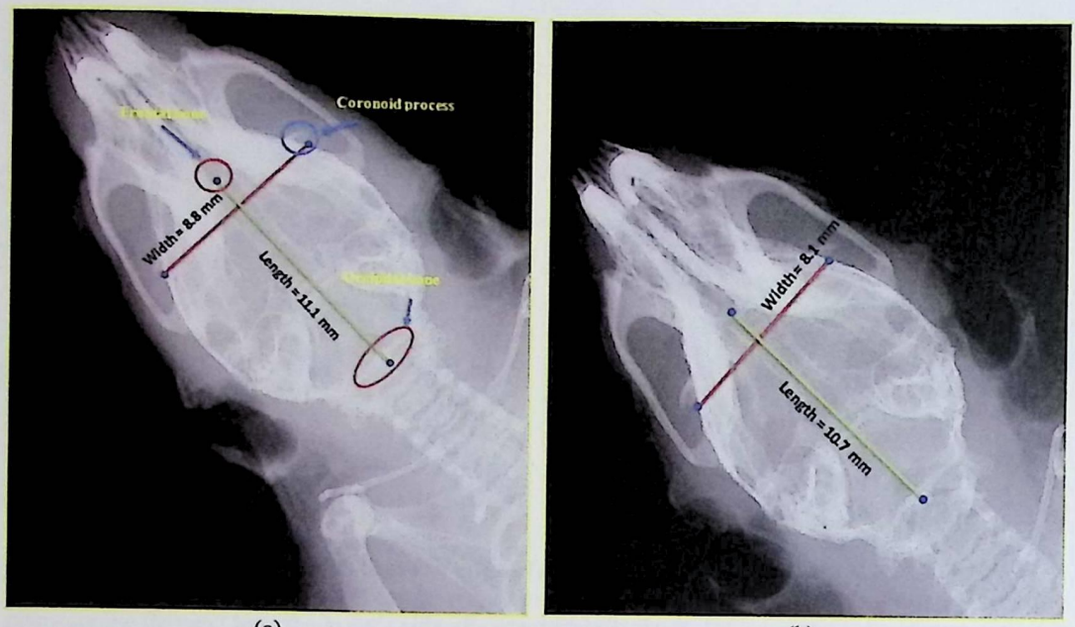
The skeleton of the mouse may involve significant changes that may relate to diseases and abnormalities, for example, abnormal spine curvature (scoliosis) and change in a skull size (Down syndrome). One element of the phenotyping pipeline is a thorough evaluation of the skeleton through systematic X-ray imaging of individual animals and manual observations of variation. The manual examination, annotations and results analysis of high-throughput X-ray images have several drawbacks. Despite the experience of manual annotators, dealing with a huge number of images could reduce the accuracy of annotation significantly due to tiredness and loss of concentration. Even in the best of its time, subjective annotation is most likely to cause *intra-observer* and *inter-observer* variation or inconsistency. In addition to the problem of accuracy, the time taken to do such manual annotation is still considerable, limiting the number of images that can be examined. An automated approach for annotating and triaging

X-ray images will reduce processing time and costs on human resources and eliminate possibilities of human errors and inconsistency in manual measurement.

1.2 Diseases and Abnormalities related to Mouse Skeleton

"A gene is a section of a threadlike double-helical molecule called deoxyribonucleic acid (DNA)" (Griffiths et al., 2005). Since the information of body functions (i.e. protein production) is encoded in a gene, any small change in the DNA of any gene could affect the health (Griffiths et al., 2005). The skeleton of a mouse could be affected by genes mutation (Grüneberg, 1955). In this thesis, our focus is on three different abnormalities in distinct parts of the mouse body as follows.

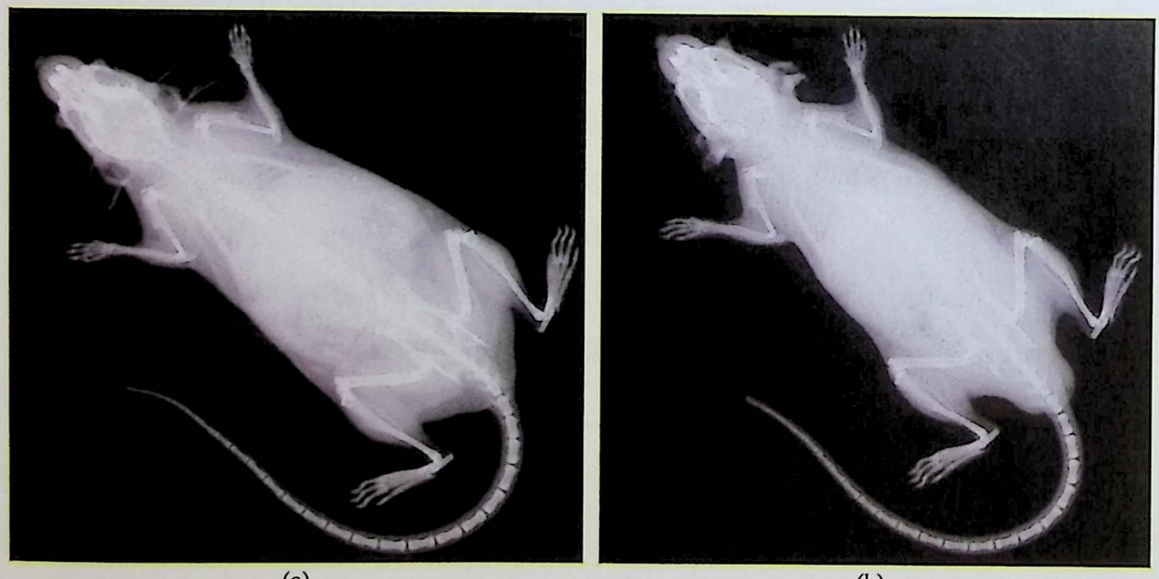
1. Skull and brain size. One of the known changes in the mouse skeleton is the change in the skull size (Johnson, 1976). The skull size is considered as an indication to change in the brain size. Because of the difficulty associated with direct measuring of brain dimensions from mouse X-ray images, specific landmarks on the skull are used for brain size estimation. The width of the brain mouse is approximated as the distance between the right and left coronoid processes of the mandible bone (blue circle in Figure 1.2 (a)) as shown in red lines in Figure 1.2. For the mouse brain length estimation, the distance between the upper edge of the frontal bone (red circle in Figure 1.2(a)) and the lower edge of the occipital bone (red ellipse in Figure 1.2(a)) is annotated as shown by green line in Figure 1.2. Two examples of normal (a) and abnormal brain size (b) are illustrated in Figure 1.2 (data is provided by Sanger Institute).
2. Scoliosis. The unusual spine curve related to vertebrae rotation is known as scoliosis (Anderson, 2007). Different studies used the animal models to study the developing of scoliosis due to genetics and other reasons (Ouellet and Odent, 2013). Figure 1.3 shows two mouse models, one with normal spine curvature (a) and the other with abnormal spine curvature (b) (data is provided by Sanger Institute).
3. Rib count. Some variations in rib count were reported due to scoliosis cases (Spencer, Gold and Hresko, 2014) . Mouse rib count is considered as one of the important symptoms that could be related to genetic backgrounds. Figure 1.4 shows normal rib count (13, 13) (a) and abnormal rib count (14, 14) (b) of mouse X-ray image (Sanger, 2013).



(a) (b)

Figure 1.2 Normal and abnormal brain size of mouse X-ray images (images are provided by Sanger Institute).

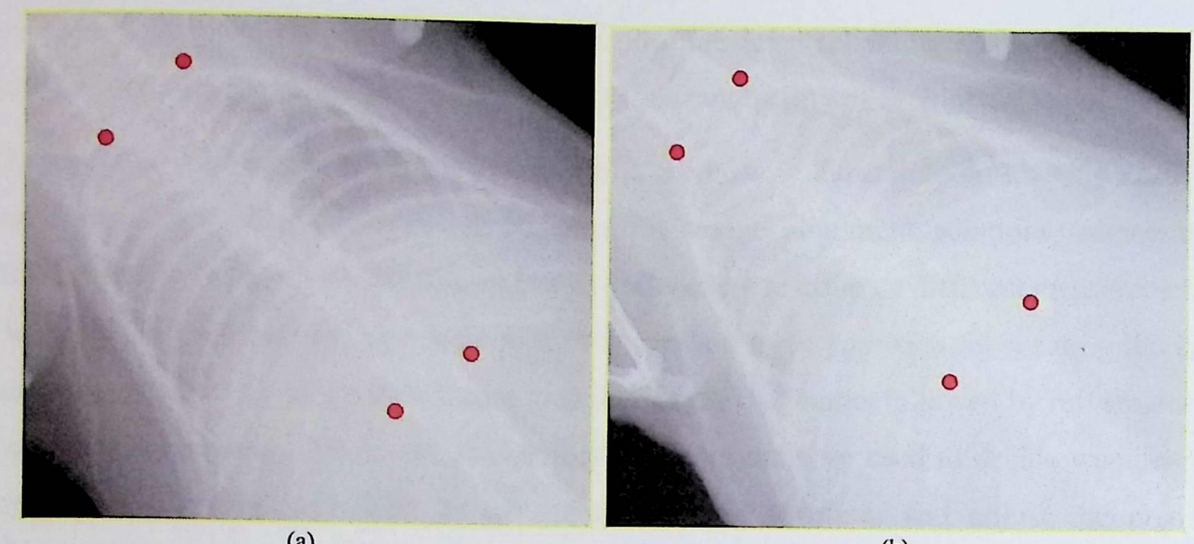
(a) Normal width and length of the mouse brain. The width of the brain is taken as the distance between the right coronoid process (blue circle) and the left coronoid process. The length of the brain is taken as the distance between the frontal bone (red circle) and the lower edge of the occipital bone (red ellipse); (b) Abnormal width and length of the mouse brain. The width and length of brain are estimated as explained in (a)



(a) (b)

Figure 1.3 Normal and abnormal spine curvature of mouse X-ray images (images are provided by Sanger Institute).

(a) Normal spine curvature. The spine looks straight with no noticeable curvature; (b) Abnormal spine curvature. The spine curves to the right side and such curvature relates to Scoliosis disease



(a) (b)

Figure 1.4 Normal and abnormal rib count from mouse X-ray images.
(a) Normal rib count from mouse X-ray image. A number of ribs on each side is 13 ribs where the red dots represent the first and last ribs in each side; (b) Abnormal rib count. The number of ribs on each side is 14 ribs where the red dots represent the first and last ribs on each side

Some examples of genes that are known to affect human health conditions relevant to the work of this thesis are:

1. Cdk5rap2. In human, Cdk5rap2 gene encodes the Cdk5 protein (regulatory subunit-associated protein 2) (Ching, Qi and Wang, 2000). A confirmed relationship between variation in Cdk5rap2 gene and change in brain structure was reported (Rimol et al., 2010).
2. Aspm. In human, Aspm gene encodes Asp protein (Abnormal Spindle Protein homolog) (Pattison, 2000). Fault in Aspm gene was linked with autosomal recessive primary microcephaly (Kaindl et al., 2010).
3. Arl4d. In human, Arl4d encodes Adp-Ribosylation factor-like protein.
4. Mcph1. Mcph1 gene mutation could lead to reduction in a brain size (Jackson et al., 2002).

1.3 Research Aim and Objectives

In this thesis, the analysis and processing of the X-ray images of the mouse's body is of the main concern. The main goal of this research is to develop fully automated solutions to annotate X-ray images of mouse's skeleton, which overcome the problems with manual annotations.

However, before we can proceed in any of the main objectives in this research, we need to deal with the low contrast issues of the X-ray images along with other challenges highlighted in the next section. Human annotators tend to have all the domain knowledge accumulated from extensive training whereas the machine does not possess that knowledge. Therefore, this thesis

argues that any credible solution should rely on multi-level reliability measures that act as quality control checks to automatically assess the output quality of each of the key stages.

The proposed annotation measures in this thesis are heavily dependent on locating accurate landmarks. Those landmarks could be used for image alignment solutions, segmenting different Region of Interests (ROIs), and as key references to estimate different measurements. The solutions proposed by the thesis follow a common main approach for locating the best landmarks. It starts by estimating initial locations of the landmarks followed by refinement of the candidate locations. Multi-level reliability measures are then used to decide whether the candidate landmarks will lead to the final optimised location, and adjust the process accordingly.

This research involves the following four main objectives.

- Objective 1. Automatic identification of specific landmark points in the mouse's brain to estimate its size in terms of width and length dimensions. Accurate mouse brain estimation results in identifying abnormalities in brain sizes.
- Objective 2. Automatic detection and segmentation of the spine of the mouse and quantifying its curvature using different angle based and index based metrics. Spine curvature quantification leads to classifying mouse spines into normal and abnormal according to their curvature measurements.
- Objective 3. Automatic detection and counting of ribs of the mouse to highlight abnormal rib count.
- Objective 4. Identifying a list of candidate genes responsible for skeletal abnormalities based on the measurements extracted from the above three objectives. For a gene mutant line (i.e. 7 mice), if the measurements of a specific organ is abnormal such as smaller or bigger brain size in the most instances of this line, it is considered as a hit gene i.e. maybe responsible for the abnormality.

In addition to the above four main objectives, the research includes other objectives such as pelvis segmentation and additional landmarks detection in the head area. All the mentioned objectives will require certain types of pre-processing and enhancements to prepare the image for annotations.

1.4 Research Challenges

There are several technical challenges in developing fully automatic mouse annotation solutions from X-ray images. Those challenges can be summarized as follows:

1- Image quality. Although the X-ray technology is considered effective and non-expensive for high throughput, it has several problems in terms of the image quality. There are several factors that significantly affect the overall quality of X-ray images namely contrast, scatter radiation and noise.

- The contrast of an object in biomedical image is related to its intensity difference from its background and surrounding tissues (Gonzalez and Woods, 2017). In certain parts of an X-ray image, the contrast may be very low which causes extra difficulties for object segmentation and landmark detection. Figure 1.5(a) for example shows the low contrasts surrounding the 12th and 13th ribs in right and left sides of the mouse body.

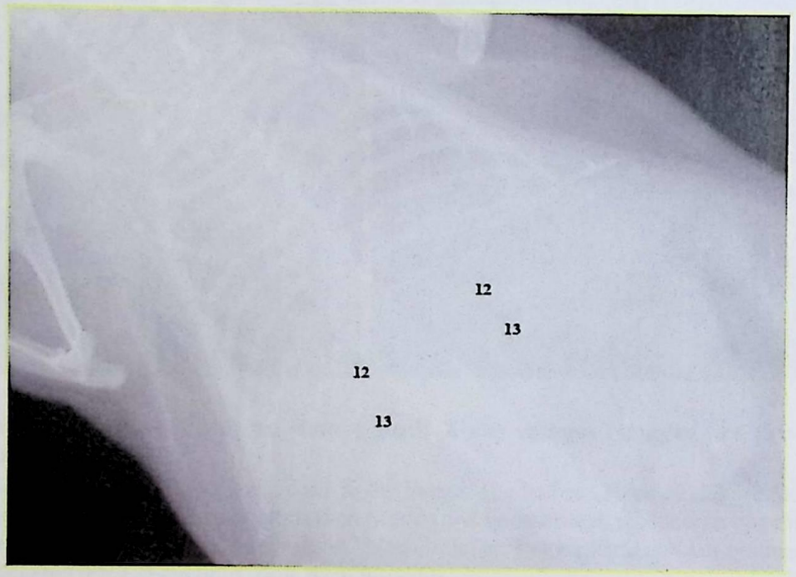
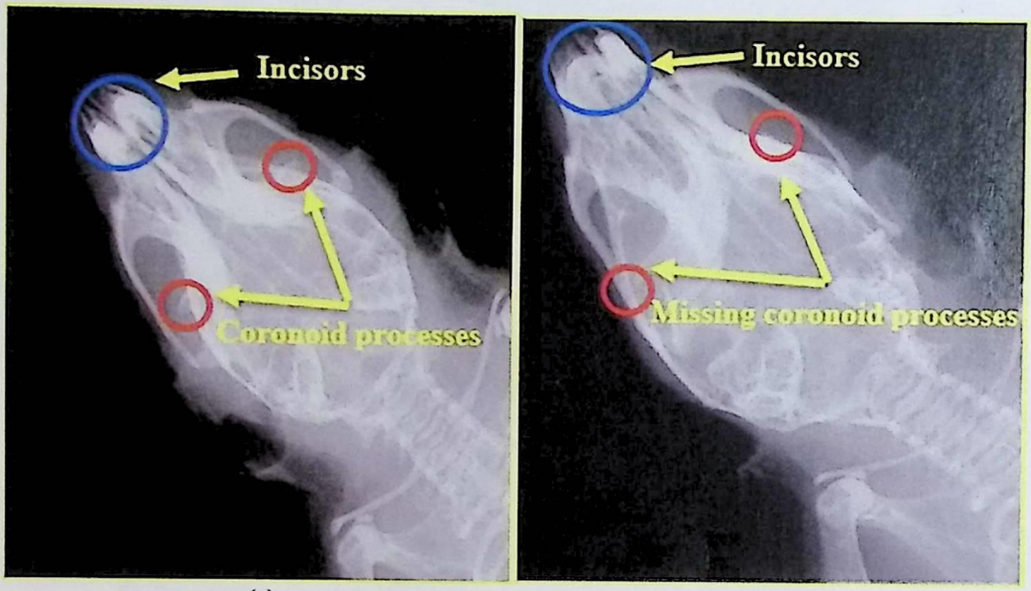


Figure 1.5 Low contrast of rib area in the mouse X-ray image (image is provided by Sanger Institute). Image shows low contrast of ribs 12 and 13 in the right and left sides of mouse rib area.

- Scatter radiation for X-ray usually happens when imaging thick object. This makes the photon beams to either attenuate or penetrate the body without any interactions (Birkfellner, Figl and Hummel, 2014) which might lead to reduce the contrast of the X-ray image.
- Noises are another well-known challenge in biomedical images. The stochastic signals which are not related to image contents are referred as noise are mainly caused by lack of X-ray dose (Birkfellner, Figl and Hummel, 2014). Increasing the number of photons for X-ray images can reduce the noise but may cause damage to tissue. In general, image noise reduces the visibility of the object especially those objects with low contrast. For our thesis, the word 'noise' or 'background noise' will be used to describe any random signal that affect the

recognition of object of interests (i.e. rib) in grey-level image. For the binary image, we will use the word 'unwanted object' to label any out-of-interest binary objects.

Some X-ray images are so poor in quality that any automatic annotations would be deemed unreliable. Figure 1.6 shows a good (a) and poor (b) quality of mouse head X-ray images.



(a) (b)

Figure 1.6 Good and poor quality of mouse skull X-ray images (images are provided by Sanger Institute).
(a)Correct position and good-quality mouse skull X-ray image. The incisors (blue circle) are identical and symmetrical. The coronoid processes (red circles) are visible and symmetrical; (b) Incorrect positioning and poor-quality mouse skull X-ray image. The incisors (blue circle) are not identical and not symmetrical. The coronoid processes (red circles) are not visible from the skull

For those images, it makes more sense to automatically exclude them for a possible manual intervention by an expert than forcing a sub-optimal unreliable annotation upon them. Therefore, it is desirable to develop annotation solutions with the capability to determine if the given annotation is unreliable.

2- Inconsistent body placement. In the data sets we are dealing with images show mouse bodies and mouse heads placed in different angles. To make a fully automatic segmentation and analysis easier, the mouse's body and head need to be aligned first. Although this added step is a non-mandatory step, it makes exploiting the symmetry of the left and right parts of the body much easier. Figures 1.7(a) and (b) show the initial position of mouse head in two X-ray images with different angles. Figures 1.7(c) and (d) show the results of alignment of (a) and (b) respectively. The angle of the head in the aligned images are very close if not identical.

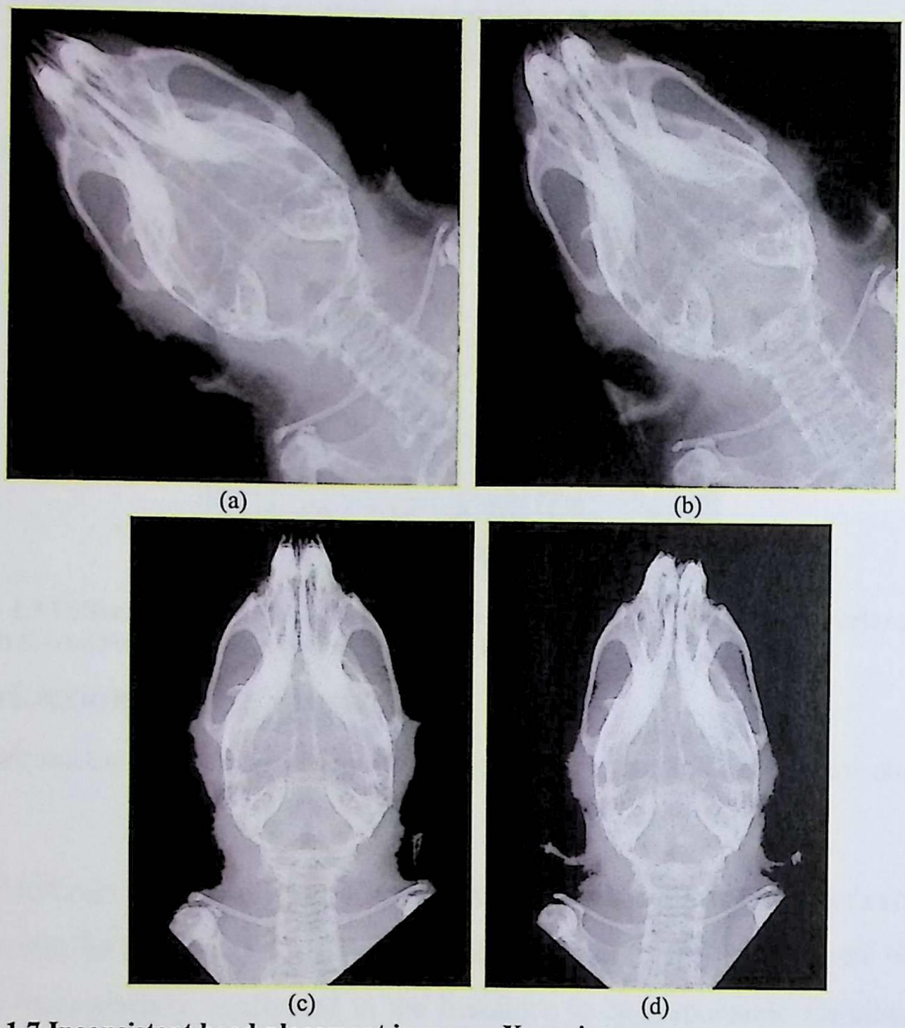


Figure 1.7 Inconsistent head placement in mouse X-ray images.
(a) and (b) show the different initial position of two mouse head X-ray images before alignment; (c) and (d) show the results of alignment for the (a) and (b) respectively to the vertical position

3- Initial segmentation of different ROIs. This research deals with different objects and bones of the mouse's body. For example, Figure 1.8 shows three different ROIs in three different areas of the mouse body in the same view. Accurate annotations of the objects of interest rely on accurate and fully automatic initial segmentation of ROI.

4- Ground truth. To evaluate our automatic methods, a robust ground truth that reports accurate measurements and associates diseases to images of the mouse is needed. Like many biomedical image applications, the ground truth that we obtained from Sanger Institute suffers from inaccuracy. Some inaccurate manual annotations exist because of the challenges in annotation of certain parts even for expert annotators (see Section 6.4 for more details).

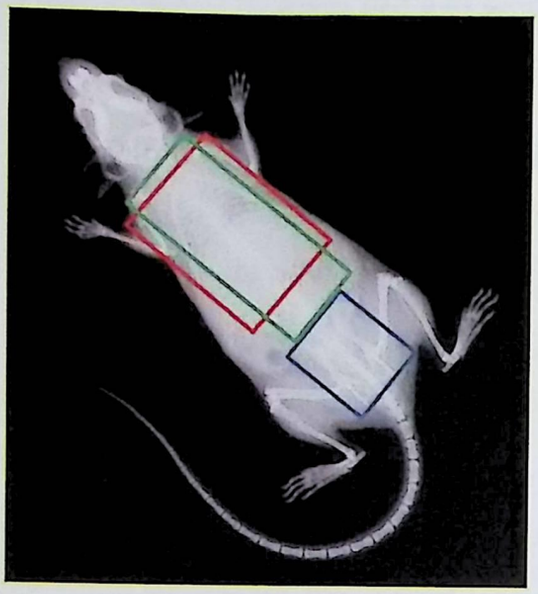


Figure 1.8 Different ROIs in full body mouse X-ray image. Spine ROI (green rectangle), rib area ROI (red rectangle) and pelvis ROI (blue rectangle).

1.5 Contributions of the Thesis

Several contributions have been achieved through this research which are summarised as follows:

1. Reliable brain size estimation of mice which leads to defining a list of candidate genes that might be responsible for abnormal dimensions of the brain. Some of the flagged genes were already confirmed in the literature to be responsible for changes in brain size.
2. Automatic detection and segmentation of a spine from full body mouse X-ray image. This accurate segmentation helps in proposing different sets of angle and index based measurements to quantify spine curvature.
3. Using spine curvature quantification to perform automatic classification of spines to normal and abnormal (Scoliotic) using cross-validation.
4. Developing rib detection and counting method to perform automatic counting of ribs. Geometrical and texture features of the rib and non-rib objects are used in a two-level classification scheme to detect ribs.
5. Developing simple yet effective methods to automatically align the mouse body and mouse's head positions in X-ray images. This alignment will make developing effective initial segmentation methods easier, which helps to crop the required ROIs in the mouse body accurately.
6. Automatic detection and segmentation of the pelvis from full body mouse X-ray image.

The contributions made on point 5 have been embedded into solutions for points 1–4 and 6. Contribution points 1, 2, 3, 4 and 6 have led to the following publications:

1- Al Okashi, O., Du, H., Selway, J., Lelliott, C., Maguire, S., Melvin, D., Langlands, K., Al-Assam, H., 2014. Automatic Ribs Segmentation and Counting From Mouse X-ray Images, in: Medical Image Understanding and Analysis (MIUA). pp. 143–148.

2- Al Okashi, O., Du, H., Al-Assam, H., 2017. Automatic spine curvature estimation from X-ray images of a mouse model. Computer Methods and Programs in Biomedicine. 140, 175–184.

3- Al Okashi, O., Al-Assam, H., Du, H., 2017. Automatic pelvis segmentation from x-ray images of a mouse model, in: Mobile Multimedia/Image Processing, Security, and Applications 2017. (Vol. 10221, p. 1022108).

1.6 Structure of the Thesis

The rest of the thesis is organised as follows. Chapter 2 presents a background to the genome projects in both human and mouse. It also gives a technical background about different image processing techniques related to this research. The main framework and methodology of our research is outlined in Chapter 3. The chapter also explains the data sets and available ground truth used in the study. In addition, the experimental protocols adopted for the research are also presented. The chapter serves as a roadmap for the research works presented in the following key chapters.

Chapter 4 describes the proposed automatic methods to detect different landmarks in the head of the mouse and then using them to estimate the brain dimensions. This chapter also explain the final outcome of the proposed methods. Chapter 5 presents an automatic method to segment a mouse spine. The segmented spine will be used to quantify the spine curvature using a set of proposed metrics. The proposed automatic method to enhance rib area and then segment the ribs is described in Chapter 6. The segmented ribs and other unwanted objects will then be checked against different properties to keep only the ribs and discard any other unwanted objects before a final counting. In Chapter 7, a collection of miscellaneous operations which are still work in progress are presented. Finally, Chapter 8 concludes the research and focus on the main contributions in the thesis. In addition, it discusses the impacts of the research and future works relating to the research.

Chapter 2: Biological and Computational Background

This chapter is devoted to the biological and technical backgrounds of the research. These fundamental knowledge must be given in order to follow the rest of the thesis. This chapter is *not* about providing a comprehensive review of the state-of-art literature on the image processing techniques for problems relevant to this research. That task is designated to each key chapter in this thesis before our solution to a specific problem is proposed. This chapter could be skipped by readers of the relevant biological or technical backgrounds.

The biological background part of the chapter aims to provide a basic understanding about the biological relevance of image annotations to be achieved through this research. This part mainly describes the human and the mouse genome projects and highlights their main findings and importance. This will lead to the justification for annotating a huge volume of X-ray images of mice, and hence the need for specially developed automated solutions for the annotation and the desire of discovering any strong links between the automatic annotation results and classification of abnormality of skeleton structures as results of gene knock-outs.

Biomedical images normally require pre-processing and enhancements so that useful features can be extracted from the image. Depending on the type of problems, segmentation or detection of certain part of the image is often essential. The technical background part of this chapter is therefore mainly concerned with image processing techniques for enhancing and segmenting images, and machine learning techniques relevant for recognizing a specific objects or landmarks from X-ray images using different features. A fundamental understanding about these techniques is therefore also essential for this research.

2.1 Biological Backgrounds

The human genome is a set of DNAs for an organism, which includes its genes. The genome involves the information required to construct and maintain the organism. Each cell in the human that has a nucleus keeps a copy of the entire genome (Griffiths et al., 2005). Genes are important to our health. It has been confirmed that some rare diseases happen because a single gene stops working (Jimenez-Sanchez, Childs and Valle, 2001). Due to such a close relation between genes and the diseases, for instance down syndrome, scientists hope to use genetic information to help in diagnosing, curing and may be preventing such diseases. Two genome projects were initiated to define the gene sequences, functionalities and their relations to

diseases and abnormalities. The first one is HGP (Green, 2015) and the second is MGP (Sanger, 2013).

2.1.1 The Human Genome Project

The HGP is an international research project to determine the sequence of the human genome and identify the genes that it contains. HGP started in 1990 and was completed in 2003. One of the ambitious goals of HGP was to provide a complete and accurate sequence of the 3 billion DNA base pairs that make up the human genome and to find all the estimated 20,000 to 25,000 human genes (Collins, Morgan and Patrinos, 2003). The successful finding of the sequence of genes has been considered as a huge scientific achievement. The success of the project leads to marvellous impact and huge value for biomedical and health care research (Collins et al., 1998).

Despite the success of HGP, the long-term goal of determining the functionality of all genes and its relation to human diseases has not been totally reached yet. Achieving such an ambitious goal requires launching different projects that will attempt to model the human diseases.

2.1.2 The Mouse Genome Project

To achieve a clear understanding about gene functionalities and its relation to diseases, mouse has been chosen as a mammalian model system for genetic research. This is mainly because the similarity between the human and mouse genome in terms of sequence and synteny (Bogue, 2003). In 2011, the International Mouse Phenotyping Consortium (IMPC) launched a 10-years project which aims to discover the functionality for every mouse gene by generating and systematically phenotyping 20000 knockout mouse strains (Brown and Moore, 2012). The first stage of the project aims to produce and phenotype 5000 mouse lines from 2011 to 2016 while the rest 15000 mouse lines should be completed by the 2021 (Adissu et al., 2014). Before the IMPC project, the WTSI started the MGP in 2006 as a one of the first program for large-scale gene function analysis in mammalian system (White et al., 2013). Later, the Sanger Institute became one of the founding members of the IMPC.

Genotype of an organism is its allelic rule (Griffiths et al., 2005). There are three types of genotype, WT (+/+) if no mutation in the gene, heterozygous (-/+) when the gene knockout is on one copy of the gene, if the gene knockout applied for both copies of the gene, this is called homozygous genotype (-/-).

Manipulation of the mouse genome involves a process called gene knock out. Gene knock out is a biological process by which an existing copy of a gene is knocked out by replacing it with

an artificial piece of DNA. The inactivated gene will lead to change in a mouse's phenotype which implies specific behavioural and other biochemical and physical characteristic changes (White et al., 2013).

A standard protocol for the characterization of the mouse phenotype, known as the core phenotype pipeline, was specified by the IMPC and endorsed by the research institutions of the consortium (Koscielny et al., 2014). One of the mandatary steps in this pipeline involves monitoring the mouse with the knocked-out genes by using X-ray imaging. The procedure of X-ray imaging the mouse model consists of several steps. First, the mouse should be anesthetized and monitored until it completely sedated. After that the X-ray image is taken and the mouse is removed from X-ray scanner. Finally, the analysis is done on the taken X-ray images. To enable a trusted assessment of a single gene knock-out, the X-ray imaging operation must be applied to at least 10 adult knocked-out animals (5 males and 5 females). The WTSI, for example, considers applying the X-ray procedure on 7 males and 7 females' mice (+ matched WT controls) for each gene knock-out. To analyse any changes on a skeleton of mouse in different ROIs of the X-ray image, 5 different mouse postures are taken. An example is shown in Figure 2.1.

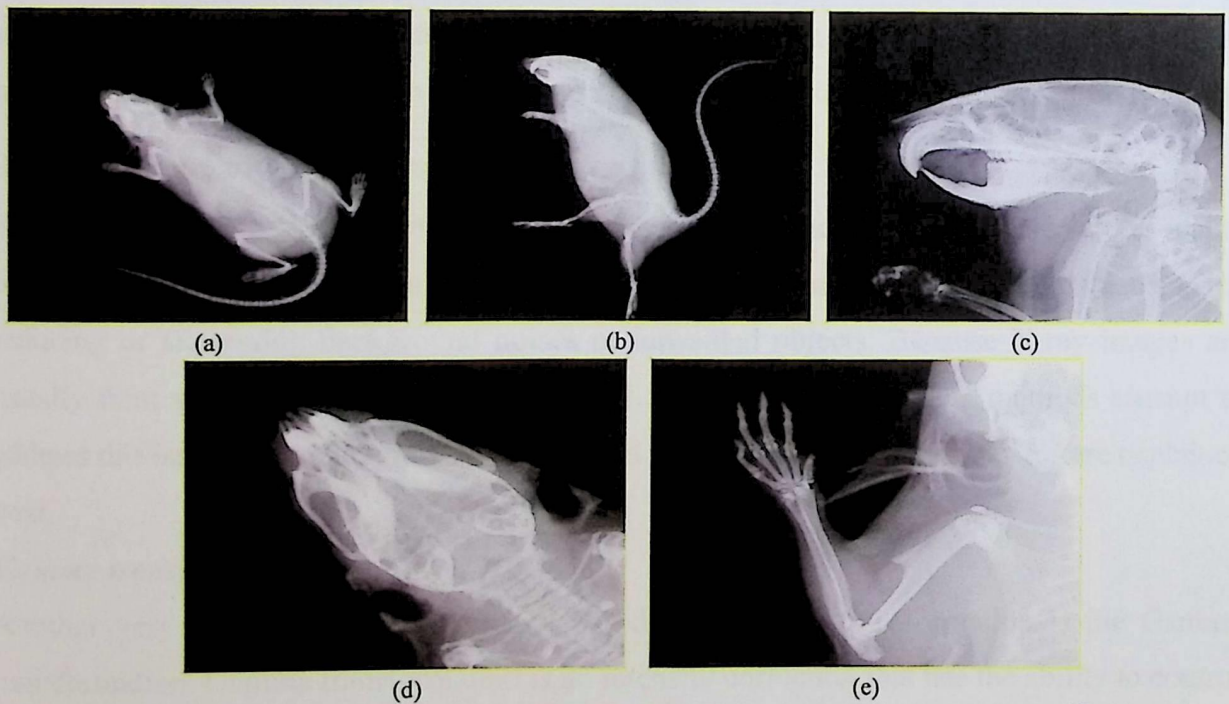


Figure 2.1 Examples of different postures of mouse X-ray images (images are provided by Sanger Institute).
(a) Full body-dorso-ventral view mouse X-ray image; (b) Full body-lateral view mouse X-ray image; (c) Close up head-lateral view mouse X-ray image; (d) Close up head-dorso-ventral view mouse X-ray image; (e) Close up left forepaw-dorso-ventral view mouse X-ray image

The substantial number of monitored lines will lead to a vast number of X-ray images which need monitoring and phenotyping. Due to the high volume of generated X-ray images from

pipelines and distinct types of assessing changes required per each image, computer-based techniques are highly desirable.

2.2 Computational Background

Since the discovery of X-ray back in 1895, medical imaging has been playing an increasingly key role in various aspects of medical practice. X-ray images, for example, can be used effectively to check the bone fractures in the hospitals. X-rays can also be used in fluoroscopy for monitoring progress during surgical operations (Birkfellner, Figl and Hummel, 2014).

Regardless of the techniques used, medical and biomedical images can serve two main purposes. First, biomedical images can be used to assist diagnoses of various diseases by identifying certain objects from the images (e.g. breast cancer). Second, biomedical images can be annotated either manually or automatically for describing certain characteristics of specific parts of the images. Although biomedical images are of very different types, they share some common problems in their practical use. In general, biomedical images may lack of clarity, have low contrast and contain several types of noise. For these reasons, to extract useful information from such images, adequate image processing techniques are needed. Overall, biomedical image processing may involve image de-noising and enhancements, image segmentation and detection of certain landmarks from the image. In the following subsections, a systematic review of the main image processing techniques related to this research is given.

2.2.1 Image Enhancements

Several algorithms and methods have been proposed to enhance digital images in either spatial or frequency domain. The enhancement mainly includes (a) increasing image contrast, and (b) reducing or supressing background noises or unwanted objects. Because X-ray images are usually faint with poor contrast (see Section 1.5), various enhancement methods attempt to address this issue. The main image enhancement techniques in different domains are explained next.

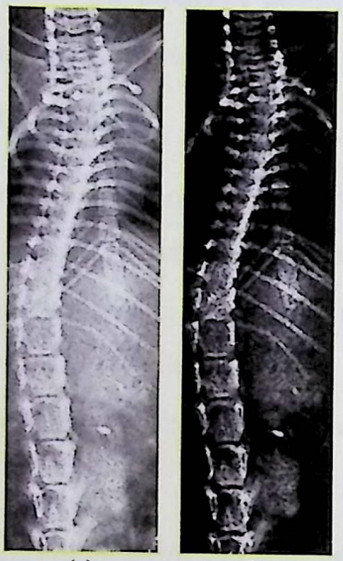
Gamma transformation

Another very popular enhancement tool based on intensity transformation is the Gamma transformation. Gamma transformation is an intensity correction that has the ability to control the brightness of image. The general formula for Gamma correction is as follows.

$$s = c\, r^{\gamma} \qquad (2.1)$$

where c and γ are positive constants, and s and r, are output and input intensity respectively.

Gamma transformation tends to reproduce the intensity of the image by mapping a narrow range of low input grey level values into a wider range of output values and vice versa. In general, the output image becomes darker when $\gamma > 1$ and brighter when $\gamma < 1$ (Gonzalez and Woods, 2017). Due to its ability to focus on a range of intensity, it could be a helpful technique in X-ray images when it is used to highlight required objects and remove other not required parts depending on their intensities. The Gamma transformation is used in this thesis to increase the contrast of the spine ROI as shown in Figure 2.2.



(a) (b)

Figure 2.2 Increasing the contrast of the spine X-ray image of mouse using Gamma transformation.
(a) Spine ROI of mouse X-ray image; (b) The result of applying Gamma transformation with gamma value of five to increase the contrast of the image

In a certain stage of spine segmentation, the spine ROI of mouse X-ray image (Figure 2.2(a)) shows high intensity of spine borders in addition to the other irrelevant structures and background noises. We can exploit the Gamma transformation with a value of γ parameter greater than 1 to darken the image. This should lead to eliminate lower intensity tissues while maintain the high intensity spine borders as shown in Figure 2.2(b).

Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast enhancement is one of the frequently used image processing function to improve poor contrast images. Histogram equalization (HE) is a non-linear technique used to enhance the contrast based on the histogram of the image (Zuiderveld, 1994).

Histogram of a digital image represents the frequency distribution of its intensity levels. For an L- level digital image with intensity levels between [0, L – 1], the histogram is given by a discrete function:

$$h(r_k) = n_k \qquad (2.2)$$

where r_k is the k^{th} intensity level, and n_k is the number of pixels in the image with the k^{th} intensity level. The histogram may be normalized as follows:

$$p(r_k) = \frac{n_k}{MN} \qquad (2.3)$$

where MN refers to the total number of pixels of an M × N image.

HE is the process that tends to span intensity over the whole grey scale. The transformation is implemented by the following equation (Gonzalez & Woods, 2017):

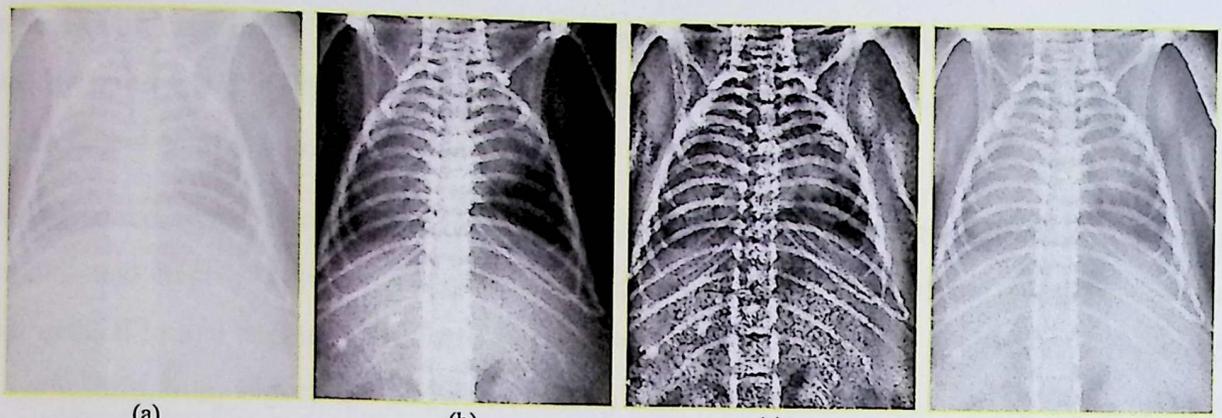
$$s_k = T(r_k) = (L-1)\sum_{j=0}^{k} p_r(r_j) \qquad (2.4)$$

where s_k is the span intensity level in the output image, $p_r(r_j)$ is the probability of occurrence of intensity level (r_j) in the input image.

Because the traditional histogram equalization applies the transformation function to the histogram of the entire image, two main drawbacks can be associated with its enhancements (Zuiderveld, 1994). The first one is that the HE may lead to improve the visibility of the noises or uninteresting areas, especially when they have a large peak in the histogram. The other problem is the missing of minor contrast differences when the number of pixels falling in a specific grey range is small.

While the basic non-adaptive histogram equalization applies the transformation function to the histogram of the entire image, adaptive histogram equalization (AHE) tends to divide the image into several regions and then applying histogram modification to each pixel based on the histogram of pixels within a region (Pizer et al., 1987). Bilinear interpolation can be then used to combine the neighbouring regions to avoid visibility of region boundaries. Despite the enhancement in image contrast over the HE, AHE still suffers from problem of over amplifying noises in homogenous regions of image.

To reduce the noises associated with applying AHE specifically in homogenous area, CLAHE was proposed by Zuiderveld (Zuiderveld, 1994). With CLAHE, the high peak in the histogram, which is caused by many pixels fall inside the same range, is clipped to allow a maximum number of pixels in each bin. To avoid change in histogram count, the clipped pixels are equally redistributed over the whole histogram. The clipping process is controlled by contrast factor parameter. Setting low value for this parameter leads to limited contrast enhancements while the higher value leads to a result equivalent to the result of AHE. CLAHE and AHE normally outperforms the basic method considerably, and are very good methods to enhance the contrast of an image by redistributing its histogram. Figure 2.3 below shows an example of applying HE (b), AHE (c) and CLAHE (d) on rib area of mouse X-ray image (a).



(a) (b) (c) (d)

Figure 2.3 Applying histogram equalization techniques on rib area of mouse X-ray image. (a) Rib area of mouse X-ray image before applying HE; (b) Result of applying HE on (a) which shows that the image contrast is non-uniform and also lack of contrast in some area of image; (c) Result of applying AHE on (a) which shows a uniform and consistent contrast of the image but with increasing of noises especially in the homogenous areas; (d) Result of applying CLAHE on (a) which shows a uniform enhancement of the image with reduction of oversaturation of the image in homogenous area

All the resulting images of applying HE (Figure 2.3(b)), AHE (Figure 2.3(c)) and CLAHE (Figure 2.3(d)) are considered better than the original image (Figure 2.3(a)) in terms of contrast. However, it seems that the result obtained from applying CLAHE is better than the others for rib area enhancement. This is due to the lack of consistency in the result of applying HE (Figure 2.3(b)) and the addition of noises as a result of applying AHE (Figure 2.3(c)). On the other hand, the result of applying CLAHE (Figure 2.3(d)) is more consistent across the whole image and less noisy because the CLAHE avoids oversaturation of the image especially in the homogenous area.

In general, adaptive and non-adaptive histogram equalization could be used effectively to enhance and pre-process biomedical X-ray images by increasing the contrast to make the required object more distinguishable from its background, making the object segmentation easier. However, the choice of suitable enhancement method is depending on the image or ROI to be enhanced.

Bilateral filter

Bilateral filter is a non-linear smoothing filter with an edge-preserving ability (Tomasi and Manduchi, 1998). The main idea behind the bilateral filtering is to combine between the domains filtering that based on geometrical closeness of pixels and ranges filtering which based on similarity function between pixels.

Combined bilateral filtering can be described as follow:

$$h(x) = k^{-1}(x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi) c(\xi, x) s(f(\xi), f(x)) d\xi \qquad (2.5)$$

with the normalization defined as follow:

$$k(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\xi, x)\, s\big(f(\xi), f(x)\big) d\xi \qquad (2.6)$$

where $c(\xi, x)$ represent the geometric closeness between the neighbourhood centre x and nearby point ξ and $s(f(\xi), f(x))$ measures the similarity between the pixels intensity for the centre x and nearby point ξ.

As explained above, bilateral filtering works by replace the pixel value at x with an average of similar and nearby pixel values. Bilateral filtering could be one of the good choices for biomedical image because it reduces the noises and at the same time preserves the important edge information. Figure 2.4 shows an example of enhancement achieved by using bilateral filtering with a rib area for mouse X-ray image.

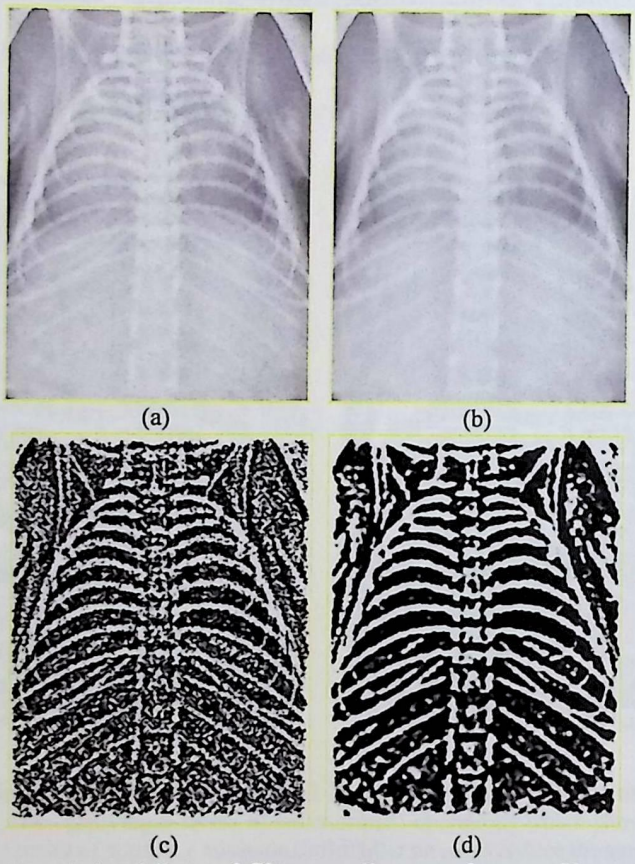


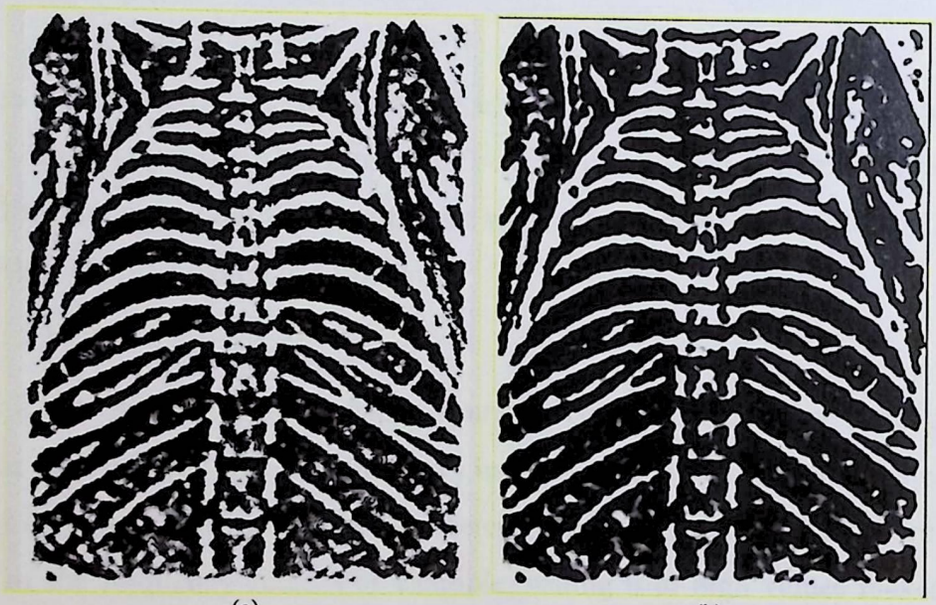
Figure 2.4 Results of applying bilateral filter on rib area of mouse X-ray image.
(a) Rib area of mouse X-ray image before applying bilateral filter; (b) Result of smoothing (a) by applying bilateral filter; (c) result of applying BHPF on a rib area before applying bilateral filtering which shows highlighting a lot of background noises; (d) result of applying BHPF on a rib area (b) after applying bilateral filtering which leads to minimizing the generated background noises

In order to see the impact of applying bilateral filter, we apply BHPF on the rib area before and after processing it with bilateral filter as shown in Figure 2.4(c) and (d) respectively. We can notice that the resulting image in Figure 2.4(d) is much better than the image in Figure 2.4(c) since the reduction of background noises obtained from bilateral filtering.

Spatial domain filter - order statistic filter

Many spatial filters can be used to minimize effects of noise and image blurring (Pitas and Venetsanopoulos, 1992). The non-linear filter principle is explained as follows. Depending on

the window size of the filter (e.g. 3 × 3) the filter scans the image and sorts the pixels masked by the window in an ascending order according to their intensity values. After that, it replaces the value of the centre pixel in the window with a new value from its neighbour values. The selected value depends on the type of filter (Maximum, Minimum or Median). The Median filter, for example, selects the median value from the sorted values of the window. The Median filter is effective in de-noising images, especially those with noises of salt-and-pepper type. For biomedical X-ray images, noise such as salt-and-pepper type is usually occurred in some stage of processing the image. So, using such statistic filters can be an important pre-processing and enhancement step before applying more sophisticated enhancement or segmentation techniques. Figure 2.5 shows an example of using minimum order filter to reduce unwanted objects from the rib area of mouse X-ray image.



(a) (b)

Figure 2.5 Enhancing the rib ROI of mouse X-ray image using minimum order filter.
(a) Rib area of mouse after applying a series of enhancement operations which shows existence of many non-rib objects; (b) The result of applying minimum order filter on (a) to reduce the small background noises

As we can see from Figure 2.5(a), the rib area contains many unwanted objects especially between the lower ribs. Minimum order filter could be effective to mitigate such unwanted objects and noises (Figure 2.5(b)).

Frequency domain filtering - Butterworth high-pass filter

One of the well-known transform to obtain the frequency domain of an image is the Fourier transform (Zadeh, 1953). Even there is no direct association between the frequency components of the Fourier transform and spatial features of an image, some broad observations can be made about their relationship (Gonzalez and Woods, 2017).

The slow varying frequency components represent the slowly varying intensity in the image. Those components are in and around the origin of the frequency transform and they are proportional to the average intensity of an image. The frequency components which are further distant from the origin of the frequency transform correspond to the faster intensity changes in the image such as edges of the objects.

Filtering in the frequency domain is one of the very beneficial filtering to enhance the image. This is based on the fact that the frequency components can provide useful information about the intensity characteristics of the image. The principle of filtering the image in frequency domain involves obtaining the frequency domain of an image, filtering it and then transforming it back to spatial domain. The process can be shown in the following form (Ballard & Brown, 1982):

$$g(x,y) = \Im^{-1}[H(u,v)\ F(u,v)] \qquad (2.7)$$

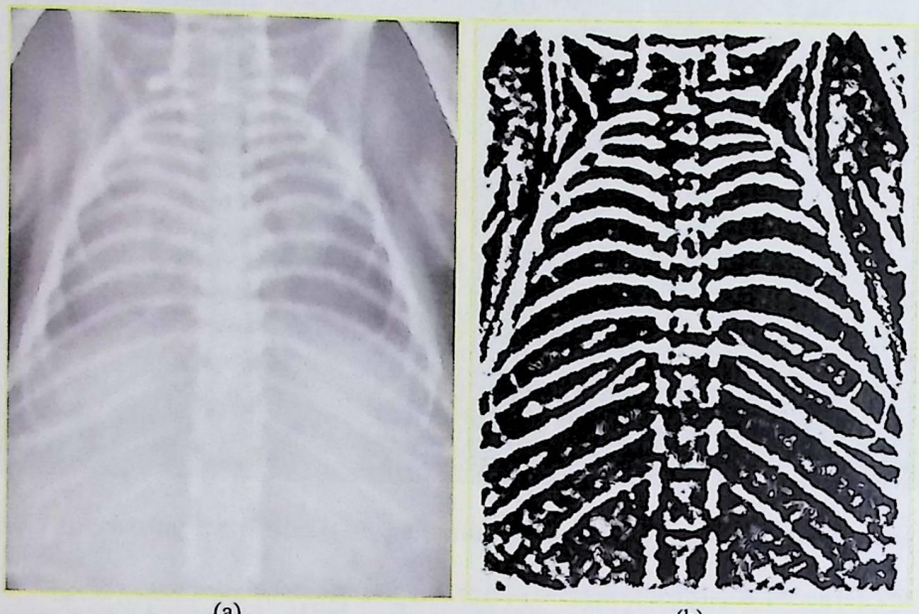
where $\Im^{-1}$ is the inverse Fourier transform, $F(u,v)$ is the Discrete Fourier Transform of input image $f(x,y)$, and $H(u,v)$ is a filter function. There are many types of filter function that could be used to filter frequency domain in different ways. One type of these filters is Butterworth High-Pass Filter (BHPF) which is defined as:

$$H(u,v) = \frac{1}{1+[D_0/D(u,v)]^{2n}} \qquad (2.8)$$

where D_0 is a cut off frequency (distance from the centre of frequency rectangle), $D(u,v)$ represents the distance between a point (u, v) in the frequency domain and the centre of the filter and n is the order of the filter. The parameter n can control the transition from low to high frequencies of the frequency filter. For example, setting n to high values makes the BHPF approach the ideal high pass filter characteristics with sharp transition from low to high frequencies. On the other hand, low values for n makes the BHPF more like Gaussian high pass filter with smooth transition from low to high frequencies.

In certain cases, the enhancement in the frequency domain is more useful than the enhancement in the spatial domain. An example of such cases is the overlapping between the target object and other objects in the background in the spatial domain, which makes it difficult to extract required objects. There are two types of filter to process images in the frequency domain: the high pass filter and the low-pass filter (Gonzalez and Woods, 2017). The high-pass filter functions attenuate low frequencies and allow high frequencies to pass. This property is essential for removing the blurring effect in an image and enhances the image details. This property makes this type of filters valuable for low contrast X-ray images. BHPF could be

exploited effectively to highlight the ribs in the rib area of mouse X-ray image as shown in Figure 2.6.



(a) (b)

Figure 2.6 Highlighting the rib from rib area of mouse X-ray image using BHPF.
(a) Rib area of mouse X-ray image before applying BHPF. The image suffers from blurring and so it is difficult to segment the ribs; (b) The result of applying BHPF on (a) which shows removing the blur and highlighting the rib objects

The rib area in Figure 2.6(a) suffers from blurring and so the ribs are difficult to be segmented using simple threshold method (see Section 6.3.3 for more details). So, using BHPF can highlighting all the ribs and diminishing the blur area which represents the low frequencies in the frequency domain of the image (Figure 2.6(b)).

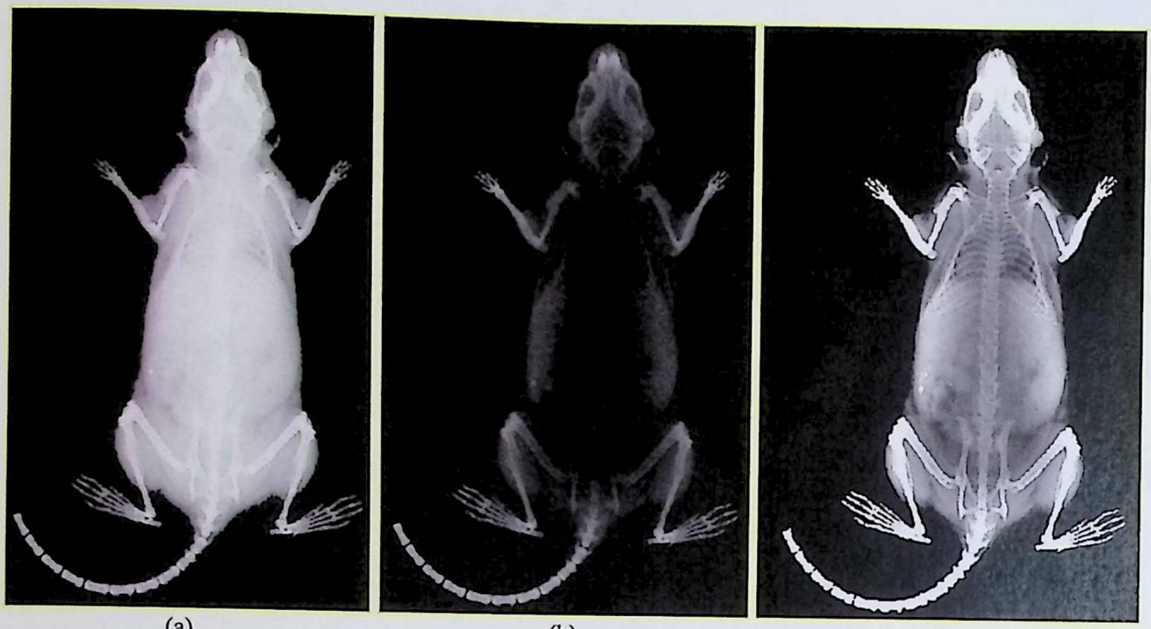
However, despite its considerable enhancements to the X-ray images, this type of filters needs an empirically setting for the main parameter such as cut off frequency to give the best result. This mean, for different quality images data set, this value should change to fit the new data set.

High-frequency emphasis filtering

One of problems of High-pass filter as described above is reducing the average intensity in the filtered image by setting the dc term (i.e. image average intensity) to zero. Such effect may not be desired for some application. The high-frequency emphasis filter overcomes this limitation by adding an offset to the high pass filter. The term of high-frequency emphasis filter is as follow (Gonzalez and Woods, 2017):

$$g(x,y) = \Im^{-1}\{[k_1 + k_2 H_{HP}(u,v)]\, F(u,v)\} \qquad (2.9)$$

where k_1 is the offset value to eliminate zero-out the dc term, and k_2 controls the contribution of high frequencies. Figure 2.7 illustrates the impact of using high frequency emphasis filtering in comparing to high frequency filtering when applying on a full-body mouse X-ray image.



(a) (b) (c)

Figure 2.7 Increasing the contrast of the limbs in X-ray image of mouse body using BHPF and HE- BHPF.

(a) X-ray image of mouse body before applying BHPF; (b) The result of applying BHPF on (a). The contrast of the image is very low because BHPF set the average intensity of the image to zero; (c) The result of applying high emphasis BHPF on (a) which shows higher contrast for the limbs of the mouse body because it prevents zeroing the average intensity of the image

Applying BHPF without emphasis on Figure 2.7(a) results in decreasing the contrast of the overall image substantially as shown in Figure 2.7(b). This results from the traditional BHPF tends to set the average intensity of the image to zero. To deal with this issue and achieve better contrast for the image, high emphasis BHPF is applied on the input image and the result is as shown in Figure 2.7(c). The offset value involves in high emphasis filtering (k_1 in equation 2.9) prevents setting the average intensity of the image to zero.

Morphological operations over binary image

Binary morphology operations refer to a set of operations that affect the shape or the structure of an object in a binary image (Matheron, 1975). They are mainly used for post-processing, pre-processing or to extract a description of objects/regions in an image. Dilation and erosion are considered as the two basic operations for other binary morphology operations. Dilation is a process of growing objects in binary images while erosion is a process of thinning objects in binary images, and both processes are controlled by a structuring element (Soille, 2004). Dilation and erosion can be expressed mathematically by the following two formulas respectively:

$$A \oplus B = \left\{ z | \left(\widehat{B}\right)_z \cap A \neq \emptyset \right\} \quad (2.10)$$

$$A \ominus B = \{ z | (B)_z \cap A^c \neq \emptyset \} \quad (2.11)$$

where A is the binary image, A^c is the complement of A and $(\widehat{B})_z$ is the structure element B after reflecting and translating it by z. The dilation of A by B defined as the set of all displacement z where A and $\widehat{B}$ overlap by at least one element. The erosion of image A by structure element B is the set of all structuring element origin locations where the translated B has no overlap with background of image A.

Two important morphology operations, opening and closing, are defined by combining dilation and erosion. Opening is an operation that has ability to smooth the contour of an object, and reduces some small unwanted objects. Opening is simply implemented by erosion followed by dilation. Closing is an operation that closes narrow gaps and fills small holes to smooth objects. Closing is fulfilled by dilation followed by erosion (Gonzalez and Woods, 2017).

An example of the applications of applying binary morphological operations in this thesis could be seen in Figure 2.8 (see Section 5.3.4 for more details). The binary closing morphology operation can be used to close the spine mask parts in Figure 2.8(a) and form a continuous spine mask as shown in Figure 2.8(b). However, using this operation with predetermined fix parameters (i.e. 40 and 90° for the length and orientation of line structure element respectively), could cause closing other unwanted parts in addition to spine parts which leads to create some big unwanted artefact as shown by red ellipse in Figure 2.8(b).



(a) (b)

Figure 2.8 Result of applying binary closing morphology operation to close the spine mask. (a) Binary spine mask segmented from mouse X-ray image. The image contains most of the parts of spine in addition to some unwanted objects; (b) Result of applying binary closing morphology operation on (a) which shows closing the parts of the spine mask and generated artefact (red ellipse)

In general, binary morphological operations can perform similarly to spatial filtering. They can be useful in a pre- and post-processing of a segmented binary mask.

Two precautions should be considered when using such operations. The first one is that they need a careful setting for their parameters (i.e. size and orientation of structure elements in open and close operations) to give the required result. Furthermore, beside their intended enhancements to some objects in the image, some other objects might receive a negative impact from applying those operations blindly on the whole image.

Morphological operations over grey-scale image

The basic morphological operations that have been defined with binary image have an extension to a grey scale image (Matheron, 1975). The erosion of image f by a structuring element b implies selecting the minimum value in the region coincident with b. At any location (x, y) of image f, the erosion is given by

$$[f \ominus b](x, y) = \min_{(s,t)\in b}\{f(x + s, y + t)\} \quad (2.12)$$

Similarly, the dilation is to select the maximum value in the window covered by a structuring element. The effect of erosion on grey scale image is to darken it while the effect of dilation is opposite to the erosion.

Similar to binary images, erosion and dilation are also used to define the opening and closing for grey scale images. Opening of the grey scale image involves erosion followed by dilation, whereas closing requires dilation followed by erosion. In general, opening tends to smooth an image by removing bright details that smaller than the structure element while closing suppresses dark details smaller than the structure element (Gonzalez and Woods, 2017). Two important transformations are resulted from above grey scale morphology. The first transformation is known as top-hat which is implemented by subtracting the opening of an image from the original image:

$$T_{hat}(f) = f - (f \circ b) \quad (2.13)$$

The other transformation is known as bottom-hat which is implemented by the closing of an image minus the original image:

$$B_{hat}(f) = (f \cdot b) - f \quad (2.14)$$

An important application of these transformations is to remove undesired objects from the input image. This is done by setting the size of the structure element in either closing or opening operations to be not fitting the size of the object to be removed. Top-hat transformation is mainly suitable for grey scale image with white foreground on a black background while the

bottom-hat transformation has an opposite application from that for the top-hat (Gonzalez and Woods, 2017). One of beneficial enhancement to increase the contrast of the image could be obtained by mixing the top-hat and bottom-hat transformations. To explain this operation, Figure 2.9 shows the enhancements of pelvis area of the mouse X-ray image using grey-scale morphology operations.

By setting a relatively large structural element, the top-hat transformation tends to eliminate most of background noises from the pelvis image (Figure 2.9(a)) and keeps very faint pelvis bones (Figure 2.9(b)). After that, adding top-hat result (Figure 2.9(b)) to the original image (Figure 2.9(a)) increases the contrast of the bones considerably but with highlighting of background noises (Figure 2.9(c)). To decrease the effect of over contrast resulting in all pelvis area, a result of applying bottom hat on the original image (Figure 2.9(d)) is subtracted from the enhanced image (Figure 2.9(c)) to produce the final image (Figure 2.9(e)). As we can see from Figure 2.9(e), the contrast of the pelvis bone is much better than the original image as shown in Figure 2.9(a).

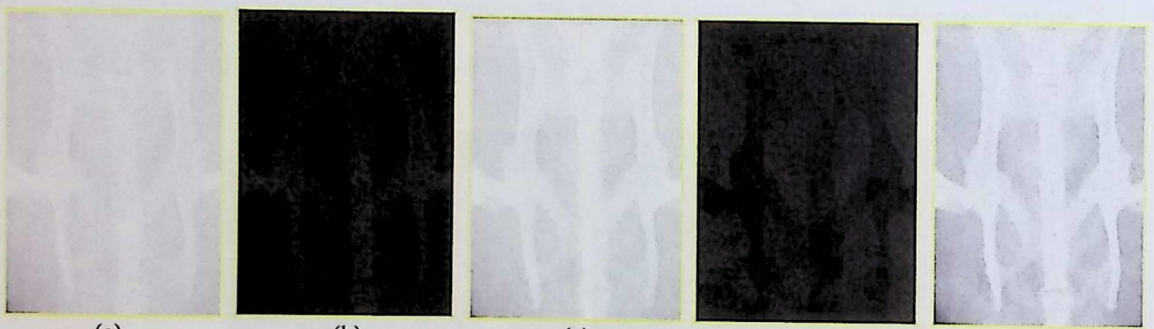


Figure 2.9 Combination of grey-scale morphology operations to enhance the pelvis area. (a) Low contrast pelvis area of mouse X-ray image; (b) Result of applying top-hat transformation on (a) which leads to remove the background blurring and keep the pelvis bones; (c) Result of adding (a) to (b) which produces a high contrast pelvis bone; (d) Result of applying bot-hat transformation on (a) which shows the inverse of applying top-hat operation; (e) Result of subtracting (d) from (c) which leads to enhanced contrast pelvis area in terms of contrast of pelvis bones

2.2.2 Image Segmentation

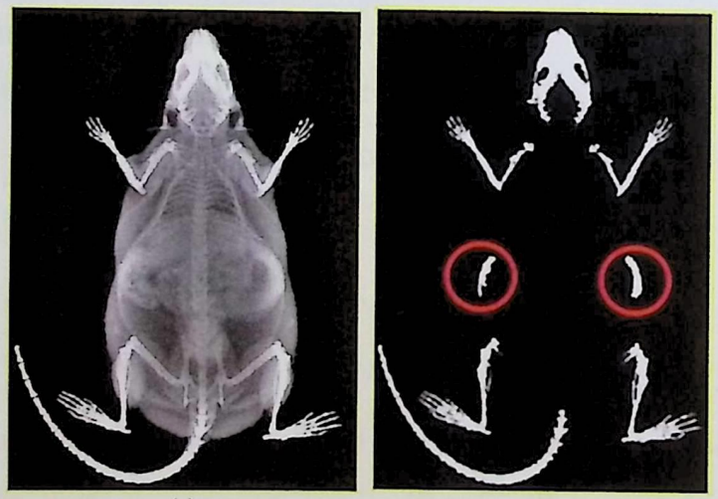
Segmentation is one of the most important topics in the image processing. Any image, especially biomedical X-ray images, will consist of many objects and regions such as the spine, ribs or skull. Segmentation aims to partition these objects or regions containing the objects according to the boundaries of the objects. So, segmentation can help the radiologist to annotate the objects. Many approaches and methods have been developed.

Classical image segmentation methods, such as thresholding, region-based and edge-based, have been effectively applied to medical image segmentation (Stolojescu-Crisan and Holban, 2013). Those methods mainly depend on pixel intensity values to perform the segmentation.

Hough transform is another type of fixed shape matching feature extraction method which could be exploited to segment an object with a specific shape. Several examples segmentation methods are described next.

Region growing

Region growing is one of the well-known segmentation algorithms. The idea of the region growing is to group pixels with similar properties into a region. This procedure can be implemented by first selecting a seed point (i.e. a pixel with an intensity value), and then including its neighbouring pixels with predefined similar properties to the seed point (Rangayyan, 2004). The region growing methods can be beneficial for segmenting certain parts of X-ray images (e.g. Limbs of mouse body) which have a good contrast. The advantage of the region growing is that its ability to segment connected ROI without producing too many unwanted objects. Region growing segmentation of limbs and skull mask (see Section 5.3.2 for more details) from the enhanced full body mouse X-ray image is illustrated in Figure 2.10. In order to segment the high intensity limbs from the full body mouse X-ray image in Figure 2.10(a), we use region growing technique with high intensity value for the seed points (i.e. 255) and the result is as shown in Figure 2.10(b).



(a) (b)

Figure 2.10 Region growing segmentation of limbs and skull mask from full body mouse X-ray image.
(a) Full body mouse X-ray image with high contrast for limbs and skull; (b) Result of applying region growing segmentation on (a). The result shows the good segmentation of all the limbs with some unwanted objects (red circles)

We can see that the region growing succeeded in segmenting the limbs and the skull with minimum number of unwanted objects as encircled by the red circles in Figure 2.10(b). As we interested in rough segmentation of the limbs, the resulting mask of Figure 2.10(b) is highly valuable for global landmarks detection (see Section 5.3.2 for more details). However, the

important and difficult point about region growing is selecting proper value for homogeneity criterion to isolate the required parts of the image.

Thresholding

Thresholding is considered as one of the simplest yet effective method of segmenting objects in image. In various stages of X-ray image processing, use of thresholding is an essential process to extract a binary mask of required object or part of the image. In the optimal case, when the object is distinguishable from its background, the selection of threshold could be the value of bottom of the bimodal grey-level histogram. However, the grey-level histogram is not always perfectly bimodal in a way that helps to select the threshold value. In the literature, there are many proposals to calculate a suitable value of threshold to discriminate between a ROI and its background automatically. Below, brief descriptions for two of very well-known thresholding methods are given.

1. **Maximum entropy method**. The maximum entropy method chooses a threshold value to achieve the maximum separation between the object and its background in the image (Kapur, Sahoo and Wong, 1985). This method was based on entropy concept. Let $p_1, p_2, \ldots, p_n$ be the probability distributions of the grey levels. From those distributions, A and B probability distributions are defined for discrete values (1 to s) and (s+1 to n) respectively. The entropies for A and B distributions can be defined as follow.

$$H(A) = -\sum_{i=1}^{s} \frac{p_i}{p_s} \ln \frac{p_i}{p_s} \qquad (2.15)$$

$$= \ln p_s + \frac{H_s}{p_s} \qquad (2.16)$$

Similarly, for B distribution, the entropy is

$$H(B) = \ln(1 - p_s) + \frac{H_n - H_s}{1 - p_s} \qquad (2.17)$$

If the summation of entropies for A and B is defined as $\psi(s)$ as follow

$$\psi(s) = \ln p_s(1 - p_s) + \frac{H_s}{p_s} + \frac{H_n - H_s}{1 - p_s} \qquad (2.18)$$

Then the optimal threshold value is the value of s that maximizes $\psi(s)$ (Kapur, Sahoo and Wong, 1985).

2. **Otsu method**. Otsu method chooses the threshold as a value that maximizes the between-class variance of the background and foreground classes (Otsu, 1979). The Otsu method is based on hypothesis that properly thresholded classes should be recognizable with respect to the intensity values of their pixels. If a normalized histogram of an image with L distinct integer intensity levels [0, 1, 2,…., L-1] is divided

into two classes (background and foreground) at a level k then the optimal threshold k^* is the threshold that maximizes between class variance (σ_B^2) as follow

$$\sigma_B^2(k^*) = \max_{0 \le k < L-1} \sigma_B^2(k) \qquad (2.19)$$

Use of adaptive threshold selection methods will enhance the overall accuracy of segmentation over fixed threshold value selection. However, thresholding segmentation still considered sensitive to change of image intensity across different images of data set. So, the obtained mask from thresholding segmentation process should considered as initial masks to help in detection of more accurate segmentation to the Object of Interest (OOI).

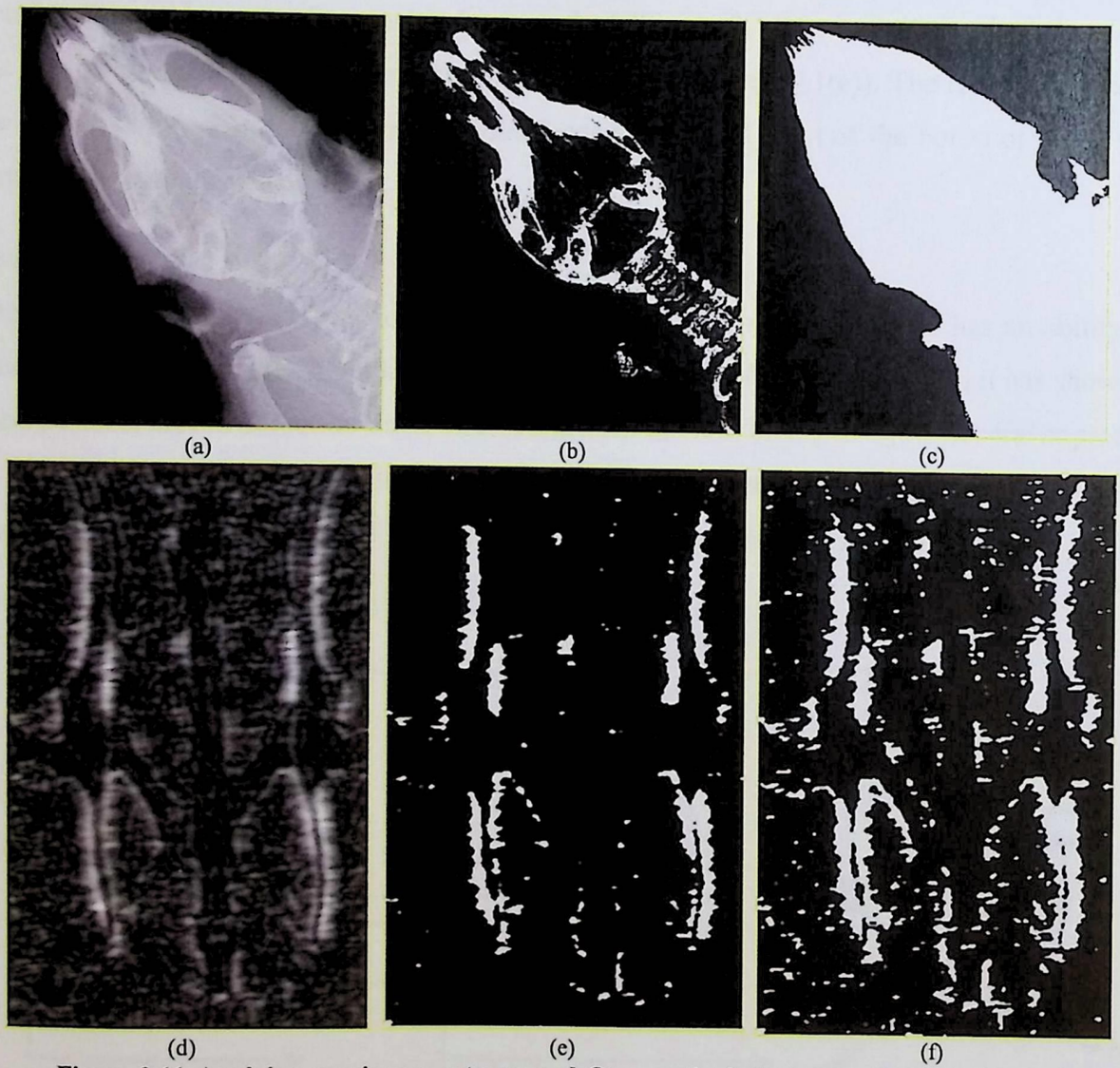


Figure 2.11 Applying maximum entropy and Otsu methods to segment various parts of from mouse body X-ray images.

(a) X-ray image of mouse head; (b) skull mask results from using maximum entropy method on (a) which leads to the removal of all the tissues and keeps the high intensity bones of the skull. This mask is considered ideal for skull modelling; (c) skull mask results from using Otsu method on (a) which segment unwanted tissues as part of the bones of the skull which leads to inaccurate skull modelling; (d) Enhanced image of pelvis of mouse X-ray image; (e) Pelvis mask results from using maximum entropy method on (a) which leads to remove some part of the pelvis bone; (f) Pelvis mask results from using Otsu method on (a) which leads to segment most of the bones of the pelvis and so considered better than the results of (e)

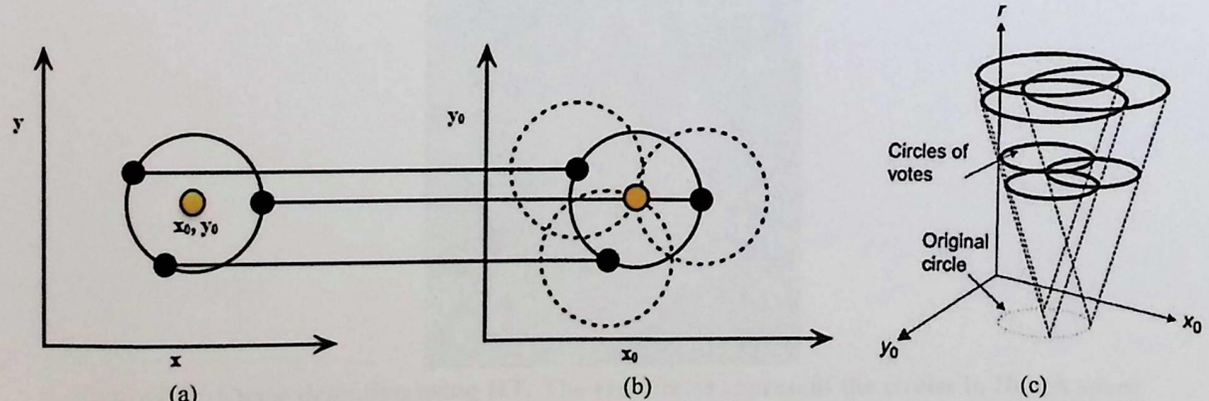
In this thesis, the segmentation is required in different proposed methods. The selection of a suitable adaptive threshold method depends on the required mask. Figure 2.11 shows the results of applying the maximum entropy and Otsu methods to produce binary masks for a skull and pelvis from mouse X-ray image. For the skull, the mask obtained by using maximum entropy method (Figure 2.11(b)) is considered better than the mask resulting from Otsu method (Figure 2.11(c)) for the purpose of skull modelling (see Section 4.3.2 for more details). This is because the resulting mask in Figure 2.11(b) has only the bones of the skull without unwanted other tissues on the contrast to the other mask resulting from Otsu method (Figure 2.11(c)). For the pelvis segmentation from the enhanced pelvis image (Figure 2.11(d)) (see Chapter 7 for more details), the mask results from using Otsu method (Figure 2.11(f)) is shown to be better than the one results from using maximum entropy method (Figure 2.11(e)). The reason for that is the Otsu method produces a threshold that leads to segment most of the bones of the pelvis mouse X-ray image.

Hough Transform

The Hough Transform (HT) (Hough, 1962) is a well-known transform that has an ability to locate geometrical shapes in an image (Nixon and Aguado, 2008). Particularly, it has shown a good performance in detection lines, circles and ellipses from binary images. General approach of HT implies mapping the points of an image from geometric space to accumulator space (Hough space). The principle of HT for circle detection is explained in the following.

A circle shape can be defined by the following equation:

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \qquad (2.20)$$



(a) (b) (c)

Figure 2.12 Circle representation in geometric and accumulator spaces for Hough transform.
(a) A circle with centre (x_0, y_0) and radius r in x-y domain; (b) The three points on the perimeter of circle in (a) generate three circles with a radius r in the 2D accumulator space. The intersection of circles in accumulator space represents the centre of the circle in the geometric space; (c) 3D accumulator space to count votes for Hough circle detection (Nixon and Aguado, 2008). The third dimension results from using different radiuses to generate the candidate circles

The formula describes the location of (x, y) on the circle with radius r and centre point (x_0, y_0). However, this equation can be visually seen as a locus of points (x, y) (Figure 2.12(a)) or as a locus of points (x_0, y_0) centred on (x, y) with radius r (Figure 2.12(b)). The second visualization of the circle formula means that each edge point on the circle could be seen as a centre of new circle with a radius r. If we extend the definition for different values of r for each edge point, we end with 3D accumulator space as shown in Figure 2.12(c). The maximum votes obtained from circles are generated in an accumulator space corresponds to the parameters of the circle in the original image.

To enable solving for the circle parameters, the parametric form of circle is used in the HT as follow.

$$x = x_0 + r\cos\theta \quad (2.21)$$

$$y = y_0 + r\sin\theta \quad (2.22)$$

Thus, the HT mapping to an accumulator space is defined by the following equation:

$$x_0 = x - r\cos\theta \quad (2.23)$$

$$y_0 = y - r\sin\theta \quad (2.24)$$

As it explained earlier, the ability of HT to detect an object with a circular shape in a binary image might be exploited very efficiently to model some part of the mouse body in X-ray images. Figure 2.13 shows an example where the HT can detect a circle to model the skull of the mouse. As we can see from the figure, the votes collected from the red circles (accumulator space) lead to detect the parameters of the circle that model the skull in the image (green circle).

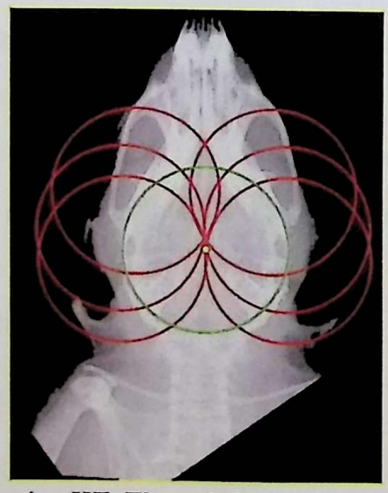


Figure 2.13 Circle detection using HT. The red circles represent the circles in Hough space which votes to select the required circle on the image (green circle) with an intersection point as a centre for the new circle (yellow point).

2.2.3 Machine Learning based Objects Detection and Classification

Machine learning is increasingly becoming one of the essential fields of general practice and activity within computer science (Rogers and Girolami, 2011). One of the recent application of different machine learning tools is in image analysis and segmentation. Using different algorithms and features extracted from the image, a model could be built to recognize a specific objects or landmarks based on extracted features by training an effective classifier for the recognition. In this section, we present two well-known machine learning approaches for classification. The first one is the Random Forest (RF) (Breiman, 1999) as one of the mostly adopted machine learning methods for classification. The other machine learning tool is the Support Vector Machine (SVM) classifier (Cortes and Vapnik, 1995).

Random Forests

RFs are an ensemble learning method based on generating multiple decision tree predictors where each tree depends on a random vector sampled independently (Breiman, 1999). For the classification problem, the output class of the RF is the class with the largest number of votes from all the trees. For the regression problem, the prediction of the observation is the average of the prediction of the trees in the model. The general steps for RF algorithm are as follow (Hastie, Tibshirani and Friedman, 2008):

1- $For\ b = 1\ to\ B$:
   a- Draw a bootstrap sample $Z^*$of size N from the training data.
   b- Induce a decision tree (known as RF tree) to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      i. Select m variables at random from the p variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two child nodes.
2- Output the ensemble of trees$\{T_b\}_1^B$.

Where B is total number of trees in the RF.

To make a prediction at a new point x:

Let $\hat{C_b}(x)$ be the class prediction of the bth RF tree.

Then $\hat{C}_{rf}^B(x) = majority\ vote\ \{\hat{C}_b(x)\}_1^B$.

Figure 2.14(a) shows a diagram for prediction process for one instance using a RF. The RF introduces the randomness in different ways. The first one is through drawing a random bootstrap sample from the training set with replacement (bagging). This leads to generate an independent data set each time by deleting some instances and replicating others. The second randomization involves choosing a random subset of variables from the bootstrapped data to build each decision tree in the forest (Witten, Frank and Hall, 2011). One of the reason to use bagging in RF is to exploit it to estimate the error of trees in the model. Because of each tree in the RF model is trained using a different bootstrap sample from the original training set, it is possible to estimate the classification error rate of the model during training time. This is done by predict the case left out during the construction of tree (Out Of Bag (OOB)) by using the same tree.

The RF were used heavily in recent years for landmarks detection from X-ray image ((Lindner et al., 2013), (Chu et al., 2014), (Ibragimov et al., 2014) and (Lindner and Cootes, 2015)). To exploit RF in the landmark detection, at the training stage, the target landmark points are labelled as positive sample (class 1) and other unwanted points are labelled as negative samples (class 0). The specific features extracted from a patch around the point sample are used to train the RF. The trained RF then could guess the class of new coming point at prediction stage depending on its features.

Support vector machine

Support vector machine is a type of non-probabilistic binary classifier which firstly presented by (Cortes and Vapnik, 1995) and since then have been used in many machine learning applications. The classification of objects in standard SVM is based on decision boundary (hyperplane). Positive class objects are lying on one side of the line while the negative class objects are lying on the other side as shown in Figure 2.14(b). Out of many possible hyperplanes that divide the two classes, there is only one that maximise the margin. The distance between the optimal hyperplane and the nearest points on either side is known as a margin (Figure 2.14(b)). To make a classification decision, the SVM is depending on the closest features to the margin which they are known as support vectors (filled circles and square in Figure 2.14(b)) (Rogers and Girolami, 2011). The class of an example is determined by which side of the hyperplane the example appears. The SVM classifier is known for working well in high dimensional feature space.

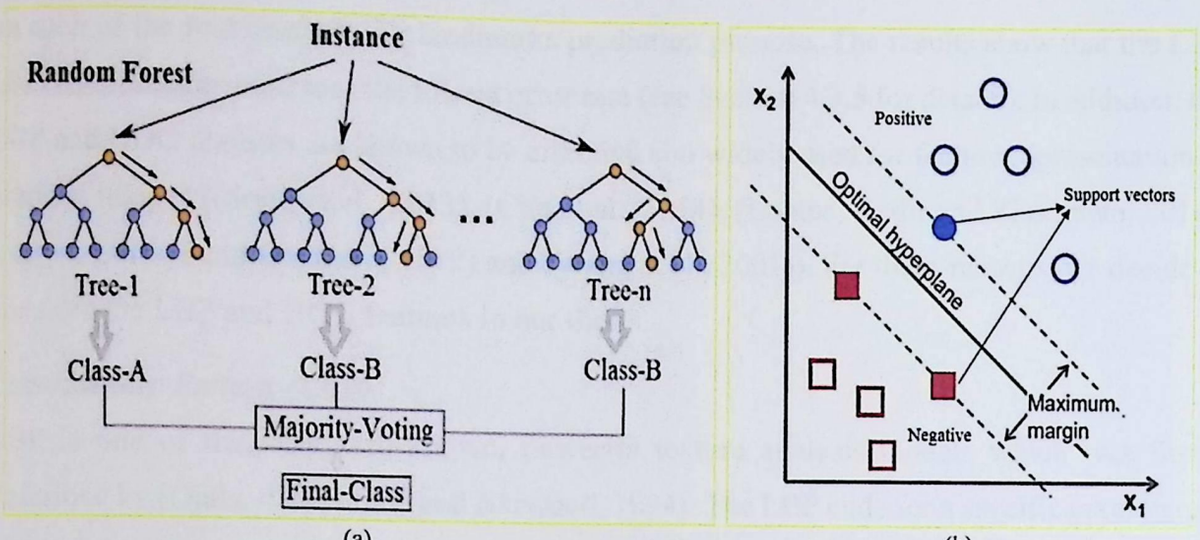


Figure 2.14 Schematic representation of RF (Breiman, 1999) **and SVM** (Cortes and Vapnik, 1995).
(a) RF prediction. Each new incoming instance is predicted by all the trees in the forest. The class of the instance is then determined depending on the majority voting from all the trees; (b) Linear SVM classifier. The optimal hyperplane (green line) is selected to maximize the margin (the distance between support vectors (filled circle and filled squares) and the hyperplane)

2.2.4 Texture-based Features

Features could be defined as any useful information to describe an object (i.e. image) and it could be used for sample classification. The feature extraction is the process of generating a set of features from an input sample (Guyon and Elisseeff, 2006). Features that describe the texture of the image could be exploited for image segmentation or pattern recognition. Several studies in the literature discussed the use of texture features. Reyes-Aldasoro and Bhalerao (Reyes-Aldasoro and Bhalerao, 2006) discussed an obtained improvement on several texture analysis techniques when they combined with feature selection method based on Brattacharyya distance. Another study was submitted by Randen and Husøy (Randen and Husøy, 1999) to compare different filters (i.e. Gabor filters and wavelet transform) for specific texture classification algorithms. They found out that the best filter for texture classification depends mainly on the texture type. However, an experiment was presented by Ojala, Pietikainen and Maenpaa (Ojala, Pietikainen and Maenpaa, 2002) showed that the Local Binary Pattern (LBP) operator outperformed the methods evaluated by Randen and Husøy.

To select suitable texture features, we invistigated two different types of texture features. LBP (Ojala, Pietikainen and Harwood, 1994) and the Grey Level Cooccurance Matrix (GLCM) (Haralick, Shanmugan and Dinstein, 1973) are chosen to represent local variations in the texture. On the other hand, the Histogram of Oriented Gradient (HOG) (Dalal and Triggs, 2005) and the Gabor Filter (Aach, Kaup and Mester, 1995) are chosen to represent variations in the texture orientation. We estimated the OOB error (see Section 2.2.3) of the RF models trained

on each of the four features for landmarks prediction purpose. The results show that the LBP and HOG features produces the lowest error rate (see Section 4.3.5 for details). In addition, the LBP and HOG features are shown to be effective and widely used for feature representation in medical images ((Song et al., 2013), (Chu et al., 2014), (Lootus, Kadir and Zisserman, 2014), (Nanni, Lumini and Brahnam, 2012) and (Wang et al., 2001)). For these reasons, we decide to consider the LBP and HOG features in our thesis.

Local Binary Pattern (LBP)

LBP is one of the very well-known, powerful texture analysis models which was firstly described by (Ojala, Pietikainen and Harwood, 1994). The LBP code for a specific pixel (target pixel) centred in a (k × k) window (e.g. k = 3) is generated by comparing its intensity value with surrounding pixels' intensity values. For centre point P and neighbour points P_x, the new value of the neighbour point after thresholding it is as follow (Nixon and Aguado, 2008)

$$s(x) = \begin{cases} 1, & if\ P_x > P \\ 0, & otherwise \end{cases} \qquad (2.25)$$

If the value of the target pixel is greater than the value of compared pixel, 1 is assigned to this pixel otherwise 0 is assigned. After setting all new values to the 8 neighbours, the LBP code for the target pixel P is derived from weighting associated with each result of thresholding as follow (Nixon and Aguado, 2008)

$$LBP = \sum_{x \in 1,8} s(x) \times 2^{x-1} \qquad (2.26)$$

Figure 2.15 shows an example of generation of LBP code.

118	190	6
69	106	110
42	31	106

(a)

1	1	0
0	161	1
0	0	0

(b)

Figure 2.15 LBP code generation. Adapted from (Nixon and Aguado, 2008).
(a) (3 × 3) window which represents intensity values from an image; (b) LBP code generated from comparing the centre value in (a) with the neighbouring values and the final code value is (10100001) where the top left is the Least Significant Bit (LSB) and the top middle is the Most Significant Bit (MSB), the decimal equivalent is 161

For a window of size (3 × 3), there should be (2^8) values in the LBP histogram. The useful extension to reduce the histogram dimension from 256 to 59 bins was proposed by (Ojala, Pietikäinen and Mäenpää (Ojala, Pietikainen and Maenpaa, 2002). This was done by allocate a bin for every uniform binary code (has at most two transitions between 0 and 1) and assign all other non-uniform pattern to a single bin. The LBP code can measure the local contrast from

the difference between points encoded as 0 and points encoded as 1. Different binary patterns of LBP correspond to basic features in the image (i.e. dark spot, bright spot, edge and end of line) (Nixon and Aguado, 2008). The ability of LBP code features to sense to change in contrast that resulting from changes in intensity values nominate it to be effective features to be used with RF for landmark detection in X-ray image.

The ability of LBP to sense the gradual intensity changes between the rib and its backgrounds could be exploited to train an effective classifier to remove non-rib objects (see Section 6.3.5 for more details).

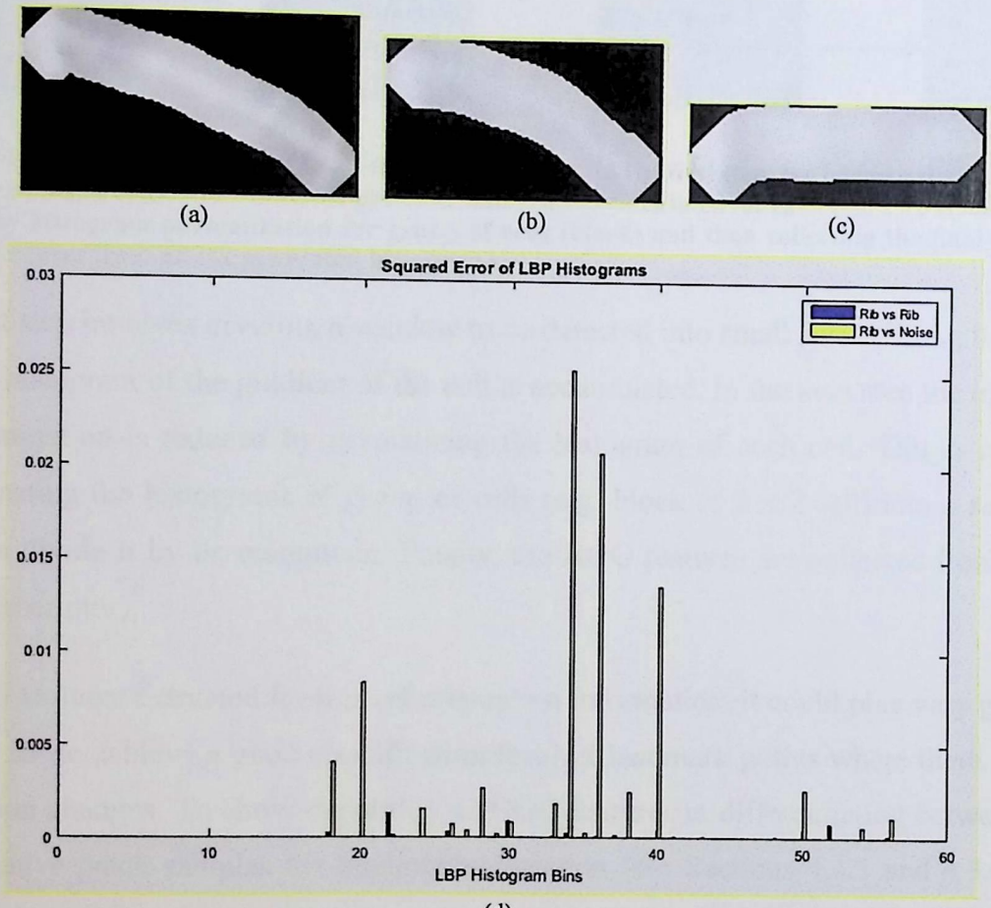


Figure 2.16 Comparision of the mean squared error of LBP histograms of rib vs rib and rib vs non-rib objects from mouse X-ray image.
(a) and (b) rib objects and their surrounding backgrounds;(c) non-rib object and its surrounding background; (d) Comparing the squared error of the LBP histograms for two ribs (blue bar) and the squared error for rib and non-rib objects (yellow bar). The results show that the squared error for rib vs rib is much less than the squared error for rib vs non-rib objects

To show the ability of LBP in sensing such changes in the image intensity, we compute the squared error of each pair of corresponding bins between the LBP histograms for two ribs as shown in Figures 2.16(a) and (b), and between the LBP histograms of a rib and a non-rib objects (Figures 2.16(b) and (c)). Figure 2.16(d) shows the comparison between the two errors and we can see that the rib versus non-rib error (yellow bar) is much higher than the rib versus rib error

(blue bar). This indicates that using LBP histogram as a feature vector can distinguishing a rib from a non-rib object.

Histogram of Oriented Gradient (HOG)

HOG is a feature descriptor that presented by Dalal and Triggs (Dalal and Triggs, 2005) with a main purpose of human face detection. The basic idea of HOG method is to extract a set of features exploiting the histogram of image gradient orientation in a dense grid. The basic steps of HOG feature extraction can be summarized in the Figure 2.17.

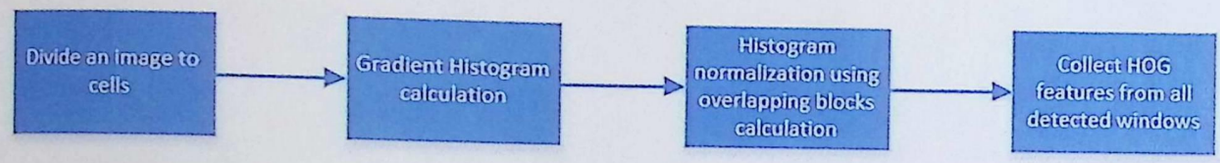


Figure 2.17 Main steps of HOG feature extraction. In the first step, the image is divided into equal-size cells. After that, the gradient histogram is calculated for each cell. This is followed by histogram normalization for group of cells (block) and then collecting the final HOG features from all the generated histograms.

The first step involves dividing a window to be detected into small size cell (e.g.8×8). After that the histogram of the gradient of the cell is accumulated. In the next step the invariance to the illumination is reduced by normalizing the histogram of each cell. This is achieved by concatenating the histograms of group of cells (e.g. block of 2×2 cell) into a single vector and then divide it by its magnitude. Finally, the HOG features are collected from all blocks (HOG descriptor).

As HOG features extracted from pixel orientation information, it could play very good role in training RF to achieve a good classification level of landmark points where there are a lot of orientation changes. To show the ability of HOG features in differentiating between positive and negative patch samples for landmarks detection (see Sections 4.3.5 and 4.3.6 for more details), we compare the squared error between their HOG features as shown in Figure 2.18. Figure 2.18(d) shows the squared error between the histogram of HOG features for reference (positive) (a) and positive sample (b) patches (blue bar). The figure also shows the squared error between the histogram of HOG features for reference (positive) (a) and negative sample (c) patches (yellow bar). The results show that the squared error for the two positive samples is much less than the squared error for the positive and negative samples. This big difference between the squared errors for the two cases confirmed the ability of HOG features to be used as a feature to classify between positive and negative patches for landmarks detection problem.

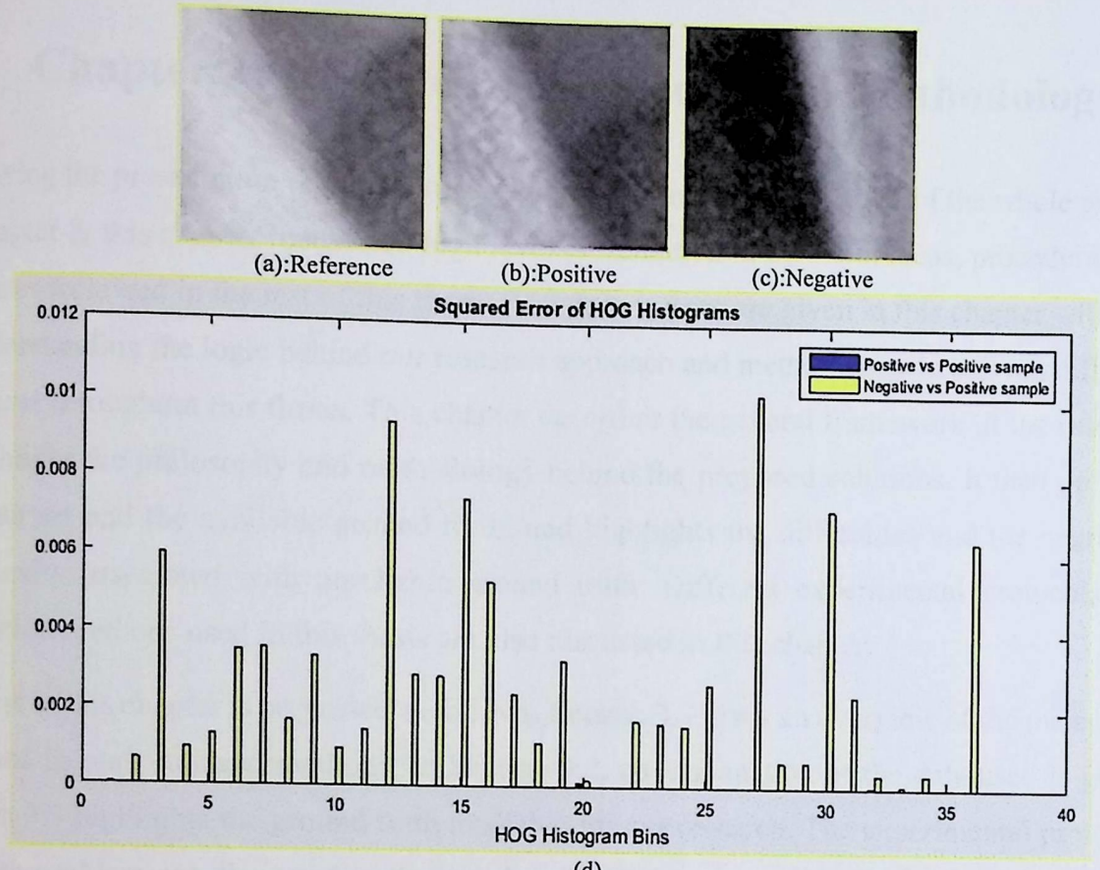


Figure 2.18 HOG features comparisons for positive versus positive and positive versus negative patch samples were sampled from mouse skull X-ray image.
(a) Reference positive patch sample extracted from width ROI of mouse X-ray head image; (b) Positive patch sample extracted from width ROI of mouse X-ray head image; (c) Negative patch sample extracted from width ROI of mouse X-ray head image; (d) Squaredd error for reference and positive patches (blue bar). Squared error for reference and negative patches (yellow bar). The results show that the squared error for the two positive samples is much less than the squared error for the positive and negative samples

2.2.5 Summary

In the previous sections, a review of biological and technical backgrounds relevant to the research carried out in the thesis was given. The first part of this chapter develops a basic understanding of the genome projects. It also highlights the needs of MGP to sort of efficient and effective automatic image annotations to deal with information extraction from huge number of generated images. The described image enhancements techniques will help to reduce the noises and enhance the contrast of the X-ray image before applying any segmentation or feature extraction methods, and therefore of an important value to this research. In addition, these techniques may play important roles in post-processing of any segmentation results before we can extract a beneficial information from it. The image segmentation methods illustrated in this chapter will be exploited in different manners to segment and focus the OOIs in our solutions. Furthermore, the explained machine learning methods along as with the described types of features will also add a significant value for finding required landmarks and classifying different objects.

Chapter 3: Research Framework and Methodology

Following the presentation of the biological and the technical backgrounds of the whole project in Chapter 2, this chapter focuses on high level presentation of the main ideas, procedures and solutions followed in the rest of this thesis. The descriptions are given in this chapter will help in understanding the logic behind our research approach and methodology in solving different problems throughout this thesis. This chapter describes the general framework of the research to highlight the philosophy and methodology behind the proposed solutions. It then presents the data set and the available ground truth, and highlights the difficulties and the degree of complexity associated with unreliable ground truth. Different experimental protocols and evaluation methods used in this thesis are also discussed in this chapter.

The rest of this chapter is organised as follows. Section 3.1 gives an overview of the main steps followed in our proposed methods. In Section 3.2, an explanation of the data used is given. Section 3.3 highlights the ground truth available for our research. The experimental protocols for each problem are discussed in Section 3.4. Finally, a summary of the chapter is given in Section 3.5.

3.1 Overview of the Research Framework

The main goal of this research is to produce fully automated solutions to annotate the X-ray images of mice and to reveal useful information from such annotations. The main stages of the generic process towards automated annotations are summarized below and outlined in Figure 3.1:

1. Automatic alignment (see Sections 4.3.1 and 5.3.1). The postures of the mouse body and the mouse head in the data set considered in this research are not consistent and involve different ranges of angles. To prepare those images consistently for other later operations, proposals are made to align the mouse head and body into a vertical position. This pre-processing step of alignment is not mandatory for the proposed solutions but it will reduce the complexity and make exploiting the body's symmetry easier.
2. ROIs segmentation/modelling (see Sections 4.3.2, 5.3.2 and 6.3.1). Since this research aims to annotate different parts of mouse skeletal structure, different ROI's around the head and body of the mouse containing the body parts should be correctly located and segmented before correct annotations are derived from the images. Depending on the

type of problem, either ROI segmentation based on global landmarks is done or the ROI is modelled with a suitable model (geometrical shape (i.e. circle) that fit on the skull of the mouse).

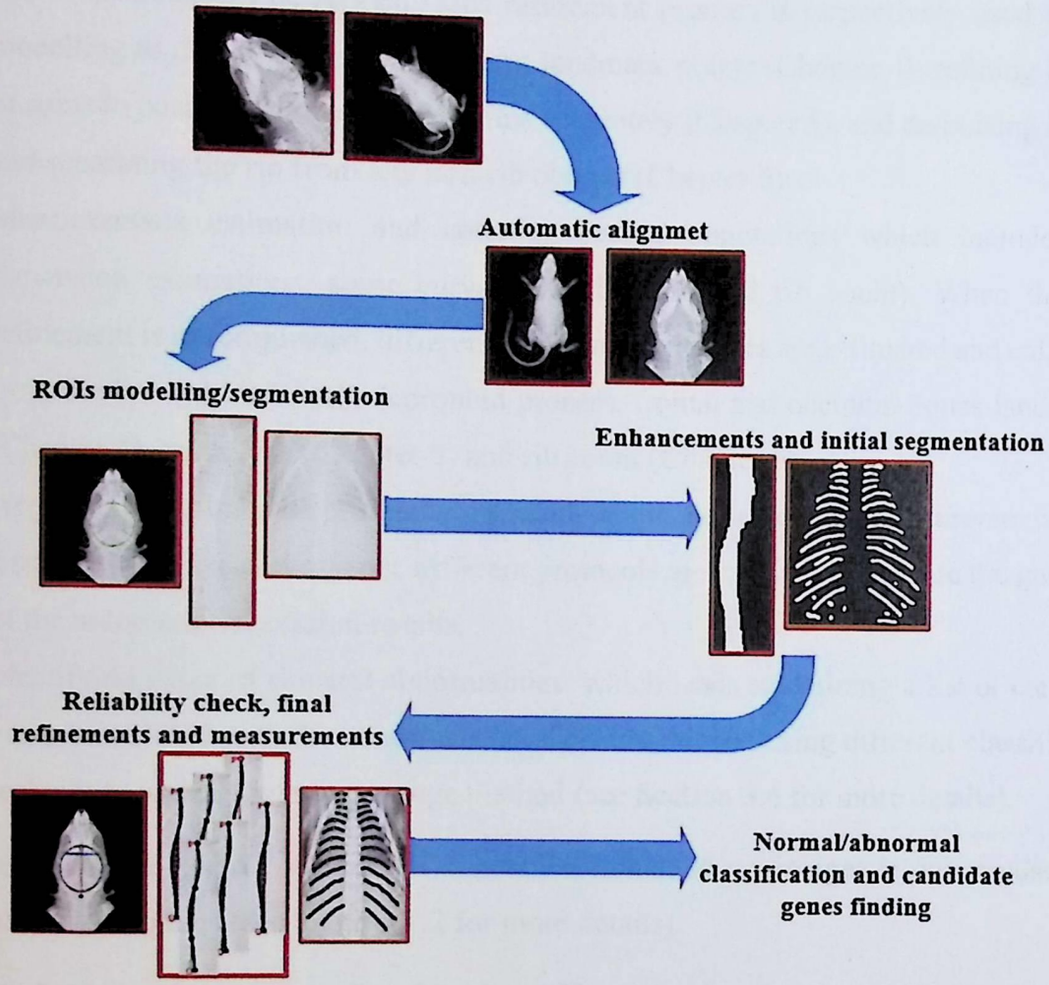


Figure 3.1 Research framework. The framework pipeline starts by automatically align the body and the head of the mouse X-ray images to the vertical position to enable modelling of the skull and ROI segmentation for rib and spine. After that the enhancements and initial segmentation operations are applied to extract the initial mask for object of interest. Next, the reliability check is done to either correct the model/mask or reject them. After that the final refinement is applied to enable measurements estimation and then classify the images to normal and abnormal and finding candidate genes.

3. Enhancements and initial segmentation. Because of the low contrast of X-ray images, the segmented ROI needs some types of image enhancements before proceeding to the initial segmentation. In the segmentation step, the initial mask of the OOI is segmented from its background.
4. Reliability measures. Depending on the type of problem, different criteria are considered from the model (Chapter 4), initial binary mask (Chapter 5) and ROI (Chapter 6) to decide if the model or the mask is reliable or not. This quality control

could even be exploited in different way i.e. to refine the mask or the model when its reliability is not satisfactory.

5. Final refinement. Once the quality control test is satisfied, the segmented objects and model are ready to be refined. This refinement process is respectively used in skull modelling to detect and refine different landmark points (Chapter 4), refining borders of spine to position the central spine line accurately (Chapter 5), and de-noising rib area and separating the rib from any non-rib objects (Chapter 6).
6. Measurements estimation and counting (actual annotations which include brain dimension estimations, spine curvature estimation and rib count). When the final refinement is accomplished, different performance metrics are estimated and calculated from the located landmarks (coronoid process, frontal and occipital bones landmarks) (Chapter 4), spine line (Chapter 5) and rib count (Chapter 6).
7. Applying experimental protocols. Depending on the annotated measurements taken from different skeletal objects, different protocols are applied to evaluate the goodness of the automatic annotation results.
8. Identifying cases of skeletal abnormalities, which leads to defining a list of candidate genes associated with the abnormalities. This is achieved using different classification techniques and the reference range method (see Section 3.4 for more details).

In this research, the overall problem of annotation mouse X-ray images is decomposed into three main sub-problems (see Section 1.2 for more details).

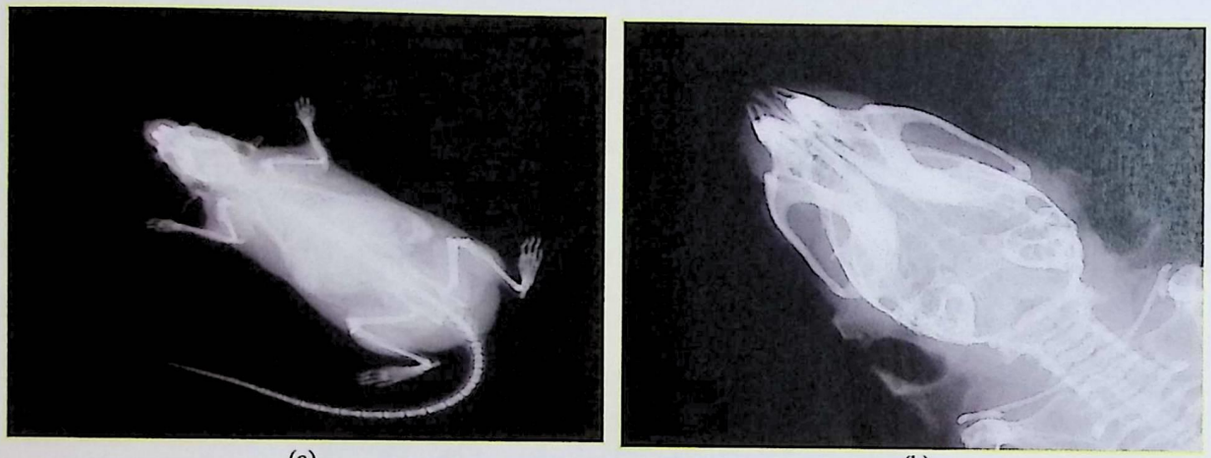
- Brain size estimation based on locating different landmark points in the X-ray images of the skull (Chapter 4).
- Spine segmentation and curvature quantification from whole body mouse X-ray images (Chapter 5).
- Segmenting and counting the ribs of the mouse from whole body mouse X-ray image (Chapter 6).

In addition to those three main problems, Chapter 7 will discuss some proposed solutions to segment and annotate different parts in the mouse body as work in progress.

3.2 Data set

The data set used in this research includes X-ray images of over 6800 mice where every mouse has 5 posture views as shown in Figure 2.1. Out of the 5 posture views, 2 X-ray images of different directions, i.e. the full body and the close-up head dorsoventral views as illustrated in

Figure 3.2, are selected in this research as they are the most suitable views for achieving the objectives of this research. The full body dorsoventral view represents the first option to estimate the curvature of the spine related to scoliosis ((Zhang et al., 2017), (Tang et al., 2008), (Kundu, Chakrabarti and Lenka, 2012), and (Sardjono et al., 2013)). This view is also more suitable than the full body lateral view to show the whole rib area (see Section 6.5.3 for more details). Furthermore, it is the only posture of the mouse that shows the pelvis. Due to its high resolution and ability to show the different ROI for different landmarks in the skull, the close-up head dorsoventral view was selected for brain size estimation object.



(a) (b)

Figure 3.2 Two mouse X-ray images views selected for different measurements extraction for the thesis.
(a)Full body dorsoventral view mouse X-ray image. This view will be used for spine curvature estimation, rib count and pelvis segmentation; (b) Close-up head view mouse X-ray image. This view will be used for brain size estimation

All images are greyscale with 256 levels of intensity and their sizes are 2368 × 2340 pixels. The data set is further divided into two sets. The first set includes 798 images of WT mice, providing the normal images of mouse skeleton. The second set contains 6,090 images of mutant mice obtained from 435 mutant lines (gene-knockout) where each mutant line has 14 images (7 males and 7 females). The scale for the conversion from pixel to millimetre for the mouse skull image adopted in this research is 0.012 mm for every pixel.

The data sets are provided by the Sanger Institute MGP (Sanger, 2013). The data are publicly available for research use and hence we follow terms and conditions about data use.

3.3 Ground Truth

In addition to the mice X-ray images, a ground truth file that contains the manual annotation for images was also provided by the Sanger Institute. The ground truth file contains a description for every mouse that determines its sex, barcode and also the genotype. The file also contains a description for the full phenotype that results from a different gene knockout.

This phenotype includes counting for different bones (i.e. ribs and vertebrae) and the status of some limbs and bones (i.e. skull, spine and pelvis) if they are normal or not.

Unfortunately, the available ground truth file does not include all the needed information for our research. For the brain size estimation, there is no recorded measurements for the mice in terms of width, length of the brain or positions of the essential landmarks. However, the Sanger institute offered us with 100 annotated images which show the manually estimated width and length of the mouse brain as shown in Figure 3.3. For the spine, the ground truth file gives an information whether the mouse has the scoliosis disease (Section 1.3) or not and also specifies the type of the scoliosis. This information is very helpful in our work, but the ground truth does not include all the measurements, for example, the degree of curvature of the spine. The last key piece of information for our research is the counting of right and left ribs. Such counting is available for every mouse in the data set, but it is not 100% accurate. The large size of the data set, the complexity of the rib count, and the low contrast of the rib area could be the reasons that cause some inaccuracy in the ground truth related to rib count.



Figure 3.3 X-ray image of mouse head with manual annotation for width and length of brain.

3.4 Experimental Protocols and Performance Evaluation Metrics

As the problems addressed in this thesis have different nature, a number of experimental techniques and evaluation metrics are used to fit the requirements of the research. In the next subsections, descriptions about the evaluation techniques are given.

3.4.1 Identifying Genetic Markers using Reference Range (RR)

In our thesis, we use the RR methodology based on percentile (White et al., 2013) to identify the genes that are associated with changes in a brain size for the mice mutant group (Chapter 4). The RR methodology establishes the normal range of WT variations. The RR is applied

separately on the measurements obtained from applying the automated method on the WT (normal) images of male mice group and female mice group to consider the gender effect. The Lower Bound (LB) and the Upper Bound (UB) of the 95% confidence interval of the WT measurements are set at the 2.5 and 97.5 percentile values respectively as illustrated in Figure 3.4 for measurements of WT male mice group.

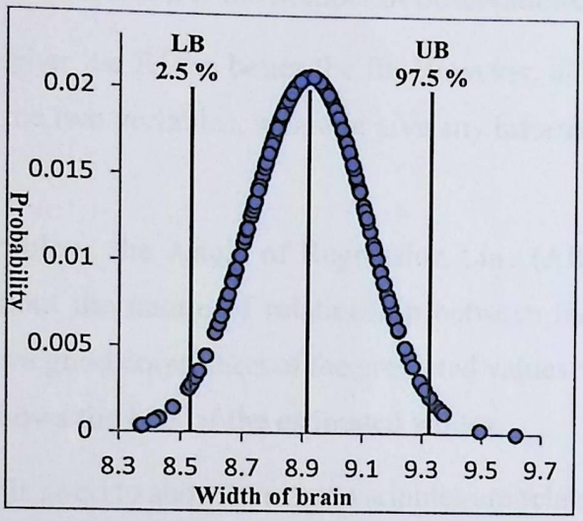


Figure 3.4 Normal distribution of width of brain measurements estimation for WT male mice. The lower and upper bounds of the RR are shown as black lines.

To decide if a given measurement for a mutant mouse is hit (i.e. it is outside the RR, which means the genes may be linked to the abnormalities), we calculate the so-called hit ratio of that mutant line. The hit ratio represents the percentage of images that falls outside the RR out of the total number of images in the mutant line (i.e. how many animals out of 7 per mutant line per gender in our case?). For example, the hit rate of a mutant line is 100% if all the 7 images of that line fall outside the RR. In this research, we set the hit rate threshold at 60% i.e. to consider a mutant line as a true hit, at least 5 out of 7 mice must fall outside the RR (White et al., 2013). It can be argued that increasing the hit rate may result in more false negative cases whereas decreasing it may lead to more false positive case. Therefore, it has been decided that 60% is a good threshold to maintain the balance between false positive and false negative cases.

3.4.2 Comparing Manual versus Automatic Measurements

To examine the correlation and the closeness between the automatic and manual measures, three different metrics are used.

Regression line

One measure to describe the relationship between the actual and prediction measures is through fitting a regression line. We can use the R^2 (coefficient of determination) statistical measure which can tell how the data is close to the fitted regression line to evaluate the correlation

between the actual and predicted measurements as illustrated in Figure 3.5. The R^2 can be calculated using the following formula

$$R^2 = \left(\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}\right)^2 \quad (3.1)$$

where x and y are two variables and n is the number of observations.

Roughly speaking, the higher the R^2 the better the fit. However, although the R^2 can explain the relationship between the two variables, it cannot give any information about the bias of the predicted variable.

To address the above problem, the Angle of Regression Line (ARL) is used. ARL gives a beneficial information about the nature of relationship between the two variables. Usually, around 45° ARL indicates a good correlation of the predicted values to the actual ones. Smaller or larger than 45° ARL shows the bias of the estimated values.

However, the R^2 measure is good to show how two variables are related but not how much their values are close or agreed (Bland and Altman, 1986). For this reason, we have decided to use the Bland Altman analysis method as explained in next section for the closeness between different measurements.

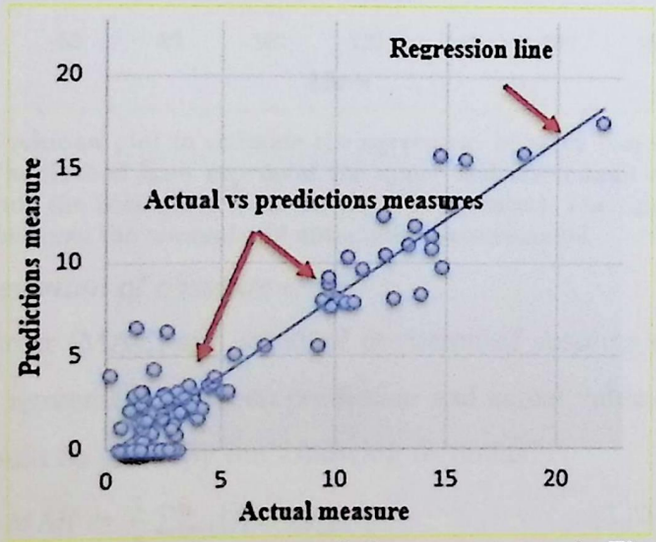


Figure 3.5 Scatter plot for actual versus predctions measurements. The regression line of the plot is shown as blue line.

Bland-Altman analysis

The Bland Altman (BA) analysis (Bland and Altman, 1986) is used to estimate the agreement between the results of the manual and automatic methods. The BA analysis is done as follow:

1. Calculate the mean and differences between the manual and automatic measurements
2. Plot the difference between the measurements against their mean.

3. Calculating the bias (mean of differences ($\overline{d}$)) and the standard deviation (σ) of the differences between the two sets of measurements.
4. Calculate the lower and upper Limit of Agreement (LoA) as follow:
   - Upper LoA = $\bar{d} + 1.96\sigma$ (3.2)
   - Lower LoA = $\bar{d}$ - 1.96σ (3.3)

For the normally distributed differences, 95% of differences lie between the upper and lower LoAs. In general, when the automatic measurements are close to the manual ones, the bias, upper and lower LoAs should be close to the zero. Figure 3.6 illustrate the BA analysis for manual versus automatic measurements for spine landmarks detection.

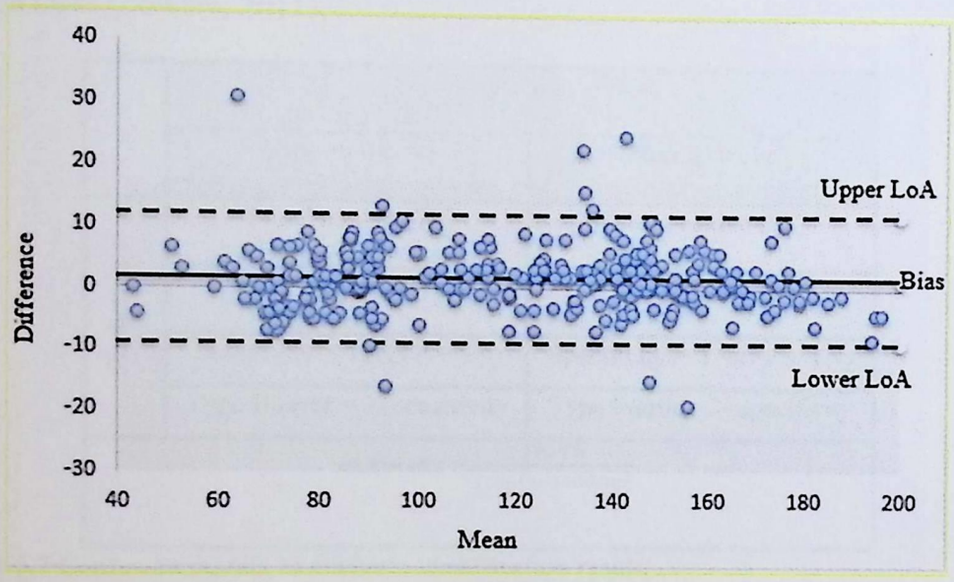


Figure 3.6 Bland Altman plot to estimate the agreement between manual and automatic measurements. The dashed lines represent the upper and lower limit of agreement. The solid line represents the bias of the data (mean of differences). The figure demonstrate a good agreement between the manual and automatic measurements.

Mean and standard deviation of absolute error

The Mean Absolute Error (MAE) and standard deviation of absolute error (σ) are used as a metric to estimate the agreement between prediction and actual values. MAE and σ for an n observations values could be given by the following formulas.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (3.4)$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((f_i - y_i) - MAE\right)^2} \quad (3.5)$$

where, f_i is the prediction value and y_i is the ground truth value.

All the above mentioned statistical measures will be used to evaluate the differences between the ground truth measurements and the automatically estimated measurements for the proposed methods in Chapters 4 and 5.

3.4.3 Statistical Measures of Classification Performance

There are different measures to evaluate the performance of the binary classification test. Sensitivity and specificity are from the well-known measures in this direction. In this thesis we use the confusion matrix as shown in Figure 3.7 to evaluate the classification results for spine classification to normal and abnormal (Chapter 5) and rib count (Chapter 6). The columns of the matrix represent the instances in the actual class while the rows represent the instances in the predicted class (Powers, 2011).

		Actual	
		Abnormal cases	Normal cases
Prediction		True Positive (TP)	False Positive (FP)
		False Negative (FN)	True Negative (TN)
		Sensitivity = TP/(TP + FN)	Specificity = TN/(FP + TN)
		Type II error = 1- sensitivity	Type I error = 1- specificity
Accuracy = $\frac{TP+TN}{Total\ population}$			

Figure 3.7 Confusion matrix to evaluate classification result.

The explanations of the terminology used in the confusion matrix are as follow.

True Positive (TP): the number of actual positive cases (abnormal cases) that classified as positive by the proposed method.

False Negative (FN): the number of actual positive cases (abnormal cases) that misclassified as negative (normal cases) by the proposed method.

True Negative (TN): the number of actual negative cases that classified as negative by the proposed method.

False Negative (FN): the number of actual negative cases that misclassified as positive cases by the proposed method.

Sensitivity (true positive rate): the percentage of positive cases that are correctly defined as positive (TP/ (TP + FN)).

Specificity (true negative rate): the percentage of negative cases and correctly defined as negative (TN/ (FP + TN)).

Type I error (false negative rate). The percentage of negative cases that are mistakenly defined as positive (FN/ (TP + FN)) (i.e. normal spine classified as abnormal).

Type II error (false negative rate). The percentage of positive cases that are mistakenly defined as negative FN/ (TP + FN) (i.e. abnormal spine classified as normal).

Accuracy. The percentage of all the cases that correctly defined either as normal or abnormal cases (Accuracy = $\frac{TP+TN}{Total\ population}$).

3.4.4 Image Classification into either Normal or Abnormal

A classification technique is needed to identify the abnormal spines from the normal ones using the obtained automatic measurements (Chapter 5). There are two main reasons which motivate the use of classification. Firstly, the manual estimation of the spine curvature is not consistent and hence there is no absolute reference to evaluate the automated measures. The second reason is related to the fact that the automatic measures may be imprecise by their nature but at least consistent. For these reasons, the goodness of the automatic measure should all depends on how accurate it is in separating the normal from abnormal spine (Scoliosis).

In general, the classification techniques tend to classify any new incoming data point to a suitable class based on the training set of data points of known classes. One of the simple and effective classification techniques is the k-Nearest-Neighbour (kNN) (Cover and Hart, 1967). kNN is a non-parametric technique that classifies the new observation based on its closeness (i.e. Euclidean distance) to the one or more (k value) of the neighbour data points of the classes in the training set. From the characteristics that makes the kNN favourite option is its simple implementation.

To validate the kNN model and assess its response to an independent data set, we propose to use leave-one-out cross validation strategy. In this strategy, the classification experiment is repeated N times for a data set with N examples. In each experiment, N-1 examples are used for training while the other example is used for testing (Stone, 1974). This type of validation allows all examples in the data set to be used for training and testing. Our proposed method in Chapter 5 will use kNN (k = 5) with leave-one-out cross validation to perform and validate the classification of data set of 525 images as shown in the flow chart in Figure 3.8.

The value of k was selected empirically after trying a range of value between 1 and 13. After obtaining the specific measure from the X-ray images, the training and testing of the proposed measurement are done by selection a same size random sample of WT and mutant mice images. After that, the leave-one-out cross validation is followed. This process is repeated 50 times with different random sample every time and the average of all rounds result is taken as the final result. The end results expressed in terms of the sensitivity and specificity.

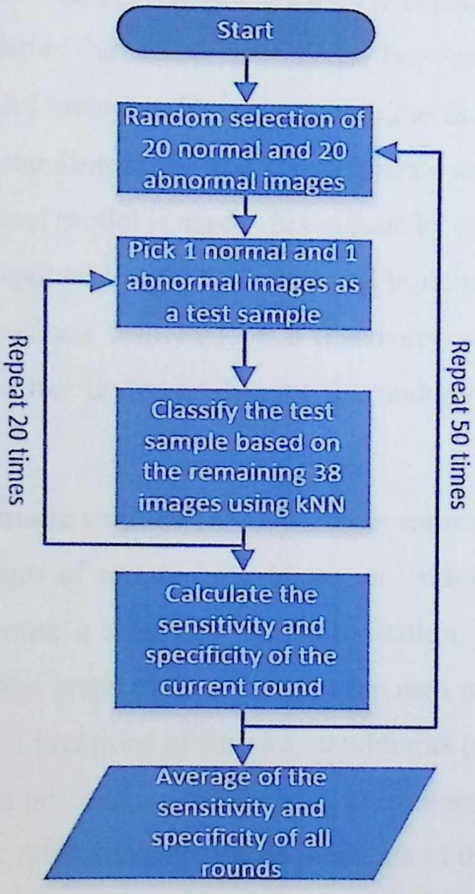


Figure 3.8 Spine image classification to either normal or abnormal using kNN classifier and cross validation. The process starts by selecting one normal and abnormal samples from random normal and abnormal sets. The kNN then classify those samples before repeating the selection of new samples. The entire process is repeated 50 times to produce the average sensitivity and specificity.

3.5 Summary

In this chapter, an overview for the research framework of automatic annotation of mouse X-ray images was presented. The framework includes the general ideas and thoughts to be detailed in the next chapters. We also described the data sets and the ground truth that our proposed solutions will be tested upon. Finally, we described the main experimental and evaluation techniques that will be used to evaluate the performance of the proposed solutions. In the next chapter, the first part of our work which includes the brain size estimation from the mouse X-ray images will be presented.

Chapter 4: Automatic Detection of Landmark Points to Estimate the Brain's Dimensions of the Mouse in X-ray Images

The estimation of brain dimensions is an important metric because it could be closely related to some diseases or abnormalities that affect the human. For example, there is a reduction in the brain size of the child with Down syndrome comparing to the normal child (Becker et al., 1991). To enable better understanding of the underlying genetic causes of brain's abnormalities and diseases in human, an animal model is used. This is done by estimating brain size of mutant mice and then discover the responsible genes of abnormal brain sizes. Using an animal model to estimate the brain's dimensions followed by a discovery of genes responsible for any changes in the size enables better understanding of the underlying genetic cause of brain's abnormalities and diseases.

This chapter presents an automatic solution for an accurate estimation of the brain dimensions from dorsoventral X-ray images of mouse head based on landmarks detection. The solution starts by alignment of the mouse's head to a vertical position. This preparation step is not mandatory, but it makes the later steps easier. The solution next models the skull of the mouse as a circle to identify the initial locations of four key landmarks (width and length landmarks) on the skull by applying image processing and analysis techniques. At this stage, the reliability of the skull model is checked. After that, the initial locations of the key landmarks are refined till the best final locations are found using a mixture of intensity-based image processing techniques and machine learning techniques. From the refined landmarks, brain dimensions are estimated in terms of relevant width and length of the skull. The estimated dimensions are then used to identify the possible genes that are responsible for changes in the brain size of the mutant mice.

The solution is designed for high-throughput processing of images, capable of handling thousands of images efficiently. The proposal recognises the degree of difficulty in fully automatic estimation of the brain dimensions from 2D X-ray images, and a quality control criterion within the solution itself is introduced to reject images in extremely rare cases where the brain dimensions cannot be correctly estimated, ensuring a good level of accuracy in the automatic estimation results.

Experimental results on the large set of X-ray head images of 798 WT and 6,090 mutant mice have demonstrated the effectiveness of the proposed solution in separating images of abnormal brain dimensions from those of normal ones. The results analysis using the reference range method led to identifying 7 genes associated with abnormal brain dimensions (Mcph1, Cdk5rap2 and Aspm genes are already known to influence the brain size while Arl4d and 'NovelGeneX' are possible novel genes). The results were then positively confirmed by the domain experts.

The main technical contribution of this chapter is the development of an efficient solution to detect and locate accurate width and length key landmarks on the skull to estimate the brain size of mice from X-ray images. The brain measurements are mainly obtained through hybridization between image processing techniques (colour intensity based approach) and machine learning based techniques in applying multi-level reliability checks.

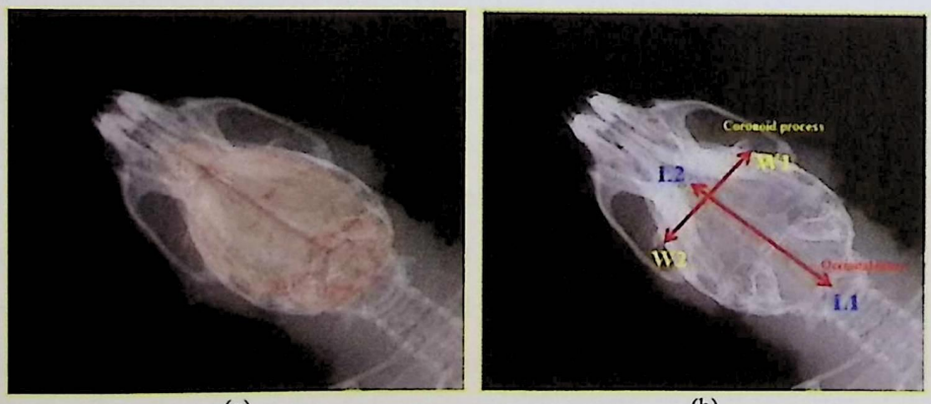
The rest of this chapter is organised as follows. In Section 4.1, the problem statement and the manual process of the brain dimensions' estimation are described. Section 4.2 reviews the literature relevant to brain segmentation and landmarks detection in human brain and head in medical images of different modalities. Section 4.3 presents the proposed automatic method and explains the adopted image processing and analysis techniques in details. Section 4.4 reports the experimental results and analysis to identify the genes associated with changes in the mouse brain size. This is followed by a discussion about the effectiveness and the efficiency of proposed solution in Section 4.5. Finally, the chapter to be summarised in Section 4.6.

4.1 Problem Statement

The recent development of high-throughput production of targeted, genetically-altered mouse lines has enabled a systematic phenotypic screening and gene discovery. Identifying genes correlated with changes in brain size is a key step towards understanding the genetic basis of neurological and developmental brain syndromes. Accurate identification of any changes in brain size is of a significant biological and clinical relevance. For example, mutations in the Mcph1 gene in human have been associated with Microcephaly, a disease defined as a head circumference is more than three standard deviations below the age-related mean (Jackson et al., 2002), which is often associated with incomplete brain development. In contrast, Macrocephaly has been reported in patients with mutations in the Kptn gene, which is also characterised by neurological disorders (Baple et al., 2014). It is likely that there are further, unidentified genes in which mutations may also cause neurological or behavioural symptoms, allied to alter brain dimensions. Therefore, the identification of genes associated with

dimensions' alterations may provide insight into both specific and generalised abnormalities in the brain.

Whilst X-ray based technologies such as Micro-Computed Tomography (MicroCT) can directly image the brain, imaging mice by this modality is limited both by the lower throughput of the current equipment and the potential higher X-ray doses and contrast agents disturbing phenotypes later in the phenotyping pipeline (Willekens et al., 2010). This is because influencing of some biological pathways in the animal which leads to possible alteration of the experimental outcomes (Willekens et al., 2010). In contrast, direct measurement of brain dimensions is very difficult from typical X-ray machines used for general skeletal analysis due to the composition of the brain making it radiolucent to X-rays. Therefore, human annotation is often used to estimate brain dimensions from the skull outline of mouse X-ray images as illustrated in Figure 4.1. Figure 4.1(a) shows the actual location of the brain in the mouse's head. The trained analyst firstly locates four key landmark points, W_1, W_2, L_1 and L_2 on the mouse skull image as shown in Figure 4.1(b). The dimensions of the brain, i.e. brain width and length, are then estimated as the distance between W_1 and W_2 and the distance between L_1 and L_2 landmark points respectively. The targeted position for W_1 or W_2 is the cross point between the skull edge line and the root of the coronoid process of the mandible of the mouse (see Section 1.2). For the length landmarks, L_1 is located at the base of the occipital bone while L_2 is located on the upper edge of the frontal spine as shown in Figure 4.1(b) (see Section 1.2 for more details).

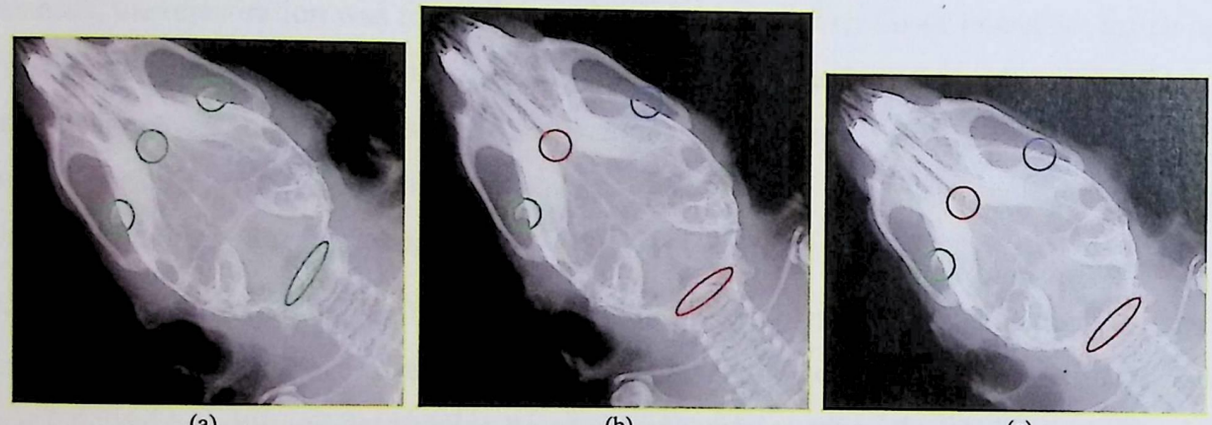


(a) (b)

Figure 4.1 Manual Annotation of mouse brain dimension landmark points (images are provided by Sanger Institute).
(a) Fusion of the MRI of a brain onto the X-ray image to show the location of brain of mouse in relative to the skull; (b) Width and length landmark points manually annotated on the mouse's head X-ray image

Automatic size estimation of a 3D object such as the brain from a 2D X-ray image is a technically challenging task due to difficulties in locating related and reliable landmarks within the image caused by colour intensity overlapping of different tissues such as bone, brain,

vessels, etc. In addition to general difficulties related to X-ray images, some specific challenges are associated with the specific width and length landmark points. The variations of size and layout of the coronoid processes across different images add an extra level of difficulty for W_1 and W_2 determination. The problem gets even worse if the coronoid processes are absent in some cases. Although L_1 landmark point is located on the base of bone, the edge of this bone is hardly identifiable in many cases. The L_2 landmark is the most difficult landmark to recognize in a mouse's head X-ray image even by human eyes. Furthermore, the bones and tissues of the neighbouring areas for both L_1 and L_2 landmarks are highly variant and inconsistent among different X-ray images of mouse's head. Individual differences of size of the mouse's heads from image to image adds another level of difficulty for fully automatic solutions.



(a) (b) (c)

Figure 4.2 Examples of good and poor ROI of the four width and length key landmarks. (a) Identifiable coronoid processes (green circles). Consistent and identifiable areas for L1 (green ellipse) and L2 (green circles) landmarks; (b) Absent right coronoid process (blue circle). Inconsistent areas for L1 (red ellipse) and L2 (red circles) landmarks. The edges of the occipital and frontal bones are not clear; (c) Very small right coronoid process (blue circle) and inconsistent areas for L1 (red ellipse) and L2 (red circles) landmarks. The edges of the occipital and frontal bones are not clear

Figure 4.2 shows some good and bad examples of the area of the four key landmarks in mouse head X-ray image. All the aforementioned challenges do not only impose difficulties for machine-based solutions but even for expert analysts with whom *inter*- and *intra*-analyst variations in manual measurement results could happen. This leads to difficulties in obtaining a solid ground truth on manual measurements to objectively assist the evaluation of the automatic solutions. Given the volume of images that are generated by the high-throughput screening, an automated solution for annotation and triaging of images for human analysis will not only reduce costs in time and human resources, but also eliminates the likelihood of human errors and inconsistency in manual measurements. Interestingly, several papers highlighted landmarks in mouse images for various skull attributes (Kawakami and Yamamura, 2008) and

(Ólafsdóttir et al., 2007), demonstrating the need for reliable analytics in model organisms such as mice.

4.2 Literature Review

In the literature, there are several publications to label and segment the brain of the mouse manually, semi-automatic or in a full automatic manner. Most of these proposals were dedicated to 3D MRI images because the ability of such imaging modality to recognize the soft tissues of the brain. Some manual and semi-automatic methods were used for labelling mouse brains ((Badea et al., 2007), (Redwine et al., 2003) and (Zhang et al., 2010)). However, the manual and semi-automatic methods are considered as time-consuming and error prone. Segmentation of the mouse brain from 3D MRI images into different stuctures was propsed in several proposals ((Ali et al., 2005), (Lee et al., 2009) and (Sharief et al., 2008)). For those methods, the registiration was firstly done using an atlas based on image intensities matching. After that, the segmentation of the brain from the image was achieved based on wraped label information about the structures. The important note about those methods is the use of single atlas to perform image registiration which has a drawbacks of bias existence that might affect the segmentation (Nie and Shen, 2013).

Nie and Shen (Nie and Shen, 2013) proposed skull stripping and ROI-labelling of mouse brain from 3D MRI images. Their method started by the registration of multi-atlas to the image. This was followed by applying deformable based segmentation to refine the labelling results using SVM classifiers that were trained on local surface patches. Finally, the atlases with better similarity to the image were allowed to contribute more in label fusions.

However, all the above proposals are designed for 3D MRI images which show high discrimination between the soft tissues of the brain and other tissues. On the contrast, the 2D X-ray image of mouse head cannot show the soft tissues of the brain (see Section 4.1) and so it is infeasible to segment the whole brain from it.

On the other hand, several existing proposals have discussed brain segmentation, dimension estimation and landmarks detection from human medical images. In general, there are several image processing applications on human brain images. It can help segment the image of brain tissue from non-brain tissues which usually required for MRI images registration. The automatic measurements obtained from the brain can also help radiologists by either comparing new patient data to existing one or for surgery planning.

In (Michalak et al., 2002), an algorithm was suggested to automatically measure the width and length of human brain in CT scan images. The algorithm involves locating the inner edge of the skull, aligning it to be vertically upright, and then estimating the width and length of the skull using the rectangle that contains the skull boundary. One drawback of this approach is that the dimensions of the brain were estimated by the direct dimensions of the skull without using any anatomical landmark points, and hence can be imprecise. Besides, CT images of brain offer more discrimination between brain and skull than full head 2D X-ray images, and hence it is a problem of lesser difficulty to solve, making the method less likely to be applied to the problem at hand.

The landmark point's detection for cephalometric analysis in human skull images attracted great deal of attention recently. Cephalometric in skull X-ray images is usually used in orthodontic diagnosis and helping in surgery planning (Miloro et al., 2014). To analyse cephalometric, it is required to localize several landmarks on the skull as well as around soft tissues.

A Pulse Coupled Neural Network (PCNN) was used in (Innes et al., 2002) to determine a region that contains the landmark point for cephalometric in X-ray images of skull. PCNN is a segmentation and edge detection method that can produce a binary image output for a target region. The proposal applied an averaging filter to reduce Gaussian noises before using PCNN to detect the required edge. However, the proposed PCNN technique needs a fine tuning for the used parameter before it can produce useful results.

It has been shown in the literature that using machine learning approaches in cephalometric landmarks detection presented promising results. (Vandaele et al., 2015), (Mirzaalian and Hamarneh, 2014) , (Chu et al., 2014) and (Lindner and Cootes, 2015) presented proposals to detect landmark points which are used for cephalometric analysis in human skull X-ray images. All mentioned proposals depended mainly on machine learning approaches using RF with voting scheme to detect the landmarks. There are some differences in applying the learning approach for every proposed method. The proposal in (Vandaele et al., 2015) localized 19 landmark points by using RF classification trees which were trained with a simple kind of features. For each landmark, a separate pixel classification model is trained to predict whether a given pixel in a new unseen image is a landmark or not. The features used to describe a pixel in this proposal were collected as a vector of simple raw intensity pixels' value around the landmarks from 6 different resolutions. To reduce the search area for landmark in the testing stage and to efficiently select the negative pixels for training, they constrained their search area

to a narrow range around the target landmark under the belief that the spatial coordinates for a specific landmark is very close across all images. In the prediction stage, the collected pixels around supposed position of the landmark were fed to the classifier and the final landmark position was selected as a median position of all pixels that were predicted with high confidence as being the landmark.

Another work in cephalometric X-ray localization was done by Mirzaalian and Hamarneh (Mirzaalian and Hamarneh, 2014). In this proposal, a set of features including LBP-based features, x-y spatially coordinates of the points, blobness, tubularness, and Zernike were used to train the RF classifier to predict the likelihood of each landmark. After that, the graphical model that encodes the joint distribution of pairs of landmarks was learnt. Finally, they proposed to use pictorial structure algorithm to optimize the cost function that consisting computed appearance models and the graphical mode. The pictorial structure requires computing acyclic graph (minimum span tree) that connecting all the 19 landmarks.

Chu et al. (Chu et al., 2014) reported a method based on the idea of using two stages to detect an initial landmark and then refine it to the final one. In the first stage, they used RF regression based on HOG features to predict the initial landmarks. To minimize area of search in test stage, they proposed two levels coarse-to-fine strategy. In the first level, they performed the search in the whole of down sampled image while in the second level they focused the search only in the ROI which resulted from the first level but with original resolution now. Because each tree in RF produces a prediction, a noisy output around the image happened. To predict an accurate position out of these individual predictions, they proposed to consider each vote as a small Gaussian distribution and then used general probability aggregating algorithm to add them all to give a probability of being a landmark for every position of the image. The second stage was based on using the initial landmarks resulted from the stage 1 to form an initial shape and then refine it using shape composition model based on sparse representation method.

Both proposed works in (Mirzaalian and Hamarneh, 2014) and (Chu et al., 2014) relied on a learnt relationship between all the landmarks to refine the final location of the landmarks. Such a refinement could affect the accuracy of the final position of some landmarks because it tends to update the position of one landmark depending on the others. On the other hand, although such procedure may decrease the accuracy for some landmarks detection, it may increase the robustness of overall detection of all landmarks. Further discussion on accuracy vs reliability is provided in Section 4.5.4.

In (Lindner and Cootes, 2015) a RF regression voting in constrained local model frame work was used to detect cephalometric landmarks. Firstly, in the training stage, RF regression model was trained for each manually annotated landmark by using Haar features. Also, manual annotated landmarks were exploited to train a statistical shape model which represents a linear model of shape variations that describes the position of each landmark. In the detection stage, Hough forest was used to initialize an estimation of the position, orientation and scale of the object of landmark in the image. Hough forests are an adapted RFs to perform a generalized Hough transform efficiently (Gall et al., 2011). After that, by using an initial estimate of the object of landmarks, features from area of interest around estimated landmark were passed to RF regression model to select a best location for a landmark based on a voting scheme. The voting from all trees for all landmarks was accumulated ultimately in a 2D histogram. By using learned shape constraints, they aimed to combine the votes in all histograms to produce the forest prediction.

The most recent work in cephalometric landmarks detection was presented by Vasamsetti et al.(Vasamsetti et al., 2017). Their proposal was based on wavelet feature template matching between input and reference images to find the ROI for different landmarks. Prior knowledge was then used to locate the landmark.

It could be noticed from the above proposals that landmarks detection in human X-ray images is a challenging problem. Several existing proposals shared a main common idea of training RF classification or regression model with a specific kind of features to predict the landmarks in a prediction stage. However, the used features and the regularization methods are different from one proposal to another.

Arguably, the literature relevant to different landmarks detection and brain segmentation based methods on different human medical image modalities is full of interesting ideas such as the skull mask and RF that can be employed in our proposal. Although there is no existing literature for the problem in mouse X-ray images, some relevant ideas from the above literature have been adopted in our proposed solutions to be explained in the following section, bearing in mind that the mouse X-ray images have their own complexities as explained in Chapter 1.

4.3 Reliable Automatic Estimation for the Mouse Brains' Dimensions

The proposed method attempts to automate the estimation of the brain dimensions by locating the same four key landmark points as trained analysts do (see Section 4.1). Our proposed method is principally based on accurate modelling of the skull. This skull model will be used

for detecting initial positions of the landmarks. The initial positions are then refined using different approaches. The method consists of six key stages.

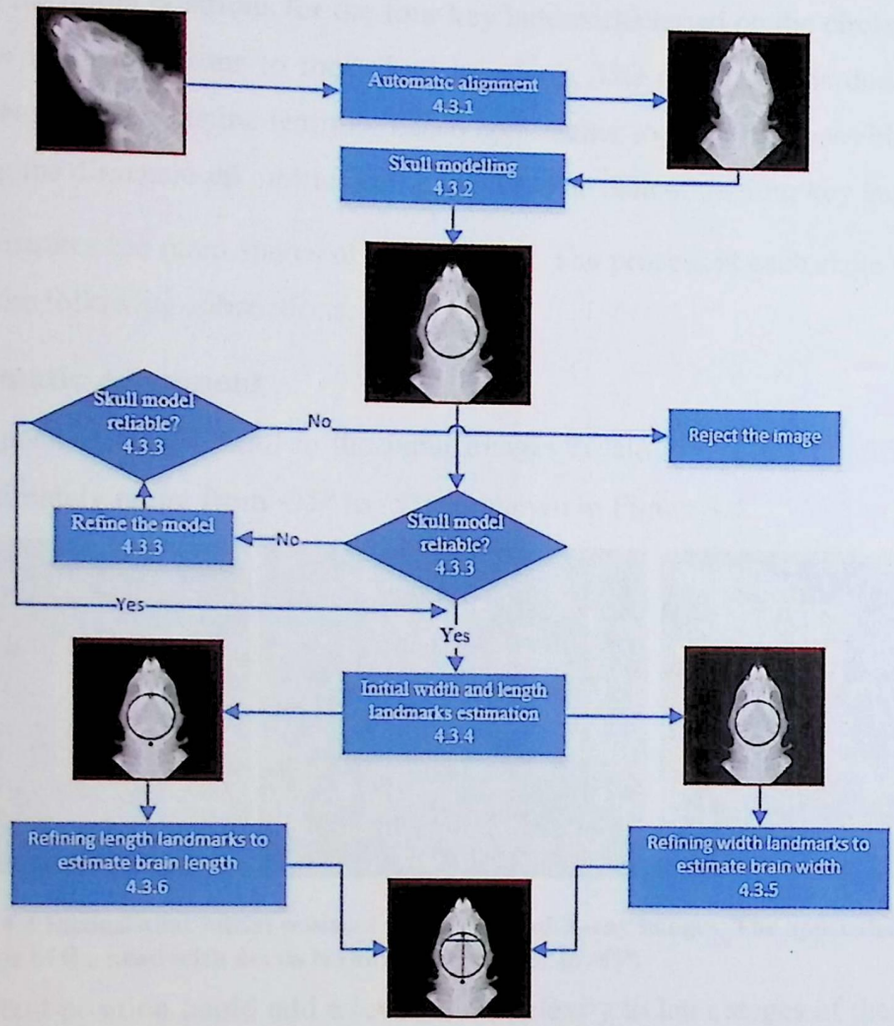


Figure 4.3 The main stages of the proposed method for mouse's brain dimensions estimation from mouse head X-ray image. The method starts by automatically aligning the mouse head to the vertical position. After that the skull is modelled using a circle. The reliability of the circle is assessed to correct or reject it. Next the circle position is used to estimate the four initial width and length key landmarks. Finally the initial locations of the key landmarks are refined to the final position.

1. Aligning the mouse head into a vertical upright position. The rotated head makes the later tasks of landmarks detection easier by exploiting the symmetry of the left and right parts of the head.
2. Modelling the skull by the best fit circle. The circle model was chosen due to its ability to approximate all the four initial landmark points.
3. Conducting a reliability check on the skull model to either accept the model as a reliable one, refining it to 'reliable standards', or -in rare cases- rejecting it and hence the image indicating that brain dimensions cannot be correctly estimated. This is to ensure that a skull model is good enough for automatic estimations and that the estimation results for the

accepted models are sufficiently accurate with a high degree of confidence. In other words, the algorithm's completeness is sacrificed for the algorithm's correctness.

4. Detecting the initial positions for the four key landmarks based on the circle model.
5. Refine the initial positions to their final locations. The refinement is done by using an intensity-based and machine learning based approaches to yield best possible results.
6. Measuring the dimensional metrics using the location data of the four key landmark points.

Figure 4.3 illustrates the main stages of the proposal. The process at each stage will be further explained in the following subsections.

4.3.1 Automatic Alignment

The original position of the skull in the input images is laid diagonally by different degrees which approximately range from -35° to -55° as shown in Figure 4.4.



(a) (b) (c)

Figure 4.4 Inconsistent initial position of mouse head X-ray images. The approximation of the angle of the head with x-axis is ranging from -35° to -55°.

This inconsistent position could add a level of complexity to later stages of the solution. The alignment of the head to upright vertical position would arguably help in reducing such complexity. To do this, the angle for rotating the skull from the current placement position to the vertical position needs to be determined. The first step of the alignment is to obtain a binary mask of the skull using Maximum Entropy threshold (Section 2.2.2). The threshold value keeps the high intensity bones in the skull and neck as a single object and eliminates other tissues and bones (Figure 4.5(a)). After that, two binary morphological operations (Section 2.2.1) are applied to close the gaps in the mask (Figure 4.5(b)) and remove objects of sizes less than the skull mask (i.e. <30000 pixels) (Figure 4.5(c)). The used size value to remove unwanted objects is selected to be big enough to remove any expected unwanted objects (red circles in Figure 4.5(b)) and at the same time much smaller than the expected size of the skull mask.

The upper leftmost and the lower leftmost points of the mask and the line between them are then determined.

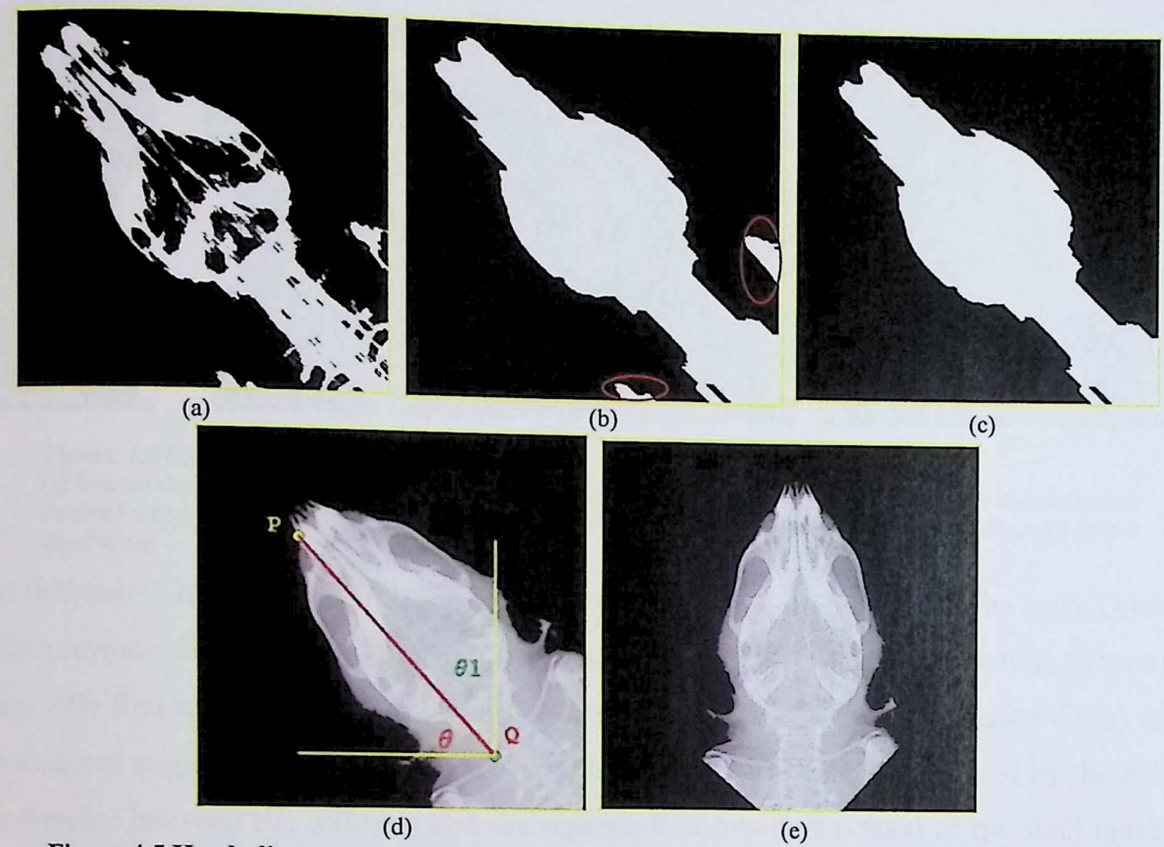


Figure 4.5 Head alignment process.
(a) Initial skull mask obtained from maximum entropy method; (b) Skull mask after applying binary closing morphology operation to fill the gaps. The red circles show the unwanted objects; (c) Skull mask after removing unwanted objects based on their sizes; (d) Detected alignment points and line superimposed on the input image ; (e) Final aligned image after rotating it by obtained rotation angle (θ1)

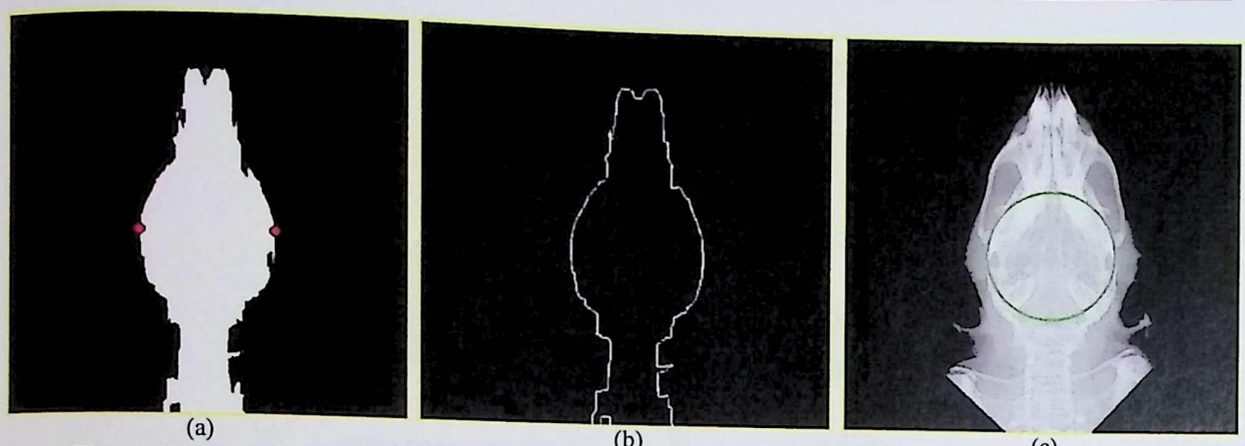
Figure 4.5(d) illustrates the two points P (P_x, P_y) and Q (Q_x, Q_y) and the red line between them superimposed over the input image. The angle of the rotation is then obtained as:

$$\theta_1 = 90° - tan^{-1}\left[\frac{(P_y - Q_y)}{(P_x - Q_x)}\right] \qquad (4.1)$$

Finally, the image is rotated by the obtained angle of rotation (θ_1) to the vertically upright position as shown in Figure 4.5(e).

4.3.2 Skull Modelling

To automatically estimate the width and the length of the brain, it is needed to locate the four key landmark points on the skull as explained in Section 4.1. It could be seen from the Figure 4.1(b) that three landmark points located approximately on the edge of the skull while the other is located above the hypothetical width line. This arrangement motivated the idea toward modelling the skull with a regular geometrical shape that could initiate a relationship with an approximation location of each of the four landmark points. A circle appears a good candidate for this purpose because of its closeness to the skull shape from one side as shown in Figure 4.1(b) and the possibility to formulate a consistent relation from the circle to all the landmark points.



(a) (b) (c)

Figure 4.6 Skull modelling using Hough transform.
(a) Rotated skull mask and the right and left most points of the mask (red points); (b) Edge of the skull mask obtained using Canny edge operator; (c) Detected skull model (green circle) using HT superimposed on the input image

The first step to locate the best fit circle is to prepare a suitable skull mask. The mask obtained in the previous alignment stage, as shown in Figure 4.5(c), is found a close outline to the skull shape. We first rotate the skull mask to the vertical position as shown in Figure 4.6(a) using the obtained angle of rotation. The initial radius of the circle is then estimated by the half of the distance between the leftmost and the rightmost points (red points) of the skull mask, as shown in Figure 4.6(a). To prepare a skull model, HT (Section 2.2.2) is applied to find the best fit circle of the skull mask. The reason behind this choice is the powerful ability of HT to detect the circle with known parameters from a binary image. The input parameters to HT function are the outer border of the skull mask (Figure 4.6(b)) which obtained by using Canny edge operator (Canny, 1986) and the candidate radiuses. Due to the importance of the best fit circle, as a step of caution, the initial radius and six other radiuses (±3 pixels of initial radius) are fed to the HT. The detected circle of the skull is as shown in Figure 4.6(c) after superimposing it on the aligned input image.

4.3.3 Reliability Check on Skull Model and Skull Model Refinement

Detected skull model is considered essential for the identification of the four initial landmark points later. For this reason, reliability measures are designed to check how reliable and accurate the position and the size of the circle are. Figure 4.7 shows the main steps of the reliability measurement and refining process of the skull model. The key features to assess the reliability of the skull model are the circle position and the length of the radius. The location of the circle is said to be accurate if the topmost point of the circle is located just above the coronoid processes and the rightmost and the leftmost points of the circle should locate on or very close to the skull edge.

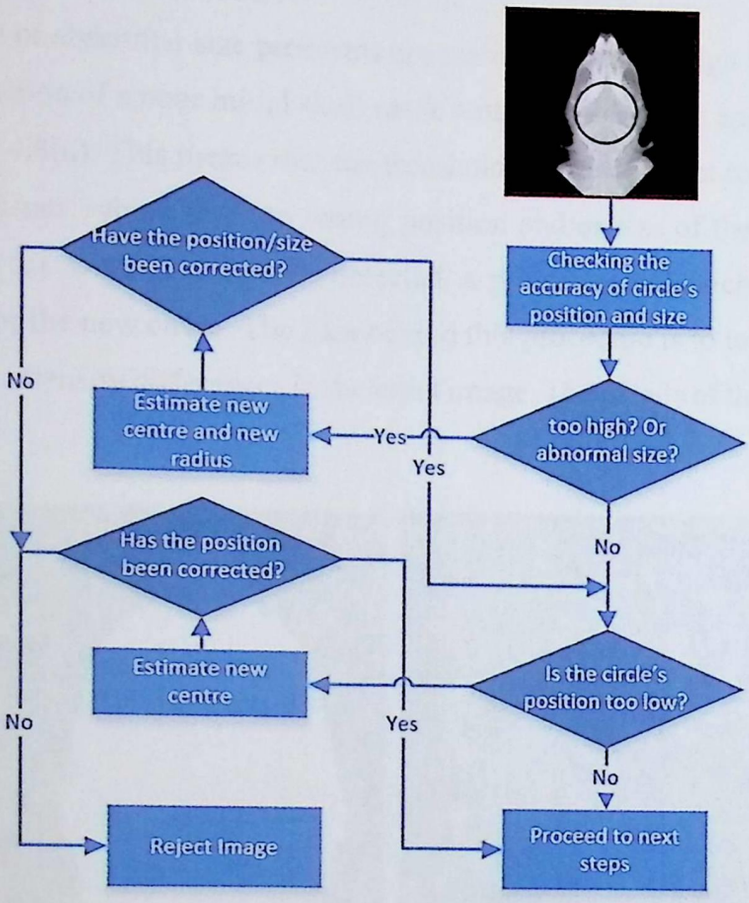


Figure 4.7 Main steps of skull model reliability check and refinement process. The process starts by checking the size and the position of the circle. If the circle is high or in abnormal size, the refinement process is done to correct it. After that, the circle's position is checked and refined if it is too low.

In particular, the accuracy is assessed by the vertical distance between the topmost point of the aligned skull mask (yellow point in Figure 4.8(c)), and the topmost point of the circle (blue point in Figure 4.8(c)). By a close examination of a random samples of images, the acceptable range for the vertical distance is estimated between 520 and 720 pixels. More about sensitivity of this choice is discussed in Section 4.5.2. Furthermore, the acceptable range for the radius length of the circle is estimated between 400 and 500 pixels. If the measured vertical distance and radius are within the above ranges, the reliability conditions are met for the circle; otherwise, the refinement process is launched.

Depending on the measured distance, if the vertical distance is below the lower bound of the acceptable range, the circle is deemed too high. If the vertical distance is bigger than the upper bound of the vertical range, the circle is considered being too low. Depending on the measured radius, if the radius length below the lower or above the upper bound of the acceptable range, the circle is deemed abnormal size. In following subsection, we consider two solutions to deal with abnormal location or size of the circle.

Circle being too high or abnormal size

The circle-too-high or abnormal size problems are mainly due to the high image intensity that results in the production of a poor initial skull mask with unwanted area and tissues segmented as shown in Figure 4.8(a). This means that the threshold value could not recognise the bone of skull from other tissues, which leads to wrong position and/or size of the circle as shown in Figures 4.8(b) and (c). Once such case is detected, a procedure is launched to estimate new centre and radius for the new circle. The idea behind this procedure is to locate the edge of the skull exploiting the intensity differences in the input image. The details of the proposed solution is as follow.

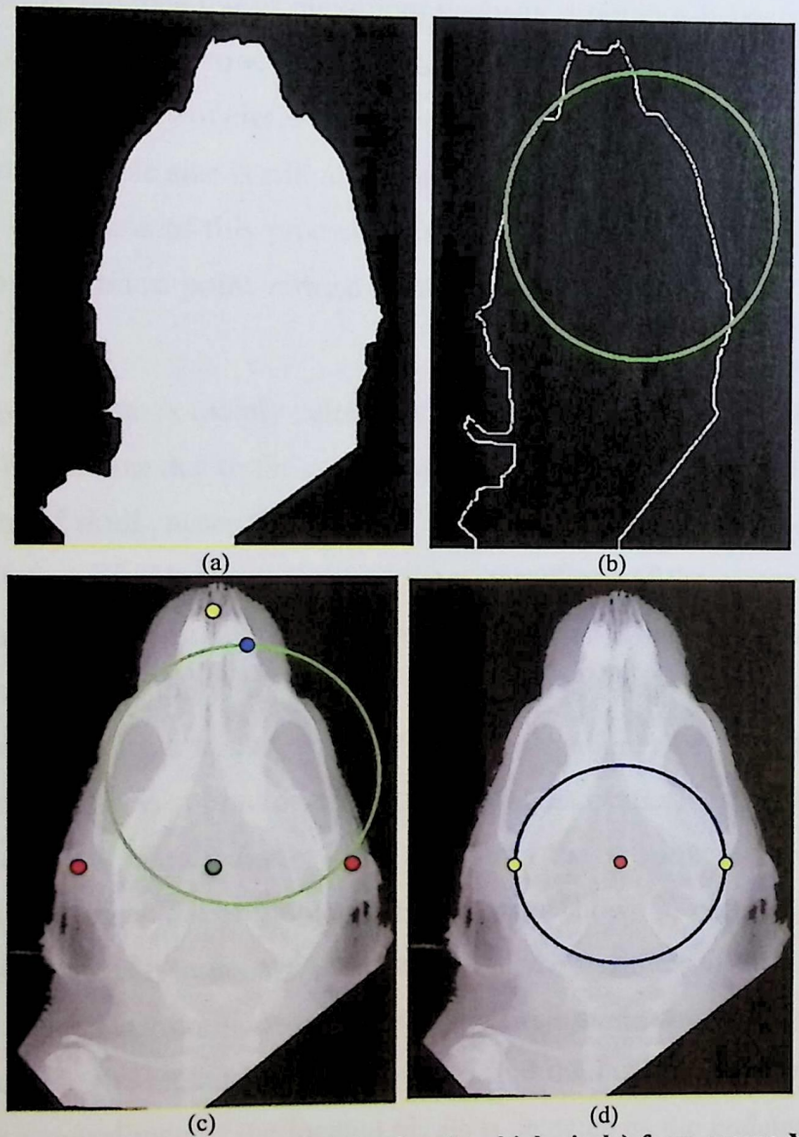


Figure 4.8 Inaccurate circle position detection (too-high circle) from mouse X-ray image and its refinement.
(a) Unsuitable skull mask with unwanted tissues segmented using maximum entropy method; (b) Border of the skull mask obtained by using Canny detector and the inaccurate detected circle (green circle) using HT; (c) Inaccurate circle position detected using HT (green circle), topmost point of the skull mask (yellow point) and right and left most points of the skull mask (red points); (d) Refined circle position to a more accurate position (blue circle)

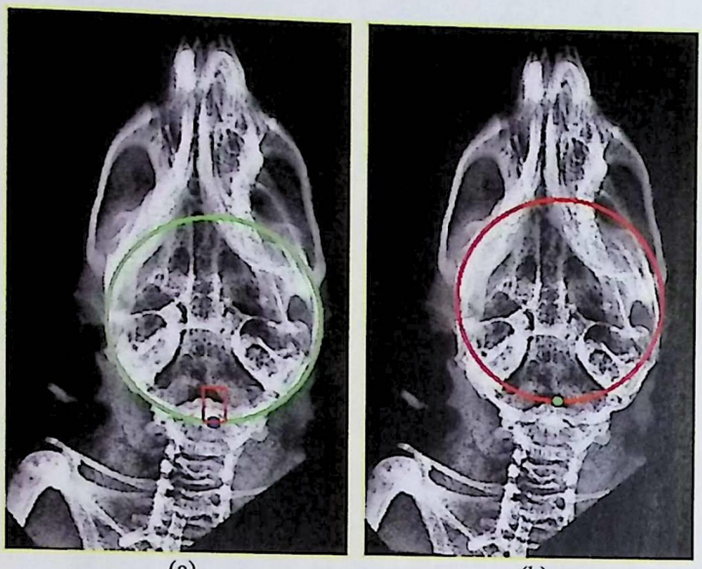
1. To estimate a new centre of the circle, the x-coordinate is selected as the x-coordinate of the topmost point of the skull mask (yellow point in Figure 4.8(c)) while the y-coordinate for the new centre is taken as the y-coordinate of the rightmost point of the skull mask (red point in Figure 4.8(c)). The estimated new centre is shown by green point in Figure 4.8(c).
2. To estimate a better radius, we exploit the fact that an accurately positioned circle should be located on the border of the skull. To find the border of the skull, we propose to analyse the intensity values of the pixels along the x-axis which pass through the new centre located in (1). The x-coordinates of the right and left borders landmarks could be detected as the location of maximum intensity difference between any consecutive pair of pixels along this row. The detected points and circle are shown in Figure 4.8(d).
3. Checking the accuracy of circle's position and radius length. If the circle position is still too high or the circle size is still abnormal, the image is rejected.

One of possible limitations of this procedure is the inaccurate estimation of the radius as a result of inaccurate rightmost point vertical location (red point in Figure 4.8(c)).

Circle being too low

The circle-too-low problem is mainly caused by the less circular abnormal shape of the initial skull mask, which is in turn due to the actual shape of the skull rather than the detected mask. The abnormal shaped skull causes the centre of the detected circle to be lower than the centre of the skull. However, usually the radius and the x-coordinate of the centre of the low circle are accurate as shown in Figure 4.9(a). Therefore, the solution proposed for tackling the circle-too-high problem does not apply. To address this problem, a simple adjustment of the circle placement is done by lifting it to new place. By our observations, the good place of the circle could be achieved if its lowest point locates at the low intensity area located above the occipital bone. To increase the discrimination of this area from the surrounding areas, the CLAHE (Section 2.2.1) is applied on the original aligned image as shown in Figure 4.9(a). To locate a point on the lowest intensity area, we propose to analyse the intensity of the pixels from a rectangle located above the current lowest point of the circle as shown in red rectangle in Figure 4.9(a). Each column in the rectangle is inspected to find the lowest intensity pixel. After that the median of the y-coordinate of the located pixels is selected as the updated y-coordinate of the lowest point of the original circle while the x-coordinate is set as the same x-coordinate for the original point. Finally, the circle position is refined to the new position as shown in Figure

4.9(b). The circle position is checked after the refinement process and if it still too low, the image is rejected due to the lack of reliability in the circle's position.



(a) (b)

Figure 4.9 Inaccurate circle position detection using HT (too-low circle) from mouse X-ray image and its refinement

(a) Initial inaccurate position of the circle (low circle) ,the lowest point of the circle (blue point) and the rectangle search area (red rectangle) above the lowest point of the circle; (b) Refined position of the circle (red circle) and the lowest point of the circle (green point)

Any rejected image due to abnormal size, too-high, or too-low circle reasons is marked as unreliable and hence excluded from any further analysis. In reality, it means that the rejected images may have to be examined by trained analysts. An example of rejected image is shown in Figure 4.10 where the initial detected circle (green circle) is too high and abnormal size because of the poor skull mask production which leads to inaccurate radius estimation (red points). However, because of the inaccurate vertical location of the right and left most points of the skull mask (yellow points in Figure 4.10), the corrected circle (blue circle in Figure 4.10) stills too low and with abnormal size and hence the image is rejected.

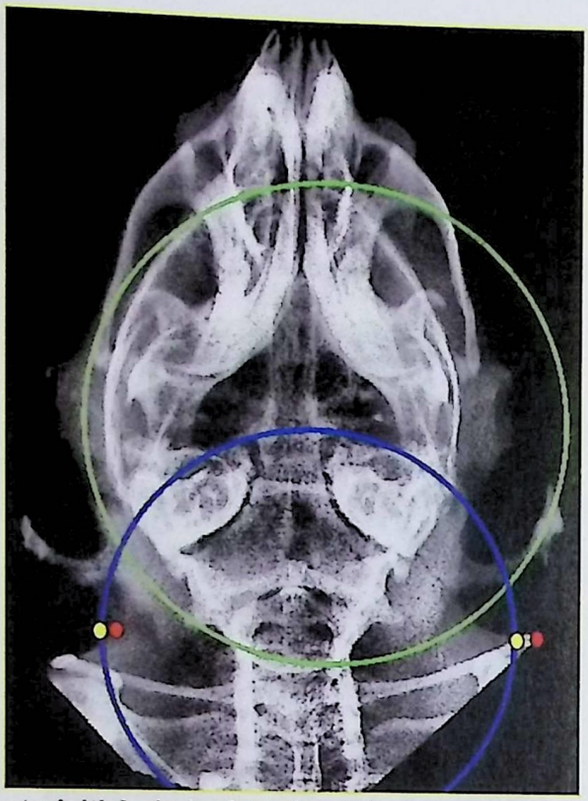


Figure 4.10 Inaccurate initial circle detection (big green circle) because of inaccurate estimation of the radius (red points) and inaccurate corrected circle (blue circle) because of inaccurate y-position of the red points which leads to low circle.

4.3.4 Initial Detection of Width and Length Landmarks

In this section, the procedures for estimating the initial locations for the landmarks based on the circular skull model are presented. Setting initial locations to the right places is very important for locating the final refined positions accurately. The initial landmark locations also helps later in determining small ROI which enables efficient searching for the final landmark positions.

Initial brain width landmarks detection

From the skull model, two observations that influence the estimation of the initial locations for W_1 and W_2 are made. The first observation is that the detected circle, when it is superimposed on the aligned image, should pass through the edge of the skull close to the coronoid processes. The second observation is that the coronoid processes in the aligned image have almost the same or very close y-coordinates. Based on these two observations, the two initial seed points are found on the circumference of the circle using the following formula to obtain x and y coordinates for every point:

$$x = r \times cos(t) + x_c \quad (4.2)$$

$$y = r \times sin(t) + y_c \quad (4.3)$$

where r is the radius of the circle, x_c and y_c are the x and y coordinates of the centre of the circle, and t is an estimated angle. To locate the initial seed points, the angle t is set to 45° for the right side and 135° for the left side respectively. These specific angles are selected from the observation of the position of the coronoid processes relative to the centre of the circle. The initial width landmarks detected points are shown in blue in Figure 4.11.

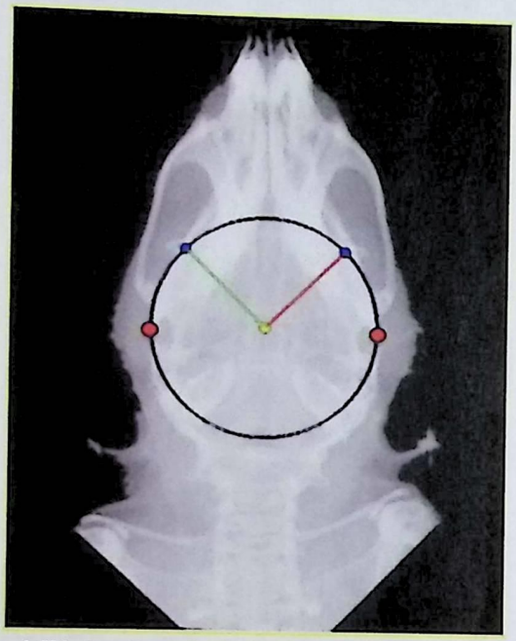


Figure 4.11 Initial W_1 and W_2 landmarks (blue points) located on the circumference of the circle using an angle of 45° for right point and angle of 135° for left point. The right and left most points of the skull mask (red points).

Initial brain length landmarks detection

Both lower length landmark L_1 and upper length landmark L_2 are estimated from different areas in terms of structure similarity and symmetry. The proposed solutions for the length landmarks L_1 and L_2 will be discussed separately next.

Detection of the initial position for L_1

The ideal L_1 landmark point should be located on the lower edge of the occipital bone as shown in Figure 4.1(b). The lowest point of the circular skull model (developed in Section 4.3.2) sits either above, below or on the bone itself which makes it unsuitable to be directly considered as the initial location for L_1. A consistent initial location for L_1 must be chosen with care.

To locate the ideal initial location of L_1, we follow a procedure similar to the one explained in Section 4.3.3 but with some modifications. A check is made on the pixel intensity value of the lowest point of the circle (yellow point in Figure 4.12). If its intensity value is very low, then the point must be located on the dark area above the bone and hence the search in the rectangle (see Section 4.3.3) is done downward for the first pixel of high intensity. Otherwise, the point

must be located somewhere on the bone and hence the search in the rectangle is directed upward for the first pixel of low intensity. This procedure resulted in locating the upper edge of the occipital bone as shown by green point in Figure 4.12 which is the desirable initial location for L_1.

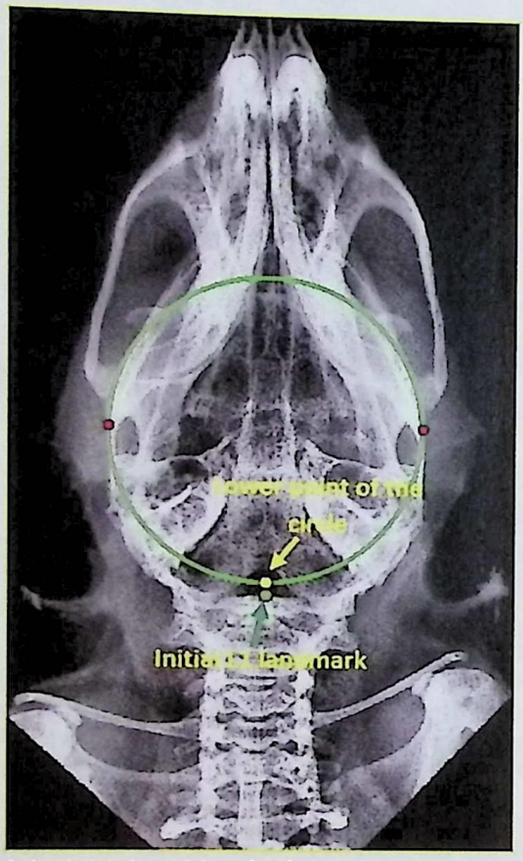
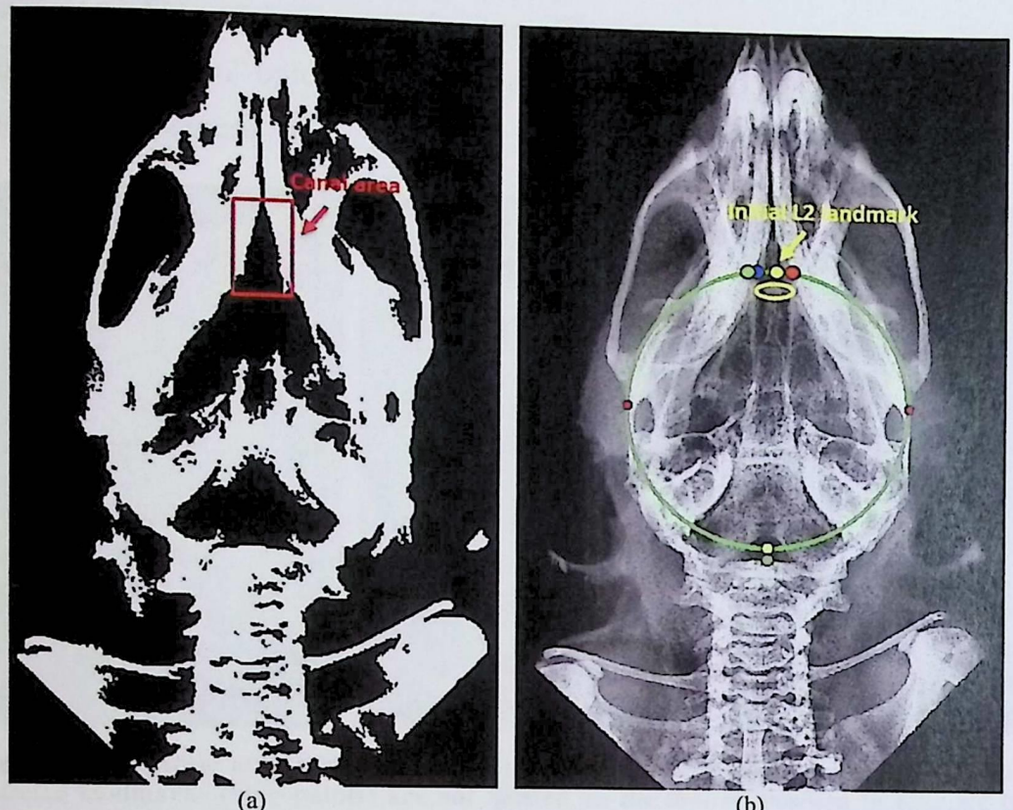


Figure 4.12 Detection of the initial position for lower length landmark (L_1). Lower point of the circle (yellow point) and the detected upper edge of the occipital bone (green point).

Detection of the initial position for L_2

The upper landmark point L_2 is not very easy to locate even for human annotators. It can be observed that the upper edge of the frontal spine is always located near the uppermost point of the detected circle and has (to some extent) higher intensity than its upper and lower neighbouring areas in the enhanced image as shown by the yellow ellipse in Figure 4.13(b). Even the topmost point of the circle is very good choice for the initial L_2, but in some cases its x-coordinate located to the left or right of the middle point above the L_2 as shown by blue point in Figure 4.13(b). This misplacement might lead to some inaccuracy in later stage. To find more accurate x-coordinate for initial L_2, we propose first to find a binary mask image which shows the black canal between the two incisors (red rectangle in Figure 4.13(a) by using Otsu method (Section 2.2.2). From the location of top of the circle point on this mask, we can determine the right and left edge of the canal as the first white pixels to the left and to the right. Finally, the x-coordinate of initial L_2 is taken as the half of the width distance between the right

and the left edge points. Figure 4.13(b) shows the right and the left edges of the canal and the initial L_2 as red, green and yellow points respectively.



(a) (b)

Figure 4.13 Detection of the initial position of the upper length landmark (L_2).
(a) Skull binary mask which shows the canal (red rectangle) near the upper length landmark point; (b) The upper edge of the frontal spine (yellow ellipse), right edge of the canal (red point), left edge of the canal (green point) and the initial L_2 landmark (yellow point)

4.3.5 Refining the Locations of the Brain Width Landmarks

In this section, the proposed techniques to refine the initial detected brain width landmarks are described. Because of the symmetry achieved from the alignment of the image and the similar structure of W_1 (right) and W_2 (left) regions, the same procedure will be adopted for both landmark points. The techniques presented here include intensity-based and machine learning based approaches. The main idea from presenting two different approaches is to maximize the effectiveness and accuracy of the final estimation by combining both approaches based on some reliability measures. After the two approaches are separately described, the proposed hybridization between the two approaches to estimate the width of the brain is presented.

Refining brain width landmark points using intensity-based approach

The initial positions for the landmark points are usually not at the desired coordinates. This is due to the variations in the location and the size of the detected skull model and the variations in the locations of coronoid processes among different images. Figure 4.14 shows several examples of initial width landmarks position (blue points). In all examples in Figure 4.14, the

initial positions of the width landmarks (blue points) are not located in the desired position of the width estimation (i.e. cross point between the skull edge line and the root of the coronoid process of the mandible of the mouse). So, for each coronoid process, a sequence of steps is taken in order to "home in" onto the desirable cross points between the skull edge line and the root of the coronoid process.

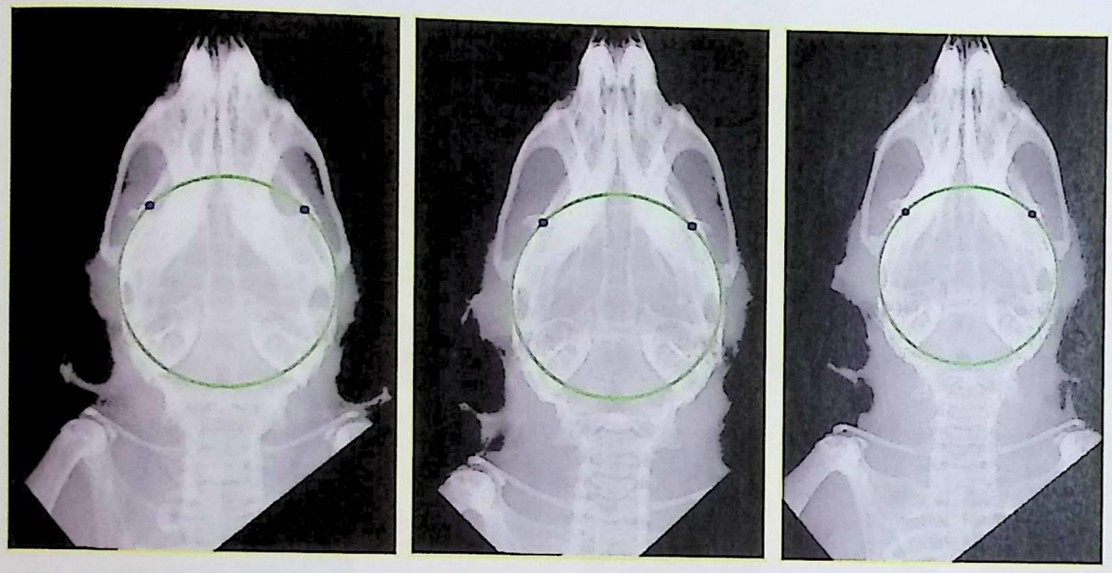


Figure 4.14 Different examples of initial width landmarks positions. In all cases, the initial detected landmarks (blue points) are not located on the desired position to estimate the width of the brain.

To refine the location of the point, the first step uses the coordinates of the initial seeding point to crop an approximated rectangle region of a certain size (e.g. 250 × 200 pixels) from an input image (red rectangle in Figure 4.15(a)) that should contain the coronoid process as shown in Figure 4.15(b). This step reduces significantly the searching area and helps getting the mask of coronoid process with few unwanted objects. The image in the cropped rectangle is first converted to a binary image as shown in Figure 4.15(c) by applying a suitable threshold value selected by the Maximum Entropy method (Section 2.2.2). Next, a mask of skull bone and the coronoid process (the first object) is obtained and other unwanted objects are eliminated (Figure 4.15(d)). The "cutting line" that separates the coronoid process from the skull is found by locating the upper and lower "cutting" points as the right topmost and the right lowermost points of the mask. An example of the "cutting line" and "cutting" points is shown by the blue line and red points in Figure 4.15(d). The angle of the "cutting line" is then calculated and a binary morphology opening operation (Section 2.2.1) is applied with the length and the angle equal to the length and the angle of the "cutting line" to produce a mask as shown in Figure 4.15(e).

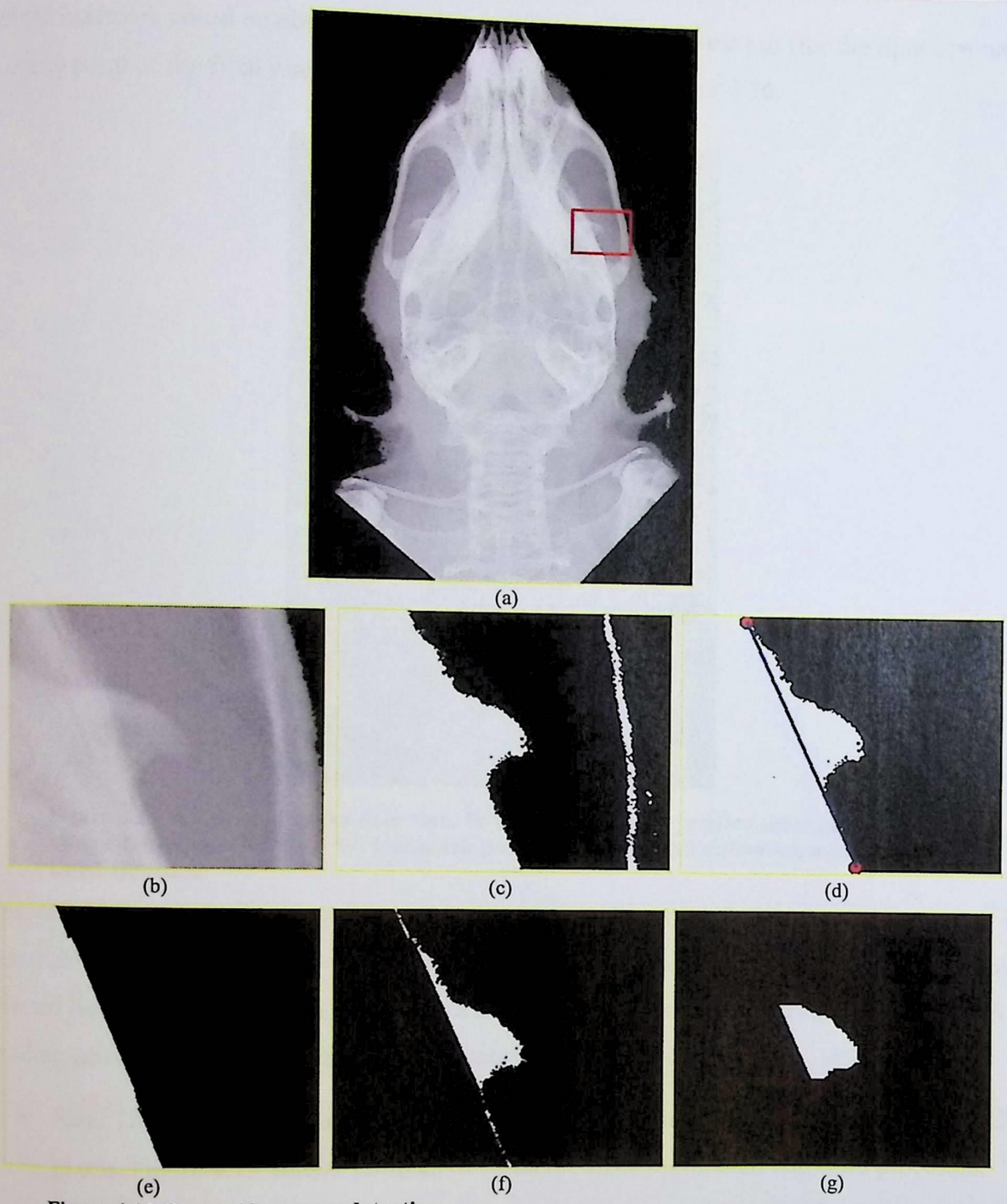


Figure 4.15 Coronoid process detection.
(a) Red rectangle automatically annotated around the right coronoid process using the initial detected landmark; (b) Cropped coronoid process ROI for the right width landmark; (c) Initial coronoid process mask obtained using maximum entropy method; (d) Upper and lower points of the mask and the line between them superimposed on initial coronoid process mask; (e) Bone mask after removing the coronoid process using binary morphology opening operation; (f) Initial coronoid process mask results from XOR operation between (e) and (d) ;(g) Final coronoid process mask after applying binary morphology opening operation on (f)

The initial mask of coronoid process (Figure 4.15(f)) is then obtained by applying bitwise XOR operation between the mask in Figure 4.15(e) and (d). Finally, binary morphology operations of opening and closing (Section 2.2.1) are applied to trim away small objects and obtain the final approximation of the mask as shown in Figure 4.15(g). The x and y-coordinates of the

refined landmark could be obtained as the coordinates of the lowest left (for the right coronoid process) point of the final mask as shown in green point in Figure 4.16.

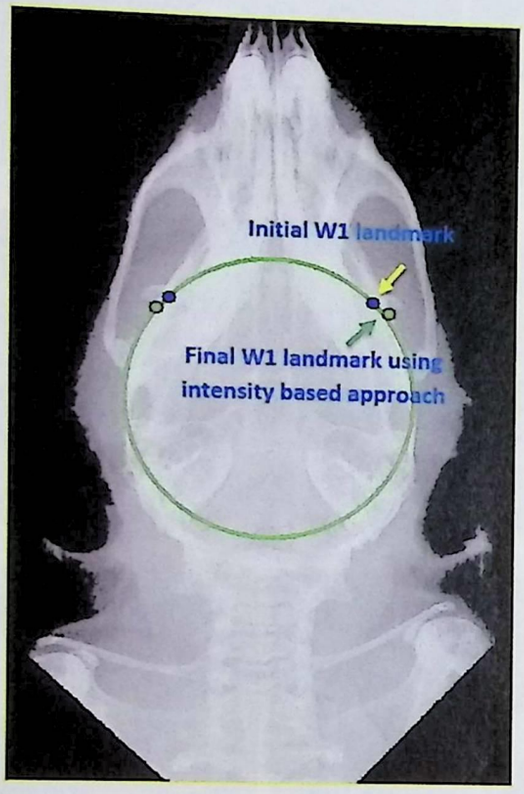


Figure 4.16 Width landmarks detection. Initial W_1 landmark position (blue point) located using circle model and final W_1 landmark position (green point) estimation using intensity based approach.

This procedure of segmenting coronoid process works only if the coronoid process is identifiable, prominent from the edge of skull i.e. the coronoid process mask can be reliably detected (see Figure 4.2(a)). By analysing a sample of images, two observations are made for deciding whether the detected coronoid process is acceptable or not:

- Size. The acceptable size of detected coronoid mask should be between 500 and 3750 pixels. Figure 4.17 shows samples of different sizes segmented coronoid processes.
- Orientation. The acceptable orientation of the detected coronoid mask should be between -35° and -75° for the right side and between 35° and 75° for the left side.

If any of the conditions mentioned above are not met, the detected mask is considered as unreliable mask and the coronoid process is marked as undetectable (see Section 4.5.1 for more details). If this is the case, the refinement will then be done by using the second approach, which is the machine learning based approach as explained in the next section.

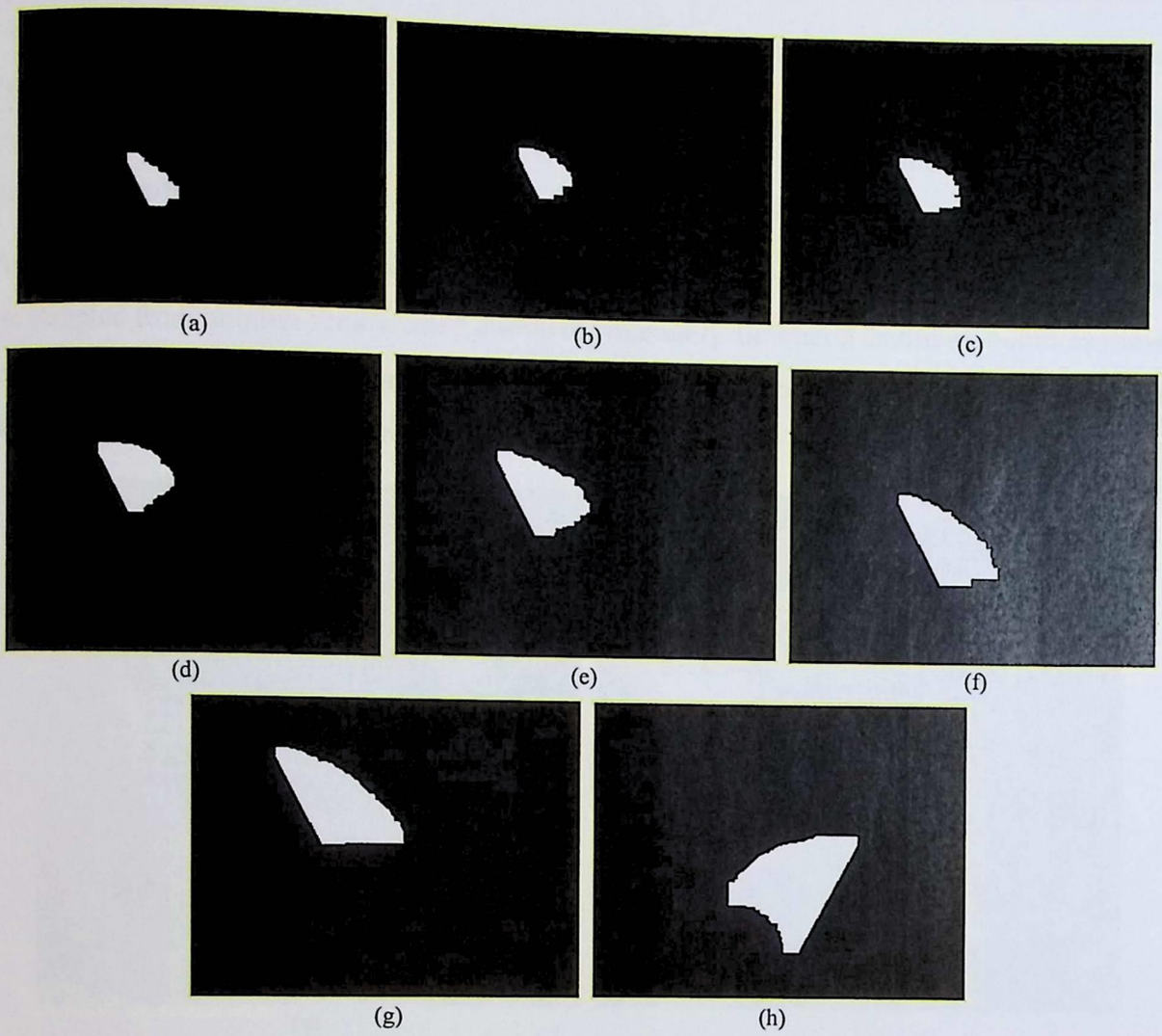


Figure 4.17 Samples of different sizes segmented coronoid processes. The accepted size for the coronoid process should range from 500 to 3750 pixels. The size of each coronoid process in the figure is as follow.
(a) 600 pixels; (b) 760 pixels; (c) 979 pixels; (d) 1492 pixels; (e) 1925 pixels; (f) 2274 pixels;(g) 2829 pixels; (h) 3740

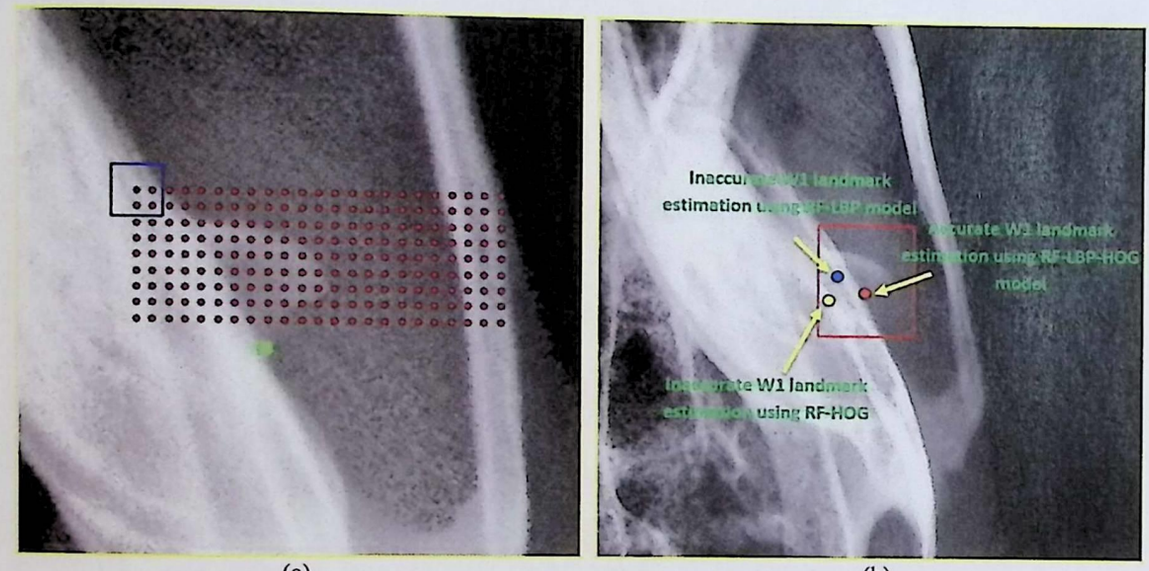
Refining brain width landmark points using machine learning based approach

The proposed width landmarks refinement in this section is based on using RF machine learning based approach (Section 2.2.3). The decision was made to use this approach because it showed good results in landmarks detection from medical X-ray images recently ((Lindner et al., 2013), (Mirzaalian and Hamarneh, 2014), (Chu et al., 2014) and (Lindner and Cootes, 2015)) The approach mainly goes through two stages, which are training the model of classification by exploiting manually annotated images and then predicting the landmark points in a new unseen image. The details of the two stages are explained as follows.

Two separate RF models for each of the two landmark points (W_1, W_2) using the enhanced image (Figure 4.9(a)), need to be trained. At the training stage, one pixel from each training image that corresponds to each landmark point is manually located. In practice, the manual landmark point is not selected down to pixel level: any points close to the "ideal" pixel e.g.

within a small window of size 6 × 6 pixels can be considered as the correct landmark as shown by a green window in a zoom-in image in Figure 4.18(a). With a high-resolution of input image and the small selected window size, the number of true positive samples is increased to a sufficient number without compromising the accuracy of the landmark point selection.

For the negative examples (non-landmark points), from every training image, a group of pixels are sampled from another area located above the manually annotated landmark points as shown in Figure 4.18(a). The positions and the dimensions of the negative sample window are determined to cover all possible types of tissues and structures expected and anticipated.



(a) (b)

Figure 4.18 Selected training pixels from W_1 ROI in training stage and W_1 landmark point location in prediction stage.
(a)Positive samples (green points) automatically annotated around the pixel selected by the manual annotator in training stage and negative samples (red points) automatically annotated from the area above the positive samples; (b)Automatically annotated testing window in prediction stage. Inaccurate estimation of W_1 landmark point (blue point) predicted by RF-LBP model, inaccurate estimation of W_1 landmark point (yellow point) predicted by RF-HOG and accurate estimation of W_1 landmark point (red point) predicted by RF-LBP-HOG model

A patch of size (33 × 33) pixels is then sampled around each positive and negative sample pixels. Figure 4.18(a) shows the patch around the first negative sample in blue. A discussion on the impact of patch size on landmarks detection accuracy is presented in Section 4.5.3.

Two types of texture features, LBP and HOG (Section 2.2.4) are then extracted from each patch. As we explained in Section 2.2.4, the RF models trained on LBP and HOG features produces lower error rate compared to other features (i.e. Gabor filter and GLCM). Figure 4.19 shows the estimation of classification error for RF-LBP, RF-HOG, RF-Gabor and RF-GLCM models (see Section 2.2.3 for details) of using different number of classification trees that trained using features collected around W_1 landmark. The LBP and HOG feature descriptors produce their features in a different manner. The LBP descriptor is powerful in recognizing the

change in contrast in a specific area. On the other hand, the HOG descriptor is focused on the orientations change in a set of pixels (Section 2.2.4).

Each model of the two (RF-LBP and RF-HOG) is trained to predict more likely position for the landmark point in a prediction stage by using all feature values (HOG or LBP) of all samples from all the training images in addition to the label of the pixel samples.

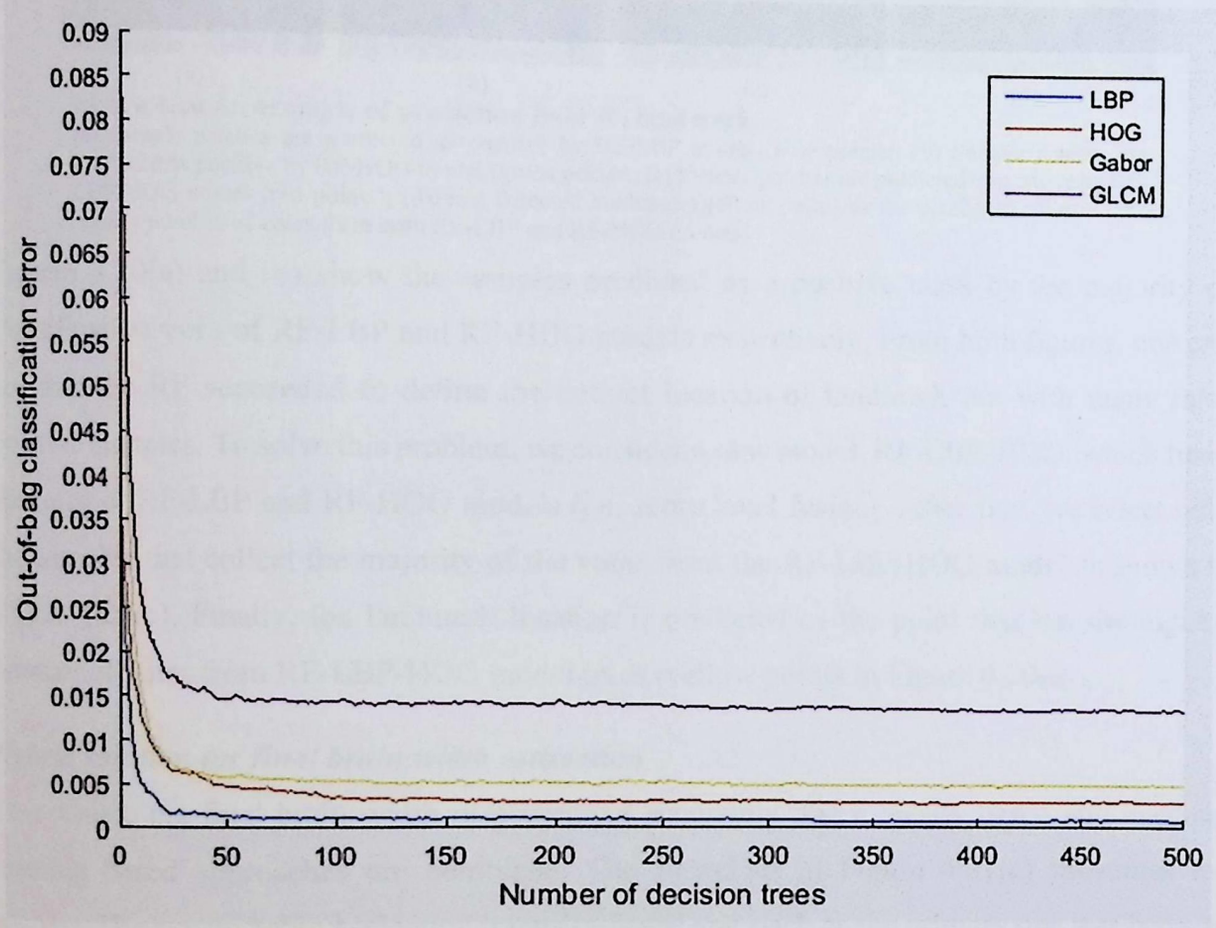
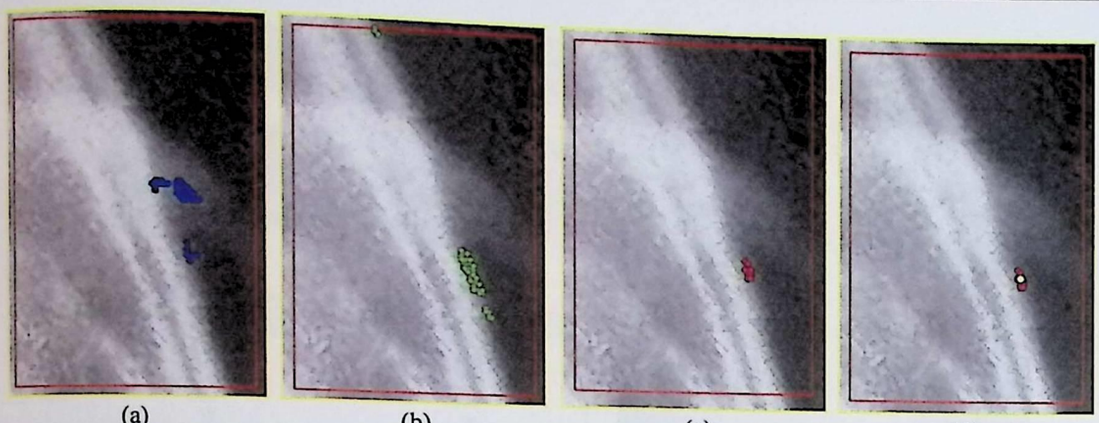


Figure 4.19 OOB classification error rate for RF-LBP (blue) RF-HOG (red), RF-Gabor (yellow) and RF-GLCM (purple) models for W_1 landmark. In general, the classification error rate of the RF model decreases when the number of trees increases until 100 trees where it starts to stabilize. The classification error rates for RF-LBP and RF-HOG models are smaller than other models (less than 0.005).

To reduce the search time, we used the location of initial width landmark point and only test the pixels within its patch as shown by red window in Figure 4.18(b).

For each pixel in the prediction window, two types of features are extracted from its patch and then predicted by both RFs models obtained in the training stage to determine its class. The final location of the landmark is decided as the pixel that has the highest sum of votes from both RF-LBP and RF-HOG models. Figure 4.18(b) shows the point predicted by RF-LBP model (blue point), RF-HOG model (green point) and the both models (red point). Another example to show how we selected the final landmark is shown in Figure 4.20.



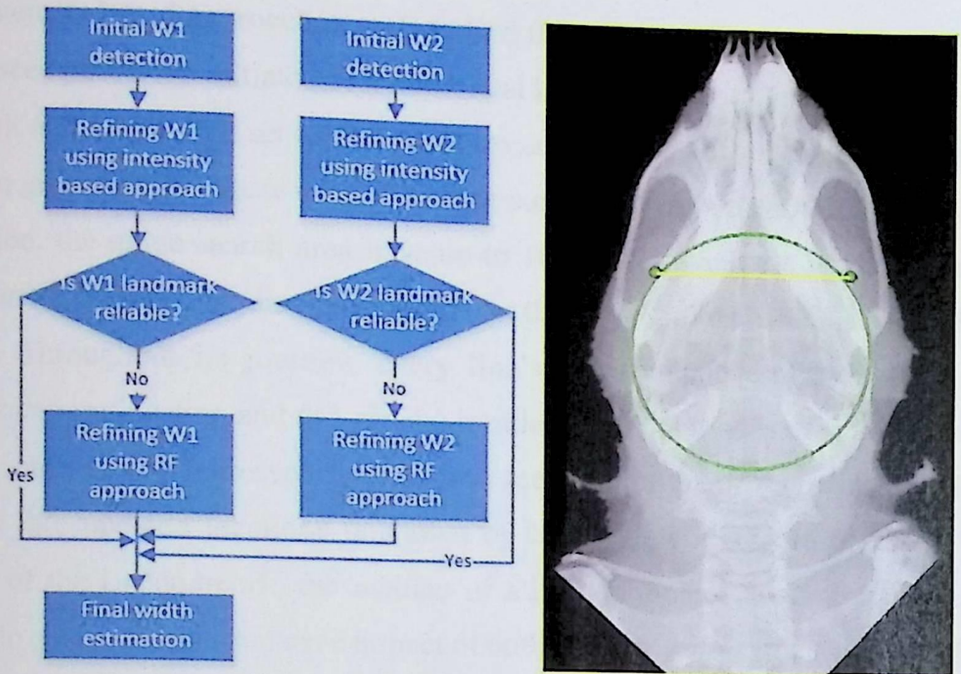
(a) (b) (c) (d)

Figure 4.20 An example of prediction final W_1 landmark.
(a) Sample patches are predicted as positive by RF-LBP model (blue points); (b) Sample patches are predicted as positive by RF-HOG model (green points); (c) Sample patches are predicted as positive by RF-LBP-HOG model (red points); (d)Final detected landmark (yellow point) as the pixel that collected the highest number of votes from both RF-LBP and RF-HOG models

Figures 4.20(a) and (b) show the samples predicted as a positive class by the majority of classification trees of RF-LBP and RF-HOG models respectively. From both figures, one can see that the RF succeeded to define the correct location of landmark but with many false positive samples. To solve this problem, we consider a new model, RF-LBP-HOG, which fuses the trees of RF-LBP and RF-HOG models (i.e. score level fusion). After that, we select only the samples that collect the majority of the votes from the RF-LBP-HOG model as shown in Figure 4.20(c). Finally, the landmark location is predicted as the point that has the highest number of votes from RF-LBP-HOG model trees (yellow points in Figure 4.20(d)).

Hybrid solution for final brain width estimation

To estimate the final brain width measurement accurately, the intensity based and machine learning based approaches are combined. The flowchart in Figure 4.21(a) illustrates the process. The reason behind selection intensity-based approach as the primary one was because this approach is faster than RF classification approach. Also, the intensity-based approach performs very well and in a consistent manner when the coronoid process is recognizable and could be detected accurately. However, the problem is mainly raised when the coronoid process is very small or not seen and so the intensity-based approach couldn't detect it accurately. When this is the case, the job of landmark point refining is sent to the RF classification solution. Finally, the width is estimated as the distance between the refined W_1 and W_2 landmark points as shown in Figure 4.21(b).



(a) (b)

Figure 4.21 Final brain width estimation from mouse X-ray image.
(a) Flow chart of final mouse brain width estimation method. The first approach to be used for the width landmark prediction is the intensity based approach. If the detection of the landmark is unreliable, the landmark is then predicted using RF approach; (b)The final W_1 and W_2 detected landmarks (green points) and the line between them (yellow line) which represents the estimation of the width of brain

4.3.6 Refining Length Landmarks and Estimating Brain Length

In this section, we propose two different approaches to refine the initial brain length landmarks of the mouse. The first approach uses different intensity-based techniques. The second approach is machine learning based on RF classification. Following final refinement of both length landmarks, a combination between the two approaches to estimate the final brain length is presented.

Refining brain length landmarks using intensity-based approach

The solutions adopted in this section to refine L_1 and L_2 landmarks are mainly based on simple intensity based approach. The idea involves setting group of initial points (i.e. line of points) near the target landmark and then those points refine themselves depending on intensity value check in neighbouring pixels in enhanced image (Birkfellner, Figl and Hummel, 2014). The full details for this procedure for both landmarks are given as follows.

Refining L_1 landmark

It could be seen from Figure 4.11(a) that the lower edge of the occipital bone, i.e. the desired location for L_1, is very faint even in enhanced image. To deal with this issue, an attempt of enhancing the ROI for L_1 is made by using morphological operations over grey-scale image as mentioned in Section 2.2.1 and the result shown in Figure 4.22. Following this enhancement,

a simple intensity based approach is used to find the points on the lower edge of the occipital bone. The seed points are initiated as a horizontal line whose centre is at the initial position of L_1 landmark and with width set to 180 pixels to ensure cover large area of the occipital bone as shown in green line in Figure 4.22. To make sure that the points of the line converge to the right location, the range search area is setup to 100 pixels along the vertical line below the initial landmark L_1 based on the observation that the target point should be somewhere within this range. Throughout its journey, every line's point calculates the intensity difference between its current location and (+5 pixels) location. The point where the maximum intensity difference obtained will represent the final new location for the line's point. The result of line converging into updated locations is shown by blue points in Figure 4.22. To find the y-coordinate of the L_1 landmark, the median of all y-coordinates of the line output points is calculated to overcome any negative impact of outliers. The yellow point shown in Figure 4.22 illustrates the final location of the refined landmark point L_1.

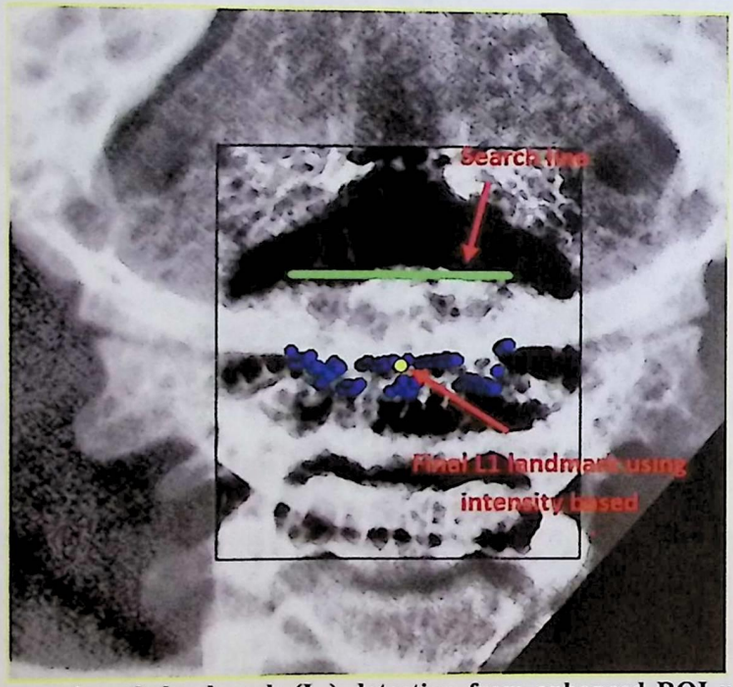
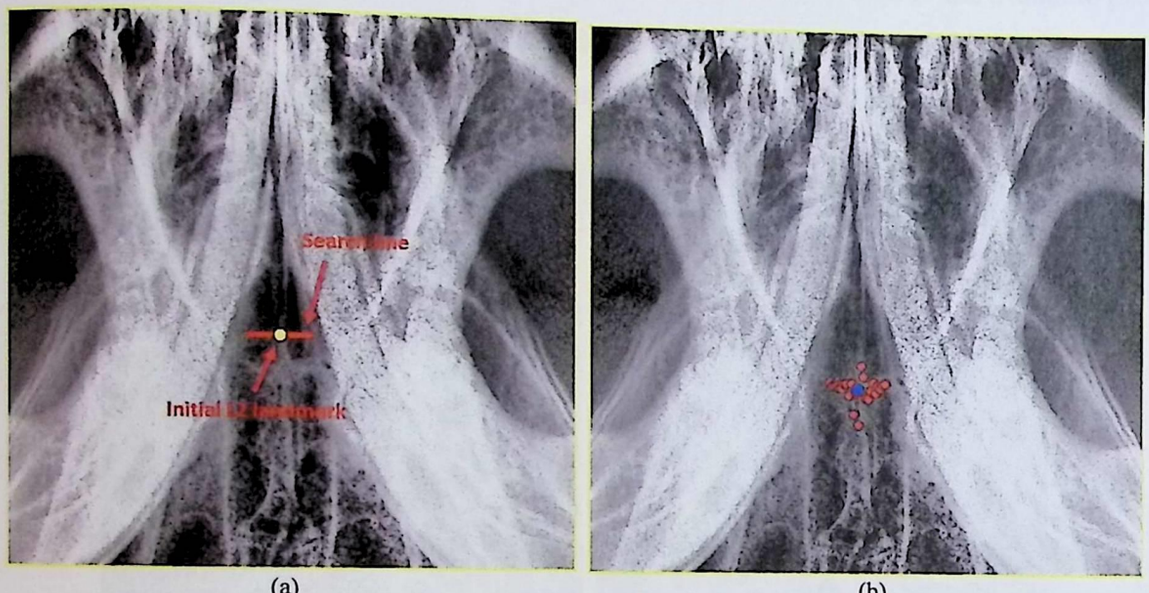


Figure 4.22 Lower length landmark (L_1) detection from enhanced ROI using intensity-based approach. Line setting (green line) by using the position of the initial L_1 landmark, migration of the points of the line (blue points) and the final L_1 position (yellow points) which located as the median of y-coordinates of the distributed points of the line.

Refining L_2 landmark

To estimate the accurate location of the upper landmark point, a process similar to L_1 refinement is applied here. Because of the complexity and the relative small size of the search area, different parameters needed to be set carefully when applying the intensity based approach. It is very important for the line to be positioned above the target landmark with a

width that fit the width of the desired area (yellow ellipse in Figure 4.13). The line is set to be centred at the initial position for L_2 landmark (yellow point in Figure 4.23 (a)) and with width equal to the distance between the right and the left edge points obtained in Section 4.3.4 as shown by the red line in Figure 4.23(a). After that, every point in the line is converged to the highest intensity point in the 100 pixels below it in the same manner for L_1 refinement. The final estimated L_2 landmark is shown in blue point in Figure 4.23(b).



(a) (b)

Figure 4.23 Upper length landmark (L_2) detection using intensity-based approach.
(a) Line setting (red line) above the target position using the initial L_2 landmark location (yellow point); (b) Points of the line after distribution (red points) and refined L_2 landmark (blue point) which found as a median of the y-coordinates of red points

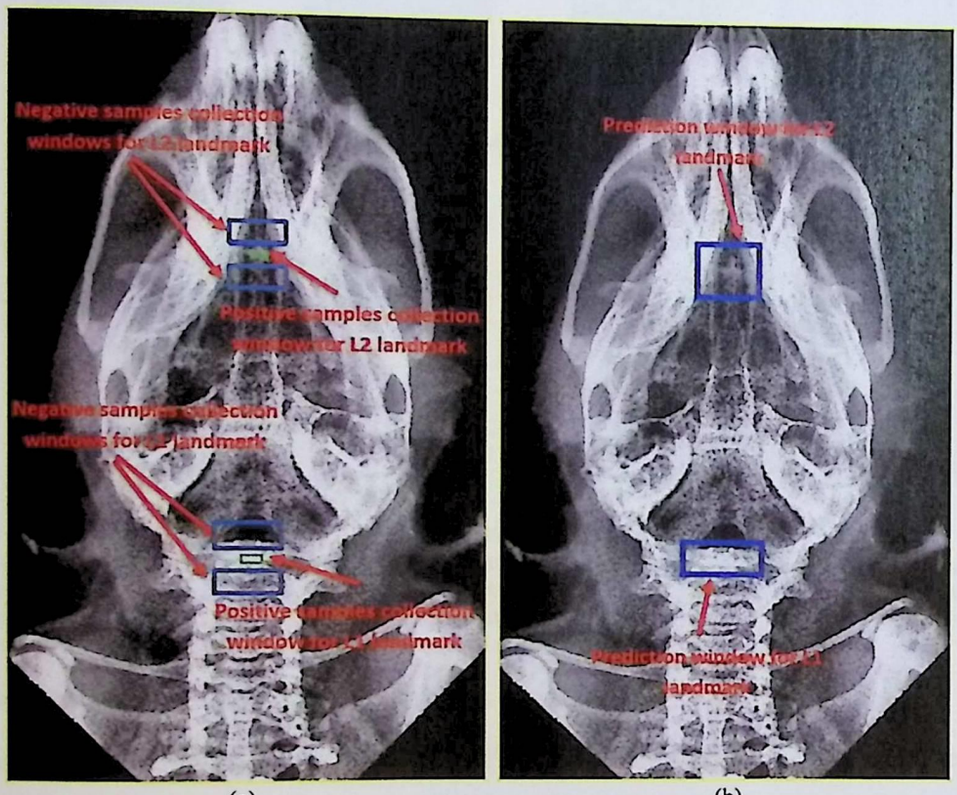
In some cases, the refined L_1 or L_2 landmarks is considered unreliable depending on some criteria (see Section 4.5.1 for details). When such unreliability case is detected, the machine learning based approach will be responsible of L_1 or L_2 landmarks refinement.

Refining brain length landmark points using machine learning based approach

In this section, the RF classification model is proposed to refine the initial detected L_1 and L_2 landmark points. The general approach application is similar to what has been applied in Section 4.3.5 but with some changes in parameters setting to suit the different in landmarks area. In the training stage, the positive samples (landmark points) are again extended by considering the annotated point by the expert annotator and any of surrounding pixels within a window of size (D × K) as shown by green windows in Figure 4.24(a) for L_1 and L_2 landmarks. For both landmark points, the number of added positive samples in the horizontal axis (K) is more than the case for width points. This because the focus here is mainly on the vertical position rather than the horizontal one, and the x-coordinate of the refined L_1 and L_2 landmarks can be taken as the x-coordinate of the initial L_1 and L_2 landmarks. The negative samples (non-

landmarks point) are sampled from two windows to above and below annotated L_1 and L_2 as shown by blue windows in Figure 4.24(a) for L_1 and L_2 respectively. The training for two models using the two extracted types of features (LBP and HOG) is done exactly the same as the training for the width landmark points.

In the testing stage, for each unseen image, the initial detected L_1 and L_2 are used to guess the search area. From the determined area, pixels are sampled and then sent to the model to predict whether it positive or negative examples. Figure 4.24(b) shows the selected windows for determined search area by using initial landmark points for L_1 and L_2.



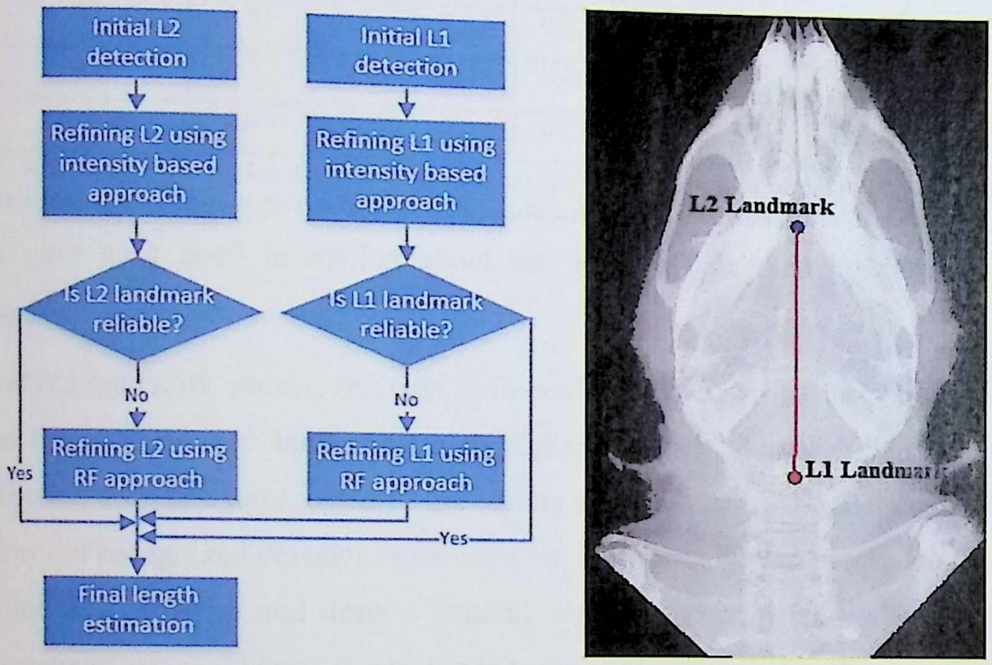
(a) (b)

Figure 4.24 Different windows in mouse head X-ray image in the training and prediction stages for L_1 and L_2 landmarks.
(a) Positive sample collection windows (green windows) and negative samples collection windows (blue windows) for L1 and L2 landmarks in the training stage; (b) Prediction windows (blue windows) for L1 and L2 landmarks in the prediction stage

Hybrid solution for final brain length estimation

As mentioned earlier, the length of the brain is estimated as the distance between L_1 and L_2 landmarks. We propose to combine the intensity based and machine learning based approaches to produce a final brain length estimation as shown in Figure 4.25(a). We decided to use the intensity-based approach as the primary approach to detect L_1 and L_2 landmarks. This is because the intensity-based approach is more accurate and efficient when the landmarks are clearly identifiable, i.e. the reliability check result is positive (see Section 4.5.1 for details).

When the landmarks are not clearly identifiable, however, the robustness of the intensity based approach drops significantly. Therefore, the texture based machine learning approach is used as a more robust but less efficient in all cases. After the final refinement of the length landmarks, the brain length is estimated as the Euclidean distance between them as shown in Figure 4.25(b). This justifies the need to fuse the two approaches in the way explained above.



(a) (b)

Figure 4.25 Final brain length estimation from mouse X-ray image.
(a)Flow chart of the final mouse brain length estimation. The first approach to be used for the landmark prediction is the intensity based approach. If the detection of the landmark is unreliable, the landmark is then detected using RF approach (b) The final L_1 (red point) and L_2 (blue point) detected landmarks and the line between them (red line) which represents the estimation of the length of the mouse brain

4.4 Experiment Results and Analysis

This section presents two sets of results. The first set is centred on comparing the automatic measurements to those obtained manually while the other is dedicated to identify genes responsible for changes in brain size.

4.4.1 Manual vs. Automatic Measurements

Due to the difficulty of labelling all images manually for evaluation purposes, random samples of 100 images were annotated by an expert to be used as a ground truth.

To assess the accuracy of the detection of the four landmark points, the Euclidean distance between the automatically detected and the ground truth landmark points was calculated in millimetres (mm). The mean (μ) and the standard deviation (σ) of Euclidean distances for each landmark point across all testing images were calculated as reported in Table 4.1.

Approach / Landmark	Intensity-based approach		RF-LBP		RF-HOG		RF-LBP-HOG	
	μ	σ	μ	σ	μ	σ	μ	σ
W_1	0.24	0.16	0.24	0.21	0.33	0.33	0.24	0.25
W_2	0.15	0.10	0.21	0.24	0.17	0.18	0.13	0.16
μ (W_1, W_2)	0.19	0.14	0.22	0.23	0.25	0.28	0.19	0.22
L_1	0.31	0.21	0.29	0.22	0.45	0.32	0.27	0.20
L_2	0.27	0.19	0.45	0.40	0.48	0.38	0.40	0.35
Average	0.24	0.17	0.30	0.27	0.36	0.30	0.26	0.24

Table 4.1 Results of all landmark points in terms of mean and standard deviation of the Euclidean distances between the automated landmark points and the ground truth

This table gave us a good indication about the accuracy obtained by different proposed approaches.

For the width's landmark points, the first note could be made from the above table that all approaches detected W_2 (left landmark) with higher accuracy than W_1 (right landmark). To understand why, a close examination of the results showed that the number of difficult cases (i.e. small or not recognized coronoid processes) in the right side are more than the left ones. To overcome this disparity and draw a general conclusion about the width key landmarks detection accuracy, we considered averaging the results of evaluating both landmarks for every method as shown in Table 4.1.

Another note could be made from Table 4.1 is that the HOG based RF model performs slightly better than LBP based RF model for the W_2 landmark detection. This is because when the coronoid process is identifiable, the orientation information plays more essential role in landmarks detection than just trivial changing in intensity values. Figure 4.26 shows some examples of identifiable and non-identifiable coronoid processes.

The table also shows that the proposed intensity-based approach works marginally better than the RF approach for the average of width landmarks detection. This confirms that the proposed intensity-based approaches are sufficiently accurate in detecting the coronoid processes in most of the cases.

However, the main source of inaccuracy in the intensity-based procedure is resulting from the lack of precision in detecting the coronoid processes. When this is the case, the RF classification approach is used to refine the final location of the width landmark.

On the other hand, the nature of difficulties in locating the length landmarks is different from the width ones. The lower length landmark L_1 is in an area where the bone edge is not clear

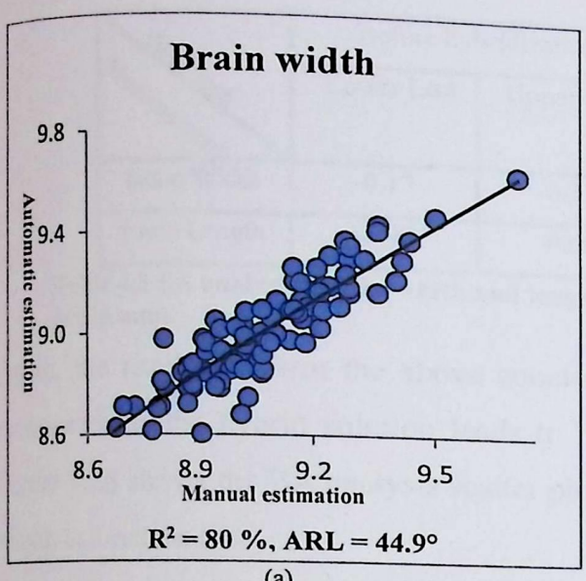
with hardly any changing happens in texture orientations. This might explain why the LBP based RF model outperformed the RF based HOG model in locating L_1. The high variations of the texture around the L_2 landmark from image to image reduces the ability of RF model to learn the L_2 landmark and thus the intensity-based approach outperform it. For the RF classification approaches for all landmarks, Table 4.1 shows that the results of detection from combining the RF-LBP and the RF-HOG models were slightly better than the results of each model individually. For the overall average for all landmarks detection, the results confirms that the intensity based approach gives the best estimation as shown in Table 4.1.



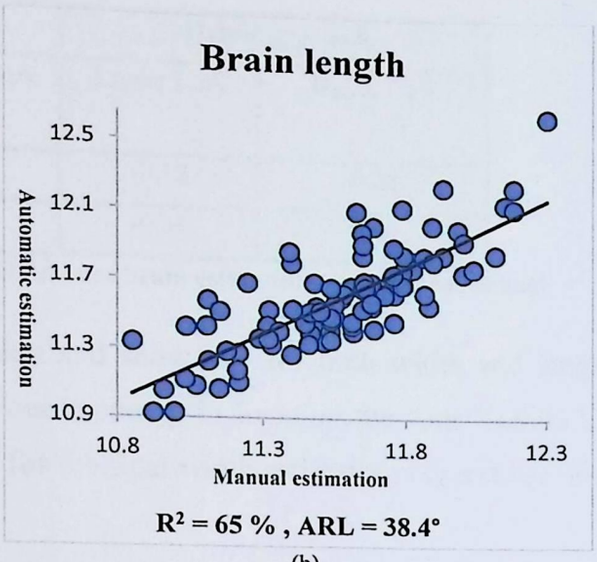
(a) (b) (c)

Figure 4.26 Examples of identifiable and non-identifiable coronoid processes from contrast enhanced mouse head X-ray images.
(a) The right coronoid process is very small and so non-identifiable while the right one is big and could be identified; (b) The right coronoid process is non-prominent from the skull and so non-identifiable while the right one is big and could be identified; (c) The right coronoid process is very small and so non-identifiable while the right one is big and could be identified

After assessing the four different automated methods of the landmarks, one needs to evaluate the proposed width and length of the brain based on the hybrid solution described in Sections 4.3.5 and 4.3.6 versus the manual measurements taken by a domain expert. In the first evaluation, we check the correlation between the manual and automatic measurements using R^2 and ARL (see Section 3.4.2 for further details). Figures 4.27 (a) and (b) show the scatter plot for the manual versus automatic measurements for brain's width and length respectively. The figures and the values of R^2 and ARL show the close correlation between the manual and the automatic estimation in most cases. However, the width still shows more consistency than the length due to the underlying difficulty in locating the length landmarks as explained earlier.



(a)



(b)

Figure 4.27 Correlation of manual versus hybrid automatic measurements estimation for the width and length of the mouse brain. All the measurements in mm.
(a) Manual versus automatic estimation of brain width measurements (blue points) and regression line (black line); (b) Manual versus automatic estimation of brain length measurements (blue points) and regression line (black line)

To further evaluate the effectiveness of the proposed hybrid solution, we compare the manual versus the automatic of the width and length measurements before and after applying the hybridisation. Table 4.2 reports the comparison results based on four parameters, namely R^2, ARL, MAE, and standard deviation of absolute error between the ground truth and the automatic measurements (σ) (see Section 3.4.2 for further details).

The table shows noticeable improvements in terms of the consistency between the manual and the automatic measurements especially for the brain's length for the reason explained earlier in Section 4.3.6.

Measures	R^2	ARL	MAE	σ
Brain width estimation (intensity-based only)	0.69	44.4°	0.11 mm	0.11 mm
Hybrid Brain width estimation	0.80	44.9°	0.09 mm	0.07 mm
Brain length estimation (intensity-based only)	0.37	31.6°	0.19 mm	0.20 mm
Hybrid Brain length estimation	0.65	38.4°	0.15 mm	0.11 mm

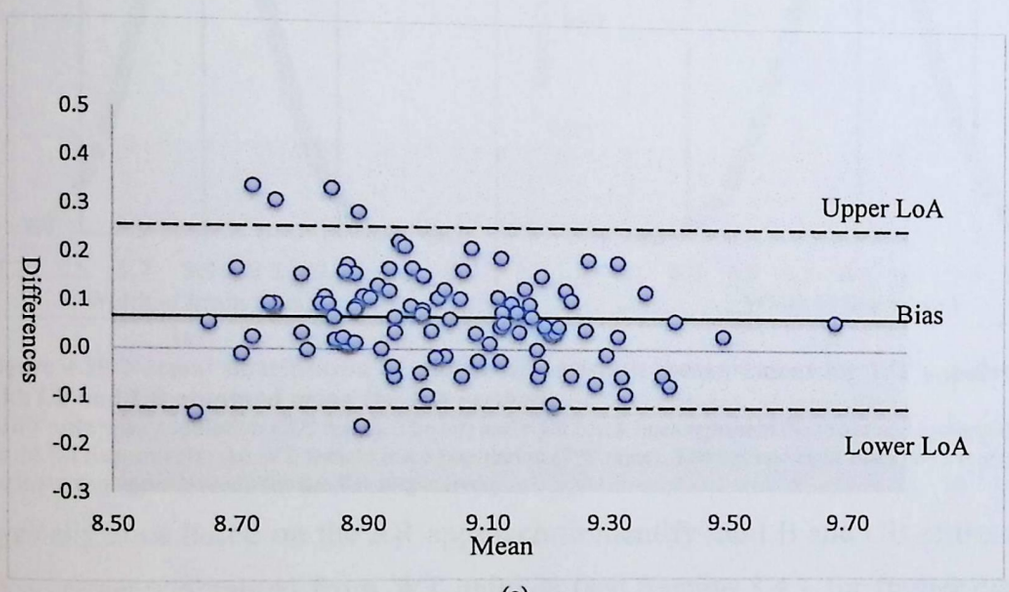
Table 4.2 Manual versus automatic width and length measurements evaluation. Table shows the correlation between the manual and automatic measurements R^2. It also shows the angle of regression line (ARL), mean of absolute error (MAE) and the standard deviation of the absolute error (σ).

However, after assessing the correlation between the manual and automatic measurements, we need to examine the closeness between them using the BA analysis (Section 3.4.2). Table 4.3 shows the lower and upper LoA for all proposed measurements.

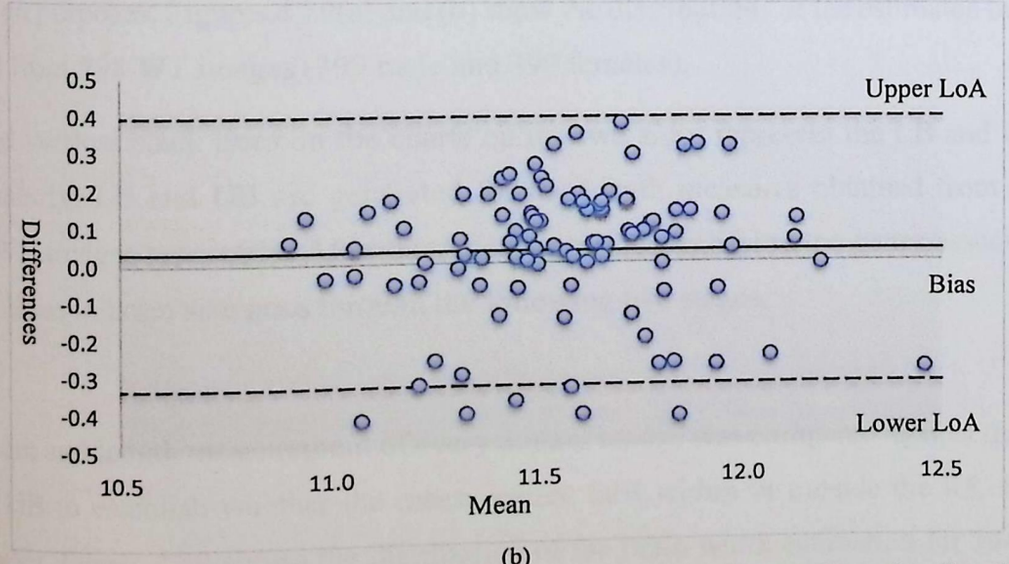
Approach / Measurement	Before hybridization		Hybrid approach	
	Lower LoA	Upper LoA	Lower LoA	Upper LoA
Brain Width	-0.17	0.33	-0.12	0.26
Brain Length	-0.45	0.60	-0.33	0.40

Table 4.3 BA analysis for the width and length of mouse brain estimation. (All measurements are in mm).

Again, the results confirm the above conclusion and show that for both width and length measurements the hybrid solution leads to better accuracy in locating the four landmarks. Figure 4.28 shows the BA analysis scatter plot for the final width estimation (a) and the final length estimation (b).



(a)



(b)

Figure 4.28 BA analysis for manual versus automatic brain width and length estimation. (All measurements are in mm). The y-axis is the difference between the manual and automatic measurements while the x-axis is the mean between them for each point.
(a) Width measurements; (b) Length measurements

4.4.2 Genes Identification

In this section, we report the results of applying the proposed method on the whole dataset of 6,888 X-ray images. The ultimate goal of the experiment is to identify the genetic basis of abnormalities in the brain size i.e. what are the genes that may lead to a bigger or smaller brains? To do this, one first needs to establish what is normal and what is not when it comes to brain size of mice for both males and females separately.

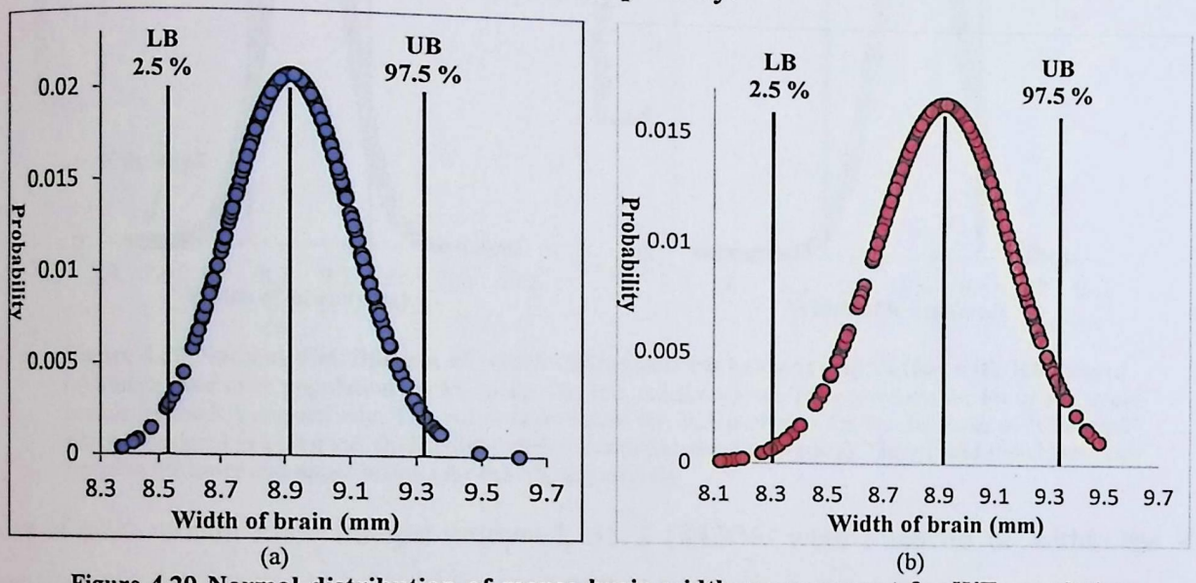


Figure 4.29 Normal distribution of mouse brain width measurement for WT population with LB and UB obtained using the RR method.
(a)WT male mice population (399 mice). The left and right black lines represent the lower and upper bounds for the RR respectively; (b) WT female mice population (399 mice). The left and right black lines represent the lower and upper bounds for the RR respectively

This is typically done based on the RR approach to identify the LB and UB of the width and length measurements obtained from WT animals (see Section 3.4.1 for further details). For illustration purposes, Figures 4.29(a) and (b) show the distributions of the estimated brain width obtained from 798 WT images (399 male and 399 females).

The solid vertical black lines on the charts on the two sides represent the LB and UB of the RR. Similarly, LB and UB are generated for the length measures obtained from male and female WT images separately. After that, the process of identifying the genes associated with abnormalities in brain size goes through the following two stages.

Stage1:

Each width and length measurement of every mutant mouse was compared against the relevant LB and UB to establish whether the measurement falls within or outside the RR. Similar to WT images, Figure 4.30 shows the distribution of the brain width estimation for 3045 mutant male mice (a) and 3045 mutant female mice (b). The vertical black lines represent the LB and

UB of the RR ([8.53 mm to 9.32 mm] male and [8.30 mm to 9.33 mm] female) which obtained from the width measures of WT male (normal) cases.

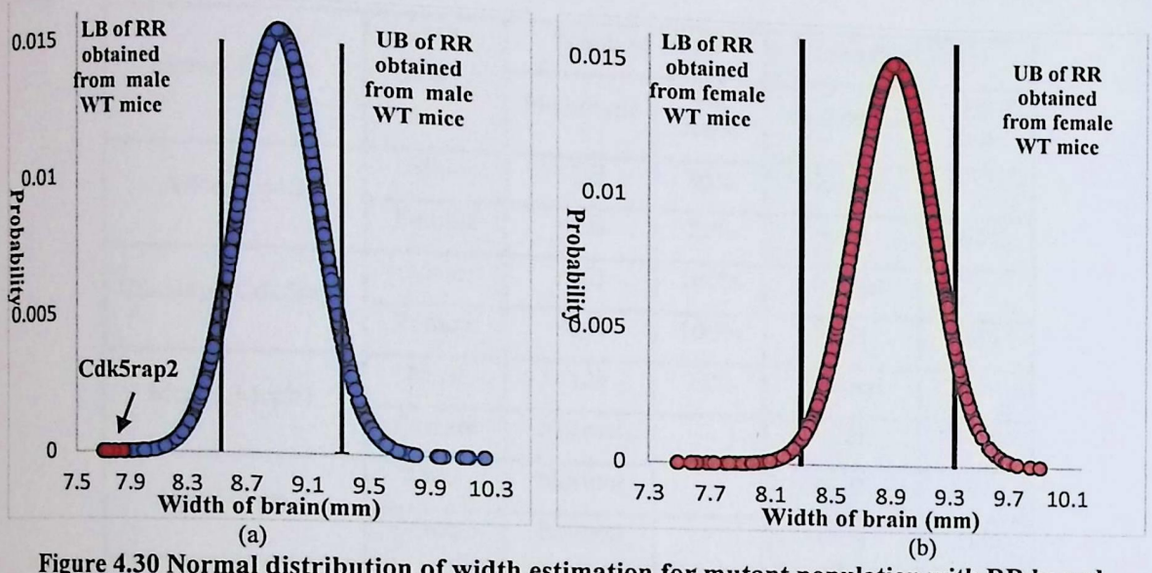


Figure 4.30 Normal distribution of width estimation for mutant population with RR bound
(a) Mutant male mice population (3045 mice). The left and right black lines represent the lower and upper bounds for the RR respectively. The red dots represent the measurements for the instances of "Cdk5rap2" which considered as a hit gene; (b) Mutant female mice population (3045 mice). The left and right black lines represent the lower and upper bounds for the RR respectively

Out of 6090 mutant mice images analysed, 5133 (84.3%) were found to be within the RR established by the WT measurements, and 957 (515 male and 442 female mice) (15.7%) of mutants fell outside the RR for both the width and the length measurements. However, it is important to highlight that not all samples that fall outside the RR are necessarily true hits (abnormal cases). Arguably the above number includes many false positive cases due to two main reasons. First is related to any incorrect landmarks detection which leads to incorrect dimension estimation. Second reason is that although the case is correctly measured as abnormal size but this case might belong to a specific gene where the total number of cases of abnormal size for it is less than the hit rate threshold (see Section 3.4.1 for further details).

Stage2:

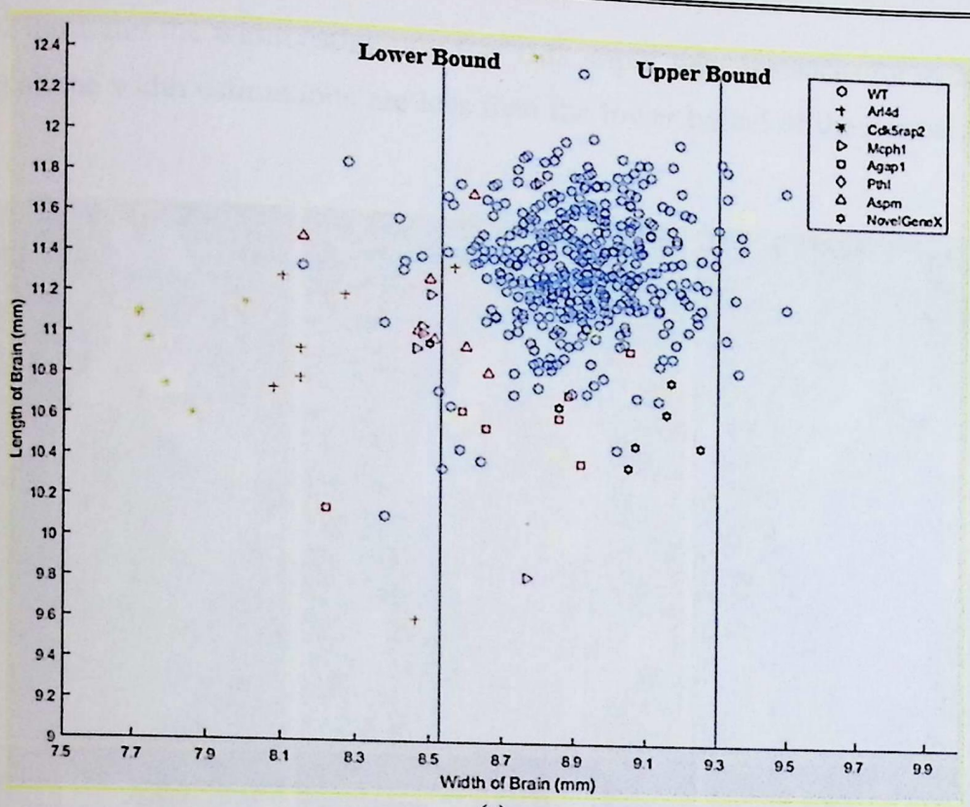
To filter out the false positive cases, we calculated the so-called hit ratio of each mutant line (see Section 3.4.1 for details). Out of the 957 outlier cases, only 73 cases were considered as true hits (i.e. hit ratio > 60%) as shown in Table 4.4 where the "-" mark is used to indicate that the ratio is below 60%. Figure 4.31 illustrates the scatter plot for widths versus lengths automatic measurements of WT male mice and the mouse instances of the seven defined genes listed in Table 4.4. Figure 4.31 also illustrates the upper and lower bound obtained using RR and the outlier measurements for the width (Figure 4.31(a)) and the length measurements

(Figure 4.31(b)). We can see that 7 out of 7 measurements of the mice instances of Cdk5rap2 gene line are outlier for the brain width estimation so it is considered as hit gene.

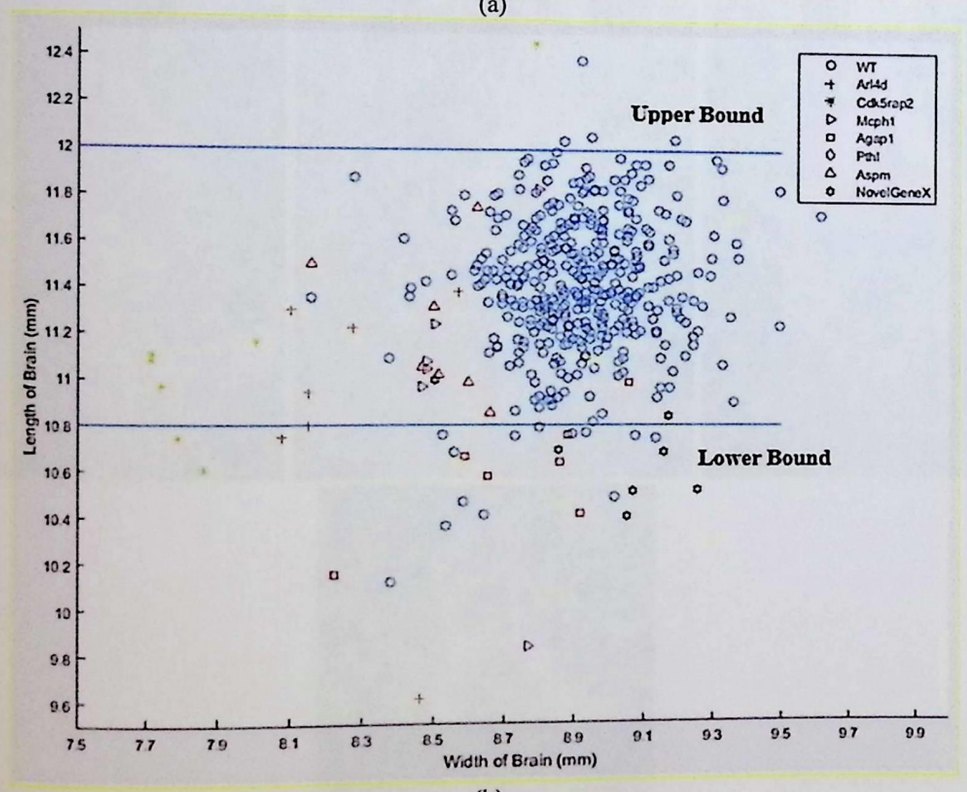
Name of allele	Gender	Width of Brain		Length of Brain	
		Phenotype	Hit Ratio	Phenotype	Hit Ratio
Arl4d/Arl4d	Male	<LB	86%	Normal	-
	Female	<LB	71%	<LB	100%
Cdk5rap2/Cdk5rap2	Male	<LB	100%	Normal	-
	Female	<LB	100%	<LB	86%
Mcph1/Mcph1	Male	<LB	71%	Normal	-
	Female	Normal	-	<LB	71%
Agap1/Agap1	Male	Normal	-	<LB	71%
	Female	Normal	-	<LB	71%
Pthlh/+	Female	Normal	-	<LB	71%
Aspm/Aspm	Male	<LB	71%	Normal	-
NovelGeneX	Male	Normal	-	<LB	71%

Table 4.4 Possible gene hits which defined using automatically estimated width and length measurements of mouse brain. The Hit ratio determine whether the gene is a hit (hit ratio >= 60%) or not (hit ratio < 60%).

The table above shows a list of 7 genes that could be potentially responsible for abnormalities in the mouse brain size. Some of listed genes are already confirmed or investigated in the literature to have a certain relationship to the change in the skull shape and thus the brain size (Tan et al., 2014). This in a way confirms the effectiveness of our developed method in identifying the list of genes responsible for changes in the brain size. Interestingly, the table shows a possible novel gene being temporarily called 'NovelGeneX', which is currently under investigation by the WTSI team. For illustration purposes, the instances of one gene (Cdk5rap2) that confirmed to have smaller width of brain are highlighted as red dots in Figure 4.30(a).



(a)



(b)

Figure 4.31 Scatter plot of the brain width-length measurements for male WT and mutant mice.

(a) Width-length measurements plot and the lower and upper bounds for the width of brain (blue lines) obtained using the RR method. We can see that 7 out of 7 of "Cdk5rap2" gene instances are outlier so it is considered as a hit gene; (b) Width-length measurements plot and the lower and upper bounds for the length of brain (blue lines) obtained using the RR method. We can see that 5 out of 7 of "NovelGeneX" gene instances are outlier so it is considered as a hit gene

Figure 4.32 illustrates the width estimation for "Cdk5rap2" gene instances for male mice. It is shown that all the width estimations are less than the lower bound of the normal range (8.53 mm).



Figure 4.32 Width estimation for "Cdk5rap2" gene male mouse instances. All the measurements are less than the lower bound of the normal range (hit gene).

Figure 4.33 illustrates the width estimation for seven WT male mice. It is shown that all the width estimations are in the range of RR (greater than 8.53 and less than 9.32 mm).

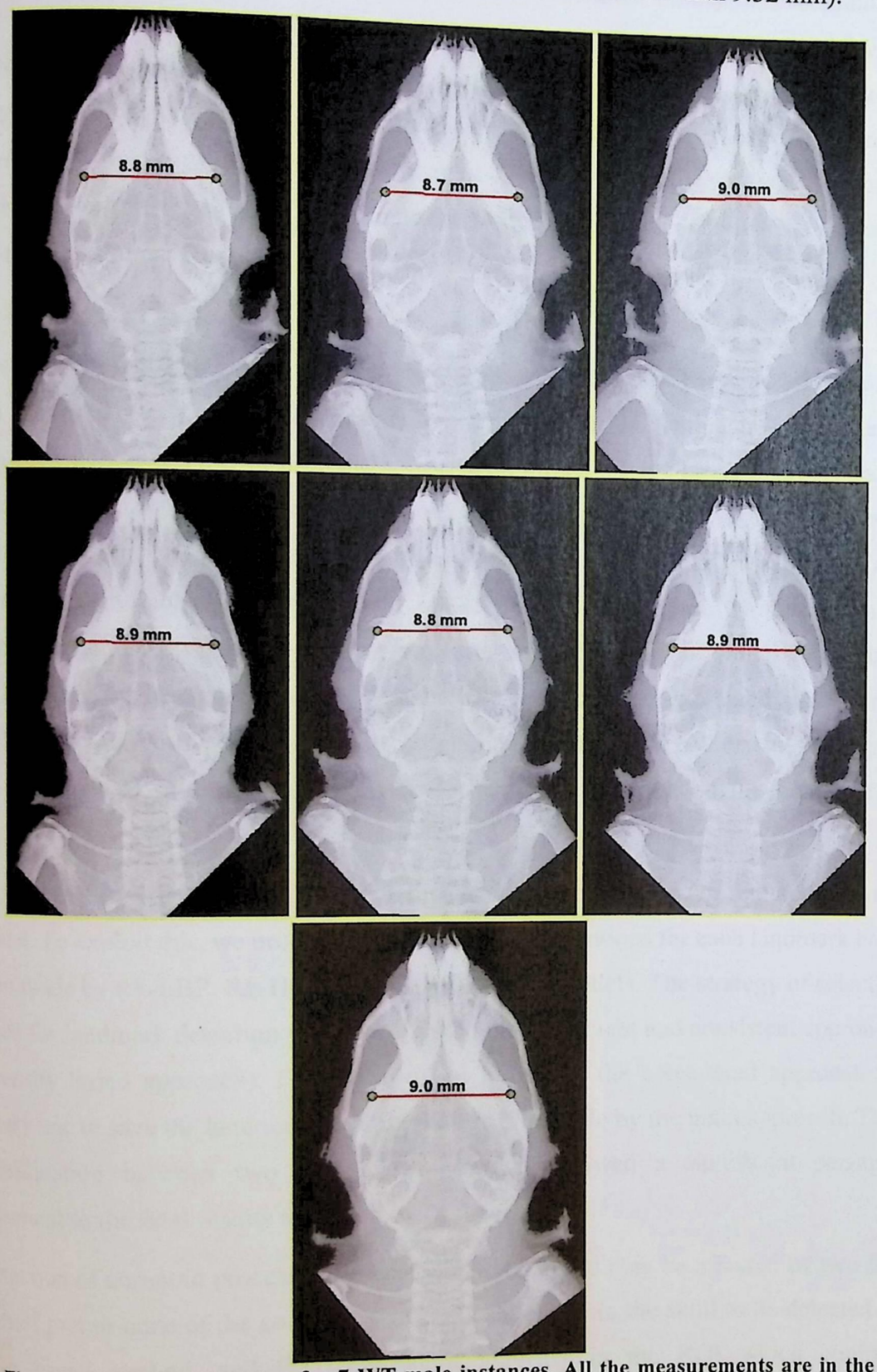


Figure 4.33 Width estimation for 7 WT male instances. All the measurements are in the normal range (i.e. normal width).

4.5 Discussion

The discussion in this section is divided into five subsections. Firstly, a discussion on the quality of the results is presented. Secondly, we discuss the parameter sensitivity of the reliability check. Thirdly, we discuss the impact of patch size around the landmark on the accuracy of its location. Fourthly, the effectiveness of proposed method when applied on different data set is discussed. Finally, we present a MATLAB tool which was designed with a Graphical User Interface (GUI) to estimate the brain size and then classify the mouse based on estimated measurements.

4.5.1 Results Discussion

The list of 7 genes along with the manual vs. automatic comparison presented in the previous section illustrates that the proposed method is effective in estimating brain's dimensions. Section 4.3.2 highlighted the fact that the correct skull modelling is a key step towards accurate dimensions' estimation. To increase the overall accuracy of applying the proposed method, a reliability criteria proposed in Section 4.3.2 was used to exclude a small percentage of images which cannot be reliably segmented. Out of 6888 images were tested, the proposed method succeeded in identifying inaccurate circle positions for 185 images of which 166 cases were correctly refined by the correction procedure in the skull modelling explained in Section 4.3.2. Only 19 images (0.28%) were finally rejected as images of "unsatisfactory quality". In other words, the proposal produced automatic measurement on 99.72% of the total number of images.

We showed that RF classification approach results depend on the type of features used to train the model. To exploit this, we produced three different estimations for each landmark based on decision made by RF-LBP, RF-HOG and RF-LBP-HOG models. The strategy of selecting the approach for landmark detection was based on the most efficient and consistent approach (i.e. the intensity based approach). Furthermore, we hybridize the considered approach with a secondary one to save the landmarks that marked as unreliable by the main approach. The idea of hybridization between two different approaches showed a significant performance improvement to the final results as shown in Table 4.2 and 4.3.

The detection of coronoid process by intensity-based method may be affected by two factors: (a) a small part or none of the coronoid process is salient from the skull to be detected by the intensity based method, and (b) inconsistent intensity in the ROI which complicates discrimination of the coronoid process from its background by using a threshold. In either case,

the correct coronoid process is not detectable at the end and the unreliability flag is raised which cause the job to be transferred to RF classification approach. Figures 4.34(a) and (b) show a case of small coronoid process (inside the red circle) and the result of thresholding it respectively. Such a very small and faint coronoid process is difficult to be recognized even by human eye. For some of such cases that not detected by intensity-based approach, the RF approach might give better estimation for the location of landmark. For example, the yellow point in Figure 4.34(c) represents the detected landmark using intensity-based approach while the red is the detected one using RF approach.

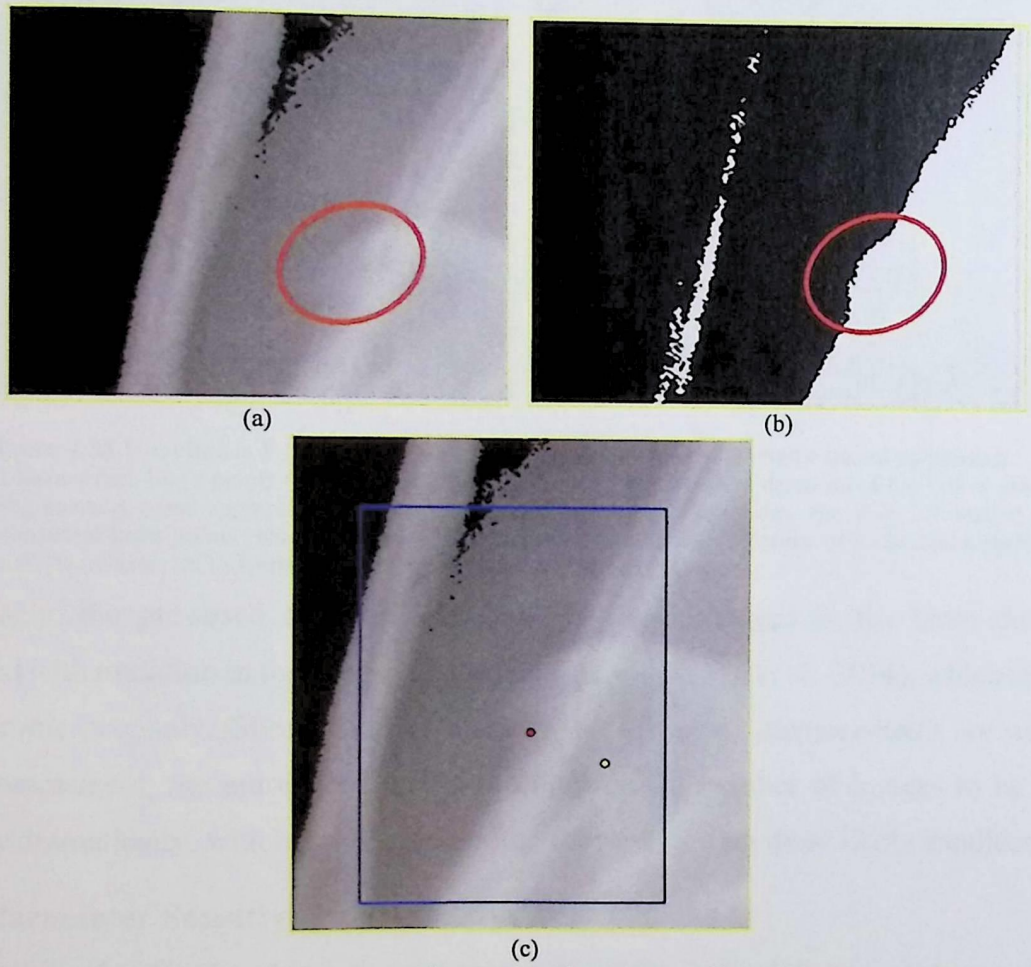
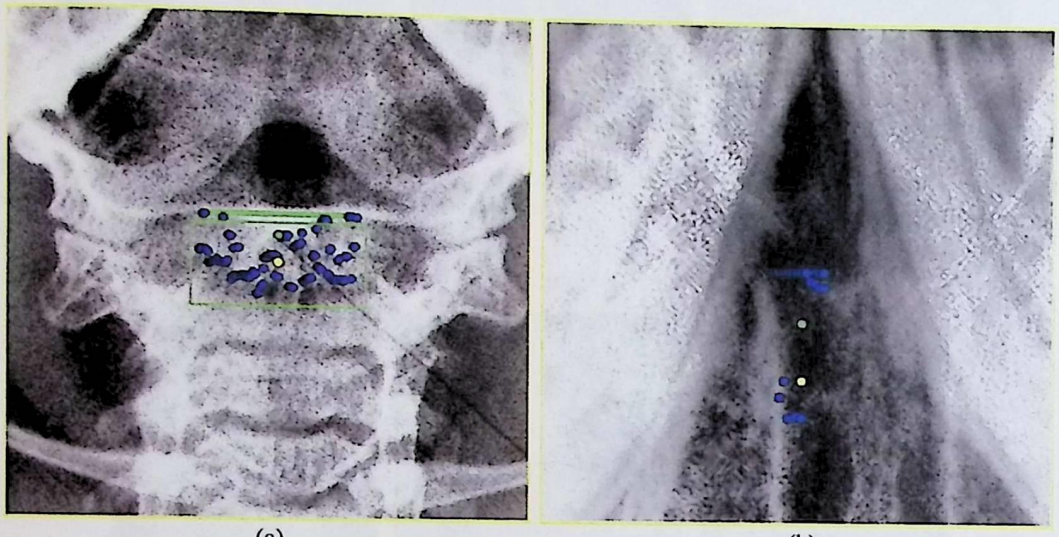


Figure 4.34 Non-identifiable coronoid processes from mouse X-ray image.
(a) Small and faint coronoid process encircled by red ellipse; (b) Mask of the coronoid process ROI; (c) Inaccurate detected width landmark using intensity based approach (yellow point) because of coronoid process mask was not detected and accurate detected width landmark using RF approach (red point)

For length landmark points, the main reason of inaccuracy is when the pixel intensity of the target bone is very low and not consistent (upper landmark L_2) or the edge of the bone is not clear (lower landmark L_1). This leads to some points of the line's point "leak out" of the related area which results in incorrect refinement (yellow points in Figure 4.35). This type of inaccuracy could be guessed by calculating the standard deviation value of the y-coordinate of the line points after distribution (blue points in Figure 4.35). It is shown empirically that if the

standard deviation greater than 15, this might signal a problem in a final refinement using intensity-based approach and then the detected landmark considered unreliable. When such unreliability case is detected, the machine learning based approach will be responsible of length landmarks refinement. Yellow points in Figures 4.35 (a) and (b) show the incorrect refinement of L_1 and L_2 using intensity based approach while the green points show more accurate prediction of L_1 and L_2 using RF approach.



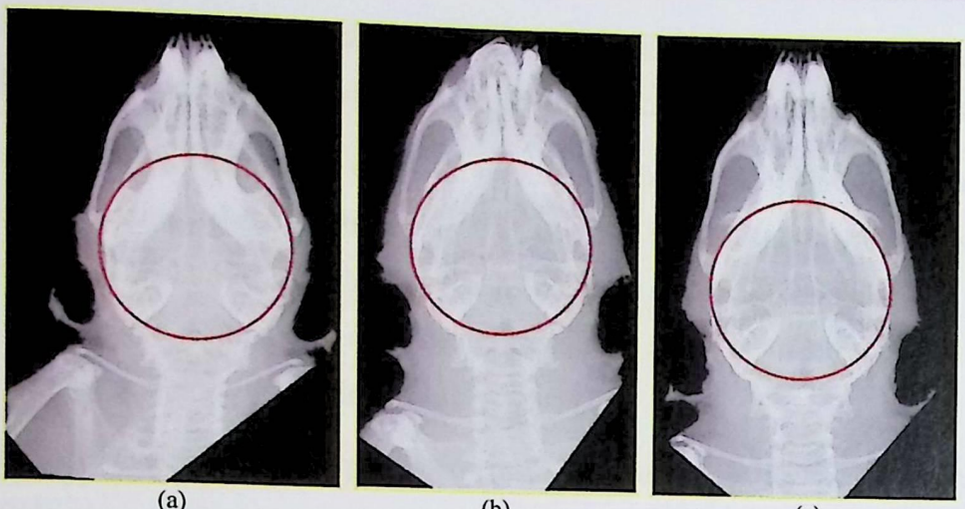
(a) (b)

Figure 4.35 Unreliable L1 and L2 landmarks detection using intensity based approach.
(a) Inconsistent line's points (blue points) distribution results in inaccurate detection of L_1 (yellow point) using intensity based approach. More accurate L_1 landmarks (green point) detection using RF model; (b) Inconsistent line's points (blue points) distribution results in inaccurate detection of L_2 landmark (yellow point). More accurate L_1 landmarks (green point) detection using RF model

Interestingly, the proposed solution was able to detect changes in the brain dimensions associated with mutation in the genes Mcph1 and Cdk5rap2 (Tan et al., 2014), which are known genes for microcephaly. Giving the fact that only the abnormal measurements are worthy for manual assessment, the automated method minimizes the number of images to be checked manually dramatically, which leads to focussing the attention on more likely candidate genes.

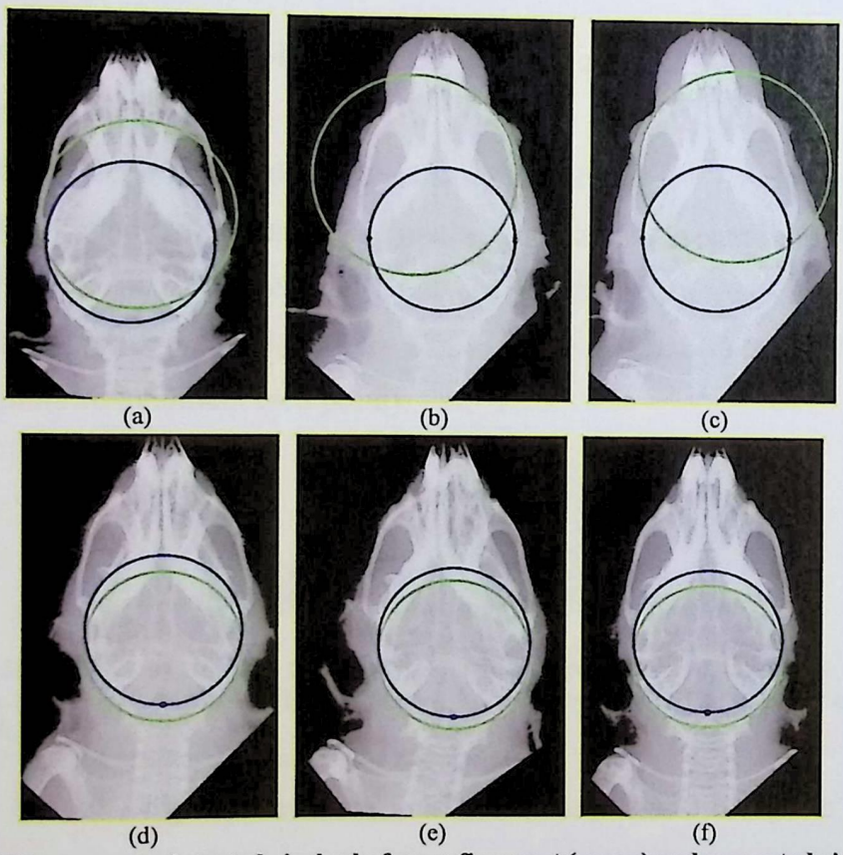
4.5.2 Parameter Sensitivity of the Reliability Checks

As we explained in Section 4.3.3, the normal range for the vertical distance between the top of the circle and the topmost point of the aligned skull mask is estimated between 520 and 720 pixels. This range was chosen to include different skull shapes and sizes. Figure 4.36 shows three different skulls that were modelled by circles with a vertical distance ranges from 542 to 717 pixels (normal range). For all the three skulls in the Figure 4.36, we can argue that the circle position is good enough to obtain search areas for all the 4 landmarks.



(a) (b) (c)

Figure 4.36 Accurate position detected circles imposed on a mouse X-ray image. The vertical distances for each circle are as follow.
(a) 542 pixels; (b) 592 pixels; (c) 717 pixels



(a) (b) (c)

(d) (e) (f)

Figure 4.37 Inaccurate detected circles before refinement (green) and corrected circle (blue) imposed on mouse X-ray images. The caption below shows the vertical distances for the circles before refinement (green circles).
(a) Too-high circle with a vertical distance of (400 pixels); (b) Too-high circle with a vertical distance of (213 pixels) 213 pixels; (c) Too-high circle with a vertical distance of (170 pixels); (d) Too-low circle with a vertical distance of (729 pixels); (e) Too-low circle with a vertical distance of (786 pixels); (f) Too-high circle with a vertical distance of (825 pixels)

As explained earlier, the circle is considered unreliable if the vertical distance is outside the range. Figure 4.37 illustrate 6 different cases with too-high (a, b and c) and too-low (d, e and

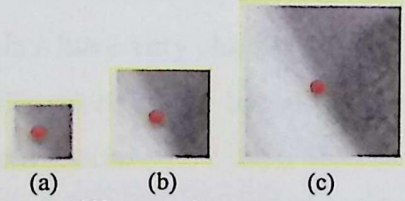
f) circles before (green circles) and after (blue circles) the proposed refinement where the caption of figure illustrates the vertical distances for each circle before the refinement.

Following observing different cases of too-high circle problem, we found that inaccurate skull mask is the key reason behind such cases. It is also noted that, the maximum value of the vertical distance for most of too-high circle cases does not exceed 400 pixels. This means that the lower bound of the normal range (520) is insensitive e.g. if we decrease it to 450, it still possible to recognize the too-high circle cases.

The value of the upper bound (720 pixels) seems effective to identify the low circle problem (Figure 4.29(d), (e) and (f)). However, it is considered somehow sensitive e.g. if we increase it to 730 for example, the image in Figure 4.37(d) will be considered in normal range circle and no refinement is done for it.

4.5.3 The impact of Patch Size around the landmark on the Accuracy of its Location

As we mentioned in Section 4.3.5, the features to train the RF model are extracted from a patch with a certain size around the annotated landmarks. The size of the patch could have an impact on the accuracy of the landmark location. To investigate this issue, we selected three different patch sizes as shown in Figure 4.38(a) (17 × 17), (b) (33 × 33) and (c) (65 × 65) pixels around the manually annotated landmark (red point).



(a) (b) (c)

Figure 4.38 Patches of Several different sizes are sampled around manually annotated landmark (red points) from W1 ROI of mouse head X-ray image. The sizes for each patch are as follows.
(a) (17 × 17) pixels (b) (33 × 33) pixels; (c) (65 × 65) pixels

As we can see from the figure, the first patch size is very small to capture enough texture information about the required landmark. The second and third patches show capture more discriminating texture-features around the landmark. Figure 4.39 shows the OOB classification error (see Section 2.2.3 for details) for the three RF models trained by LBP features extracted from the 3 different patch sizes (Figure 4.38).

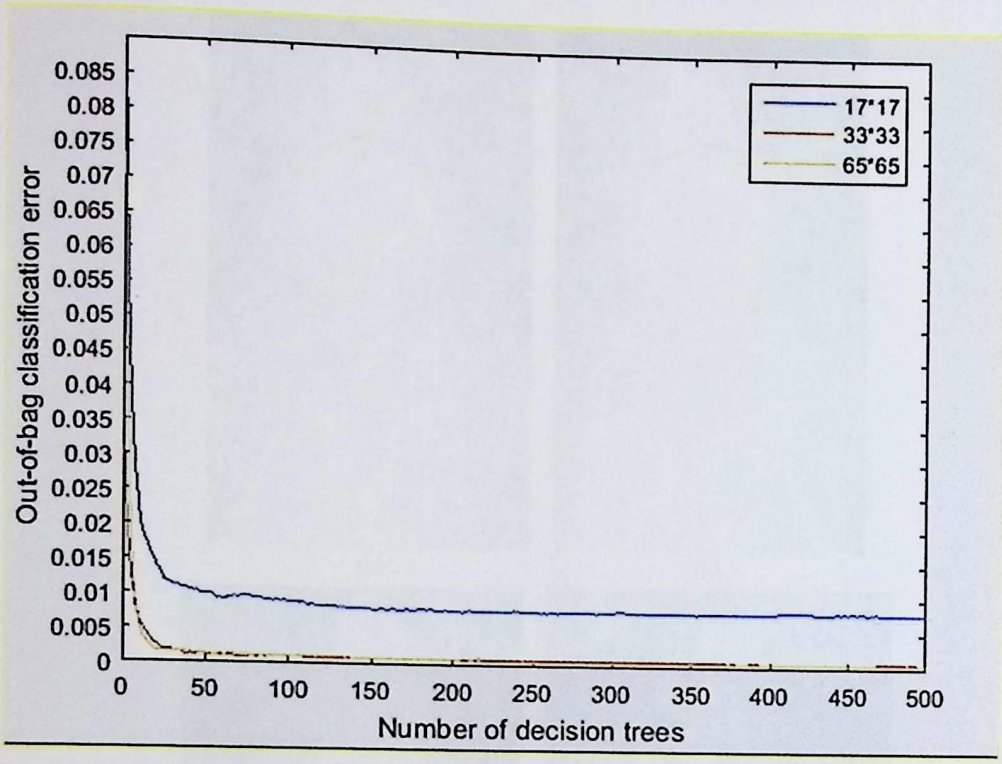


Figure 4.39 OOB classification error for three different RF-LBP models were trained by features extracted from patches of different sizes ((17 × 17), (33 × 33) and (65 × 65) pixels). The lowest error rates are associated with the RF trained with the features extracted from the patches with (33 × 33) and (65 × 65) pixels size.

The figure shows that the OOB classification error for RF-LBP model (17 × 17) pixels is higher than the other two sizes. It also can be seen that the error rate is very close between the models trained by features extracted from the patches of sizes (33 × 33) and (65 × 65) pixels. Following this analysis, we decide to select the size of the patch for width landmark detection to be (33 × 33) pixels. Because of both patches have very close error rate, we prefer the smaller patch size, (33 × 33) pixels, over (65 × 65) pixels patch because it leads to more efficient RF model especially in terms of features extraction.

4.5.4 Effectiveness of Proposed Method when Applied on Different Data set

In this section, we discuss if our proposed method could be applied on a different data set. For this purpose, we consider several samples of skull mouse X-ray images which produced by the MRC Harwell. It can be seen in Figures 4.40(a) and (b) that there are 3 key differences from the images of our data set: 1) the resolution of the image is less than those images produced by Sanger Institute, 2) the image include large parts of the arms and the rib area, and 3) the quality of the image in general is lower than our images in terms of clarity of required landmarks. To see how our method performs on the new samples, we run it firstly without any modification. We also consider using the intensity based method to locate the four landmarks. Figure 4.40(a) shows the result of modelling the skull and the four obtained landmarks for one sample.

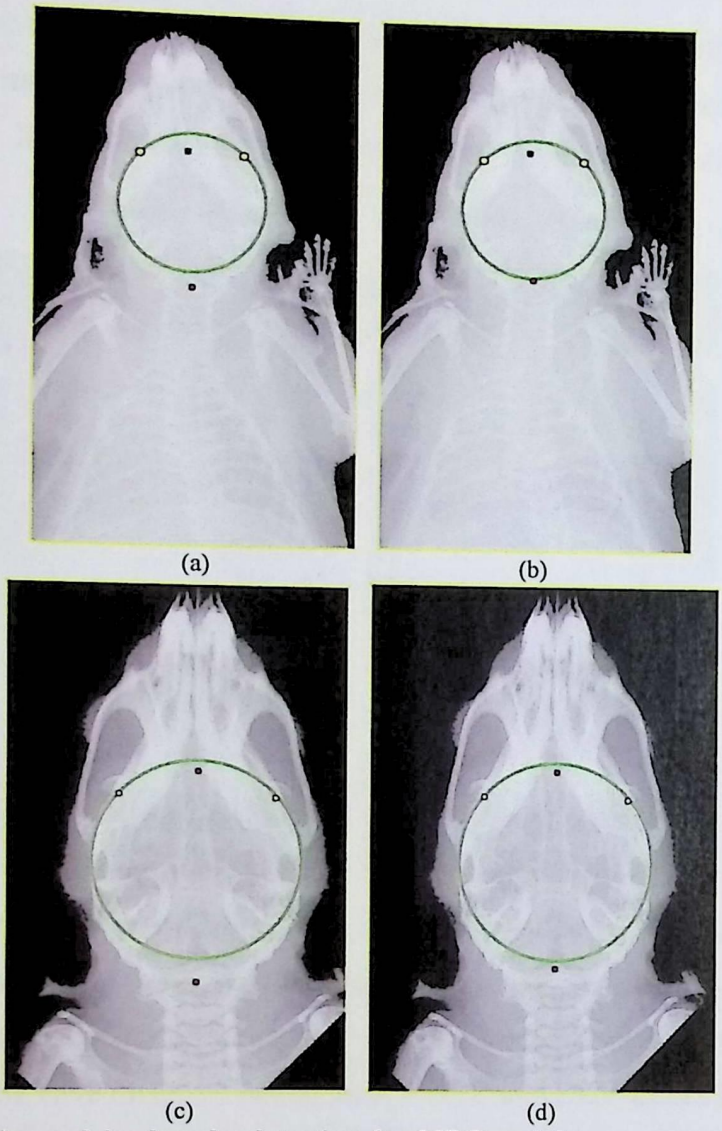
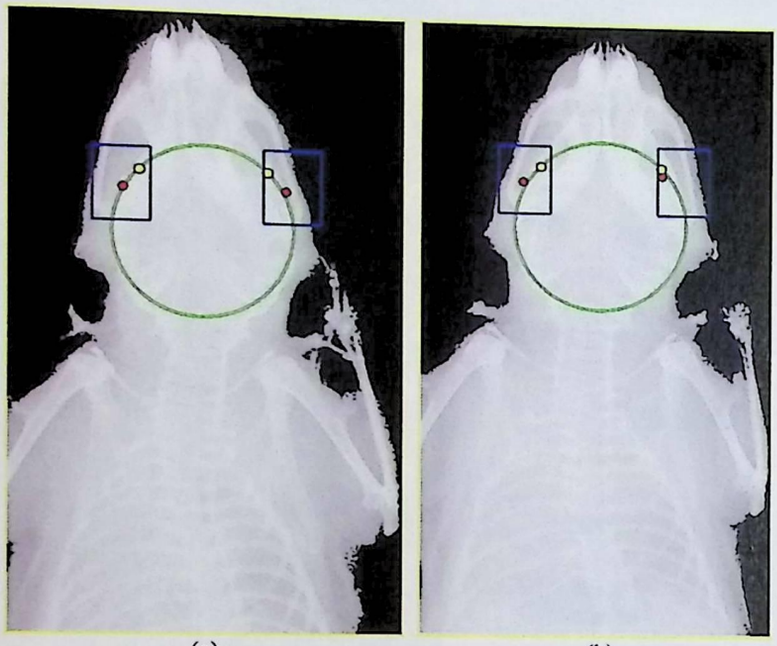


Figure 4.40 Circles and landmarks detection for MRC and Sanger mouse head X-ray images.
(a) MRC mouse head X-ray image annotation using original parameters designed for Sanger Institute data set; (b) MRC mouse head X-ray image annotation using tailored parameters; (c) Sanger mouse head X-ray image annotation using original parameters; (d) Sanger mouse head X-ray image annotation using tailored parameters

Subjectively, we can say that the initial position of the circle (see Section 4.3.2 for details) is located in a good position. In addition, W_1 and W_2 landmarks (see Section 4.3.5 for details) are not located accurately because of the coronoid processes for MRC images are not recognized. For the length landmarks, the accuracy is low. This is mainly because the radius of the search area selected to refine the initial detected landmark to the final position is not optimised for the new images. Figure 4.40 shows one example from MRC data set ((a) and (b)) and Sanger data set ((c) and (d)) before and after changing the parameters for the lower length landmark refinement. We can see that the accuracy of lower length landmark detection (red point in Figure 4.40(b)) for the new data set is enhanced while it decreased for the image from the old data set (red point in Figure 4.40(d)).

To investigate how the machine learning based solution is responding with the new data set, we used the same RF model that we trained it for our initial data set to predict the position of the width landmarks. Red points in Figure 4.41 show the detected points using machine learning approach.



(a) (b)

Figure 4.41 Width landmarks detection from MRC mouse head X-ray images using RF model.

(a) and (b) The width landmarks (red points) are predicted using RF model from the search area (blue boxes)

Initial results show the possibility of detecting the width landmarks for some cases. However, because the two datasets have different resolutions, search area (blue boxes in Figure 4.41) for the MRC dataset may need to be fine-tuned. To conclude, the proposed method showed possibility of modelling the skull from the new data set using HT which could help in locating different search areas. However, the limitations and inaccuracies that could happen in landmarks detection are mainly due to the different in images resolution and quality between the two data sets.

4.5.5 MATLAB Tool for Brain Size Estimation

MATLAB® was used to design a tool with a GUI as shown in Figure 4.42. The tool offers option to select the type of the mouse to be measured as WT or mutant mouse. Another option is to choose the gender of the mouse. The two options are essential to enable the tool making the final decision about the class of the mouse. After that, the user can load the mouse head image and produce "*Measurements Estimation*" to automatically locate the four different width and length landmarks and the distances between them. The obtained distances represent the width and length estimation of the brain and will be displayed by millimetres. Depending on

the estimated dimensions from the tested mouse and the RR calculated from the WT data set, the tool classifies the brain size as either "*Smaller*" or "*Bigger*" than the normal size, or "*In Normal Range*".

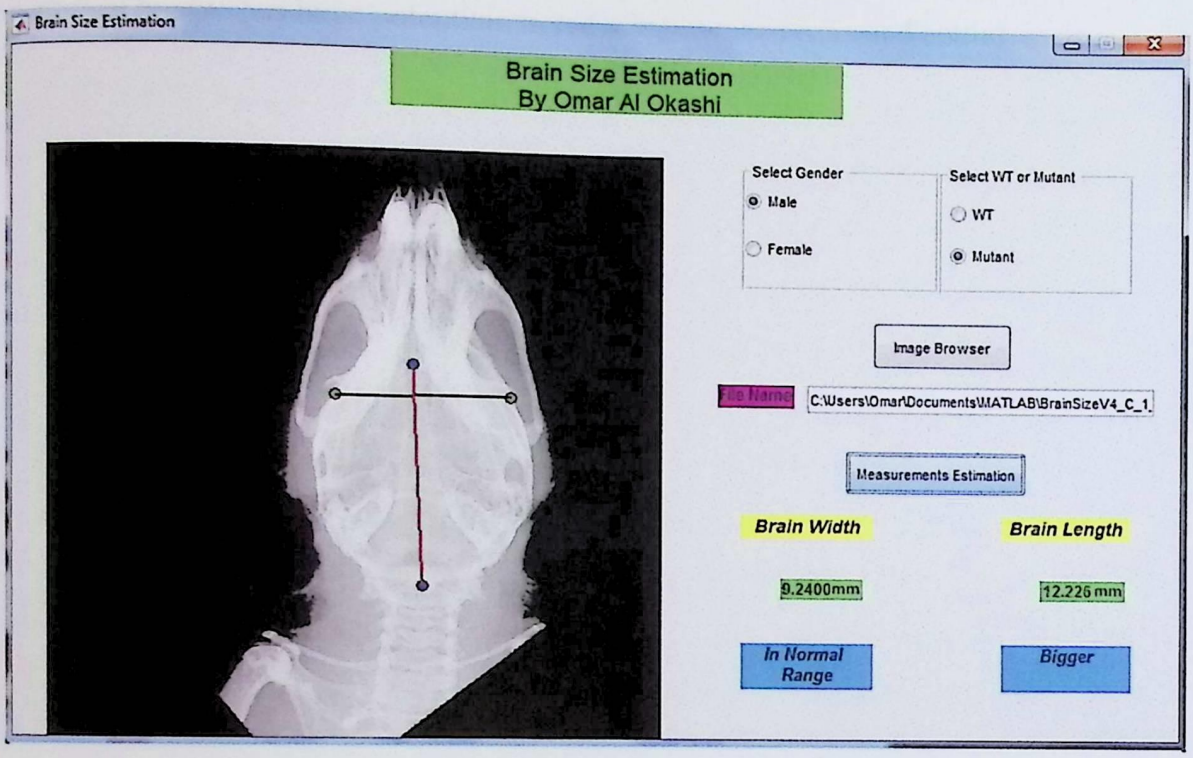


Figure 4.42 Snapshot of the brain size estimation MATLAB tool. The GUI enable the user to select the image using "Image Browser" button. In addition, it offers determination of the gender and whether the mouse is mutant or WT. The final results are expressed in terms of the measurements in mm and whether the measurements are in normal or abnormal range.

4.6 Summary

Accurate estimation of the brain size from high-throughput X-ray images of mutant mice is essential for understanding the genetic basis of brain abnormalities. We argued that the problem of landmark detection in medical and biomedical images is still an open problem where a one-fits-all solution does not exist, and that the machine learning solutions can assist landmark detection.

This chapter described a fully-automated algorithm that models the skull and locates different landmark points to estimate the width and length of the brain in X-ray images. The proposal combined the intensity-based and machine learning based techniques to locate the accurate final locations of the landmarks which directly influence the accuracy of brain dimension estimates. The estimation results produced by the proposed method succeeded in identifying a list of genes that might be associated with the changes in the brain dimensions. Some of these genes were already known to cause Microcephaly, which demonstrates the effectiveness of the proposed solution. At the same time, the proposed method helped in isolating a potentially

interesting unknown gene (NovelGeneX), further indicating the potential of the proposed method. Furthermore, the evaluation in terms of the manual versus automatic measurements illustrated it viability. The chapter finally argues that the automation of landmarks from biomedical images has huge potentials in reducing human errors and cutting costs.

Chapter 5: Automatic Spine Segmentation and Curvature Quantification

After estimating the brain size and finding a list of possible genes that affect the brain size, the next body part to consider is the spine. The shape of the spine is prone to changes (i.e. changes in its curvature) as a result of different diseases. Automatically generated metrics to accurately assess such changes in the shape of the spine from medical images is thus of a great importance.

This chapter proposes a fully-automated solution to segment the whole spine in X-ray images of mice and then quantify its curvature with both angle-based and index-based metrics. The proposed solution consists of different stages to be applied on the segmented spine ROI from the aligned full body mouse image. The alignment of the body to upright position should reduce the complexity of the solution. The binary spine mask will then be segmented from the enhanced spine ROI. Because the segmented mask could affect the accuracy of all later stages, some criteria should be checked to decide the goodness of the mask. Upon this check, the image is either proceeded to next step, refined, or rejected. Extracting smooth contour from the spine mask is also essential to extract an accurate spine central line before using it to generate six different metrics to quantify the curvature of the spine. The fully automatic process aims at reducing time and cost for high throughput situations on the one hand, and producing consistent quantification measurements without *inter*- and *intra*-observer variations of the manual measurement on the other. To our knowledge, our work is the first attempt that deals with fully automatic spine segmentation and curvature quantification for mouse X-ray images.

The automatic solution must meet the following technical challenges:

- The solution must produce an accurate segmentation of the spine by overcoming difficulties relating to low contrast within the ROI and blurred spine borders.
- The solution must accurately locate important landmarks, i.e. the end and the apical vertebrae for accurate curvature quantification.
- The solution must produce angle-based and index-based metrics to quantify the curvature of the spine.

The effectiveness of the curvature metrics extracted from the X-ray images of normal and abnormal spines will be then analysed. The experimental results show that some of the automatic measures are close to and consistent with the best manual measurement results by

annotators. The test results also demonstrate that most of the curvature quantification metrics produced by the proposed solution are effective in distinguishing abnormally shaped spines from the normal ones.

The rest of this chapter is organised as follows. Section 5.1 describes the problem of spine curvature and explains the main measures to quantify the curvature. In Section 5.2, we review the existing solutions for the spine segmentation and curvature quantification in the literature. Section 5.3 presents the proposed solution in terms of the detailed operations and justifies why these operations are necessary. In Section 5.4, we present the experimental results and analyse the effectiveness of the proposed solution. Section 5.5 discusses the results, different possible solutions to deal with the problem, source of inaccuracies and the implementation of a software tool for researchers based on the proposed method. Finally, the summary of the chapter is presented in Section 5.6.

5.1 Problem Statement

The spine is a main structural component of a mammal's body skeleton. It has the functions of containing the nerve cord, supporting the body and protecting the thorax and abdomen areas of the body. This important skeleton component normally curves naturally. However, certain conditions, which may well be genetic, may cause the spine severely curving towards the left or to the right side of the body away from the normal position (Lonstein, 1999). Such deformity of the spine can seriously affect the body's normal movements and functions. Therefore, appropriate quantification of spine curvature serves the purpose of detecting if a spine is normally shaped or severely curved, and if it is latter, how severe the abnormal curvature degree is. In practice, precise quantitative evaluation of the curvature serves as the first step toward diagnosing spinal diseases such as scoliosis.

Measurement of spine curvature related to scoliosis is usually extracted from the dorsoventral view of medical images ((Tang et al., 2008), (Zhang et al., 2009), (Kundu, Chakrabarti and Lenka, 2012) and (Sardjono et al., 2013)). Different ideas have been attempted to estimate the curvature of the spine either as an angle or as an index value. Vrtovec, Pernuš and Likar (Vrtovec, Pernuš and Likar, 2009) reviewed five main proposals in quantifying the spinal curvature from 2D dorsoventral X-ray images of humans ((Cobb, 1948), (Diab et al., 1995) , (Chen, Chen and Chiou, 2007) , (Ferguson, 1930) and (Greenspan et al., 1978)), as illustrated in Figure 5.1. With all the five methods being applied to the same spine model, all the angle-based measures as shown in Figure 5.1(a), (b) and (c) are fundamentally the same, and the angle-based measure illustrated in Figure 5.1(d) is in fact the complement of the other angle-

based measures. The difference lies in the method of computation. Greenspan et al (Greenspan et al., 1978) described a measure of the spine curvature as an index rather than an angle as shown in Figure 5.1(e). Among all the five proposed measures, Cobb angle was formally adopted as the *de facto* standard for scoliosis by Scoliotic Research Society (SRS) in 1966 (Vrtovec, Pernuš and Likar, 2009). Using any methods for accurate quantification of spine curvature manually, as depicted in Figure 5.1, requires accurate locations of end and apical vertebrae, their centres and/or their endplates.

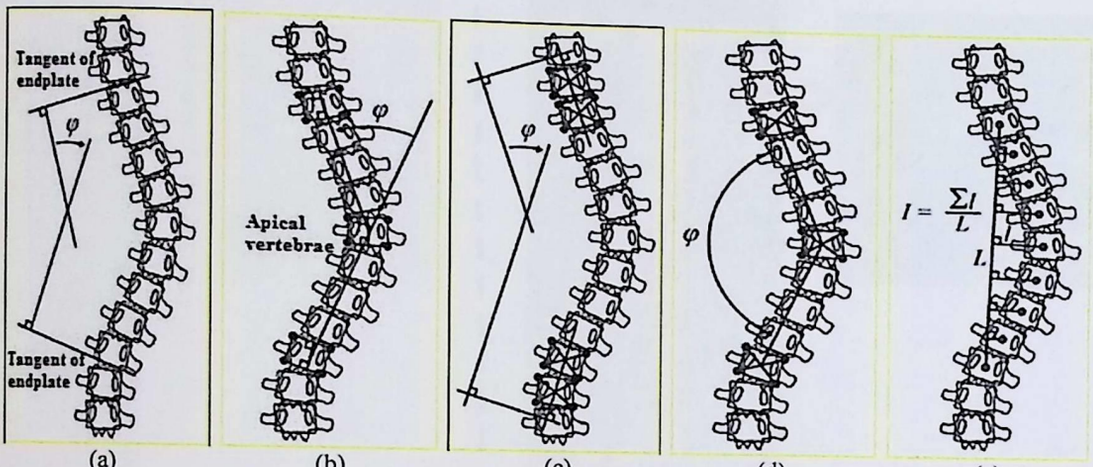
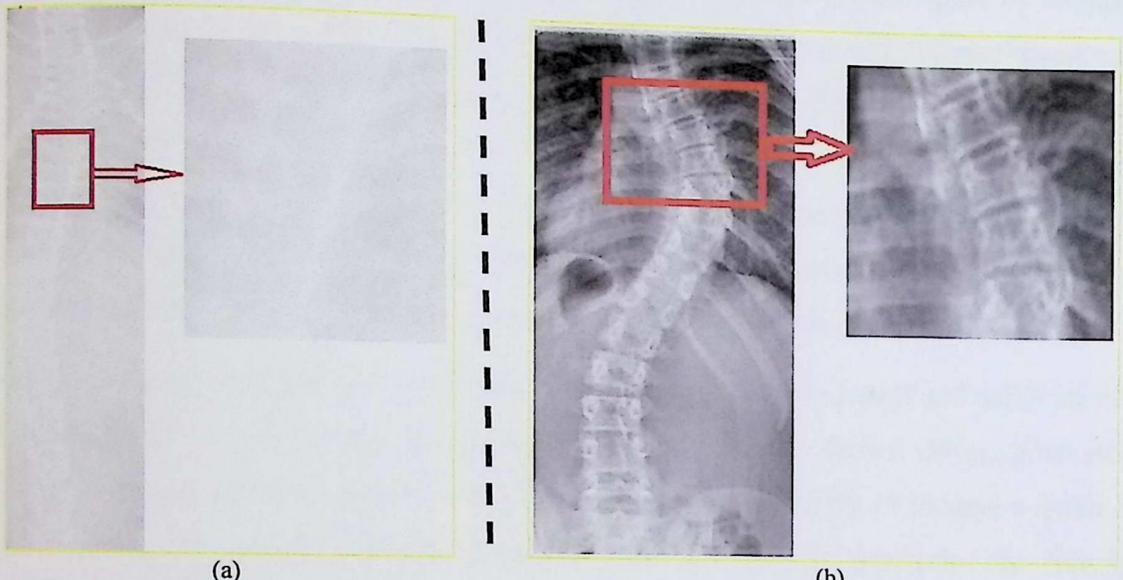


Figure 5.1 Spinal curvature measurements, adapted from (Vrtovec, Pernuš and Likar, 2009) (a) Cobb method: estimating the curvature as the angle φ between the two tangents to the upper and lower endplates of the upper and lower end vertebra respectively; (b) Diab et al. method: estimating the curvature as the angle φ that is formed between the intersection of lines extended from both end vertebrae and intersect at the centre of apical vertebrae. The centre of each vertebra is determined by the intersection of lines orthogonal to the upper and lower endplates; (c) Centroid method: estimating the curvature as the angle φ between the lines passed through the two upper and through the two lower vertebral centroids. The centre of each vertebra (centroid) was found as the intersection point of two lines extended between the opposite corners of the vertebrae; (d) Fergusson method: estimating the curvature as the angle φ between the two lines that extended from centres of the end vertebrae to the centre of the apical vertebra. (e) Greenspan index: estimating the curvature as the index ($I = \frac{\Sigma l}{L}$) where L is the length of vertical line extended between centres of upper and lower end vertebrae and l is the length of each orthogonal line between the centre of each vertebrae in spine curve and the vertical line

The low quality of mouse's spine images adds extra difficulties in automated solutions for X-ray images for mice than those for humans. The X-ray images generated by the MGP tend to show many blurry borders around the vertebrae, which makes it difficult even for domain experts to precisely locate the borders. This means that manual measurement of spine curvature will encounter *inter-observer* and *intra-observer* variations. An automated computer-based solution will overcome this problem by always giving a deterministic and consistent result. However, such solutions still face difficulties in accurate segmentation of vertebrae and identification of the end and apical vertebrae. We believe that the main cause of the problem is that the full body mouse X-ray images are not optimised to zoom in a particular part of the body (i.e. the spine), resulting in poor contrast in areas for vertebrae and endplates, especially in the upper part of the chest area. Figure 5.2 shows an example X-ray image of a mouse spine

structure (a) in comparison with that of the human spine structure (b). The figure has shown the difference in terms of contrast and brightness between the human and mouse spine X-ray images as highlighted by the red boxes. It is clear that the level of contrast between the vertebrae and the backgrounds for the mouse image is much lower than that for the human image. Finding a novel solution to overcome the difficulties is the focus of our proposed solutions.



(a) (b)

Figure 5.2 Comparison between spine X-ray images of mouse and human where the red boxes zoomed into the upper parts of the chest area.
(a) X-ray image of mouse spine which shows unclear vertebrae and endplate especially in the upper part of the spine as illustrated by red box; (b) X-ray image of human spine which shows clear vertebrae and endplate

5.2 Literature Review

In the literature, the work dedicated to the mouse spine, especially in terms of automatic segmentation and curvature quantification of mouse spine from X-ray images, if not at all non-existent, seems very limited. However, several works about spine and vertebra enhancements, detection and segmentation from different medical imaging modalities for humans have been reported. In the existing work, there are several goals behind segmenting and labelling the vertebrae from the human spine images such as estimating the curvature of the spine and diagnosing vertebrae fractures.

Peng et al. (Peng et al., 2005) addressed the case of whole spine detection from MR images. The proposal was based on intensity profile generation from the 2D sagittal slice that shows the largest number of disks. From the intensity profile, the centres of N intervertebral disks were located. By applying a small window around initial located centre, the pixels that belong to this disk can be determined, and then more accurate centre and boundary of disks were

determined. Based on located disk locations, the vertebrae locations were annotated and then segmented using the Canny edge operator.

Moura et al. (Moura et al., 2007) analysed an X-ray image for vertebrae detection in patients with scoliotic spines. Firstly, they built a multiresolution Gaussian model to process the image at lower resolution which helps in omitting small details and facilitating the detection of the spine. Their method next determined the approximate location for the spine by counting the number of bright pixels in every column and consider the column with highest counting as the spine column. This is then followed by isolating the spine from any attached bones and detecting the vertebrae location along X and Y directions using hierarchical and progressive threshold analysis. Unfortunately, no experimental results were presented in (Moura et al., 2007) to evaluate the effectiveness of the method. Column-wise counting of the brightest pixels also seemed a crude way to locate the spine, and computationally expensive.

Gabor filters for edge detection with different angles were used to detect and delineate plateaus of lumbar vertebrae in X-ray images (Ribeiro et al., 2010). The original image, processed with a bank of 180 Gabor filters where the Gabor filters were rotated by 1° to span angular degree from 1° to 180°. A semi-automated procedure was proposed to determine the four lumbar vertebrae manually from the original image and the output of Gabor filters. ROIs resulting from the semi-automated procedure in addition to the result of manual delineation of the plateaus and Gabor magnitude response were used to train the neural network to estimate the overlap between the manual delineation and the result of automatic detection. Al-Helo et al. (Al-Helo et al., 2011) proposed a method to segment a lumbar vertebra from CT scan images. Their method started by finding the location of each intervertebral disk to determine the centre of each human lumbar vertebra in a CT scan image. Active Shape Model (ASM) was then used to detect a rough contour for each vertebra. The initial contour located by ASM was then fed to Gradient Vector Flow to find more accurate edges and concavities. Both (Ribeiro et al., 2010) and (Al-Helo et al., 2011) focused on finding the fine boundaries for vertebrae in a specific part of the spine.

Lootus, Kadir and Zisserman (Lootus, Kadir and Zisserman, 2014) proposed a method to detect and label vertebrae in lumbar MR images. Their proposal was based on training Deformable Part Model using HOG features to annotate candidate boxes around each vertebra after searching all the MR sagittal slices. Next, a suppression algorithm was implemented to remove most of the false positive boxes. The remaining vertebrae candidates' boxes were then used as an input to the graphical model to localize and label more accurate position for each vertebra.

Aubert et al. (Aubert et al., 2016) proposed a learning based method to detect a spine and pelvis location in 2D radiograph images. In the training stage, they sampled 100 patches around each annotated landmark (of vertebrae and femoral heads) and then used the intensity levels of those patches as features to train the Deep Neural Network (DNN) to predict the displacements between the patch and the ground truth landmark. In the detection stage, the Statistical Shape Model (SSM) was translated on the suitable coordinates in the input image based on the observation that the pelvis centre location is defined as a point where the sum of rows and columns intensities was maximal in the lower half of the image. An initial position of patch for every landmark was estimated via the SSM and then converged with the predicted displacements from the DNN to an accurate position of the landmark.

Several researches in the literature consider implementation of an automatic and semi-automatic techniques to estimate the curvature of the spine as the Cobb angle or as an index in the human X-ray images. In addition to the survey by Vrtovec, Pernuš and Likar (Vrtovec, Pernuš and Likar, 2009) as mentioned before, a more recent review but with fewer details was conducted by Janumala and Ramesh (Janumala and Ramesh, 2015).

Tang et al. (Tang et al., 2008) described an automatic method to generate and evaluate an index to assess the spinal curvature scoliosis in an X-ray chest image. They worked out the central curve of the spine and then found the best fitting line of this curve. The index was expressed as the Root Mean Square Error (RMSE) between the central curve and the best fitting line. To find the threshold value that discriminate between normal and abnormal spines, an expert was used to read the resulting indices from group of images and decide the index of the normal spines.

Zhang et al. (Zhang et al., 2009) developed a method that used Fuzzy Hough Transform (FHT) to detect the slopes of the end vertebrae from which the Cobb angle was calculated. Manual pre-processing operations include cropping of the ROI, adjusting the contrast of the image and selection of end vertebrae were applied first. After that, edges of the upper and lower vertebrae were detected using the Canny edge operator. Those detected edges were then fed to FHT and the candidate peaks from the Hough space were selected based on the vertebral shape prior. Finally, the Cobb angle was calculated as the sum of the angles of both end vertebrae.

Kundu, Chakrabarti and Lenka (Kundu, Chakrabarti and Lenka, 2012) proposed another method for performing the Cobb angle measurement. As a pre-processing step, they manually selected the vertebra ROIs, followed by de-noising the image by using Non-Local Euclidean

Trimmed Mean and enhancing its contrast by using histogram equalization. After that, the Canny edge detection operator based Otsu's threshold was used to detect edges from the images of the pre-processed ROIs. Finally, HT was executed for vertebrae slopes detection and then the Cobb angle was measured from those slopes. Another proposal to estimate the Cobb angle from an X-ray image was presented by Samuvel et al. (Samuvel et al., 2012). The centre of each vertebra was initially placed manually and then updated by a vertebrae mask that finds the best fit of the mask on the vertebrae through changing its angle, scale and size. Once the best score for the mask was found, the angles between the lines connecting the centres of the vertebrae and the horizontal x-axis were found. Finally, the Cobb angle is calculated as the difference between the maximum and minimum angles.

Sardjono et al. (Sardjono et al., 2013) have proposed a fully automatic Cobb angle determination method by using a modified Charged Particle Model (CPM) and different curve fitting methods. At first, they found the edges of the spine by using the CPM. Next, they used different curve fitting through obtained particles to make better alignment for spine edges. Then they divided the spine into three or two parts depending on the number of curvature in the spine. For each part, they found the angles for each curve fitting method as the angles between the lines connected group of particles and the x-axis. After that they determined the Cobb angle(s) as a difference between the most extreme angles for consecutive parts.

The most recent work to estimate the Cobb angle measurement from radiograph images was presented by Zhang et al. (Zhang et al., 2017). They used pixels' intensity levels extracted from vertebral patches to train a DNN to detect the borders of the end vertebrae. Cobb angle could be then calculated using the slopes of the end vertebrae. Their method was not fully automatic where the user is prompted to manually select the end vertebrae during test stage. However, they showed that their results are comparable to manual estimation of the Cobb angle with MAE less than 3°.

Another recent work presented by Wu et al. (Wu et al., 2017) in which BoostNet was used to automatically annotate spinal landmarks from scoliotic spine. The main idea of their work was to use boost layer of the BoostNet to enhance the feature space by reducing outlier features. They also proposed to improve the regression accuracy through using a prior output layer that capture the dependency information between the spinal landmarks.

The literature as described above has some interesting ideas and thoughts to segment the spine and quantify its curvature. However, there are a number reasons that make the presented

solutions not applicable 'as is' to automatically estimate the Cobb angle for mouse curvature quantification.

- Human X-ray images tend to be localised, i.e. it is optimised to cover a particular part of the body under investigation while this is not the case for the mouse images. This makes the clarity of X-ray images of human spine structures tends to be much better than the X-ray images of mouse spine structures obtained from the MGP. This means that many existing solutions may not be applicable in coping with the less clear images of mice.
- Most existing solutions are semi-automatic where manual intervention is required (Zhang et al., 2009) , (Kundu, Chakrabarti and Lenka, 2012), (Ribeiro et al., 2010) ,(Samuvel et al., 2012) and (Zhang et al., 2017)). This type of intervention is not feasible in processing large volume of images in a high throughput situation. Besides, all existing solutions only produce one type of spine curvature metric either as the Cobb angle or as an index, but not both.
- Most of the solutions for spine segmentation and/or curvature quantification are vertebra based solutions. As we explained in Section 5.1, the vertebrae in mouse spine of X-ray images, especially the upper ones, have unclear borders compared to human vertebrae. This fact makes using vertebra based solutions for curvature quantification infeasible for spine mouse X-ray images. More details about suitability of using vertebrae-based solutions on our data set will be discussed in Section 5.5. However, there are two exceptional solutions based on full spine segmentation instead of vertebra to quantify the curvature of the spine. The first one is the index based proposal presented by (Tang et al., 2008). An interesting idea in this solution is the use of the central line of the spine to estimate its curvature. However, no details were given about the method used to obtain the spinal curvature line. The other solution is the one presented by (Sardjono et al., 2013) to find an initial landmarks of the spines' borders and then using different curve fitting solutions to refine the borders. The use of curve fitting solutions is one of interesting idea that help to fit more accurate whole spine segmentation.

5.3 The Proposed Method

In this section, we propose a fully automatic method to segment a spine from the mouse X-ray image and quantify its curvature using different metrics. Our proposed solution is based on accurate extraction of the outer borders of the spine which is made possible by a sequence of image enhancement techniques and reliability measures check. A single central spinal line

could be then extracted from the outer borders, and several critical landmarks can then be located on this spinal line. Those landmarks are responsible for generating several angle and index based measurements to estimate the curvature of the spine. The seven main stages of the proposed solution are illustrated by the flowchart in Figure 5.3.

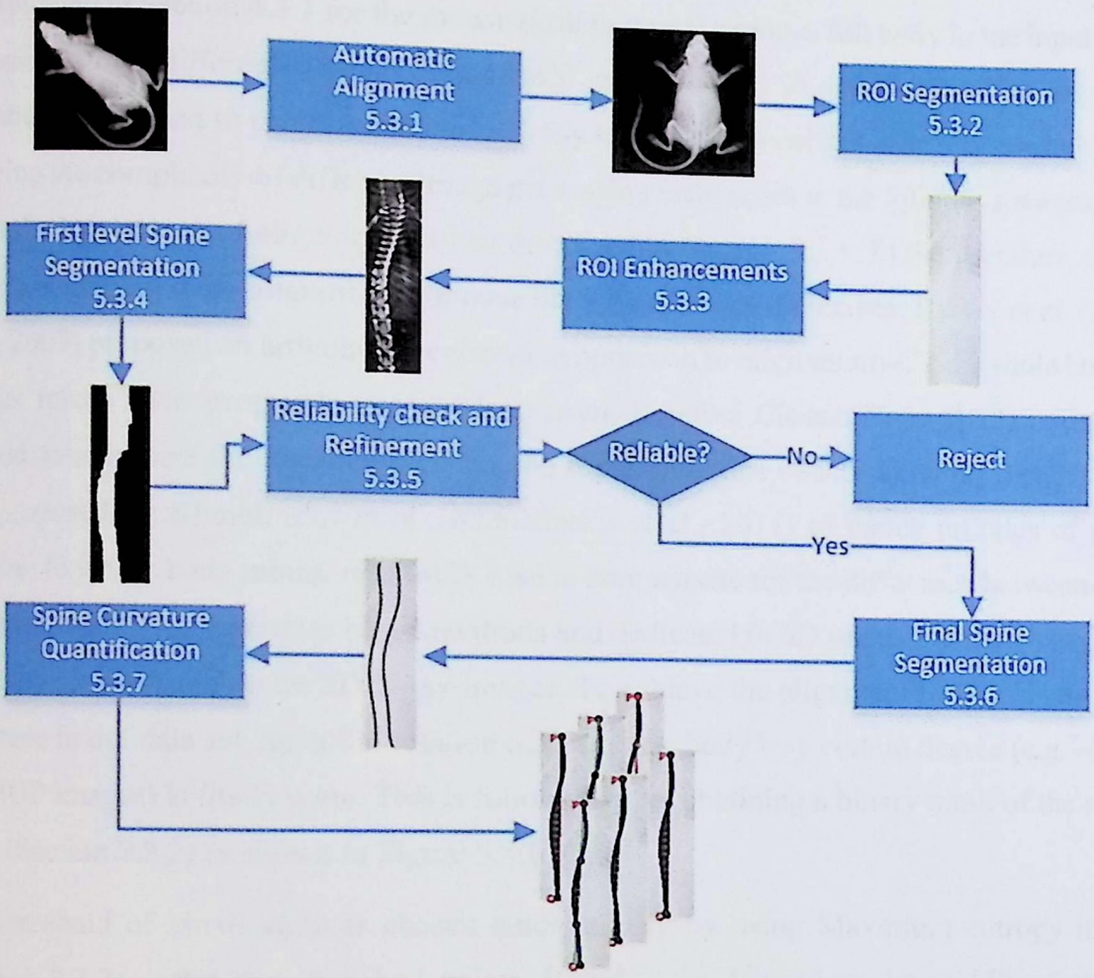


Figure 5.3 Main stages of spine segmentation and curvature quantification from mouse X-ray image method. The method starts by aligning the image vertically and then cropping the spine ROI. This is followed by enhancement and segmentation of spine. At this stage, the reliability check is done to either accept, refine or reject the segmented spine. Finally, the curvature quantification of the spine are estimated as a six angle and index based metrics.

The process starts by aligning the full body of the mouse to the vertical position. Bringing the mouse body in all the images to a known range of vertical angles will reduce the complexity of the next stages especially in terms of defining critical landmarks. Next, the ROI, i.e. the spine area, is segmented from the rest of the mouse body. This is followed by enhancing the ROI to enable the first level segmentation of the spine from the ribs and other unwanted objects within the ROI. At this stage, it is important to check the reliability of the spine mask segmentation and refine/reject the image accordingly. Once the spine mask passes the reliability check, the second level spine segmentation is done to obtain a fine contour of the

spine. The method finally derives angle and index based measurements metrics from the final segmented spine after locating the crucial landmarks. The following sub-sections will explain each step in more details.

5.3.1 Alignment for Full Body Dorsoventral Input Image

As explained in Section 4.3.1 for the mouse skull image, the mouse full body in the input image is positioned in different degrees around 45° as shown in Figure 5.4(a). The alignment operation is intended to position the mouse body to a vertical position. The alignment helps in reducing the complexity of different image processing techniques in the following stages. More about the importance of alignment will be discussed in Section 5.3.5. In the literature, several proposals discussed the alignment of mouse body position in 3D images. Baiker et al. (Baiker et al., 2007) proposed an articulated registration approach to align micro-CT of whole body 3D data of mice. Their proposal was based on using Iterative Closest Point (ICP) registration method to minimize the distance between the representations of the skeleton. Another work was proposed by Khmelinskii et al. (Khmelinskii et al., 2011) to match an atlas of mouse skeleton to whole body mouse micro-CT scan to compensate for the difference between scans. However, both works are atlas based methods and dedicated to 3D micro-CT scan images and thus they are not suitable for 2D X-ray images. To achieve the alignment for 2D X-ray image of mouse in our data set, an initial rotation of the mouse body by a certain degree (e.g. -45° for the MGP images) is firstly done. This is followed by an obtaining a binary mask of the rotated body (Section 2.2.2) as shown in Figure 5.4(b).

The threshold of binarisation is chosen automatically by using Maximum entropy method (Section 2.2.2). After that, two key points, P and Q are located as the middle of the two horizontal lines that pass approximately through the centre of the head and the centre of the body of the mouse respectively as illustrated in Figure 5.4(b). Heuristically, for the images of the MGP, the two horizontal lines (red lines in Figure 5.4(b)) are set to be 250 and 1000 pixels below the topmost point of the mask (green point in Figure 5.4(b)). However, any vertical position for the lines on the upper and lower parts of the mask below the topmost point could be used for the alignment procedure. If the angle θ between the vertical line passing Q and the line formed by P and Q as shown in Figure 5.4(c) is less than 2°, initial rotation by -45° is considered to be satisfactory and no further rotation is required. Otherwise, the mouse body position in input image as shown in Figure 5.4(a) is rotated by the angle equal to (θ + -45°). This check is done to increase the efficiency of the process. The final aligned image is shown in Figure 5.4(d). The main concepts of the alignment of the full body in terms of selecting two

points and calculating the slope between them is similar to what has been done in the alignment of head in Chapter 4. However, the difference here is that the lower alignment point is selected to be in the middle instead of the lowest point of the body. This selection could be more suitable because the presence of the limbs and tail might complicate setting the point in the lower part of the body.

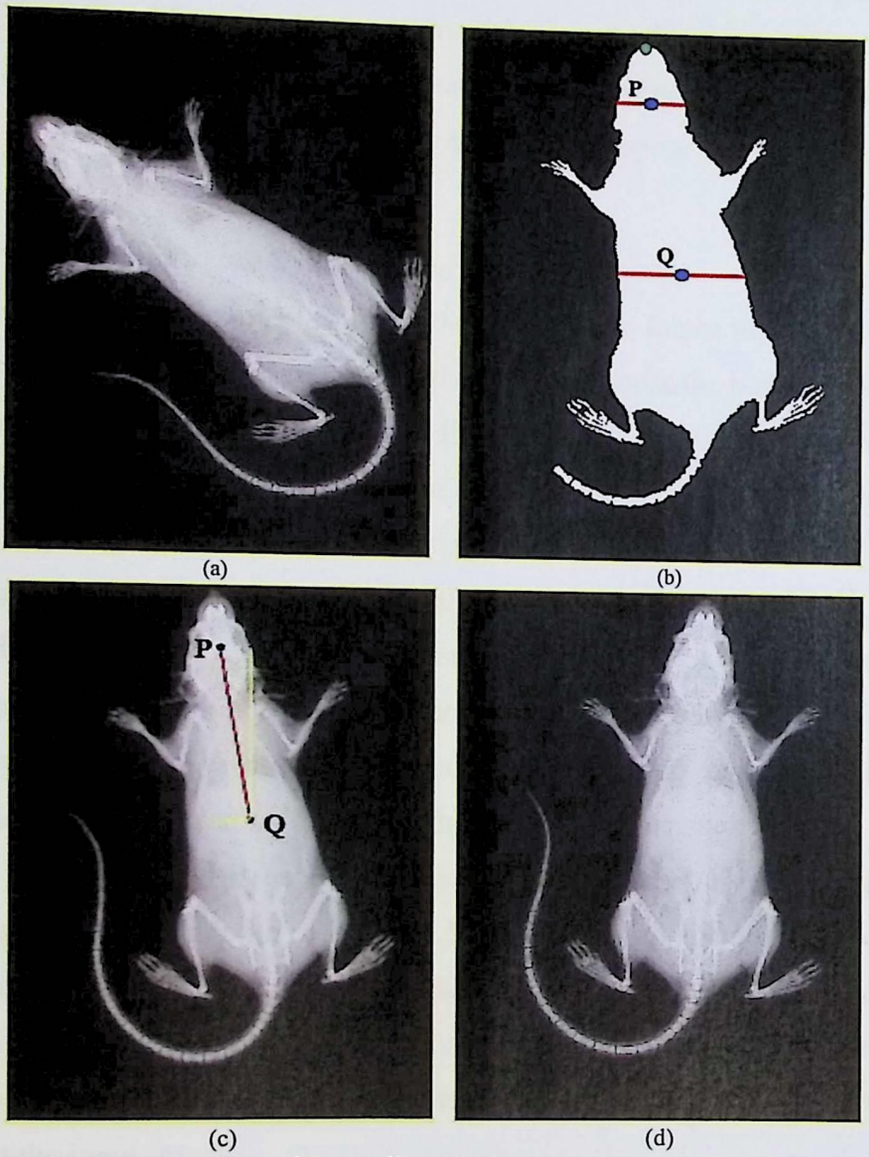


Figure 5.4 Full body mouse X-ray image alignment process.
(a) Full body mouse X-ray input image before alignment; (b) Rotate of (a) by (-45°) and then convert it to a binary image. Annotation of the topmost point of the mask (green point), the middle of the horizontal distance (red line) at the head of the mouse (P) and the middle of the horizontal distance (red line) at the abdomen of the mouse (Q) blue points on the mask ; (c) P and Q key landmarks and the line between them (red line) superimposed on the rotated input image; (d) Rotating the input image by the angle (-45 + θ) where θ represents the angle between P-Q line and y-axis line

We argue that although the alignment may not be a mandatory step for the proposed method, it does decrease the level of the complexity in the later stages such as cropping of ROIs.

5.3.2 Spine ROI Segmentation

This segmentation stage is essentially about correctly locating the coordinates of the spine ROI from the full body mouse X-ray image, i.e. a complete spine area from the bottom of the skull to the top of the pelvis must be correctly isolated from the whole image. With this crucial first step, the later stage for correctly identifying the spine contour becomes not only easier but also more efficient. It is realised from our observation of the images that the coordinates of the ROI can be estimated from the location of some bones near to the spine (i.e. Limbs and pelvis). We will refer those locations as *global key landmarks* and we propose the following approach to locate them.

Global landmarks detection using a mask based approach

This approach is based on obtaining a limbs-and-skull mask to locate the required landmarks. To do so, we propose to increase the contrast of the limbs. Because the bones of the limbs show sharp edges in an input image, using high-pass filtering could help in highlighting them. The BHPF (Section 2.2.1) with a relatively small cut off frequency (10) can to some extent do the job as shown in Figure 5.5(a). However, totally removing the average intensity of the image by BHPF will lead to a very poor contrast of the limbs to be segmented later. Instead, we propose to use High-Frequency Emphasis Filtering as described in Section 2.2.1 and the result as shown in Figure 5.5(b). Using High-Frequency Emphasis Filtering instead of BHPF only will considerably increase the contrast of the limbs. After that, the region grow segmentation technique (Section 2.2.2) is used to extract high contrast objects such as limbs and skull as shown in Figure 5.5(c). The consistent and known locations of the limbs helps in preparing the final limbs-and-skull mask as shown in Figure 5.5(c) by removing the unwanted objects in the middle and lower of the body. The resulting mask (Figure 5.5(d)) could help to obtain four global key landmark points: the right most point of the left arm, the left most point of the right arm, and the top most points of both right and left legs. Figure 5.5(e) shows those landmarks superimposed on the input image.

We will further discuss the possibility of using RF based alternative approach to detect the global landmarks in Section 5.5.3.

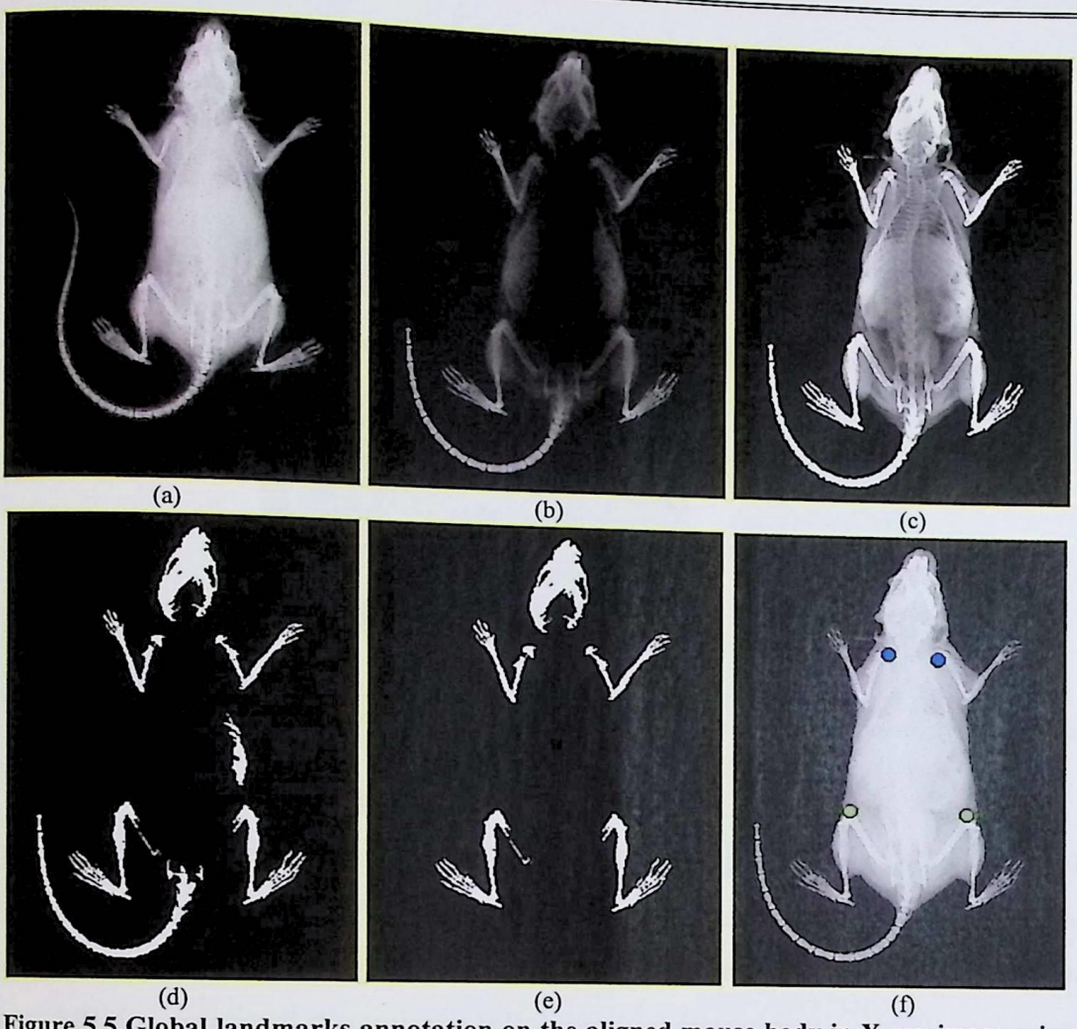
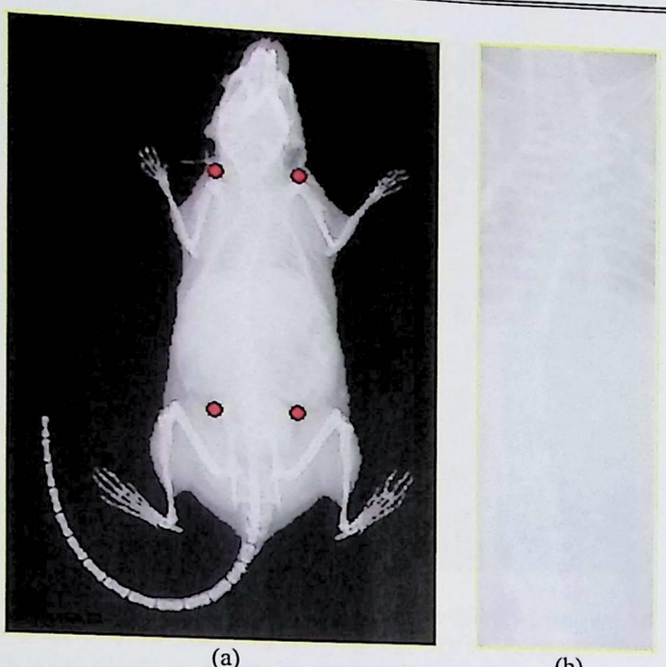


Figure 5.5 Global landmarks annotation on the aligned mouse body in X-ray image using mask based approach.
(a) Aligned mouse body X-ray image; (b) Result of applying high-pass filter on (a) which shows very low contrast for the limbs; (c) Result of applying emphasis high-pass filter on (a) which leads to high contrast limbs; (d) Result of applying region-growing segmentation on image in (c); (e) Resulting Limbs-and-skull mask after removing unwanted objects and the tail; (f) The four annotated global landmarks superimposed on the input image. The global landmarks are obtained from the Limb-and-skull mask in (e)

Once we locate the global key landmarks, we can use them to estimate the coordinates of the spine ROI. To approximate the width of the spine ROI without losing any part of spine, we decide to use the x-coordinates of the upper two global key landmarks (blue points in Figure 5.5(f)). Even the width seems wider than the width of the spine, those points are chosen to guarantee including all area of the spine especially in the cases where there is a severe spine curvature to the left or right. For the length of spine ROI, the upper point is selected as the y-coordinate of the upper global landmark -50 pixels. For the lower point, the y-coordinate of the lower right global landmark is used. This selection will cover the area of the interest. Figures 5.6(a) and (b) show the superimposing of the four points for the ROI on the input image and the segmented spine ROI respectively.



(a) (b)

Figure 5.6 Spine ROI segmentation using global landmark points.
(a) Annotated landmark points of spine ROI on the aligned mouse X-ray image (red points); (b) Segmented spine ROI based on global landmark points

5.3.3 Spine ROI Enhancement

This step aims to prepare the image for accurate spine segmentation through enhancing the contrast of the segmented ROI. As shown in Figure 5.6(b), the cropped ROI may have very poor contrast that affects the accuracy of the spine segmentation. To deal with this problem, the cropped ROI is enhanced through the following two operations:

1- Enhancing image contrast by applying the CLAHE (Section 2.2.1) twice on the segmented ROI of Figure 5.6(b). To get a satisfactory level of contrast enhancements, we set the number of tiles for CLAHE to be (4 × 2). This selection considers the size of the image and the relative small width of ROI. The effect of this operation is shown in Figure 5.7(a).

2- As we can see from Figure 5.7(a), the spine borders have a relatively high intensity values than other objects in the image. To highlight high intensity borders of the spine and eliminate low contrast irrelevant structures, we apply Gamma transformation (Section 2.2.1) with a relatively high value for γ parameter, (5), and the result is as illustrated in Figure 5.7(b). As we can see from Figure 5.7(b), a lot of unwanted tissues and background artefacts has been diminished. However, if we increase the value of γ to 10, we can even reduce more of unwanted structure and tissues as shown in Figure 5.7(c). This of course will be accompanied by decreasing the contrast of the spine which makes segmenting it in the next step more difficult.

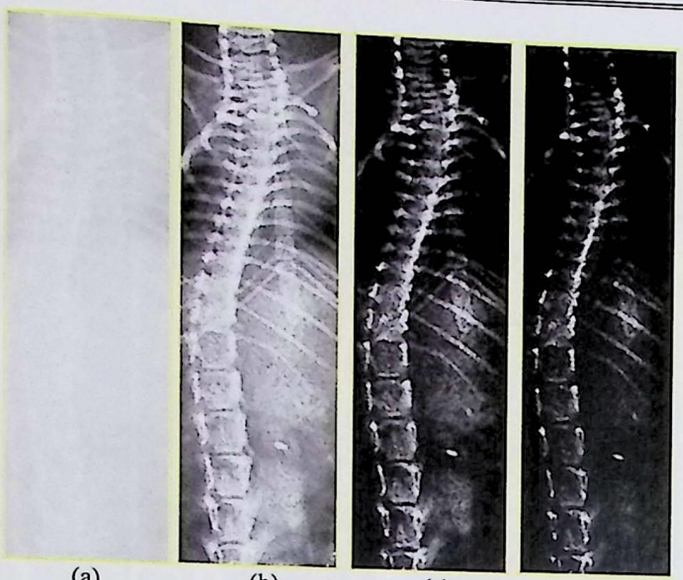
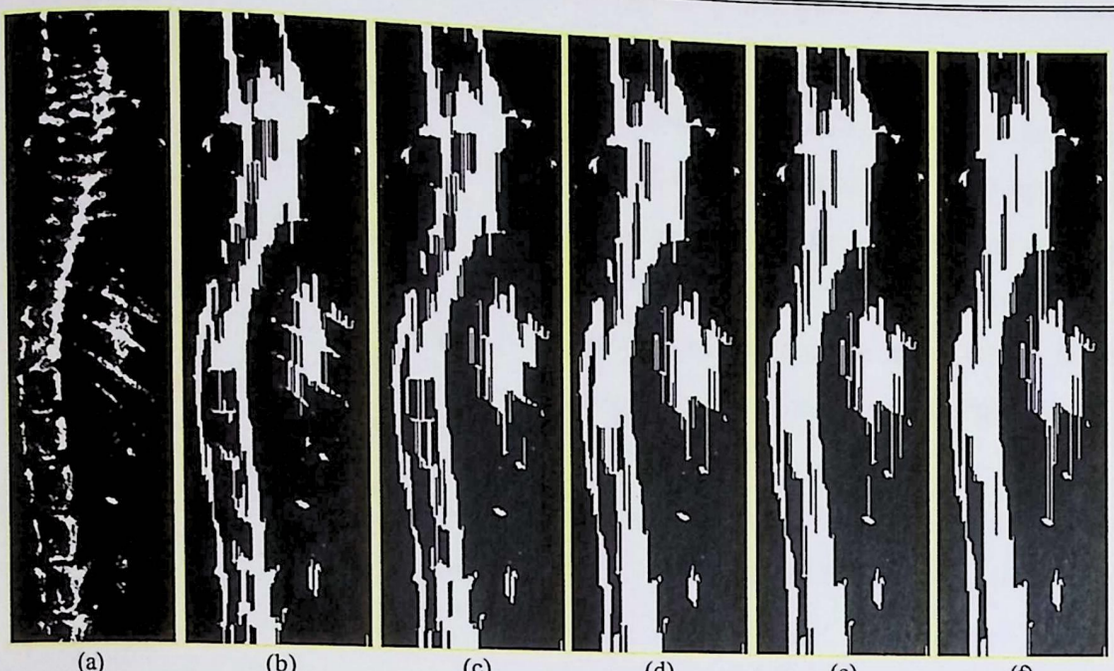


Figure 5.7 Enhancing the spine ROI cropped from mouse X-ray image.
(a) Input image of mouse spine ROI; (b) Result of applying CLAHE twice on (a), which leads to enhance the overall contrast of the spine object in the image; (c) Result of applying Gamma transformation with gamma value of (5) on (b). The result shows highlighting the borders of the spine and diminishing other unwanted objects; (d) Result of applying Gamma transformation with gamma value of (10) on (b), which shows reducing more unwanted tissues and noises but also affecting the contrast of the spine borders

5.3.4 First Level Spine Segmentation

The resulting image from the previous stage as shown in Figure 5.7(c) still has some background noises and artefacts besides the body of the spine. In this step, we aim to segment an initial binary mask of the spine as a one connected object. The importance of this step is coming from its direct effect on the final refinement of the borders of the spine and so the estimation of the curvature. The initial spine segmentation starts by thresholding the resulting image of the enhancement process (Figure 5.7(c)) using Otsu method (Otsu, 1979). The resulting image contains most of the spine objects in addition to some unwanted objects (i.e. part of ribs) as shown in Figure 5.8(a).

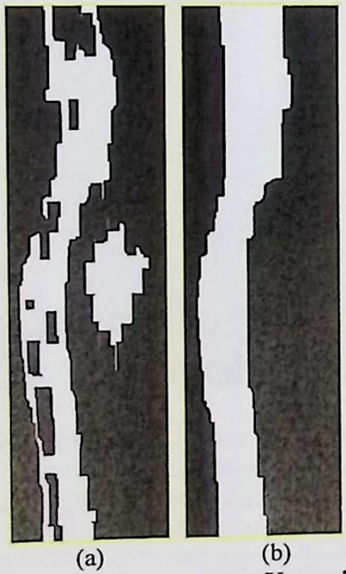
To connect all the pixels of the spine together, we apply vertical and horizontal closing morphology operations with a structure element length of 40 and 10 respectively. Figure 5.8 shows the results of using different structure element length values to close the spine mask vertically. The figure demonstrates that the length range from 30 to 70 pixels for the line structure element could work fine to close the spine mask parts. However, the value around 40 pixels (Figure 5.8(c)) could be the best value because it closes most parts of the spine mask without attaching unwanted objects to the spine mask.



(a) (b) (c) (d) (e) (f)

Figure 5.8 Sensitivity of length of structure element to close the spine mask.
(a) Result of thresholding the enhanced spine using Otsu method, which contains most of the spine objects in addition to some unwanted objects; (b) Result of vertical closing of the image in (a) with line structure element of length 30 pixels, which shows closing most of spine objects without attaching unwanted objects to the spine; (c) Result of vertical closing of the image in (a) with line structure element of length 40, which shows closing most of the spine objects without attaching unwanted objects to the spine; (d) Result of vertical closing of the image in (a) with line structure element of length 50 pixels; (e) Result of vertical closing of the image in (a) with line structure element of length 60 pixels; (f) Result of vertical closing of the image in (a) with line structure element of length 70 pixels, which shows closing most of the spine objects with attaching unwanted objects to the spine

After removing the small objects, the resulting image is as shown in Figure 5.9(a).



(a) (b)

Figure 5.9 Initial segmentation of spine from mouse X-ray image.
(a) Spine mask after applying binary morphology operations of closing and removing small objects; (b) Spine mask after removing irrelevant objects and filling the holes

At this stage, the irrelevant objects are usually existed in addition to the spine mask. We analyse the properties of spine mask and irrelevant objects from a group of images. Following this analysis, we suggest using some geometrical features to recognize and remove the unwanted objects as follow:

- Length of objects. The length of the unwanted objects is shorter than the spine mask (i.e. < 400 pixels). Figure 5.10 shows different samples of spine mask where the length of irrelevant objects is less than 400 pixels in all the cases as illustrated by the red lines.
- Distance from the right or left borders of the image. The x-coordinate of the right or left sides of object's bounding box is close to the right or left borders of the image (i.e. < 50 pixels). Figure 5.10 shows different samples of spine mask where the distance between the irrelevant objects and the image border is less than 50 pixels in all the cases as illustrated by the yellow lines.
- Distance from the upper or lower borders of the image. The y-coordinate of the upper or lower sides of object's bounding box is further from the upper or lower borders of the image (i.e. > 100 pixels).

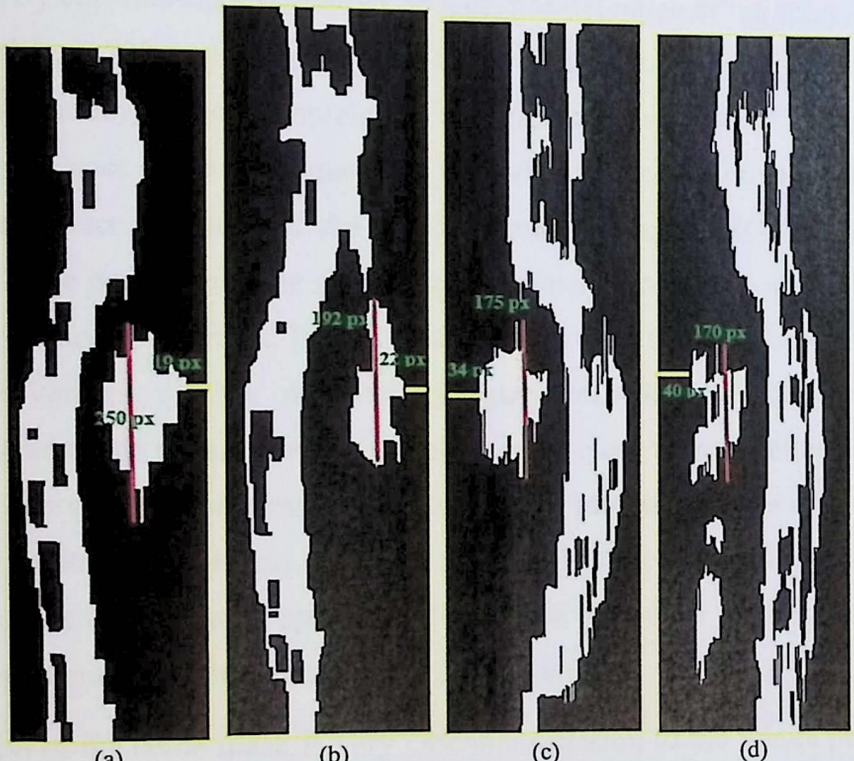


Figure 5.10 Some geometrical properties of unwanted objects near the spine mask. The red line represents the length of the unwanted object while the yellow line represents the distance between the unwanted object and the nearest border of the image as follow.
(a) Length of unwanted object is 250 pixels and its distance to the border is 19 pixels; (b) Length of unwanted object is 192 pixels and its distance to the border is 22 pixels (c) Length of unwanted object is 175 pixels and its distance to the border is 34; (d) Length of unwanted object is 170 pixels and its distance to the border is 40 pixels

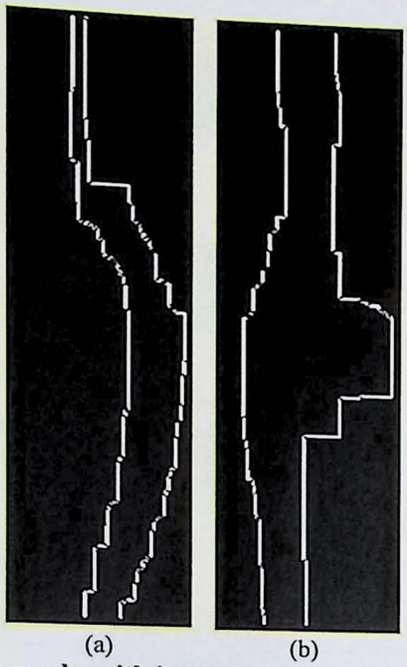
By creating some rules depending on the above features, we can exclude irrelevant objects with the spine mask. The final step to produce an initial spine mask is to use vertical and horizontal closing and filling holes morphological operations (Section 2.2.1) as shown in Figure 5.9(b).

5.3.5 Reliability Check and Refinements of the Segmented Spine

The resulting spine mask from the previous stage could be distorted because of attaching unwanted objects to it. Such inaccuracies in the segmented spine mask need to be identified and, if possible, recovered from. Due to the importance of the spine mask segmentation to all later stages, some criteria are selected to estimate the spine mask quality. This reliability check leads to one of three outputs: 1) reliable mask (no actions required), 2) require improvement which might lead to either output 1 or output 3, and 3) reject the image for lack of reliability. The list of the underlying criteria that are used to estimate the quality of the spine mask can be summarised as follows:

1. The length of the spine mask. If the length of the resulting mask is shorter than the minimum expected length, the mask is considered as unreliable and the image is rejected. By considering a sample of images, normal range of the spine length could be estimated (e.g. 700 to 1000 pixels for Sanger images). Cases of short masks happen mainly when they are not complete due to extremely low contrast of the image.
2. Number of pieces of the spine mask. In optimal cases, the resulting mask should be one continuous piece. Because of the lack of contrast in spine ROI for some images or missing some parts of a spine mask during segmentation stage, sometimes we end up with partitioned spine mask. If such case is discovered, the suggested solution is to use adaptive vertical closing morphology operation (Section 2.2.1) with a length proportional to the distance between the pieces to close them together.
3. The thickness of the spine mask. Due to the low intensity values in some parts of the spine or existence of objects (i.e. ribs) which might be connected to the spine mask, some regions of the binary spine mask might be thinner or thicker than the expected normal spine mask as shown by Figure 5.11(a) and (b).

The thickness of the spine mask is very essential criterion to the whole procedure, so we will discuss how to find out and how to correct this problem in the next subsection.



(a) (b)

Figure 5.11 Unreliable spine masks with irregular thickness problem.
(a) Borders of spine mask with a thin upper area. This case usually happens because of missing parts of spine mask during enhancement process; (b) Borders of spine mask with a thick middle area. This case usually happens because of attaching unwanted parts to the spine mask during enhancement process

Identifying and correcting irregular thickness of spine mask

By observing a range of different samples, we found that the normal distance between any two opposite points on the right and left spine borders should be in a range between 50 and 90 pixels. Any distance out of this range is considered as unreliable and needs to be dealt with accordingly. We propose a procedure to identify such inaccuracies when they occur and then try to fix them. Figure 5.12 shows a flowchart for the proposed procedure. Some explanations about the solution above are given next. The first test in the process checks if the horizontal distance between every pair of opposite points on the right and left borders is in the normal range or not. If it is, no further refinement is needed, and the spine mask is passed to the next stage of the process. Figure 5.13(a) shows the right and left borders' pixels of binary spine mask superimposed on the input image. The figure indicates areas where the distances are within the normal range (marked by green arrow lines) and areas where the distances are out of the normal range (marked by blue arrow lines). If any out-of-range distance is detected, some mask refinement is then required, i.e. to bring the "out of line" point back into line by off-setting the point's x coordinate value with the normal range of 70 pixels.

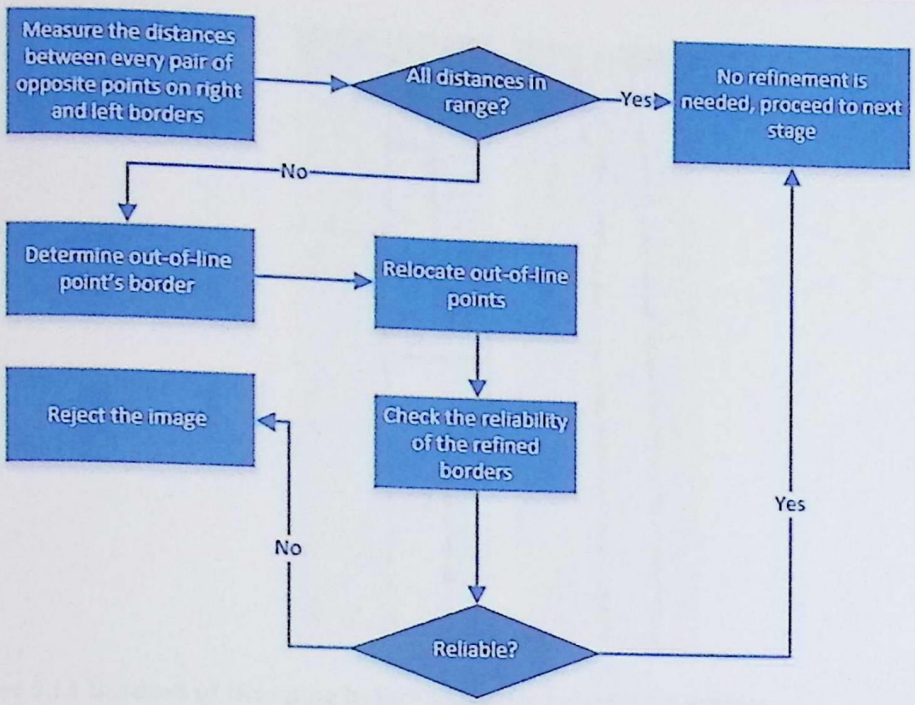


Figure 5.12 Process of discovering and refining of irregular thickness of mouse spine mask. The process starts by measuring the distances between every pair of opposite points on the right and left borders. If any irregular distance is discovered, a relocation of the out-line point is done. After that the reliability of the refined borders is done to decide whether it is accepted or not.

However, we need to determine which point of the pair that is in fact "out of line". We therefore apply the following procedure to the points of both borders:

1. Locate the pair of "out of line" points.
2. For each point, obtain the x-coordinates of the point and 50 points before and 50 points after it along the border.
3. Calculate the standard deviation for obtained x-coordinates.
4. Consider the side with the largest standard deviation as the side where the "out of line" point occurs.
5. Relocate out-of-line point to the new position (x-coordinate) 70 pixels away from the other point of the pair. Figures 5.13(a) and (b) show the spine border before and after refinement respectively.

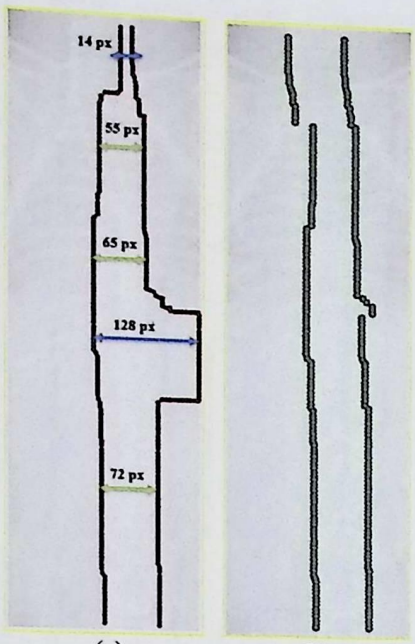


Figure 5.13 Borders of the spine before and after refinement process.
(a) Border of the mouse spine before refinement with thickness problem (blue arrows) and normal distances (green arrows). The irregular thickness is either large (128 pixels) or small (14 pixels); (b) Result of refinement process where the distances between every pair of opposite point are in the normal range (50 to 90 pixels)

The standard deviation is chosen as a good indicator because the points at the "out of line" positions have longer distance from the mean point. Using 50 points before and after is based on the fact that "out of line" points are locally defined in a local area rather than globally defined for the whole spine. In limited cases, the standard deviation check as described above could not recognize the border with "out of line" points. This mainly happens when the variations of the x-coordinates of pixels on the correct border is more than the variations for opposite pixels in the border with "out of line" points. This case will lead to inaccurate borders' pixels relocation as shown in Figure 5.14(a). To filter such inaccurate results, we propose the following procedure for each border.

1. Find the maximum values of the absolute differences between the x-coordinates of each pair of consecutive points on each border (one maximum value on each side).
2. If both maximum values are greater than a predetermined threshold (i.e. 30 pixels), label the image as unreliable. The value of the threshold is selected empirically to filter out the cases with inaccurate refinements while not affect the correctly refined images.

Figure 5.14 shows an example of different refined images. The values of maximum differences of x-coordinates for right and left borders after refinements are associated with each example.

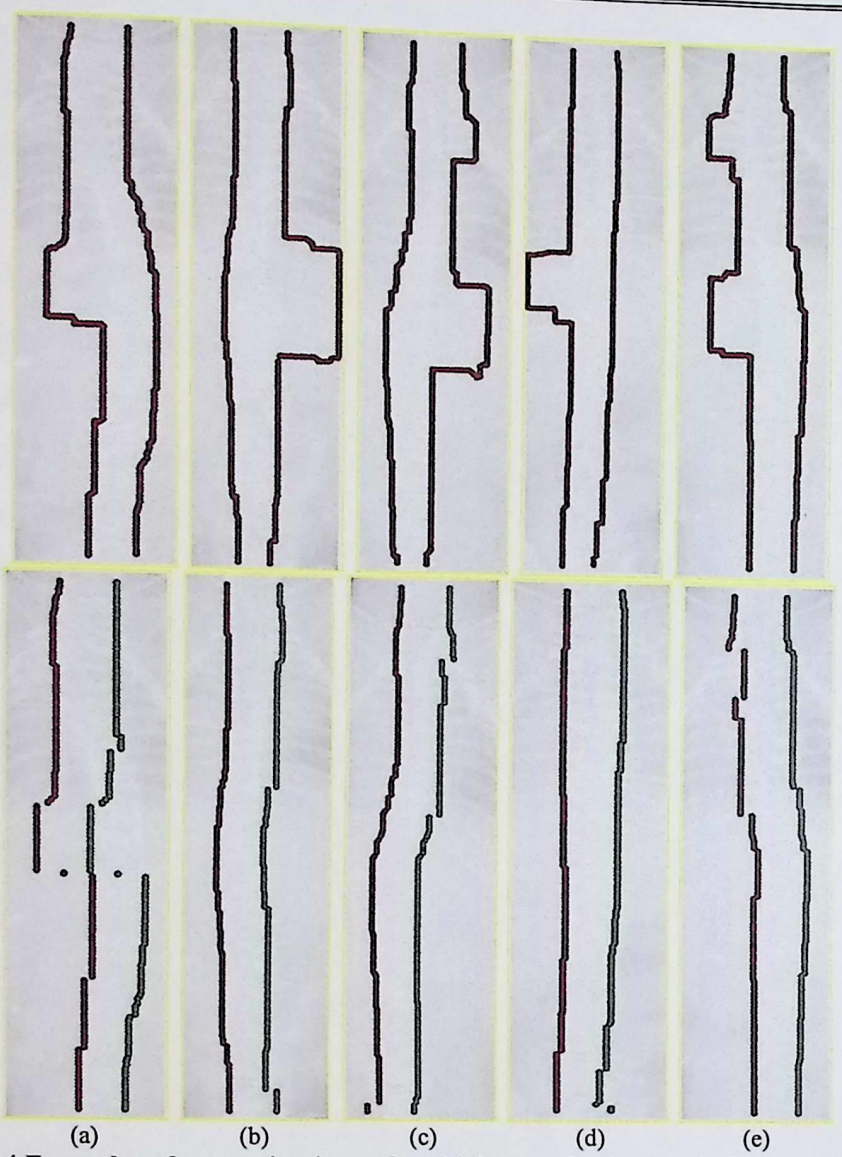


Figure 5.14 Examples of correcting irregular thickness of spine mask. The upper image in the column represents the spine borders before refinement and the lower one is the result of refinements. The right and left maximum differences for the borders after refinements are as follows.

(a) Inaccurate refinement of spine borders with right and left maximum difference is (37, 37). If both right and left maximum values are greater than 30, the mask is considered unreliable; (b) Accurate refinement of spine borders with right and left maximum difference is (16 , 5); (c) Accurate refinement of spine borders with right and left maximum difference is (4 , 1); Accurate refinement of spine borders with right and left maximum difference is (d) Accurate refinement of spine borders with right and left maximum difference is (19 , 3); (e) Accurate refinement of spine borders with right and left maximum difference is (9 , 21)

The reliability check and refinement process for thickness of the spine mask is dependent on two parameter settings. The first one is the range of the acceptable distance while the second is the "default" thickness after relocation process. Because of the thickness is different if it is taken from different locations on the spine, we decide to make the acceptable thickness range for the spine mask wide enough (50 to 90 pixels) to cover different locations of the spine. For the thickness after refinement, we set it to be in the middle (70 pixels) between smaller and larger expected thickness for spine. This value will set the border of the spine mask to fit either on the border or out of the actual border location in the input image.

We argue that our main goal in this stage is not to get a very precise location for each point on the borders but to bring them as close as possible to the correct location. The next stage can then take the current positions of the pixel's borders as an input to refine them to more accurate locations.

5.3.6 Second Level Spine Segmentation

The first level spine segmentation and refinement process is based on binary mask of the spine. The mask is meant to cover the entire area of the spine. However, as we can see from different examples in Figure 5.14, the spine mask is still a rough estimation of the spine contour due to the thresholding, morphology operations and estimation of the thickness of the spine mask applied in previous sections. This means that the spine contour boundaries are imprecise, which affects the precision of the spine curvature measurement later. The goal of second level spine segmentation is to start from the output of the refinement process and then obtain more accurate and smooth outlines of the spine contour. In the next subsections, we will introduce two possible approaches to further refine the spine contour and then discuss how to select the most suitable one for our method.

Intensity based refinement of spine mask borders

In this approach, the main idea is to refine the location of all or some of points of the mask spine borders to more accurate location. To perform this, we tend to exploit the intensity difference between the border of the spine and the background in the enhanced image (Figure 5.7(c)). The proposed procedure of the refinement is as follow:

1. Select number of points from the border of the spine mask and superimpose them on the enhanced image (Figure 5.15(a)).
2. For every point on the left spine border of the mask, we horizontally search 40 pixels towards the right for the pixel with the maximum intensity value above predefined threshold (i.e. 95). If such a pixel is found, it is taken as the new spine border pixel to replace the original border pixel from the spine mask; otherwise the original point of the mask is removed (i.e. the pixel intensity value is set to 0). We repeat the process for every point on the right spine border, but instead of searching towards the right, the searching is towards the left. Figure 5.15(b) shows the refined location in yellow. The idea of removing the points that fail the above location refinement procedure is a key point of this procedure. This leads to removing most of out-of-place points which leads to more accurate fitting in the next step.

3. After refining all the points, we are fitting them using polynomial fitting of degree 6 to find the best fitting smooth contour (Figure 5.15(c)) (Sardjono et al., 2013).



(a) (b) (c)

Figure 5.15 Intensity based refinements of the spine borders.
(a) Superimposing of points (red points) obtained from the borders of the spine mask on the enhanced image; (b) Refining of the points of the spine mask (red points) to the new location (yellow points); (c) Polynomial fitting of refined points

This method works fine for the locations where the borders in the enhanced image have high intensity. However, one of issues about this approach is one direction search (toward the borders of the spine). This option is selected because the thickness of the spine mask is made wide enough to ensure all the pixels of the spines' borders sit outside the spine. So, there is no point to search both directions. The second issue is the sensitivity of selecting the length of search radius. Large value might update the position to wrong location inside the spine while small value might not work if the point is distant from the border. Empirically, we found that setting this parameter to 40 will lead to a good refinement. Although this approach needs large number of points to work well, the key advantage of it is that increasing the number of points will not have a noticeable negative impact on the efficiency due to the simplicity of the underlying stages.

Texture based approach to detect points on the spine borders

Texture based landmarks detection could be one of the candidate solutions to detect the points of the spine borders. We decide to investigate how the RF model (Section 2.2.3) could refine the points for this problem. We are encouraged to use RF model by the fact that we can use the current points of spine borders as a seed for a small search area. This can offer a good chance for more accurate and efficient refinement of points. The procedure of using RF models is applied in similar way as described in Section 4.3.5. Figure 5.16(a) shows one example of

training image where from each manually annotated point (red point), one patch as a positive sample (black box) and two patches as negative samples (red boxes) are determined. The extracted LBP and HOG features from the patches of 100 training images are then used to train two universal RF models.

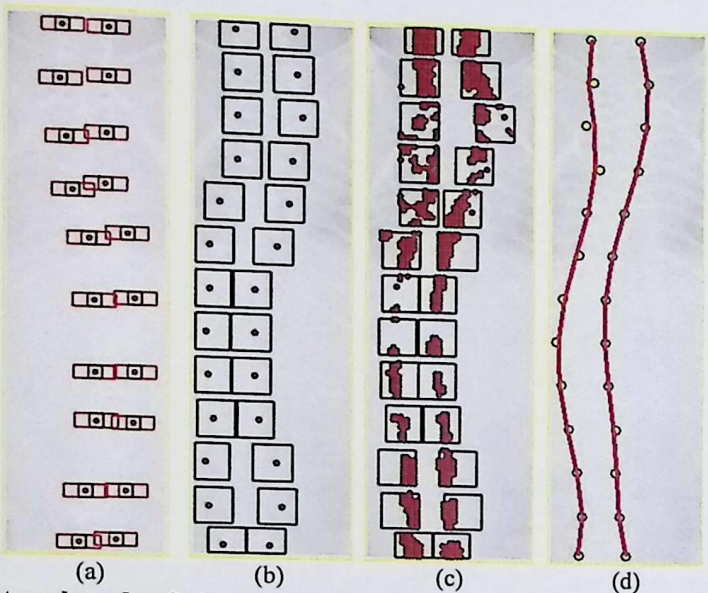


Figure 5.16 Texture based refining of spine borders.
(a) Example of training image, the positive (black boxes) and negative (red boxes) patches are sampled around manually annotated points (red); (b) Sampled points from borders of the spine mask (red points) and the rectangle search areas (black boxes); (c) Pixels are predicted as positive by RF-LBP model (red points); (d) Detected landmarks (yellow points) as a median of x-coordinates of all the pixels predicted as positive and fitting borders of the spine (red)

The classification error rate (Section 2.2.3) of the RF-LBP and RF-HOG models for different number of trees is shown in Figure 5.17. The figure demonstrates that the classification error rate for the RF-LBP model is less than the one for the RF-HOG model. This could be explained in the sense that the borders of the spine shows changes in the contrast from surrounding area with no clear changes in orientation. For this reason, we decide to use the RF-LBP model to predict the landmarks on the spine borders.

In the prediction stage, the determination of the search areas on the spine borders (black boxes in Figure 5.16(b)) is done based on number of points (i.e.13) from each border of the binary spine mask (red points in Figure 5.16(b)). The biggest area of the search box is set to be toward the border of the spine to reflect the fact that the seed points usually located outside the spine. The class of each sample is then predicted by using the RF-LBP model. Figure 5.16(c) shows the pixels that predicted as positive for each search area. As we can see from Figure 5.16(c), the most of predicted points in the search area located precisely on the borders of the spine. So, using the median of the x-coordinates of predicted points to locate the final point could give an accurate detection by getting rid of all the outlier points. Figure 5.16(d) shows the final points

as yellow. The y-coordinate here is not of the concern so we can set it to be the y-coordinate of the sampled points. After obtaining all the landmarks, we fitting them using polynomial fitting to get the smooth contour for the spine as shown in Figure 5.16(d).

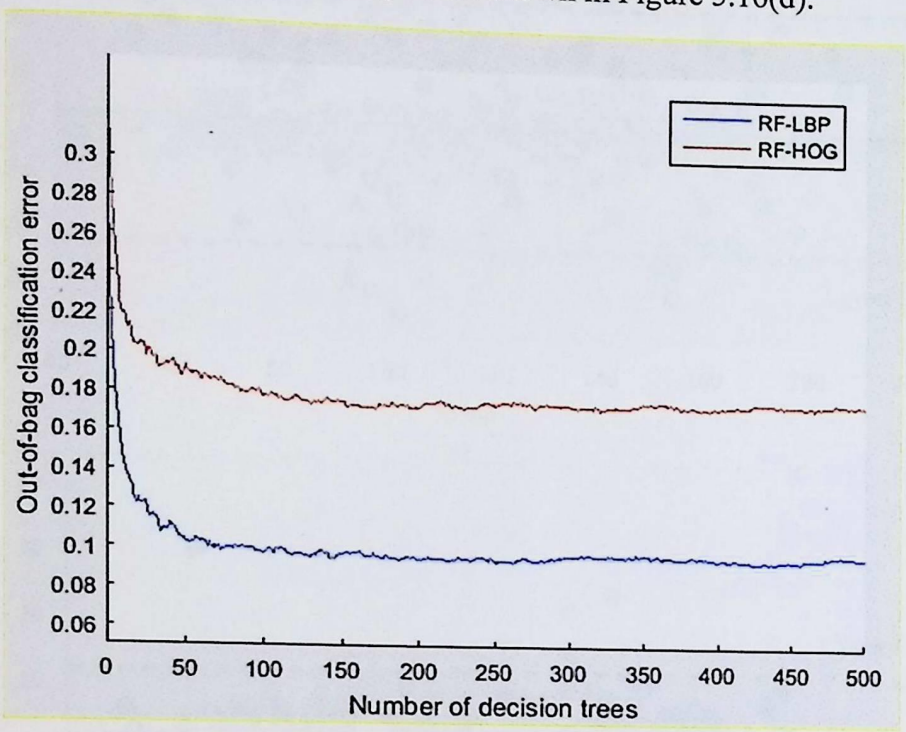


Figure 5.17 Classification error rate for RF-LBP (blue line) and RF-HOG (red line) models. The error rate is calculated across different number of trees in the RF models.

Intensity based versus texture based refining for borders of the spine

As we see in the previous two subsections, we can estimate the contour of the spine using either intensity or texture based approach. Here, we want to investigate the strengths and limitations of both approaches.

The first issue to consider is how accurate each approach to locate the points on the spine borders. To do this, we manually annotate 260 points from various locations on both right and left borders (13 points on each side) from 10 different images. After that we implement both approaches on those images to locate the points on the spine borders.

For comparison purpose, instead of deleting the points that exceeded the search threshold for the intensity based method, we decide to keep them in their original locations. This is mainly to compare 13 points on each side for both methods. To measure the closeness between the manually annotated points and the points located by each approach, we used the BA analysis (Section 3.4.2). The results of analysis for both approaches are shown in Figure 5.18(a) for intensity based and (b) for texture based approach.

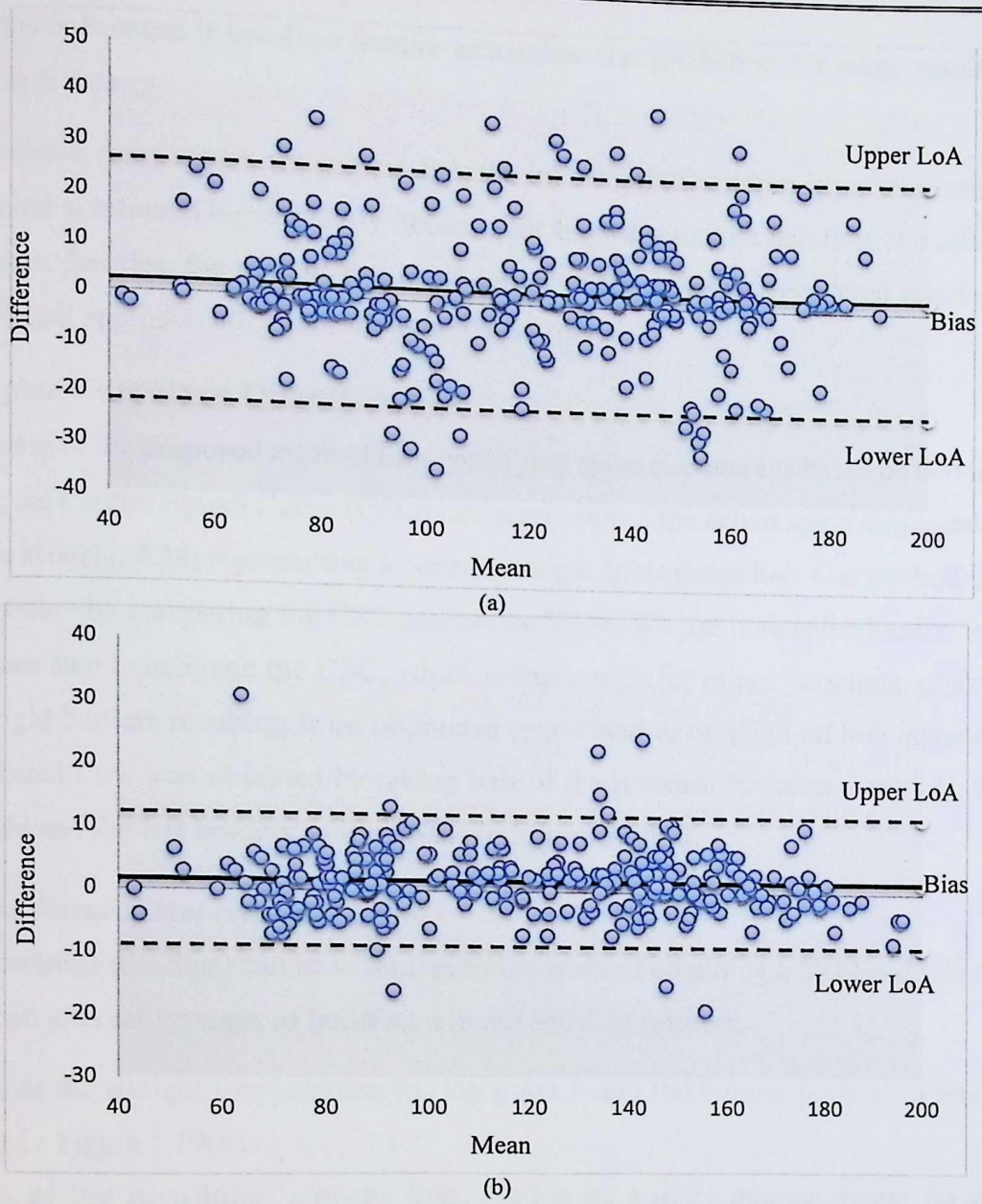


Figure 5.18 BA analysis for intensity and texture based landmarks detection versus manually annotated landmarks.
(a) Intensity based method. The result shows that many detected point has relative high difference (-21 to 26 pixels) from the manually annotated points; (b) Texture based method. The result shows that most of detected points has small difference (-9 to 12 pixels) from the manually detected points

It can be argued that the texture based approach is more accurate with LoA range from -8.9 to 12.1 pixels and most of the points located inside the range. The LoA for the intensity based is bigger with range between -21.6 to 26.4. However, we should highlight that the final fit of spine borders for the intensity based approach could be highly enhanced by increasing the number of points and using the option of deleting outlier points.

The second important issue to consider is the efficiency of each approach. The intensity based approach was shown to be more efficient even when increasing the number of points. However, the situation is not the same for the texture based approach even with much fewer number of

points. This is because it involves feature extraction and prediction for many pixels in each point's search area.

From the above comparison, we can say that the intensity based approach is more suitable for our proposed solution. This is mainly because for high throughput solution, the efficiency is critical issue. Besides, the overall fitting of border of the spine is considered satisfactory for intensity based approach after excluding outlier points.

5.3.7 Spine Curvature Quantification

The final step of the proposed method for quantifying spine curvature is based on two important elements: the Central Spinal Curve (CSC) line representing the actual spine shape and a Spinal Reference Model (SRM) representing a normal straight spine shape line. Our method quantifies spine curvature by comparing the CSC against the SRM. We use both spine borders located in the previous step to estimate the CSC, which compensates for minor variations caused by the left and right borders resulting from imprecise segmentation of spine on low quality images. The estimated CSC was obtained by taking half of the distance between every pair of points on the right and the left borders as illustrated in Figure 5.19(a).

Estimating linear spinal reference models

We acknowledge that there can be several possible ways of obtaining a SRM and will therefore propose two alternative ways of building a linear SRM as follows:

1. SRM_1 as the straight line between the top point P and the bottom point Q of the CSC, as shown in Figure 5.19(b).
2. SRM_2 as the best-fitting straight line, $f(x) = a_0 + a_1 x$, among points on the CSC calculated by the least square method as follows:

$$a_1 = \frac{(n\sum xy - (\sum x)(\sum y))}{(n\sum x^2 - (\sum x)^2)} \quad (5.1)$$

$$a_0 = \frac{(\sum y - a_1(\sum x))}{n} \quad (5.2)$$

where x and y are the coordinates of each point on the CSC. An example SRM_2 is shown in Figure 5.19(c).

While SRM_1 seems a nature and simple way to estimate a reference model, it is heavily dependent on the accuracy of locating the two points P and Q. SRM_2, on the other hand, relies on all the points on CSC to estimate the reference model and hence should be more stable even the locations of P and Q are not very precise. The idea of using best fitting line to estimate the

curvature of the spine was considered in the literature to estimate human spine curvature (Tang et al., 2008).

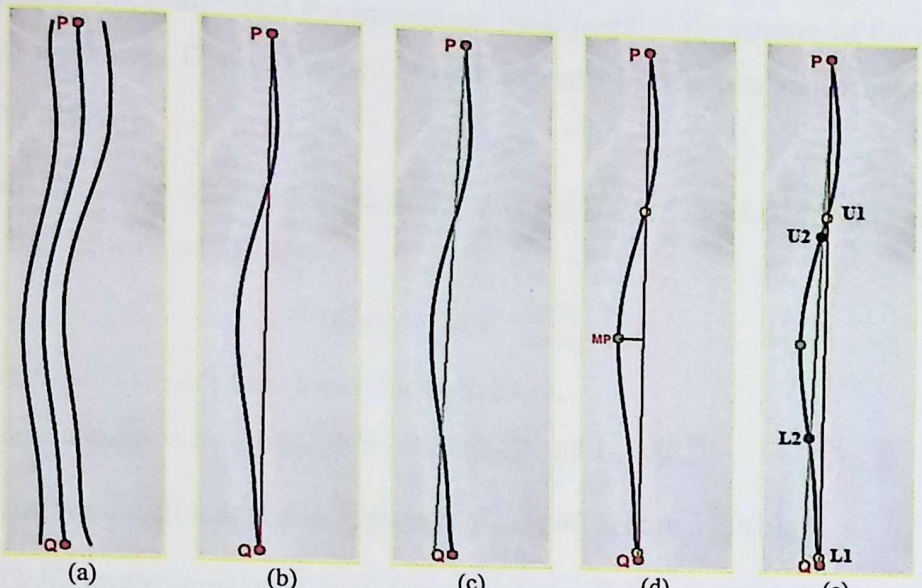


Figure 5.19 Locating CSC, SRMs and Landmark Points.
(a) Estimated CSC in blue and right and left spine borders in black; (b) SRM_1 as the red straight line between the top point P and the bottom point Q of the CSC; (c) SRM_2 as the green best-fitting straight line among points on the CSC; (d) Maximum peak point (MP) annotated in green on CSC;(e) Points of intersection between CSC and SRM_1 (U_1 and L_1 yellow points), and between CSC and SRM_2 (U_2 and L_2 black points)

Locating the landmark points

Based on the CSC and the two possible SRMs, our proposed method further identifies five key landmarks as the intersection points between the CSC and the SRMs as follows.

1- The maximum peak (MP) point on the CSC. This is achieved by finding multiple local maximum peak points and further selecting from these points the one that has the maximum distance between the CSC and SRM_1, as shown in Figure 5.19(d). We use SRM_1 rather than SRM_2 to find the MP because it usually maximize the distance to the biggest curvature area. This will lead to locate the MP landmark accurately on the main curve of the spine.

2- The upper and lower points on SRM_1 and SRM_2 intersecting with the CSC. Those points are determined as the first intersection points with the CSC above and below the maximum peak point. Figure 5.19(e) shows the intersection points U_1 and L_1 derived from the SRM_1, and the intersection points U_2 and L_2 derived from the SRM_2. There are subtle differences between the locations of intersection points depending on the reference model used.

Quantifying spine curvature

The Cobb angle measurement of curvature replicates the manual process by using the upper and lower intersection points and the maximum peak point as the centres of the start, the end, and the apical vertebrae. Due to the use of two different SRMs, two angle measurements can be derived as follow.

- $Angle_1$: the acute angle which formed between U_1-MP and L_1-MP lines and calculated from the formula as follow:

$$Angle_1 = \tan^{-1}\frac{m_1 - m_2}{1 + m_1 \times m_2} \qquad (5.3)$$

  Where m_1 and m_2 are the slopes for both lines.
- $Angle_2$: calculated as $Angle_1$ but for U_2-MP and L_2-MP.

$Angle_1$ and $Angle_2$ are illustrated in Figure 5.20(a) and Figure 5.20(b).

Besides the Cobb angle measurements, we also derived distance-based index measurements based on the difference between the CSC and the SRM used. A distance-based index can be obtained by using the RMSE as shown below:

$$\text{RMSE} = \sqrt[2]{\frac{(\sum_{i=1}^{n}(X_i - XR_i)^2)}{n}} \qquad (5.4)$$

where X_i represents the x-coordinate of the point i on the CSC, and XR_i represents the x-coordinate of the horizontally projected point on the SRM, and n is the number of the points on the CSC. Depending on the chosen SRM and selection of the starting and ending points on the CSC, the RMSE value can be calculated in different ways, which results in a number of index measurements:

- SRM_1 based indices:
  - SRM_1_IndexA: RMSE between the whole CSC and SRM_1, as illustrated in Figure 5.20(c);
  - SRM_1_IndexB: RMSE between the segment of the CSC bounded by the key points U_1 and L_1, and SRM_1, as illustrated in Figure 5.20(d);
- SRM_2 based indices:
  - SRM_2_IndexA and SRM_2_IndexB: similar to the above indices but using SRM_2 instead of SRM_1 as shown in Figures 5.20(e) and (f).

It can be argued that the accuracies of the angle based measurements are more dependent on the precision of locating the three key landmark points than the index based measurements.

This may make the index-based measurements more consistent and stable than angle-based measurements.

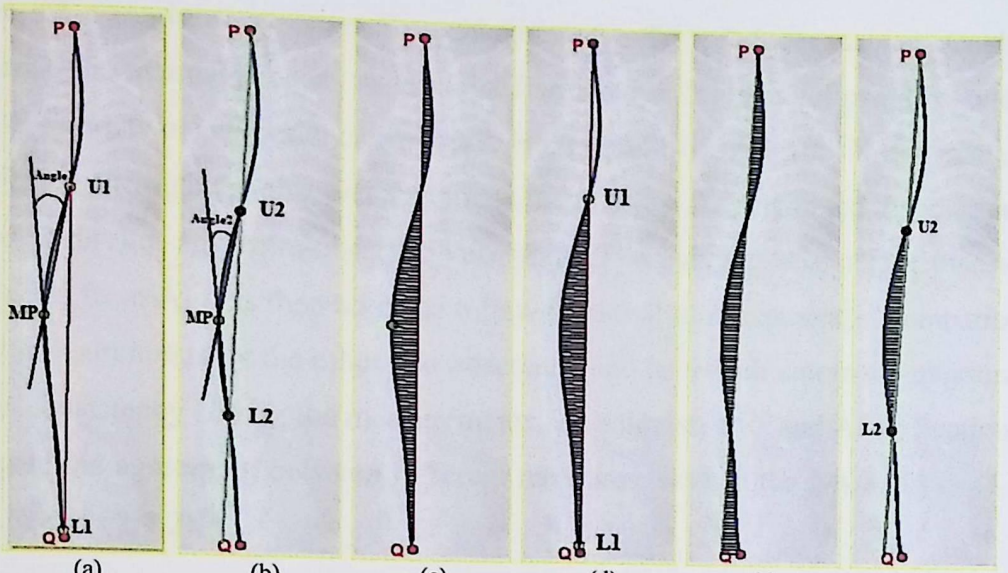


Figure 5.20 Angle based and distance-based index measurements.
(a) $Angle_1$ as the angle formed between line passes through U_1 and MP points and line passes through L_1 and MP points; (b) $Angle_2$ as the angle formed between line passes through U_2 and MP points and line passes through L_2 and MP points; (c) SRM_1_IndexA as the RMSE between the whole CSC and SRM_1; (d) SRM_1_IndexB as the RMSE between the segment of the CSC bounded by the key points U_1 and L_1, and SRM_1; (e) SRM_2_IndexA as the RMSE between the whole CSC and SRM_2; (f) SRM_2_IndexB as the RMSE between the segment of the CSC bounded by the key points U_2 and L_2, and SRM_2

5.4 Experiment Results and Analysis

This section presents two sets of experiments that are conducted to evaluate our proposed method. The first set aims to evaluate the consistency and closeness of the automated angle-based measurement results to manual measurement results of spine curvature. The second set aims to evaluate the effectiveness of all the automatic curvature measurements in accurately identifying cases of scoliosis through classification.

For the first experiment set, 100 full body X-ray images of mice in the dorsoventral view were randomly selected (one image per mouse) from a collection of 6000+ images of mice obtained from WTSI. The dataset consists of 75 images of WT (normal) mice and all the images of mutant mice with gene knocked-out known to be linked with some form of scoliosis which defined and labelled by domain experts at the WTSI (25 images in total). Among the 25 images of mutant mice with scoliosis, there are 22 images of the Thoracic Scoliosis cases, 2 images of the Lumbar Scoliosis cases, and 1 image of Thoracolumbar Scoliosis. For the second experiment, we increase the size of WT images pool to be 500 in total to include more natural variations.

5.4.1 Manual versus Automatic Measurements

Before we compare the automatic measurements to the manual ones, it is necessary to establish the degree of variations in manual measurements. To produce manual Cobb angle measurements, the manual process as explained in Section 5.1 was followed by one expert observer (A_1) and two inexperienced observers (A_2 and A_3). All 100 X-ray images were measured for observing inter-observer variations. In addition, A_1 performed the measurements 4 times for observing any intra-observer variations. The average of the four measurement results obtained from A_1 was then taken as referred manual measurement in comparison with the manual measurements by the other two observers, and later with automatic measurements. To examine consistency among the measurements, we calculated R^2 and ARL (Section 3.4.2). To investigate the agreement between different measures, we use the BA analysis (for more details see Section 3.4.2).

Table 5.1 shows the evaluation results for manual measures using R^2 and ARL on the whole dataset, mutant only subset, and WT only subset. Table 5.2 shows the lower and upper (LoA) bounds obtained from applying BA analysis for manual measures on the whole sample dataset, mutant only subset, and WT only subset.

Methods Used (Manual Measurements)	Data Set of both Mutant and WT Mice		Data set of Mutant Mice only		Data Set of WT Mice only	
	R^2	ARL	R^2	ARL	R^2	ARL
A_1 Trial Average	0.92	44.4°	0.82	39.78	0.32	34.1°
A_1 Trial Average vs A_2 Trial	0.86	38.9°	0.57	30.70	0.03	8.20°
A_1 Trial Average vs A_3 Trial	0.85	42.0°	0.74	35.88	0.15	12.6°
A_2 vs A_3	0.79	38.2°	0.40	30.28	0.14	26.7°

Table 5.1: Intra-Observer and Inter-Observer variations in the manual measurements of spine curvature Cobb angles using R^2 and ARL

Methods Used (Manual Measurements)	Data Set of both Mutant and WT Mice		Data set of Mutant Mice only		Data Set of WT Mice only	
	Lower LoA	Upper LoA	Lower LoA	Upper LoA	Lower LoA	Upper LoA
A_1 Trial Average	-2.4	2.3	-2.3	2.4	-2.5	2.1
A_1 Trial Average vs A_2 Trial	-3.9	4.4	-6.1	4.6	-3.0	4.0
A_1 Trial Average vs A_3 Trial	-4.6	3.1	-4.3	3.5	-4.8	2.9
A_2 vs A_3	-6.1	4.1	-6.5	7.2	-5.5	2.6

Table 5.2: Intra-Observer and Inter-Observer variations in the manual measurements of spine curvature Cobb angles using BA analysis.

The results presented in both tables confirm that that inter- and intra-observer measurement variations do exist in terms of both consistency and closeness of measurements although most of the measurements from the three observers are correlated. As expected, the intra-observer variations are smaller than the inter-observer variations (the first row of each table vs. the other

rows in each table). The inter-intra-observer variations are the highest for WT mice where the spine should look almost straight and thus manually locating the landmarks would have a greater deal of variations, particularly for inexperienced observers. Keep in mind that the scale for mutant data set measurements are higher than the one for the WT data set measurements. There seems also a certain degree of consistent bias where inexperienced observers tend to measure the angles larger than those by the expert observer, specifically for the WT mice.

Although more about inter- and intra-observer variations can be further discussed, our interest lies in the closeness and consistency between the manual and automatic measurements. Table 5.3 and 5.4 show the comparison results of automatic vs. manual angle-based measurements results using R^2 and ARL (Table 5.3) and the lower and upper LoA bounds and the bias (Table 5.4) obtained from applying BA analysis (Section 3.4.2) on the whole dataset, mutant only subset, and WT only subset. Figure 5.21 shows the scatter plot and the regression line for reference manual measure versus $Angle_1$ and $Angle_2$. Figure 5.22 shows the scatter plot for BA analysis for reference manual measure versus $Angle_1$ and $Angle_2$

Methods Used (Automatic vs Manual Measurements)	Data Set of All		Data set of Mutant Mice		Data Set of WT Mice	
	R^2	ARL	R^2	ARL	R^2	ARL
$Angle_1$ vs. A_1 Trial Average	0.91	40.2°	0.87	45.4°	0.17	24.5°
$Angle_2$ vs. A_1 Trial Average	0.89	40.0°	0.75	40.2°	0.21	31.4°

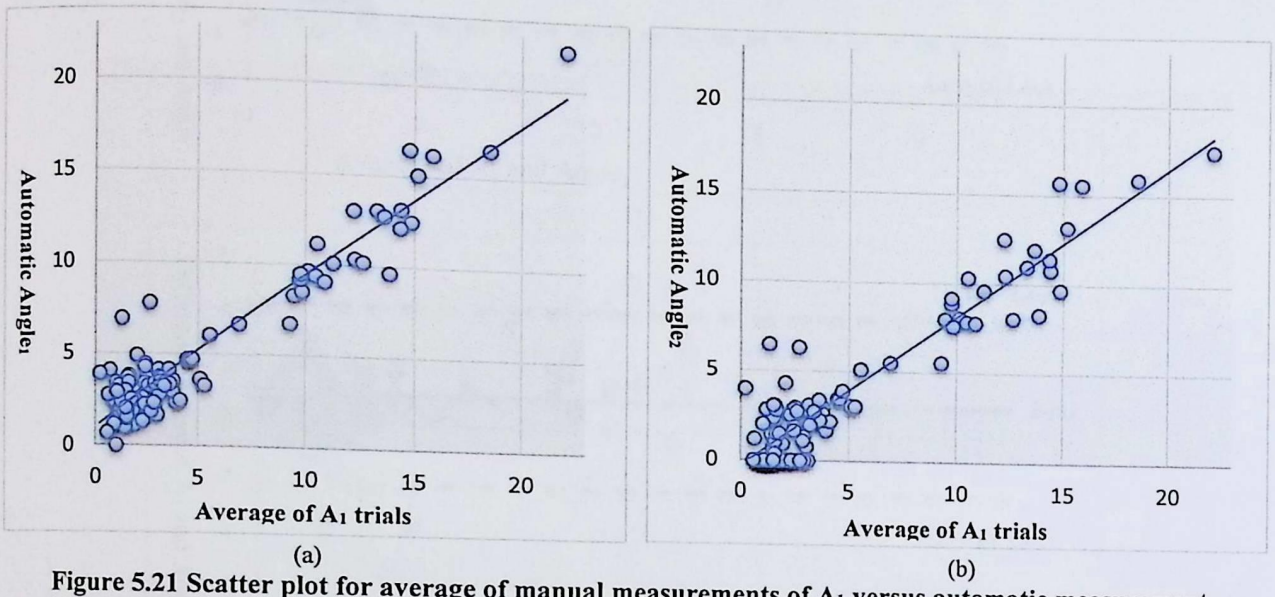
Table 5.3 Automatic versus manual angle-based measurements of mouse spine curvature using R^2 and ARL.

Methods Used (Automatic vs Manual Measurements)	Data Set of All			Data set of Mutant Mice			Data Set of WT Mice		
	Lower LoA	Upper LoA	Bias	Lower LoA	Upper LoA	Bias	Lower LoA	Upper LoA	Bias
$Angle_1$ vs. A_1 Trial Average	-3.2	2.7	-0.23	-1.5	3.5	0.98	-3.2	2.76	-0.63
$Angle_2$ vs. A_1 Trial Average	-2.2	4.3	1.08	-1.05	5.5	2.23	-2.2	3.6	0.69

Table 5.4 BA analysis for automatic versus manual angle-based measurements of mouse spine curvature.

From the R^2, upper and lower LoA reported in Tables 5.3 and 5.4, it is evident that both automatic angle measurements are close to and highly correlated with the manual measurements by the expert observer. $Angle_1$ measurements are more close to the manual measurements than $Angle_2$ measurements especially for the whole data set and mutant mice. This give an indication that using SRM_1 to locate the curvature landmarks is more natural than using SRM_2. We can also note from Table 5.4 that the bias for both $Angle_1$ and $Angle_2$ is

increased for mutant data set with advantage for $Angle_1$. This indicates that whenever there is a detectable curvature in the spine, both $Angle_1$ and $Angle_2$ tend to give higher estimation of angle than the manual estimation. However, because of limitation of mutant data set (25 images) this conclusion yet to be confirmed.



(a) (b)

Figure 5.21 Scatter plot for average of manual measurements of A_1 versus automatic measurement obtained by using automatically obtained measurements $Angle_1$ and $Angle_2$
(a) Average of manual measures taken by A_1 observer versus automatic measurement obtained by using $Angle_1$; (b) Average of manual measures taken by A_1 versus automatic measurement obtained by using $Angle_2$

The ARL value of 45.42° indicates that $Angle_1$ measurements are highly correlated to the manual measurements for mutant mice with scoliosis symptoms. Although the relative variations between the automatic and the manual measurements are large for the WT mice, it should be recognised that the real differences between the measurements are small. In addition, the variations are comparable in terms of scale with inter-observer variations shown in Table 5.1.

Since the manual measurements of spine curvature index are not available, it is not feasible to compare the closeness of the automatic index measurements with manual measurement.

It should be realised that testing only the correlation and closeness of automatic measurements to the manual measurements is insufficient when there is no solid ground truth and when manual measurements have their own variations. The effectiveness of the automatic measurements must be judged when they are used in diagnosing scoliosis cases.

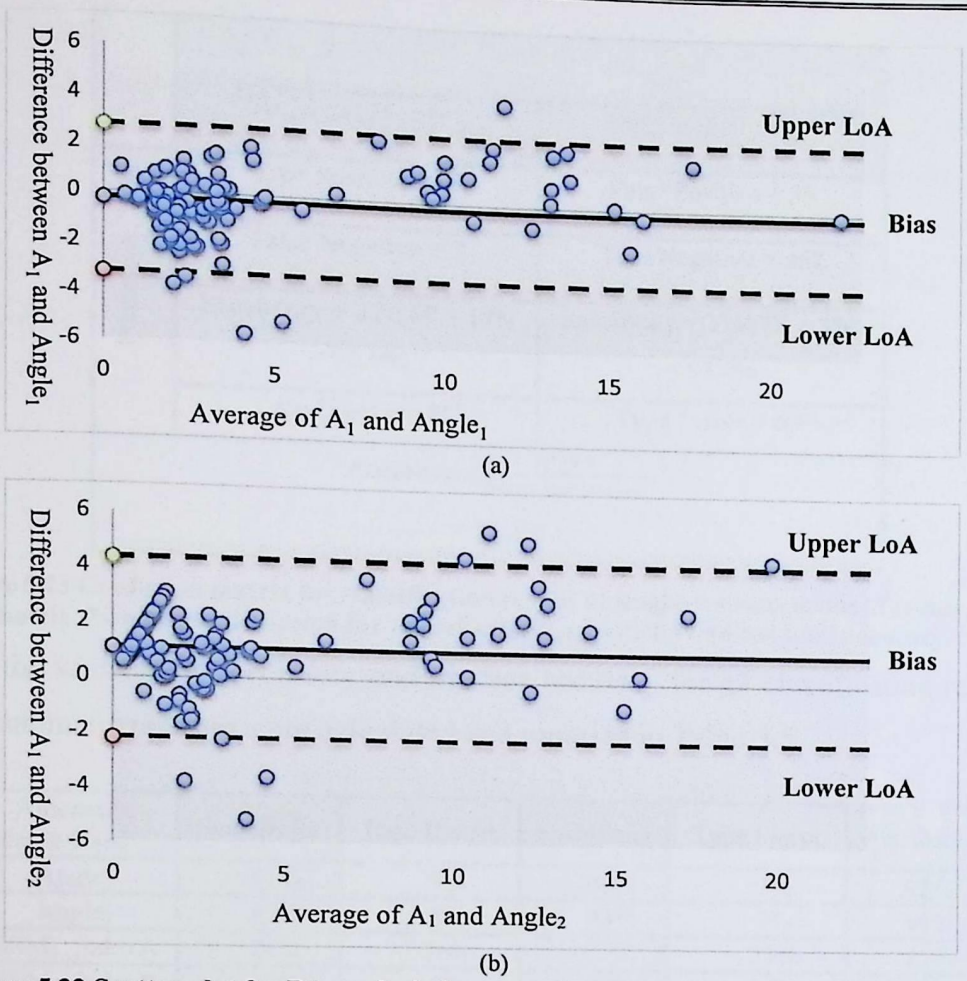


Figure 5.22 Scatter plot for BA analysis for manual measure versus $Angle_1$ and $Angle_2$
(a) Average of automatic measurements (Angle1) and manual measures taken by A_1 observer versus difference between them; (b) Average of automatic measurements (Angle2) and manual measures taken by A_1 observer versus difference between them

5.4.2 Automatic Identification of Scoliosis Cases

The aim here is to identify abnormal spines (Scoliosis) from the normal ones using the proposed angle and index based measurements. After obtaining the spine curvature estimation metric from the X-ray images, we classified it to normal and abnormal using the kNN classifier and leave-one-out cross validation as explained in Section 3.4.3. Figure 5.23 shows the confusion matrix calculation for $Angle_2$ measurements which includes the sensitivity, specificity and the total accuracy calculation.

	Actual	
	Scoliosis (25 cases)	Normal (500 cases)
Prediction	True Positive = 23	False Positive = 33
	False Negative = 2	True Negative = 467
	Sensitivity = TP/(TP + FN) = 92.0%	Specificity = TN/(FP + TN) = 93.4%
	Type II error = 8%	Type I error = 6.6
Accuracy = $\frac{TP+TN}{Total\ population}$ = 93.3%		

Figure 5.23 Confusion matrix for classification results of $Angle_2$ measurements of mouse spine. The matrix shows the calculation for the sensitivity, specificity and the total accuracy.

Similarly, the sensitivity, specificity and average accuracy for all classification results of all proposed automatic measures are calculated and reported in Table 5.5.

Automatic Metric Used	Sensitivity	Type II error	Specificity	Type I error	Accuracy
$Angle_1$	91%	9%	93%	7%	92%
$Angle_2$	92%	8%	93%	7%	93%
SRM_1_IndexA	84%	16%	77%	23%	81%
SRM_1_IndexB	95%	5%	84%	16%	89%
SRM_2_IndexA	91%	9%	87%	13%	89%
SRM_2_IndexB	90%	10%	90%	10%	90%

Table 5.5 Classification results using automatic spine curvature measures of mouse spine X-ray image.

In general, the results in the table show that all proposed automatic measures perform well in classifying scoliosis cases but SRM_1_IndexA performed marginally worse than the others. For cases of normal spine, the SRM_1 performed marginally worse than that for the scoliosis cases. Some normal spines are mistaken as scoliosis. This is because SRM_1 is formed depending on uppermost and lowermost points only. So, any inaccuracy in locating one or both of points might lead to maximizing the distances with the CSC.

Overall, $Angle_1$, $Angle_2$, and SRM_2_IndexB performed better than other measures. The results also indicate that measurements obtained from SRM_2 performed marginally better than SRM_1. This seems to confirm the earlier stated hypothesis of better promised performance of SRM_2 due to the fact that SRM_1 is determined by only two points whereas the SRM_2 is determined by all points on the CSC.

A close examination of angle based and distance based measurements shows that in some cases when one measure misclassifies the image, the other one gets it right. This can be exploited by applying a simple majority voting decision-level fusion. So, we decide to label the spine based on votes out of all 6 types of measure. The results of applying this decision level fusion (as shown in Table 5.6) demonstrate that the fusion between the different measures add a slight enhancement of total accuracy over the top accuracy in Table 5.5.

Automatic Metric Used	Sensitivity	Specificity	Accuracy
Decision level fusion between all 6 measures	95%	92%	94%

Table 5.6 Classification results using majority voting decision-level fusion among all the six measurements.

It is desirable to know the relationships between the automatic angle-based measures and the index-based measures. Table 5.7 shows the degrees of correlation between each of the 4 index-based measures with the $Angle_1$ measure. It is not surprising to see that the correlations between the angle-based measures and index-based measures are stronger for the whole data set and mutant subset, and weaker but still substantially for the WT subset because of the greater deal of variations in locating the landmarks for normal spine.

Methods Used (Index Measurements vs $Angle_1$ Measurement)	Whole Data Set of Mice (R^2)	Data Set of Mutant Mice (R^2)	Data Set of WT Mice (R^2)
SRM_1_IndexA	0.85	0.72	0.34
SRM1_IndexB	0.91	0.81	0.50
SRM2_IndexA	0.93	0.80	0.58
SRM2_IndexB	0.90	0.84	0.47

Table 5.7 Correlation between $Angle_1$ and Index-based measures. The results show high correlation for whole data set and mutant data set and low correlation for the WT data set.

It is also intriguing to establish what would be the Cobb angle threshold to separate normal from abnormal spines when manual and automatic measures are used in classifying the cases. We used a One-Rule classifier (equivalent to a one level decision tree) on the known cases and obtained the Cobb angle threshold of 5.57° for the manual A_1 average measures. The classifier also established the Cobb angle threshold of 6.04° for the automatic $Angle_1$ measure and 5.10° for the automatic $Angle_2$ measure. Both thresholds for the automatic measures are close (< 0.5°) to that for the manual measure.

It is important to note that the above results provide only a proof-of-concept. Since the proposed method was only tested on a relatively limited data sample and the differences between the performances of different measures are only marginal, it is not feasible to draw a solid

conclusion about the best measures. Both angle-based and index-based measures have their own merits.

Because of the limitation of the mutant cases with scoliosis disease in our data set, we could not recognize specific genes that might lead to change in the curvature of the spine and satisfy the 60% hit ratio threshold. However, the proposed method showed the ability to recognize the normal from abnormal cases depending on the spine curvature. Following the definition of abnormal cases, we can recognize two genes, namely 'ROSA26' and 'Sel1l' that could lead to changes in the curvature of the spine. Each of those genes has two cases which were recognized as abnormal curvature by the proposed method. Figure 5.24 shows several normal curvature spine cases and the abnormal curvature spine cases for the two candidate genes. Further investigation is needed to confirm the findings.

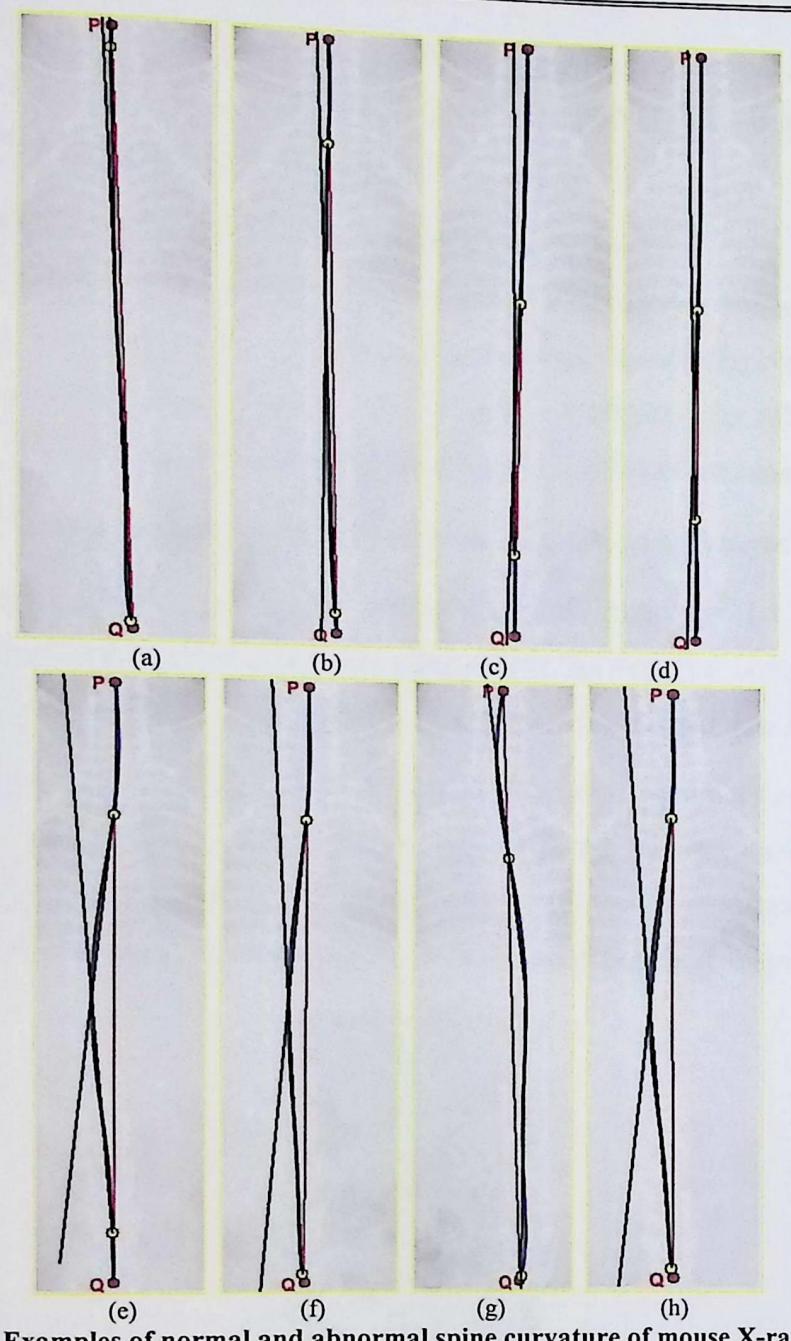


Figure 5.24 Examples of normal and abnormal spine curvature of mouse X-ray images.
(a), (b), (c) and (d) Normal spines which classified as normal by the proposed method; (e) and (f) Abnormal spine curvature resulting from 'ROSA26' gene knockout and they are defined as abnormal by the proposed method; (g) and (h) Abnormal spine curvature resulting from 'Sel1l' gene knockout and they are defined as abnormal by the proposed method

5.5 Discussions

In this section we present different discussion areas related to the proposed solutions. Firstly, the literature review section showed that vertebrae based solutions to estimate the spine curvature has been widely used on human X-ray images but are they effective on mice image? Section 5.5.1 addresses this question. After that, a discussion on the global landmarks detection for spine ROI segmentation using machine learning approach is presented in Section 5.5.2. This is followed by a discussion on scoliosis being 3D curvature estimated by 2D image and

the reasons behind possible inaccuracy in classifying normal and abnormal spines. Finally, the last section presents a brief description on the developed MATLAB tool designed for spine curvature estimation to be used by other researchers.

5.5.1 Investigation of the Vertebrae based Solutions

As explained in Section 5.2, several existing proposals rely on manual/semi-automatic annotation of two vertebrae (i.e. the start and end of the spine curvature) before estimating the Cobb angle from human spine images. This section investigates the effectiveness of two approaches to segment the mouse spine vertebrae based on RF and intensity profile.

Random Forest based approach to detect the top and bottom spine's vertebrae

This solution is based on using two RF regression tree models (Section 2.2.3) trained with the original pixel intensities extracted from patches around the top and bottom vertebrae. The distance from the centre of each patch to the centre of vertebra is used as a label of the patch.

At the detection stage, the trained models should be able then to predict the distance between the centre of any patch near the vertebra and the centre of vertebra itself. The U_2 and L_2 landmarks (explained in Section 5.3.7) are selected as an initial seed points to initiate the patches because they are close enough to the top and bottom vertebrae respectively.

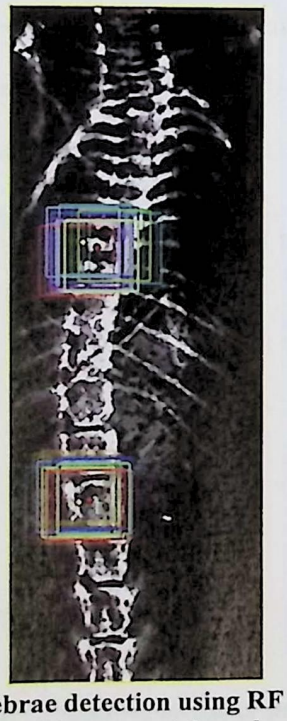


Figure 5.25 Top and bottom vertebrae detection using RF based approach. Several patches automatically set near the upper and lower vertebrae based on U2 and L2 (red points) to predict the centre of the vertebrae (yellow point).

Figure 5.21 shows the U_2 and L_2 landmarks in red and the patches around each of them. The extracted intensities from each patch are then fed to the trained RF models to predict a new

centre for the patch (toward the centre of the vertebra). The median of the coordinates of predicted centres from different patches is then used to estimate the coordinates of the actual centre of vertebrae as shown in yellow points in Figure 5.25.

Initial experiment showed that locating upper vertebra is not reliable but it has some promising results for the lower part which can be exploited in different problems. As we can see from Figure 5.25, even with the enhanced image, there are no clear boundaries between the upper vertebrae and they have different shape from the lower vertebrae. We believe that such challenges can highly complicate the segmentation of vertebrae of mouse spine with good accuracy. This observation suggests that the option of vertebrae segmentation to estimate the curvature is not the favoured option. However, the appearance of lower vertebrae (Lumbar vertebrae) shows a slight better appearance than the other part of spine. For this reason, in the next section, we will investigate another solution that focus on lumbar vertebrae detection.

Intensity profile based approach to detect the lumbar vertebrae

The second proposed approach to detect the vertebrae is relying on the intensity profile of the spine to recognize the vertebrae position. The idea was presented in (Peng et al., 2005) to segment the vertebrae from human spine MR image. The method aims to exploit the low intensity of the disk of vertebrae to approximate the vertebrae locations by generating an intensity profile for the spine ROI. To exploit the difference between the lower vertebrae and their disks, we use the enhanced image from Figure 5.7(c).

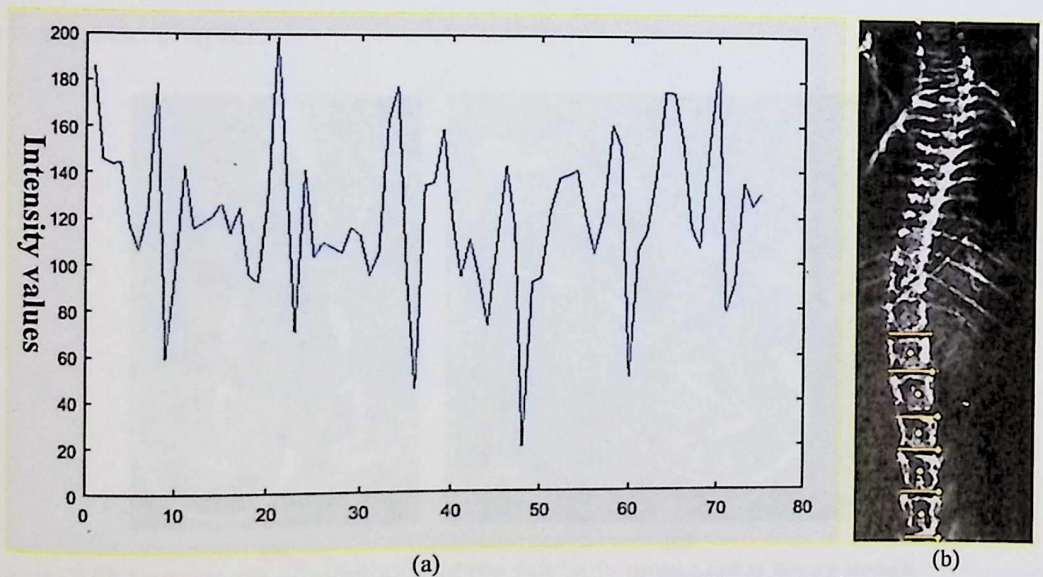


Figure 5.26 Lumbar vertebrae detection using intensity profile approach.
(a)Intensity profile for spine ROI (lumbar vertebrae). The local minima represents the possible location of the disk; (b) Detected lumbar vertebrae. The local minima from the profile are used to determine the location of each disk (yellow line) and then the borders and the centres of the vertebrae (yellow points) are estimated

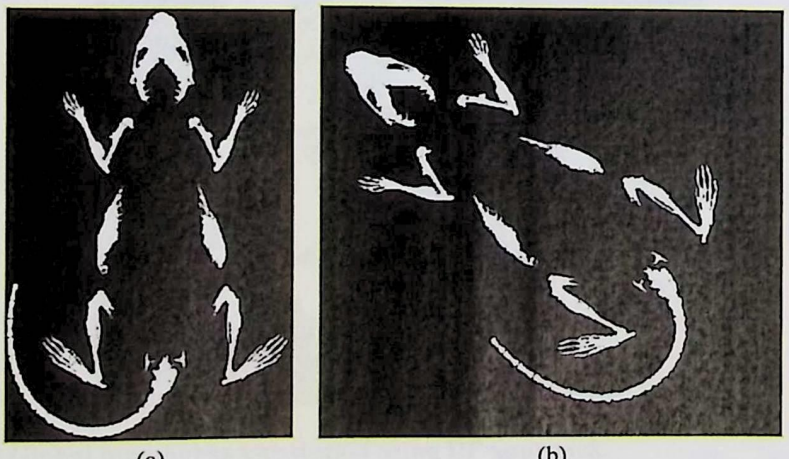
Because of the small size of the upper vertebrae and their complex structure along with the fact that the X-ray image covers the whole body, the borders or disks between the vertebrae tend to be unclear (Figure 5.26(b)). As the clear separation between the vertebrae is a requirement for the intensity profile method, we argue that the upper vertebrae are not a good target for the method. For this reason, we decide to implement this idea on the lower vertebrae of the spine to include all the lumbar vertebrae. The procedure starts by generating the intensity profile of the ROI area as shown in Figure 5.26(a). After that, the local minima in the intensity profile (disk location) are exploited to estimate the borders and centres for the lowest 5 vertebrae (Figure 5.26(b)).

Although the intensity profile is not ideal to estimate the spine curvature, initial investigations show promising results for lumbar vertebrae segmentation. For the MGP, the accurate segmentation of lumbar vertebrae could offer a valuable information and measurements that might underpin the genetic basis of lumber-based diseases.

As a conclusion, the difficulty involving in vertebrae segmentation from mouse images due to the reasons mentioned above confirms the effectiveness of spine curvature estimation based on the whole spine segmentation described in this chapter.

5.5.2 Importance of Alignment of the Mouse Body

The alignment of mouse full body was presented as the first stage in our proposal (Section 5.3.1). It forms the base for all later stages for spine segmentation and also rib detection as we will see in the next chapter.



(a) (b)

Figure 5.27 Importance of alignment of the full body mouse from X-ray image.
(a)Limbs-and-skull mask segmented from aligned image. It shows the possibility of exploiting the symmetry of the right and left parts of the body; (b) Limbs-and-skull mask segmented from not aligned image

We argued that the alignment is not a mandatory step but it can reduce the complexity in some stages. To show that, we prepared a limbs-and-skull mask for a mouse with and without

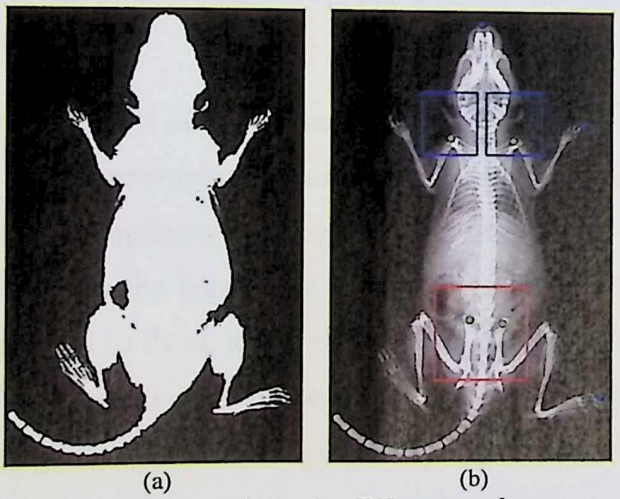
alignment as illustrated in Figure 5.27(a) and (b) respectively. It can be argued that using the rotated mask can make exploiting the symmetry of the left and right body parts easier.

5.5.3 Global Landmarks Detection Using Machine Learning

In Section 5.3.2, we discussed how to locate global landmarks for ROI segmentation using mask based approach. In this section, a discussion about locating global landmarks using a texture based RF approach is presented. In addition, a brief comparison between the two approaches will be given.

Global landmarks detection using texture based Radom Forest

The global landmarks detection from mouse body could be dealt with as a machine learning problem. Can RF model (Section 5.3.6) be exploited effectively in this context? The motivation to use RF models to detect the global landmarks is of two folds: 1) it did show good accuracy in landmarks detection from X-ray images of mouse head (Chapter 4), and 2) the area of the landmarks seems to have distinct texture from the surrounding areas. Similar to the previous approach, we select the two shoulder plates as upper global landmarks. For the two lower landmarks, we decide to locate the upper most points of the pelvis bones (Figure 5.28(b)) for two reasons. Firstly, this choice of points shows the power of machine learning over the mask based approach in detecting the landmarks from different parts of the mouse body. It can be also argued that the y-coordinates of the top most points of the pelvis and the lower global landmark points (Section 5.3.6) should be the same or very close.



(a) (b)

Figure 5.28 Global landmarks annotation using RF approach.
(a) Binary mask of mouse body in X-ray image. This mask is used to estimate the initial search area for the RF approach; (b) Upper search areas (blue rectangles), lower search area (red rectangle) and annotated global landmark (green points) using RF-HOG model

For each of the landmark, a RF-HOG model was trained using 100 images using similar procedure as explained in Section 4.3.5. In the prediction stage, a binary mask of the whole

mouse body (Figure 5.28(a)) is used to estimate the search areas (rectangles in Figure 5.28(b)) for the landmarks. Finally, the four global landmarks are predicted using RF-HOG model as shown by green points in Figure 5.28(b).

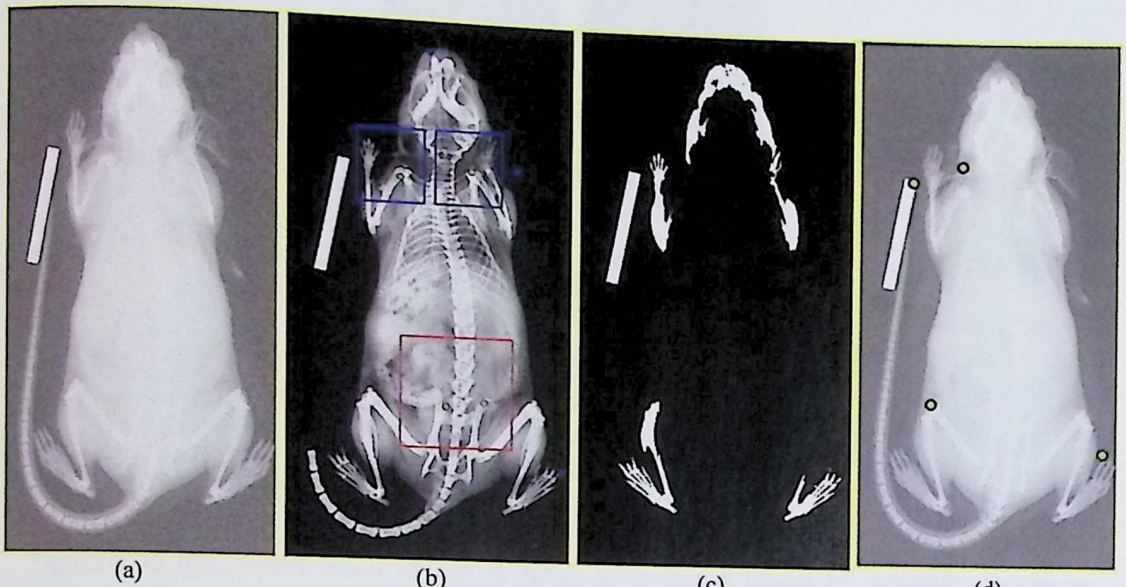
Mask based versus texture based Radom Forest for global landmarks detection

Following presenting the two methods to locate the global landmarks, we provide a general discussion in terms of accuracy, efficiency and robustness.

- Accuracy: The mask based approach is heavily dependent on the quality of limbs-and-skull mask i.e. it can detect the global landmarks accurately if the mask is segmented correctly. However, the texture based RF landmarks detection could be affected by other factors such as the search area. If the search area is allocated in the wrong place, there is no chance to locate landmark accurately. Moreover, the classifier might produce a lot of false positive samples if the search area is relatively big.
- Efficiency: The mask based approach has no complex operations or feature extraction. This is why it is more efficient than the RF approach. Even we use RF approach to detect only 4 landmarks, it still needs more computational time than the mask based approach because of relative big search area.
- Robustness: Like any heuristic approach, the mask based approach needs pre-setting for several thresholds and parameters to work correctly on a specified data set. Apart from the training operation which is needed, the texture based RF approach needs only a determination of the search area, and hence a better robustness than the mask-based solution.

A good case to investigate how both approaches are responding in different circumstances when applying them on an example images from different data set. Figure 5.29(a) shows an input image from a completely different source, the MRC Harwell (MRC Harwell, 2017). As we can see from the figure, the image of the mouse different resolutions and different mouse placement to those in our data set provided by Sanger Institute. To apply the texture based RF as explained above, we need a slight change in the procedure of upper search areas determination to locate them correctly. Apart from this, we apply the same trained models were developed for our data set and the results as shown in Figure 5.29(b). For the mask based method, we segment the limbs-and-skull mask without any change on our method as shown in Figure 5.29(c). It can be seen that the produced limbs-and-skull mask is not as we expect and hence the global landmarks annotation are not accurate (Figure 5.29(d)). To obtain a suitable limbs-and-skull mask for the new data set (i.e. similar

to Figure 5.5(e)) we need to change many parameters and adding an extra steps to cope with the changes in the image.



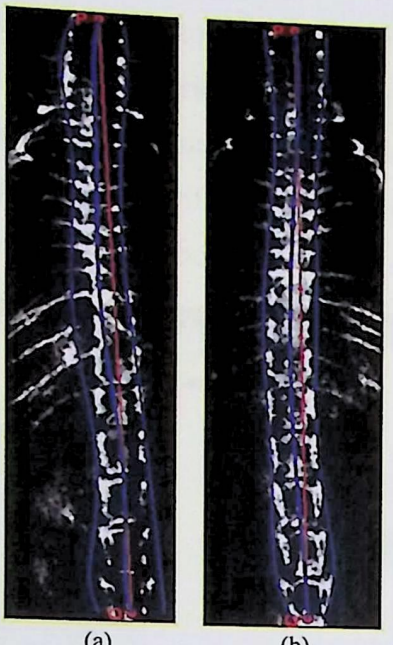
(a) (b) (c) (d)

Figure 5.29 Applying machine learning and mask based approaches on image from MRC Harwell (image is provided by MRC Harwell).
(a) Full-body mouse image from MRC Harwell; (b) Global landmarks (green points) detection using RF-HOG model; (c)Limbs-and-skull mask segmentation of input image in (a); (d)Global landmarks detection (green points) using mask based approach. The detection of landmarks is inaccurate

To sum up, it is be argued that the accuracy of the mask based approach is consistent on the same dataset. In addition, the mask based approach is more efficient but less robust than the texture based RF approach. As we are dealing with thousands of images, we argue that efficiency could be more important than robustness.

5.5.4 Sources of Inaccurate Normal/Abnormal Spine Classification

Scoliosis is defined as a 3D spinal deformity (Rajiah, 2015). Therefore, in addition to using traditional Cobb angle measurement to quantify the curve of the spine, vertebral rotation is another important factor to study of scoliosis. The comparison study in (Lam et al., 2008), showed the limitation of radiograph images to show of the spinous process of vertebrae with larger rotation. Therefore, the accuracy of Cobb angle measurement may be decreased when there is a distortion in a spinous processes tip. As a result, a decision on identifying scoliosis from 2D X-ray imaging for spine is far from being 100% accurate by both machine and human experts. As we can see from Tables 5.5 and 5.6, there are some false cases. A close investigation of the images that were repeatedly misclassified shows that false positive cases are mainly due to the spine in the original images having a mild curvature. Such a mild curvature may increase as the result of inaccurate fitting of one or both borders. Figure 5.30(a) shows an example image which labelled as normal by an expert and misclassified as abnormal by our method (Type I error).



(a) (b)

Figure 5.30 Misclassification of mouse spine from X-ray image by automatic proposed method.

(a) Mouse spine X-ray image labelled as normal by the manual annotator while classified as abnormal by the proposed method (Type I error). This is mainly because of existing of mild curvature in the spine; (b) Mouse spine X-ray image labelled as scoliotic (abnormal) by the manual annotator while classified as normal by the proposed method (Type II error)

On the other hand, some cases were subjectively labelled as scoliotic by an expert while they looked closer to normal images in terms of curvature degree. Figure 5.30(b) shows a positive case which was misclassified (Type II error). Such cases are difficult to identified correctly because they either misclassified by the expert or they considered as scoliosis for a different reason which not recognized by 2D X-ray images as explained earlier.

We argue that if the spine looks straight, there is a very little chance to be mistakenly classified as false positive, and the same argument is true for the images with a recognizable curvature.

5.5.5 MATLAB tool for Spine Curvature Estimation

A toolbox containing the proposed method for spine curvature quantification has been implemented using MATLAB, and is now publicly available[1]. The tool returns all six possible automatic measurements for selected image as illustrated in the screenshot in Figure 5.31. The user has an option to select and browse an image from a specific folder by using "*Image Browser*" button. The selected image will then have displayed on the GUI as shown in the Figure 5.31. Once the image is loaded, the curvature quantification of the six-different metrics is done and the different metrics are shown after pressing "*Estimate Metrics*" button. Furthermore, the figure for each metric calculation method is displayed to give the user an idea

[1] http://www.buckingham.ac.uk/research/appliedcomputing/sce-tool

of how the measure is calculated. The output from different metrics gives the user an opportunity to compare between different metrics in terms of the method of the calculation and the result of curvature quantification. The quantities are then saved in a separate excel file. The final important button, *"Classify Image"*, will display the class of image as "*Normal*" or "*Abnormal*" based on kNN classification method.

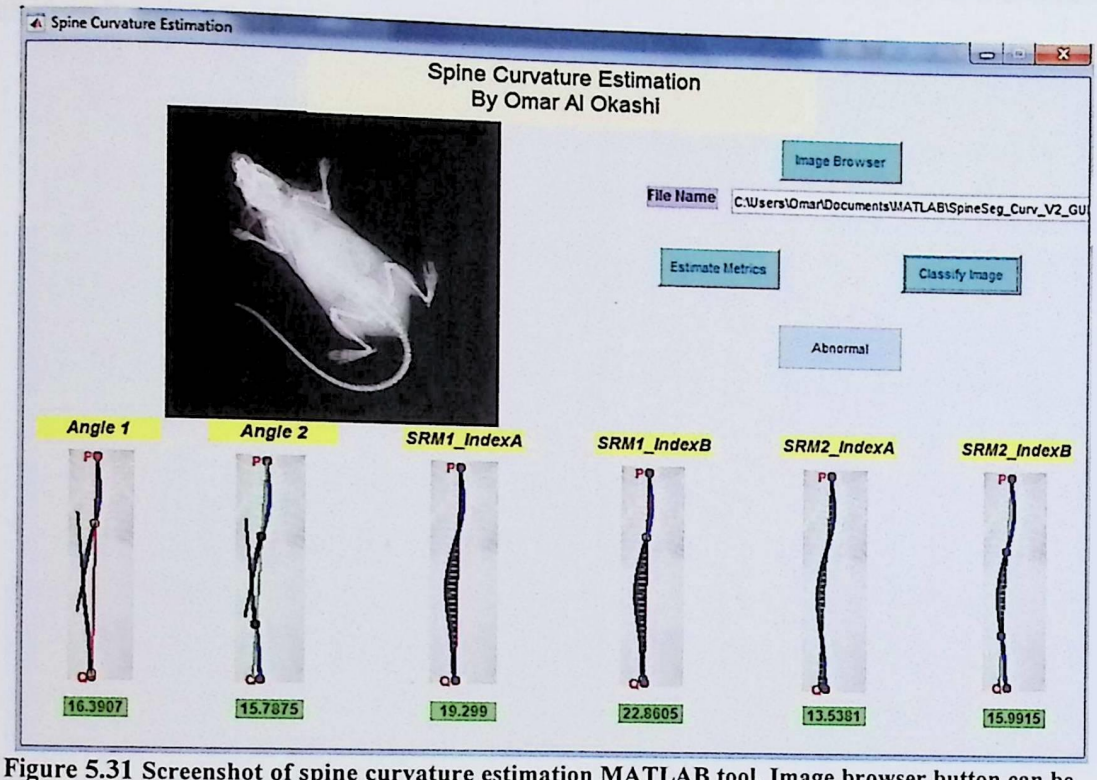


Figure 5.31 Screenshot of spine curvature estimation MATLAB tool. Image browser button can be used to browse the image file. To estimate the metrics, "Estimate Metrics" button is used. After that, the "Classify Image" button shows whether the spine is normal or abnormal.

5.6 Summary

Spine curvature estimation is essential to evaluate and quantify the degree of spine deformity in diseases such as scoliosis. In this chapter, we proposed an automatic method for spine curvature estimation from X-ray images of mice using different angle and index based measurements.

The principal idea of our solution is based on initial segmentation of ROI of spine by using the limbs-and-skull mask for this purpose. After that, the reliable fine contour of the spine is obtained by applying image enhancement and refinement techniques. This helps in extracting an accurate spine line which in turn leads to obtaining correct landmarks to quantify the curvature. Accurate spine contour allows locating accurate location of landmarks that govern the curvature estimation. The proposed method is augmented with some reliability checks to exclude certain images which may not produce accurate results.

The performance on a sample of images showed good results, in both the closeness to manual measurement and the classification evaluation, clearly demonstrating the promises and potentials of the proposed method. The decision level fusion of angle based measures with index-based measures could further enhance the classification results. As a side-product of our work, we produced a software framework for the proposed method and made it available for researchers in the community.

Chapter 6: Ribs Detection and Counting from Mouse X-ray Images

After the segmentation and quantification of the spine in the previous chapter, the next task is to annotate the X-ray images of mouse body to produce a correct count of the ribs. The abnormal number of the ribs is said to be related to some form of genetic abnormalities (Wattanasirichaigoon et al., 2003). For example, human congenital abnormalities can lead to additional ribs and potential nerve damage to the arms (i.e. thoracic outlet syndrome), whereas loss of ribs can lead to damage within the thoracic cavity due to a lack of protection (Wattanasirichaigoon et al., 2003).

This chapter proposes a pattern recognition based solution for the rib detection and counting based on texture and geometrical features of the ribs from the dorsoventral view X-ray images of full mouse body. The rib count encounters a list of problems which need to be solved. Because we consider a full body mouse image, we need firstly to segment the rib ROI to avoid existence of irrelevant body parts. For some images, obtaining accurate rib count is infeasible due to extremely low-quality image or some limitations in the proposed method. We need to reject such images in early stage using some criteria based on spine mask segmented from the previous chapter. For the images that pass this reliability check, we need to increase their contrast in order for rib objects to be easily identified. Once the rib masks obtained and smoothed, all the background objects resulting from imaging 3D objects using 2D techniques and other irrelevant objects have to be filtered away before the rib can accurately counted. Two-Level classification scheme using texture and well-known geometrical features of rib is exploited to do this job.

The rest of this chapter is organised as follows. Section 6.1 explains the problem statement and outlines the difficulties of manual rib count. The literature relevant to segmentation methods of ribs in human X-ray images are reviewed in Section 6.2. Section 6.3 describes the proposed automatic solutions to detect and count the ribs. The results of rib count are reported and analysed in Section 6.4. A discussion about difficulties and source of errors in the proposed solution are presented in Sections 6.5. Finally, the summary of the chapter is presented in Section 6.6.

6.1 Problem Statement

Rib count refers to the accurate count of the number of ribs on both sides of the spine from an X-ray image of a mouse. Automatic rib detection and counting are concerned with effective machine based solutions to accurately count the number of ribs within a rib cage. Manual counting of the ribs is considered as a time consuming and tedious process. This is because many ribs (~ 26) need be count in each image, and there can be a huge number of X-ray images that are continuously generated from the problem domain (i.e. MGP). In addition, the low contrast of certain ribs near the abdomen area requires a great concentration of the mind and a good level of experience. Figure 6.1 shows an example of manual counting of ribs on the both sides of a mouse in X-ray image. The count of 13 ribs on each side indicates that the mouse is of normal type.

Automatic detection and counting of the ribs from full body dorsoventral X-ray images faces the following challenges.

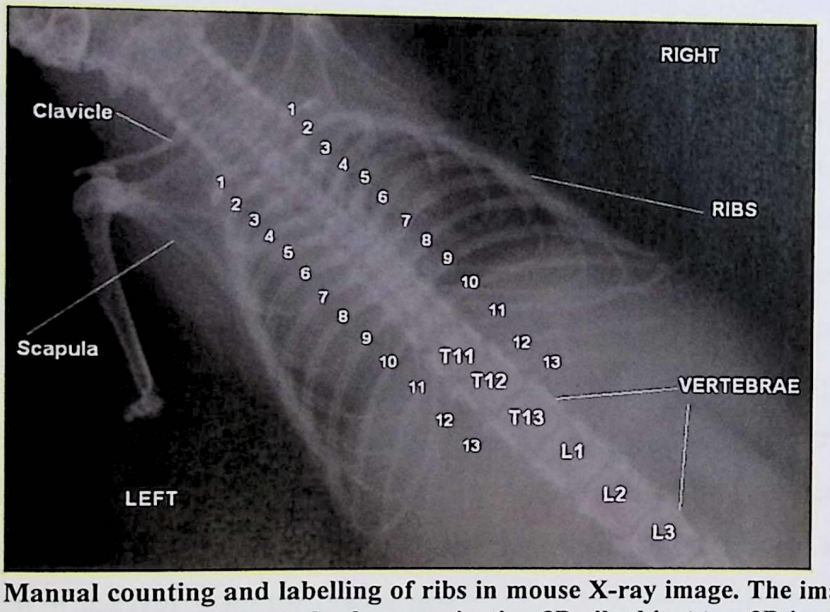


Figure 6.1 Manual counting and labelling of ribs in mouse X-ray image. The image shows the overlapping rib shapes that results from projecting 3D rib object to a 2D image space.

First, the source images in our study are full-body images and hence accurate segmentation of ROI, i.e. the rib areas, is an essential and critical pre-processing step. Second, certain ribs can be difficult to distinguish even for humans due to the lack of contrast between the ribs and the backgrounds, i.e. subtle differences between intensity values of nearby pixels. For instance, in Figure 6.1, ribs 12 and 13 on each side have much poorer contrast than other ribs. Third, since a rib is a curve shaped 3D object, projecting such a 3D object to a 2D image space by X-ray will lead to many overlapping rib shapes for most, if not all, ribs especially at the edge of the rib cage, adding confusions and hence difficulties in counting the ribs correctly. This is evident

from the rib shapes inside the X-ray image in Figure 6.1. Furthermore, the inherent noises in X-ray images (see Section 1.3 for more details) presents technical difficulties in the ribs identification.

6.2 Literature Review

Similar to the situations as stated in Chapters 4 and 5, the research regarding the automatic rib detection and counting from images of mouse is very limited. Most relevant existing work has been about the rib segmentation and measurements from images of humans for different medical modalities. In the existing work, we can recognize three main purposes of ribs segmentation from human images. The first purpose is to enhance the lung visibility by removing the segmented ribs as unwanted objects. The second purpose is to use the rib cage as a reference to determine the locations of other soft tissues in the chest. The rib measurements and fracture diagnosis are the final purpose of ribs segmentation.

In this review, we categorize the existing works according to the used medical modality into CT-scan and X-ray imaging as follow.

A framework to segment and label ribs in 3D computed tomography scan images was proposed by Staal, van Ginneken and Viergever in (Staal, van Ginneken and Viergever, 2007). Firstly, the ridge voxels of the ribs were detected by applying a selected threshold and followed by a ridge detection method that considers a rib as a 1D curve structure. Then the detected ridge voxels were grown into a primitive after examining certain conditions for candidate voxels. The detected primitives were then classified to either foreground (ribs) or background by using a spin-glass classifier with local and global features. From a primitive, a centre line was then constructed by using the same ridge detection method for ridge voxels. Those centre lines were then labelled heuristically, and finally a seeded region growing algorithm was used to perform a full segmentation of the ribs from the centre lines.

Klinder et al. (Klinder et al., 2007) suggested a model-based method to segment and label the rib cage from human chest CT scan. Triangulated surface models of all 24 ribs were initially constructed from patient CT scan images through an interactive segmentation process. After that, the models were fit to their corresponding anatomical objects by using a shape constrained model in modelling an initial rib cage. To determine the proper position of the final model, ribs were detected by applying a radial cylindrical ray search in every image slice and observe the pattern of the grey value which should be high at the start and end of the ribs and low elsewhere.

was used to reduce the noise of the lung fields. After that the Otsu segmentation method was used to detect a lung parenchyma. Finally, ribs were denoised using the wavelet denoising method and morphology operations such as corrosion and expansion. However, the final result shows connected ribs as well as non-rib objects resulting from chest tissues and other organs, which makes it unsuitable for automatic counting of the ribs.

The work proposed by Lee et al. (Lee et al., 2012) aimed at locating the ribs to supress them from lung area at a later stage. To determine the boundaries of the lung, they proposed to use ASM. Their solution was based on using Generalized Hough Transform (GHT) but with a local sampling scheme i.e. generating a different rib template for each processed image. A sequence of enhancements operations (bilateral filter and histogram equalization) and Sobel operator to detect the border of ribs were applied before the rib tablemate was selected from the low-edge image. After that, GHT followed by a Gaussian for smoothing, Otsu for thresholding, and opening and closing operations was used to obtain a rib boundary.

Li, Luo and Hu (Li, Luo and Hu, 2015) presented a method that segments the ribs from X-ray radiograph of the human chest. Their proposal started by detection the lungs boundaries using ASM so all later operations were applied on segmented lungs fields. After that, the median filter was used to smooth the image before it was subtracted from the original one to obtain a detail image that contains the ribs. At this point, a small threshold was applied to segment the ribs. The resulting binary image was then smoothed using a small sliding window that either set their pixel's value to 0 or 1. To reduce unwanted objects and separate connected ribs in the binary image, a multiple template matching model was used to extract the rib area. After that, they used a graph-based method to extract the real rib structure and remove non-rib tissues. This method was based on thinning all the structures in the binary image to one-pixel width. The centre line of the structure was then found as a shortest distance between the two extreme points using Dijkstra algorithm. SVM with five geometrical features that describes the rib was then employed to build a rib model to eliminate any centre line that does not fit the model. To refine the result, a trained decision tree classifier was tested by the geometrical features of existing ribs and their count. The output of the classifier is how many missing rib or extra non-rib objects existed.

Li et al. (Li *et al.*, 2015) used the rib centre lines located by their previous solution to determine the border of the ribs with the aim to supress them to achieve more visibility to nodule in the lungs. For this purpose, they used the rib centre line to separate the lung field in sub regions where each region includes one rib. After that, a rib model was built to describe the grey-level

intensity in the rib region by using PCA and point distribution model. Finally, the rib model together with the calculated background intensity were used to supress the rib texture from the lung field.

In (Cong, Guo and Li, 2016), a method was proposed to segment the ribs from human radiograph images with intention to remove them later to enhance the lung visibility. The pre-processing step in their proposed method used Gaussian filter to enhance the rib contrast and bilateral filter to reduce the noises from the input image. To obtain the lower boundaries of the ribs, a Sobel edge enhancement filter was applied. After that, the GHT was used with nine manually created rib models to locate the lower boundaries of the ribs. Orthogonal lines to every point in the lower boundary were arranged column by column to generate two-dimensional image where the rib was brought to approximately horizontal posture. Finally, they proposed a bilateral dynamic programming algorithm based on cost function that incorporated the width of the ribs and the smoothness of the rib boundaries to estimate the parallel boundaries of each rib.

All the previous four proposed methods were focused on the nine most visible ribs on each side that overlap with the lung in human X-ray images, so they were not targeting at the detection the whole set of possible ribs. Because of some common properties of X-ray images of human and mouse (i.e. lack of contrast), some enhancements and pre-processing methods discussed can still be valid for enhancing the X-ray images of mouse. From those techniques are the using of bilateral filtering to reduce the background noises and histogram equalization to increase the image contrast. For the initial segmentation of the ribs, most of the work cited in the literature used a fixed threshold value to binaries the image. However, we can argue here that any threshold value will not help in obtaining initial rib mask in case of mouse X-ray image (to be explained in Section 6.3.3). The reason behind is that the quality of mouse X-ray images in our data set is much poorer than the human chest X-ray images. Also, the lower ribs appear very thin. The contrast of the region is so low that the ribs cannot be easily detected by any threshold value.

Besides the enhancements and pre-processing methods as described above, two main approaches for rib detection are in the existing literature. The first one is the rib template matching to locate the rib location while the other is using machine learning to separate ribs from non-rib objects. Using rib template matching in our case will not be very useful because of the wide variety of shape, size and orientation of lower and upper ribs across images. However, using classification is still an interesting idea that may help ribs recognition in our

problem domain. In addition, some ideas in graph-based methods can also be useful for extract real rib structure at the post-processing stage of our work.

In general, several ideas and techniques proposed in the literature are found to be potentially useful for our proposal for rib detection and counting for mouse. However, since there are no reports on utilising these ideas and techniques in directly addressing the problem of rib segmentation and counting in biomedical mouse images, the effectiveness of these techniques is yet to be seen.

6.3 The Proposed Method

The complex nature of the rib area in the X-ray images requires a hybrid solution that combines different techniques to segment and count the ribs. After considering the alignment of the mouse body described in Chapter 5, we present our solution as a sequence of six stages as shown in Figure 6.2.

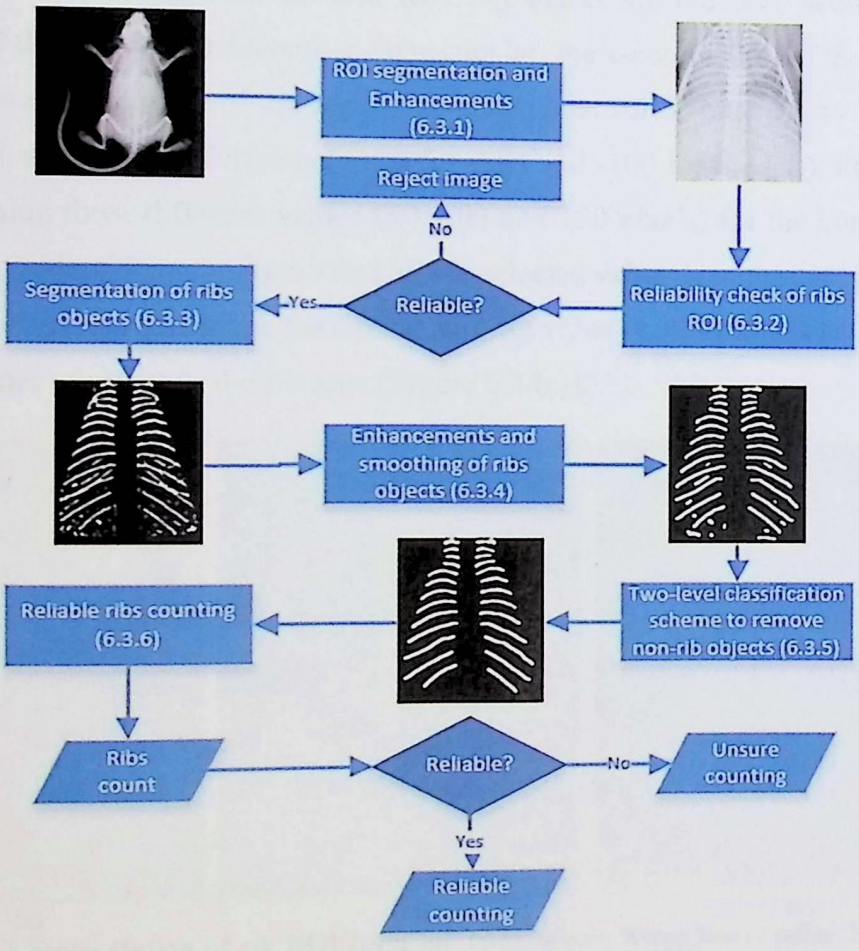
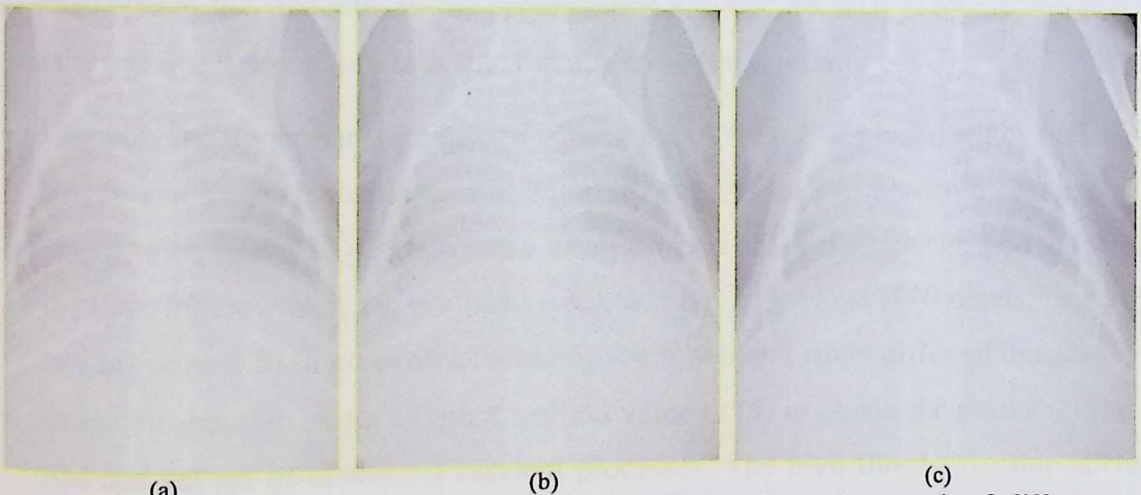


Figure 6.2 Main steps of rib detection and counting method. Method starts by cropping and enhancing of the ribs ROI from the full body mouse X-ray image. This is followed by a reliability check for the ribs ROI. After that, segmentation and post processing operations are done to obtain the rib masks. Finally, the two-level classification scheme is applied to remove irrelevant objects and then counting the ribs.

The solution starts with segmentation and enhancements of the rib area (Section 6.3.1). Next, we perform a reliability check to evaluate the goodness of the spine mask obtained from Chapter 5 before using it to disconnect the ribs from the spine (Section 6.3.2). After passing the reliability check, a three-step process (Section 6.3.3) is followed to obtain the candidate ribs. The resulting image is then further processed by morphological operations to enhance and smooth on the candidate ribs (Section 6.3.4). The process also removes some non-rib objects from the image. After this process, a two-level classifier using texture and geometrical features is used to further exclude non-rib objects (Section 6.3.5). At the final stage, the remaining rib objects are counted, and the counting result is returned (Section 6.3.6).

6.3.1 Segmentation and Enhancements for Rib ROI

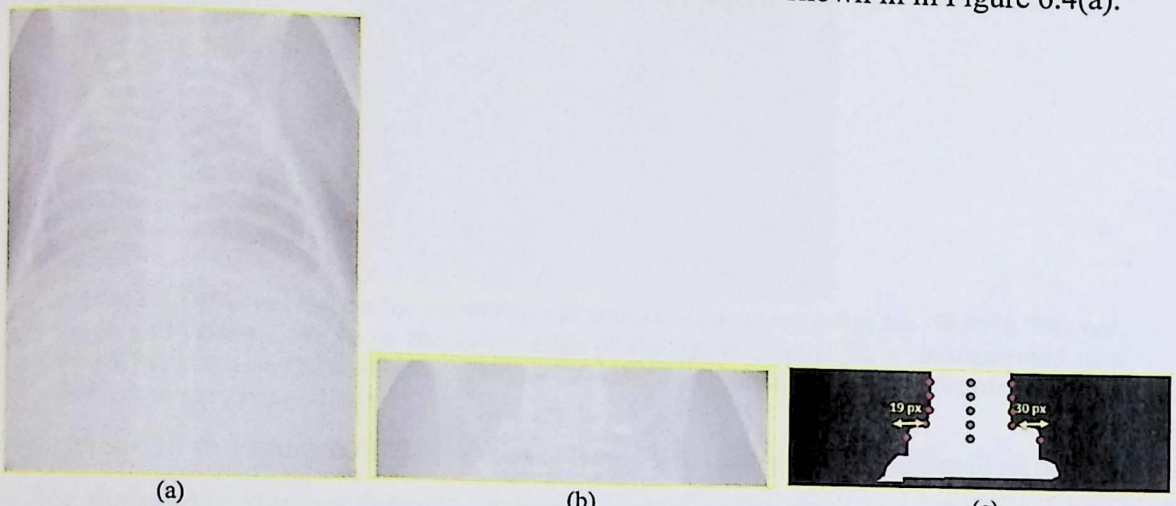
The proposed solution uses the global landmarks obtained from limbs-and-skull mask as explained in Section 5.3.2 to prepare for the rib ROI segmentation. In order not to miss any possible ribs, the coordinates of the ribs ROI landmarks are prepared according to visual observations of the images we encounter. In particular, the x-coordinates of the four points are estimated as the x-coordinates of the upper two global landmarks (blue points in Figure 5.5(f)) with horizontal shifting of 100 pixels (+100 for right and -100 for the left). Figure 6.3 shows the effect of using three different values (80, 100 and 120 pixels) for the horizontal shifting parameter. It is evident from the figure that all the selected values lead to cropping the ribs ROI correctly. However, increasing the horizontal shifting value could lead to including unwanted objects (i.e. bones of arms) in the rib area (Figure 6.3 (c)).



(a) (b) (c)

Figure 6.3 Segmentation of rib ROI from full body mouse X-ray image using 3 different horizontal shifting values.
(a) Segmented rib ROI using 80 pixels for horizontal shifting; (b) Segmented rib ROI using 100 pixels for horizontal shifting; (c) Segmented rib ROI using 120 pixels for horizontal shifting. Increasing the value to 120 pixels could lead to include some unwanted objects in the rib area

Estimating the y-coordinates of the landmarks is less trivial. For the two upper landmarks of rib ROI, setting their final y-coordinates to the y-coordinates of the upper two global landmarks can guarantee the enclosure of all the ribs with some part of cervical vertebrae. The y-coordinates of the lower two landmark points is set to 500 pixels below the y-coordinates of the upper landmark points. By observation, this length guarantees involving all the ribs with a small margin below the final ribs. The segmented rib ROI is shown in in Figure 6.4(a).



(a) (b) (c)

Figure 6.4 Estimation of the accurate upper cut point for rib ROI.
(a) Initial segmented rib ROI before refining upper cut; (b) Cropped upper part of the rib area which includes the first upper rib; (c) The binary mask of (a). The red points represent the edge of the binary mask while the green points represent the middle of the mask

However, the upper area of the ROI contains the clavicle bones and the processes of the cervical vertebrae as we can see from Figure 6.4(a). There is a high probability for such objects to be segmented in a later stage and wrongly considered as ribs. The confusion could happen because of the small size of the first ribs so the accurate upper cut of the ROI is highly required. We propose the following procedure to refine the upper cropping point.

- Cropping the upper part of the segmented rib area (first 120 pixels) which includes the first two or three ribs as shown in Figure 6.4(b). As we can see from Figure 6.5, any length value above 80 pixels to crop the upper part of the rib area can lead to including the first rib. However, we prefer to use a relative high value (120 pixels) (the yellow line in Figure 6.5) to guarantee including the first upper rib in different images.
- Image binarisation using a high threshold value (215) to obtain the masks for the first ribs and get rid of unwanted cervical processes. The high threshold value is chosen because the upper ribs usually have higher intensity values than the cervical vertebrae processes.
- Applying binary morphology operations, i.e. vertical closing and horizontal opening, to obtain a binary mask as shown in Figure 6.4(c).

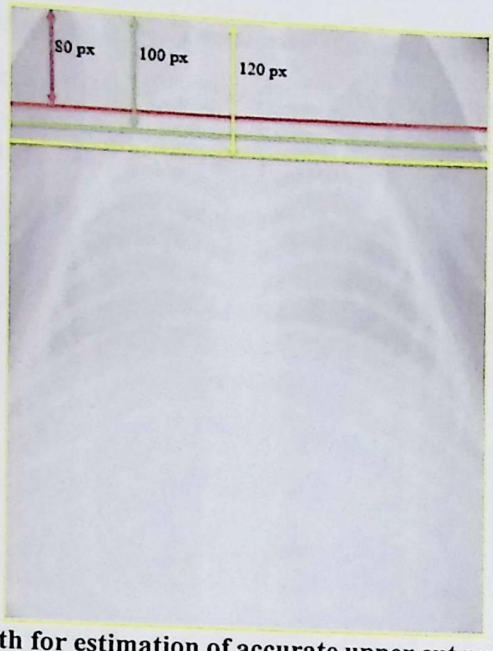


Figure 6.5 Different length for estimation of accurate upper cut point for rib ROI. The red, green and yellow lines show the lower edge of the upper part of rib area segmented with length 80, 100 and 120 pixels respectively. All the three distances lead to include the upper first rib in the cropped part.

As we can see from Figure 6.4(c), the mask should ultimately show the position of the first rib as a first sharp edge starting from the top point of the mask. Exploiting this observation, we aim to locate the ideal cut-off point for the upper boundary of the ROI as follow. We start from top middle point of the mask (green point in Figure 6.4(c)) to horizontally search toward left and right for the first black pixel encounters (red points in Figure 6.4(c)). This operation continues downward until the difference between the x-coordinate of the current edge point and the previous edge point (the yellow line in Figure 6.4(c)) is large enough (> 10). This mean the first upper rib has been found and so the cut-off point is determined at this location.

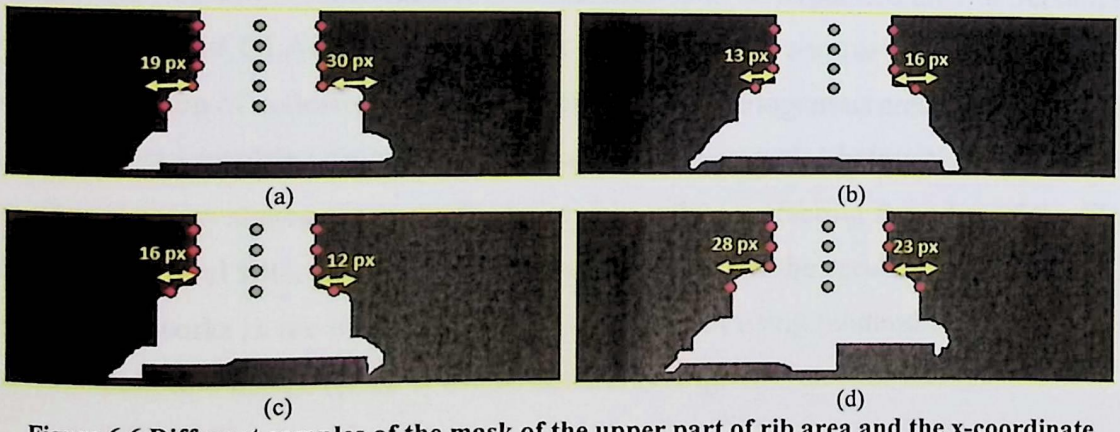
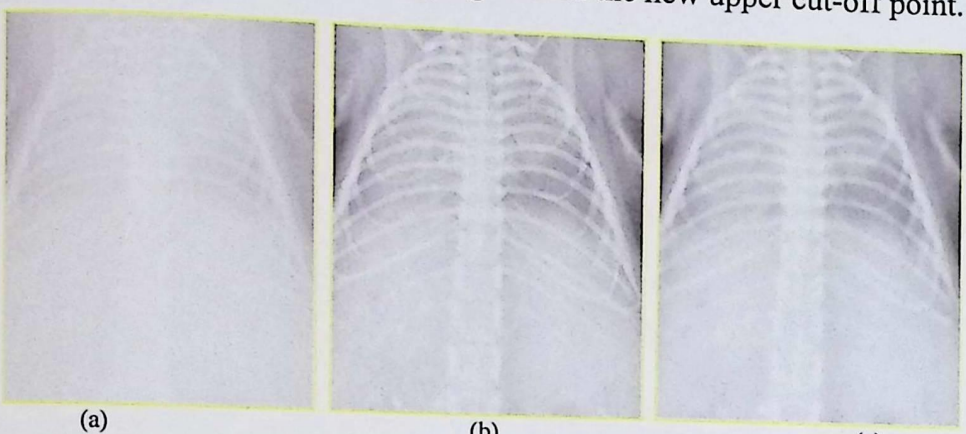


Figure 6.6 Different samples of the mask of the upper part of rib area and the x-coordinate difference values that enable detection of the first rib location. All the difference values for the right and left ribs are greater than 10 pixels as follows.
(a) (19,30); (b) (13,16); (c) (16,12); (d) (28,23)

Figure 6.6 shows four different samples of the mask of the upper part of rib area associated with the value of the difference between the x-coordinate of the last edge point and the point

before it (the yellow lines in Figure 6.6). The figure demonstrates that the difference value that leads to determine the y-coordinate of the first rib is always larger than 10 pixels.

Figure 6.7(a) shows the rib area after cropping it from the new upper cut-off point.



(a) (b) (c)

Figure 6.7 Rib ROI enhancements.
(a) Rib area after refining upper cut. This leads to removing the upper cervical vertebrae and start from the first rib; (b) Rib area after applying CLAHE on (a) to increase its contrast; (c) Rib area after applying the bilateral filter on (b) to smooth it

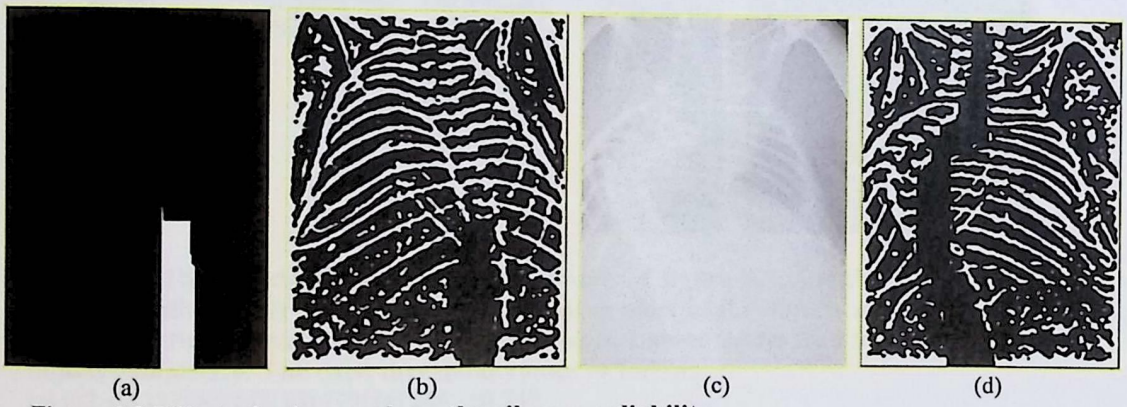
The segmented ROI as shown in Figure 6.7(a) may be of very low contrast that may lead to some ribs, particularly the lower ones, being missed by the later steps. To increase the ribs contrast and prepare them to be detected in the next stages, two image enhancement techniques are applied on the rib ROI. Firstly, we apply CLAHE (Section 2.2.1) to increase the rib contrast as shown in Figure 6.7(b). Using of histogram equalization to enhance the human rib radiograph image was presented by Lee et al. (Lee et al., 2012). The enhancements expected from CLAHE operation could be affected by the how many regions to divide the image before applying the histogram equalization (number of tiles parameter). However, we set this parameter to (8 × 8) and the rationale of this selection is to be explained later in Section 6.5.2. The good property of CLAHE is that it tends to increase the contrast to an acceptable level without proliferation of undesired noises especially in the homogenous area. At this stage, some background noises resulting from the soft tissues in between and below the ribs exist which might affect the next stages. To smooth the image without affecting the edge of the ribs, we decide to use bilateral filter as described in Section 2.2.1 and the result is as shown in Figure 6.7(c). Previous works in the literature showed the value of using bilateral filter to enhance the human rib radiograph image ((Lee et al., 2012) and (Cong, Guo and Li, 2016)).

6.3.2 Reliability Check of Rib ROI

Before proceeding to the following steps, it is important to conduct a reliability check upon the segmented ROI image. One of the important factors that can affect the accuracy of the rib detection and segmentation is the spine. The effect of the spine on the rib area could be

explained for two reasons. The first reason is that the spine mask plays an essential role in separating rib area (see Section 6.3.3). Inaccurate spine mask may cause some errors in rib detection and counting. The consequences of an unreliable spine mask have been discussed in the previous chapter in the context of spine curvature estimation (see Section 5.3.5 for more details). The second reason is that extreme spine curvature may affect the rib detection. Therefore, we derive the following schemes to exclude (reject) images as unreliable images:

- Short or undetected spine mask. If the spine mask is marked as shorter than the normal range for any reason (Section 5.3.5), the image is rejected as unreliable. Figures 6.8 shows a short spine mask (a) leads to unseparated rib area (b) (Section 6.3.3) to right and left ribs.
- Extremely curved spine. If the curvature estimation of the spine is very big (Section 5.3.5), the image is rejected as unreliable. In this case, the spine mask itself may be considered as acceptable but the consequence of extreme curvature spine could affect the rib objects considerably as shown in Figure 6.8(c). Figure 6.8(d) shows the rib area of Figure 6.8(c) after applying the spine mask. The value of the spine curvature to reject the rib area image is selected following observation of various examples of curvy spine.



(a) (b) (c) (d)

Figure 6.8 Effect of spine mask on the rib area reliability.
(a) Unreliable short spine mask due to missing upper parts of it; (b) Rib area after masking it by the short spine mask in (a). The short spine mask cannot separate the rib area to right and left rib properly; (c) Rib area with extremely curved spine; (d) Mask of rib area in (c) after applying spine mask on it. We can see that most of ribs are distorted because of the extreme curvature of the spine

6.3.3 Segmentation of Rib Objects

The goal of this stage of the process is to filter out ribs by separating them from the spine and the rib cage. This complex task is performed through three sequential steps as explained next.

Rib isolation from its background using BHPF

The first step is to extract all or at least most of the rib objects from the image through binary masking. The most existing works reported in the literature did this step by using a fixed threshold value ((Staal, van Ginneken and Viergever, 2007), (Li et al., 2010), (Lee and Reeves,

2010), (Cui et al., 2012), (Li, Luo and Hu, 2015)). For the data sets we deal with, simple thresholding is ineffective because of the poor contrast between intensity values of rib objects and their surrounding tissues as illustrated in Figure 6.7(c). A fixed threshold value normally confuses rib objects and other tissues. Figure 6.9 supports this observation by showing the effects of thresholding at different values to the segmentation outcomes. The segmentation results either show no clear rib objects or merged rib objects, none of which are desirable.

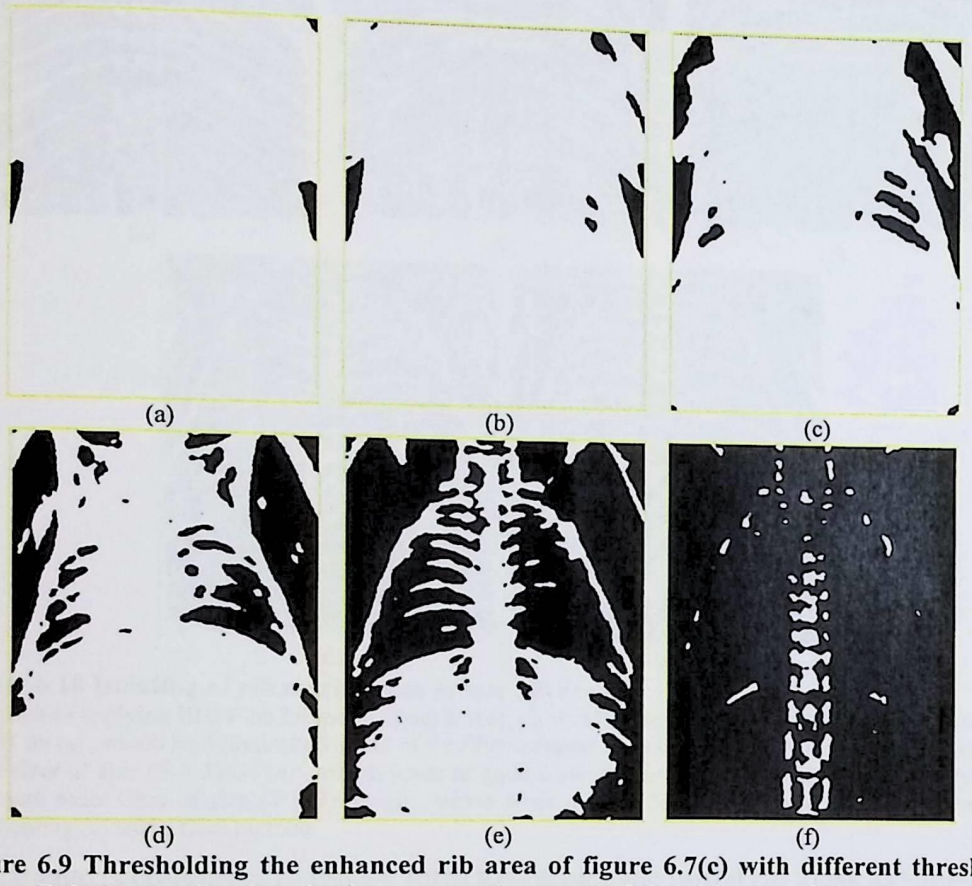


Figure 6.9 Thresholding the enhanced rib area of figure 6.7(c) with different threshold values. Relative small values lead to segmenting most of the objects and tissues and so no clear separation between rib and non-rib objects. Larger values lead to diminishing the rib objects completely. The used threshold values as follows
(a) 102; (b) 127; (c) 153; (d) 179; (e) 204 ; (f) 230

For this reason, we decide to use BHPF of first order as described in Section 2.2.1 to highlight the rib objects. We set the cut off frequency parameter to be 50 to achieve a good level of isolation from background and at the same time not to lose important information from the rib objects as shown in Figure 6.10(a). Further discussion of the selection of cut off value is presented in Section 6.5.2. In addition to the required details of the image, the filter also highlights different non-rib objects in the background. To mitigate these small non-rib objects, a minimum order filter (Section 2.2.1) is required. To decide the suitable window size for the filter, we apply three filters with different sizes ((3 × 3), (5 × 5) and (7 × 7)) and the results are as shown in Figures 6.10(b), (c) and (d) respectively. It is evident from the figure that the filters

of window sizes of ((5 × 5) and (7 × 7)) gave better results than the filter of (3 × 3) window size. We select to use the filter with a window size of (5 × 5) because it reduces the noises and doesn't affect the rib objects (i.e. making the ribs very thin).

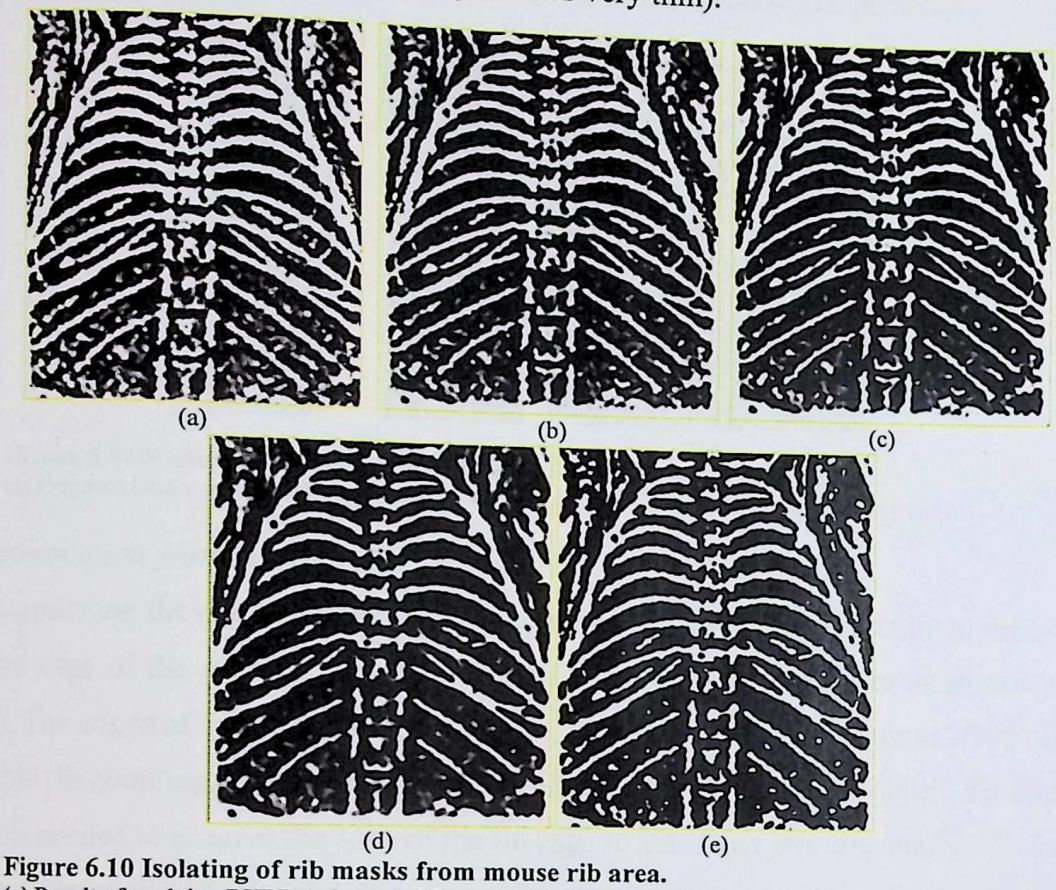


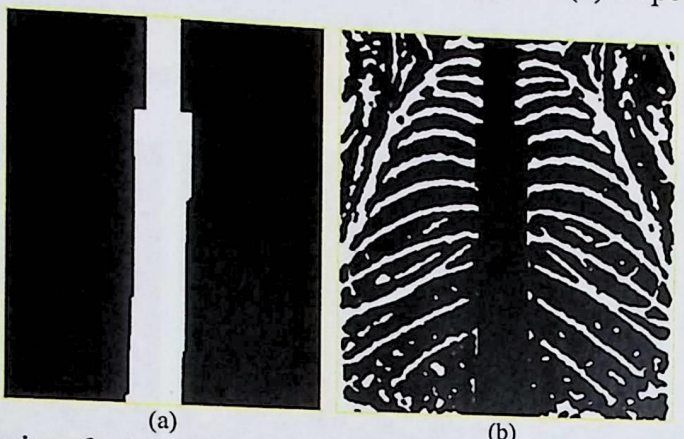
Figure 6.10 Isolating of rib masks from mouse rib area.
(a) Result of applying BHPF on image of figure 6.7(c); (b) result of applying a minimum order filter of size (3 × 3) on (a), which leads to limited effect of the filter on noise reduction; (c) result of applying a minimum order filter of size (5 × 5) on (a), which leads to good level of noise reduction; (d) result of applying a minimum order filter of size (7 × 7) on (a), which leads to good level of noise reduction; (e) result of thresholding (c) using Otsu method

Finally, the image is converted to binary form to produce the initial rib mask as shown in Figure 6.10(e).

Rib segmentation from the spine

The resulting image from the previous step is still too "noisy" in the sense that it contains too much non-rib objects. Therefore, these unwanted objects must be filtered out before any meaningful counting can be performed. Spine, as shown in Figure 6.10(e), is one of the biggest structures that complicates rib detection. It also has a direct impact on the reliability check of the rib area as we discussed in Section 6.3.2. We propose to use the binary spine mask which obtained in the previous chapter as a mask to clear the spine from the middle of the ROI. However, the spine mask needs some pre-processing before it is applied. Firstly, the spine mask should be cropped to fit the size of the rib area. Another important pr-eprocessing step is to shrink the upper quartat of the spine mask. This is required because by our observation, the

upper three ribs are usually small. This is making them prone to be masked out by the spine mask espicially if the upper area of spine mask is thick. The prepared spine mask and the rib area after removing the spine are shown in Figures 6.11(a) and (b) respectively.



(a) (b)

Figure 6.11 Exclusion of spine object from the rib area using spine mask.
(a) Prepared binary spine mask to be applied on rib area; (b) Rib area after masking the spine

Rib segmentation from the rib cage

After segmenting the ribs from one end by excluding the spine, it is important to segment them from the edge of the rib cage which is connected to the ends of the ribs as shown in Figure 6.11(b). The edges of the rib cage are the results of projecting 3D objects onto 2D X-ray image. To enable rib counting, we need to represent each rib as an independent object. To achieve this goal, it is needed to remove the edge of the rib cage to avoid any jointing among ribs from their end. To remove the edge of the ribcage, we proposed a linear function that initiates from the end of the top rib towards that of a lower one. The following procedure is applied on each rib side to trim the rib cage edge.

1- Top most point determination. Locate the topmost point as the middle topmost point of the spine mask previously prepared for rib area (Figure 6.11(a)).
2- Seed points determination. Locate two seed points for the right and left sides with same y-coordinate of the topmost point. In order to select the suitable value for the horizontal distance from the topmost point, we test three different distances (50, 60 and 70 pixels) as shown in Figures 6.12(a), (b) and (c) respectively. We found that the small value (i.e. 50 pixels) can satisfy removing the most object of the rib cage but with removing large part of the ribs (Figure 6.12(a)). On the other hand, the large value (i.e. 70 pixels) leads to remove minimum part of the rib cage (Figure 6.12(c)). To compromise between the two values, the value of the horizontal distance was set to 60 pixels such that satisfies removing the most object of the edge of the rib cage with a minimum part of ribs (Figure 6.12(b)).

3- Starting from the seed point and considering all the rows downward to the lower border of the image. For each row, all the pixels in the row toward the border of the image are set to zero. However, to follow the slope of the ribcage frame, the start point of the column in the current row is increased by 3 pixels each 10 rows. Red lines in Figure 6.12 show the slope of the trimming lines for both right and left ribs.

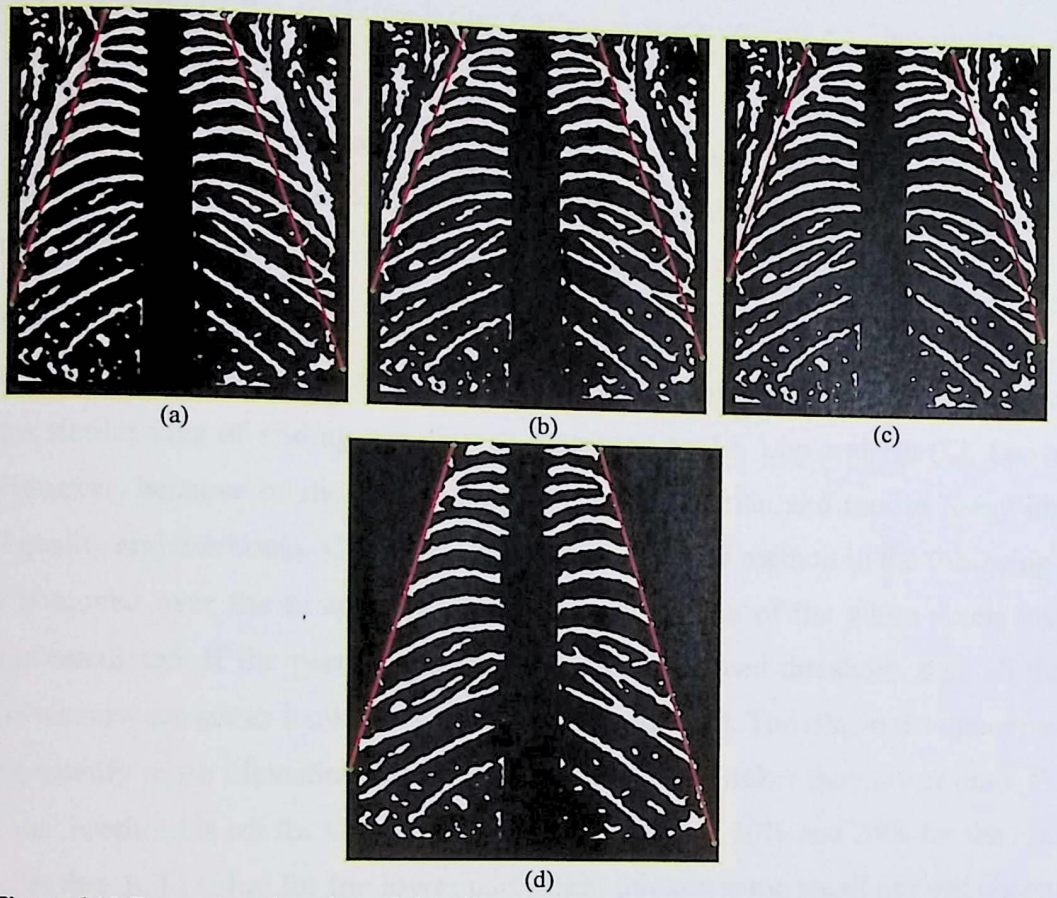


Figure 6.12 Removing the edge of the ribcage.
(a) Setting the horizontal distance for seed points (green) to 50 pixels from the topmost point of the spine. This leads to remove the edge of the ribcage but with large parts of the rib objects; (b) Setting the horizontal distance for seed points (green) 60 pixels from the topmost point of the spine. This leads to remove the edge of the ribcage with minimum parts of the rib objects; (c) Setting the horizontal distance for seed points (green) 70 pixels from the topmost point of the spine. This leads to remove minimum of the edge of the ribcage; (d) Removing the edge of the ribcage using the horizontal distance of 60 pixels

This function trims most of the noisy edges of the ribcage and separate the ribs as shown in Figure 6.12(d). It also removes all the non-rib objects to the right or left of the slope lines. Although crude in cutting off parts of ribs, as we are only concerned with rib count as opposed to rib length measurement, the function is simple and effective for separating the edge of the ribcage from disjointed ribs, and removing many non-rib objects in the border regions of the image. After removing most parts of the ribcage edge, there still exist non-rib objects and connected ribs which may further confuse rib count. Those problems will be considered, and solutions explained in the next section.

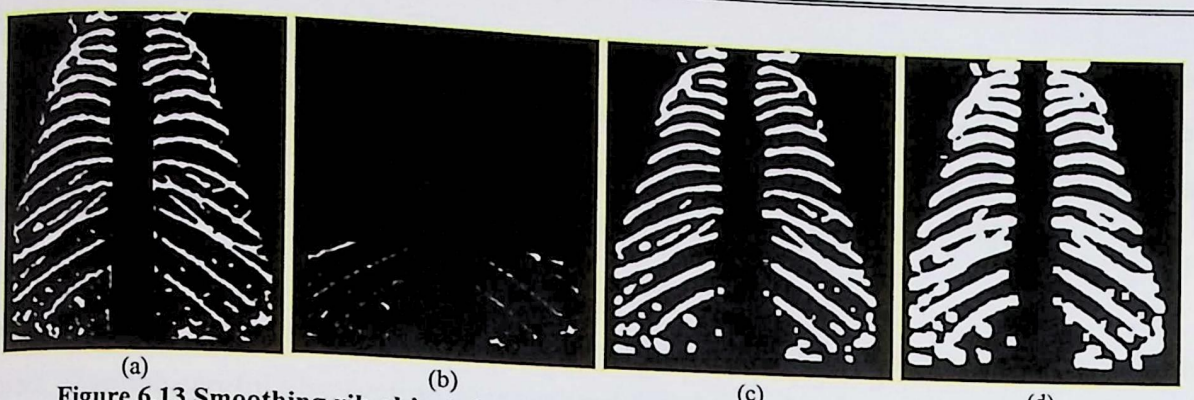
6.3.4 Enhancements and Smoothing of Rib Objects

At this stage, some further processing are needed to enhance the overall rib area and rib objects. The enhancements involve smoothing the mask of the rib objects, merging of rib mask, separating connected ribs and enhance the shape of the rib mask. The ultimate goal of this stage is to obtain smooth rib masks which cover of the actual rib boundaries in the input image. Correct identification of the rib boundaries further ensures correct rib classification and thus correct rib count at a later stage.

Because of the differences in characteristics between the upper and lower ribs, we consider dividing the image to upper and lower ribs from the middle point in the rib area. In the following we describe four proposed operations to enhance the rib objects.

Smoothing rib objects

In this stage, a procedure is used to smooth the shape of a rib object. To apply the smoothing, we used a similar idea of sliding window as presented by Li, Luo and Hu (Li, Luo and Hu, 2015). However, because of the difference between the human and mouse X-ray images in terms of quality and thickness of the rib objects, we tailor the method in the following way. A window is moved over the binary image and the percentage of the white pixels inside this window is calculated. If the percentage is above predetermined threshold, then all the pixels inside the window are set to 1 otherwise the pixels are set to 0. The ribs in the upper part of the image are usually more identifiable, easier to segment and thicker than lower ones. For those reasons, the threshold is set for the upper part of the image is 40% and 20% for the lower part. The smaller threshold value for the lower part might thicken some small non-rib objects, but it will help to connect broken parts of a very thin rib mask and so preserve the rib in the process of "unwanted objects filtering" in later next stages. To decide a proper window size for smoothing rib objects filter, we test three different sizes to enhance the rib area in Figure 6.13(a).



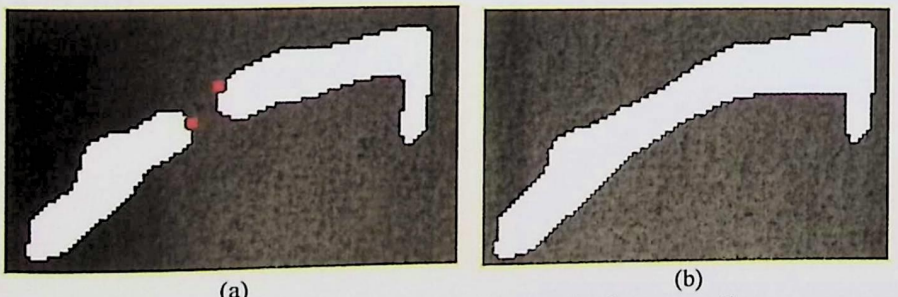
(a) (b) (c) (d)

Figure 6.13 Smoothing rib objects in mouse rib area.
(a) Rib objects before smoothing process; (b) Rib objects after smoothing the area with window of size (5×5 pixels); (c) Rib objects after smoothing the area with window of size (10×10 pixels); (d) Rib objects after smoothing the area with window of size (15×15 pixels) which leads to thickening the rib and non-rib objects.

It is found that using small window size (5×5 pixels) will lead to diminishing most of the rib objects in the upper and lower rib area (Figure 6.13(b)). On the other side, using larger window size (15×15 pixels) results in thickening the rib and non-rib objects considerably and thus merging ribs together (Figure 6.13(d). A window size between the two aforementioned sizes, (10×10 pixels), could give good smoothing to the rib objects without diminishing or thickening them as shown in Figure 6.13 (c).

Rib merging

In some cases, the rib object is divided into two or more pieces especially in the lower ribs. This case will lead to possibility of removing those parts as unwanted objects. The procedure of rib merging is based on the horizontal and vertical distances between the nearest points in two objects as shown by red points in Figure 6.14(a). By observation, we found that if both distances are small, those two objects are most likely belonging to the same rib. After obtaining the two objects that met above criteria, we apply two adaptive closing binary morphology operations with length equals to the distance between them and orientation equals to the orientations of each of them.



(a) (b)

Figure 6.14 Divided rib mask merging using adaptive closing operation.
(a) Divided rib mask into two parts with the close edge points of each part (red points); (b) Rib mask after merging its parts using binary closing operation with length equal to the distance between the two objects and orientation equal to the orientations of the two parts

To guarantee effective merging of the two objects, we set the orientation of each closing operation to the orientation of each of objects and the result is as shown in Figure 6.14(b). However, despite the effectiveness of this procedure in merging parts of a rib, there are unavoidable consequences of merging some unwanted objects to a bigger object.

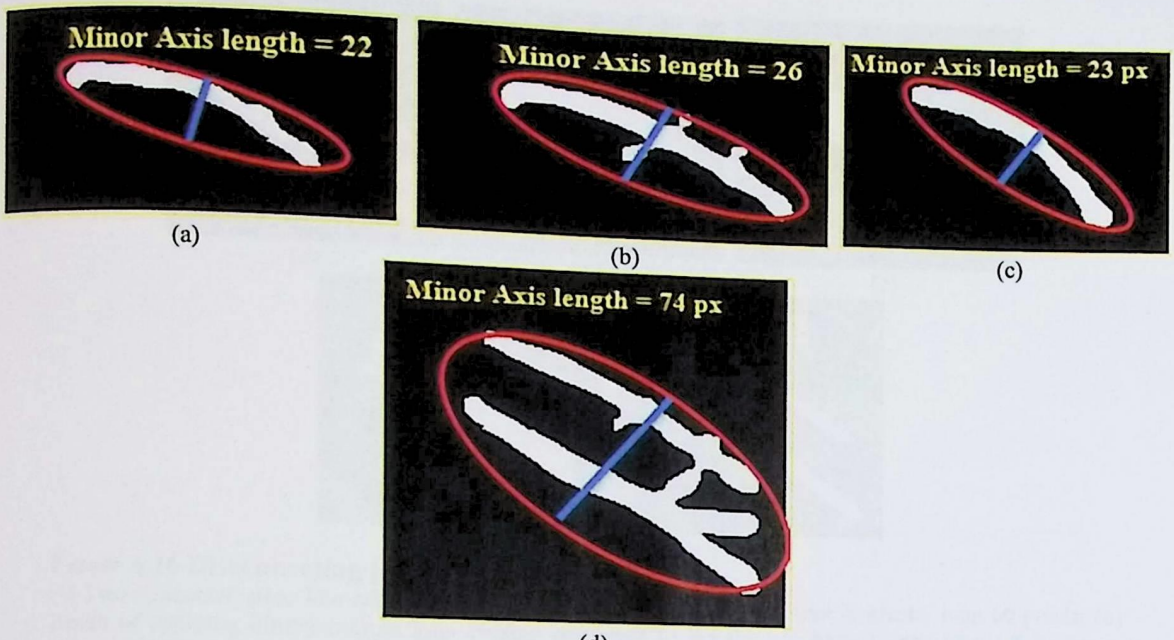
Separating joint ribs

One of the common problems in the rib area is the connected of two or more ribs as a one object. Through careful observation, it is realised that the connection exists either in the middle of the ribs because of the project of the of the rib curves from the ventral side or at the end because of unremoved segments of the ribcage. The first case usually happens to the lower ribs while the latter to the upper ribs. Both situations affect the accuracy of the rib detection and counting. Here we will describe two procedures, one dealing with the connections in the lower ribs and the other the connections in the upper ribs.

- Separating joint ribs in the lower rib area

  The main idea behind this procedure is to form a mask equivalent to the total gap between the two joint ribs. This gap mask is then used to mask out any unwanted objects existed between the two ribs. The details of this procedure is as follow.

  1. Examine the likelihood of connected "joint ribs". First, we find the smallest fitting ellipse shape of the whole object. We then use the minor diameter of the ellipse to verify if the object implies joint ribs. If the minor axis of the ellipse around the object is greater than 60 pixels, this object is likely to be joint ribs. Figure 6.15 illustrates several examples of single and connected rib objects with the ellipse around each one. The minor axis length of the ellipse around the single ribs (Figure 6.15 (a), (b) and (c)) is usually small and less than 60 pixels while it bigger (greater than 60 pixels) for the joint ribs (Figure 6.15 (d)).



(a) (b) (c) (d)

Figure 6.15 Several examples of minor axis length of ellipse around single and connected rib masks.
(a) Single rib mask (minor axis length is 22 pixels); (b) Single rib mask (minor axis length is 26 pixels); (c) Single rib mask (minor axis length is 23 pixels); (d) connected ribs mask (minor axis length is 74 pixels)

2. Separate "joint ribs". First, a binary vertical closing morphology operation (Section 2.2.1) is performed to fill the gaps between the joint ribs to form a solid background as shown in Figure 6.16(b). The bitwise XOR operation between the original "joint ribs" object (Figure 6.16(a)) and the solid background (Figure 6.16(b)) is applied and the result as shown in Figure 6.16(c). This is followed by a binary closing morphology operation to find the gap mask between the two actual ribs as shown in Figure 6.16(d). We then apply the bitwise AND operation between the original "joint ribs" object (Figure 6.16(a)) and the complement of the gap mask (Figure 6.16(d)) and the result is as shown in Figure 6.16(e).

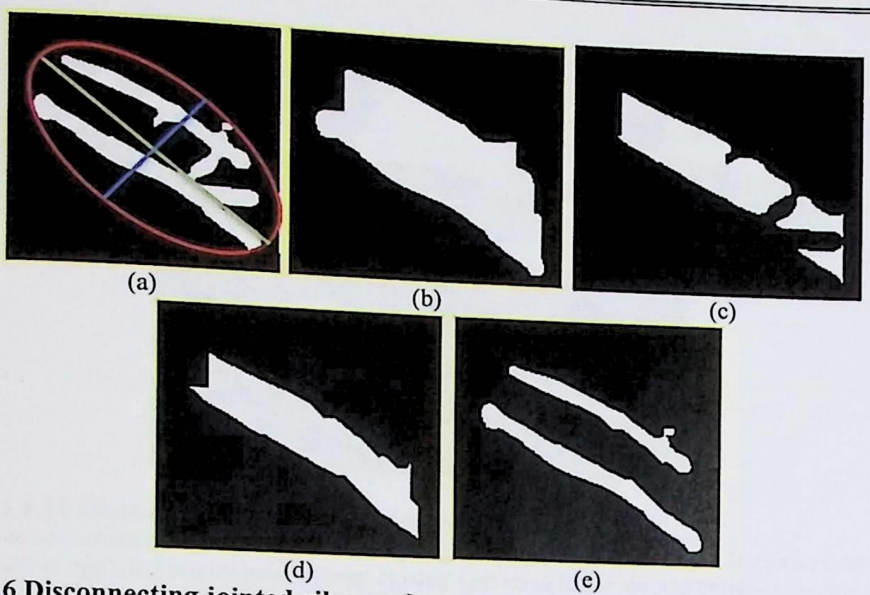


Figure 6.16 Disconnecting jointed ribs mask.
(a) Two connected ribs. The minor axis length (blue line) of the red ellipse is greater than 60 pixels; (b) Result of applying binary morphology closing operation to fill the gap between the ribs; (c) Result of applying XOR operation between (a) and (b); (d) Result of applying binary morphology closing operation on (c); (e) Result of applying AND operation between (a) and complement of (d) to disjoint the ribs

This proposed procedure helps in disjointing some of joint rib cases to enhance the final shape of the rib masks. However, this procedure is designed with thresholds that best fit the most cases of the two joint ribs in the lower rib area. So, the procedure may not work if the joint ribs involve more than two ribs or if the two joint ribs size and minor diameter length do not fit the set thresholds.

- Separating joint ribs in the upper rib area

  Similar to the upper procedure, this one is also aim to extract and exploit the gap mask to disjoint the ribs. Because this procedure is specific to the joint ribs from the end and not from the middle, some modifications are done from the above procedure as follows:

  1. Repeat the steps for the first procedure but with following changes:
     - The threshold to detect the joint ribs (Figure 6.17(a)) in Step 1 is set to 30 pixels.
     - Skip the binary closing operation after XOR operation in Step 2. Figures 6.17(b) and (c) show the resulting solid background and the gap masks.
  2. Determine the cut point as the start point of the distant object as shown by red arrow in Figure 6.17(c).
  3. Cut the connected ribs from the determined cut point (the red line in Figure 6.17(d)) and the result as shown in Figure 6.17(e).

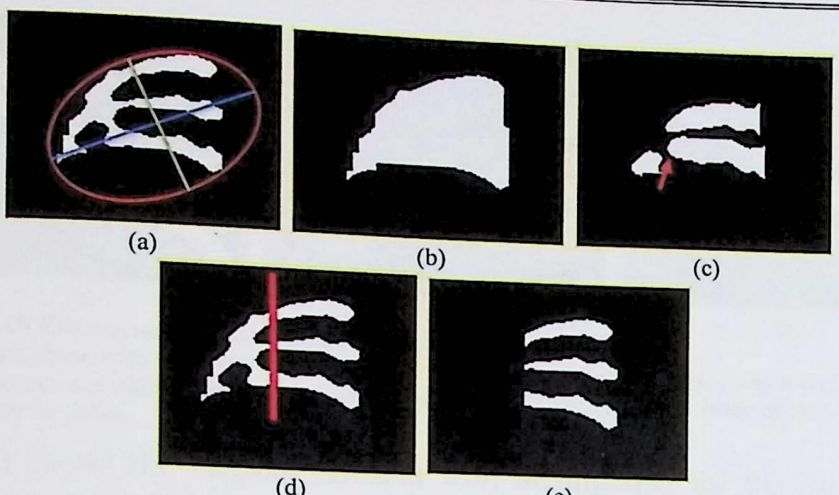
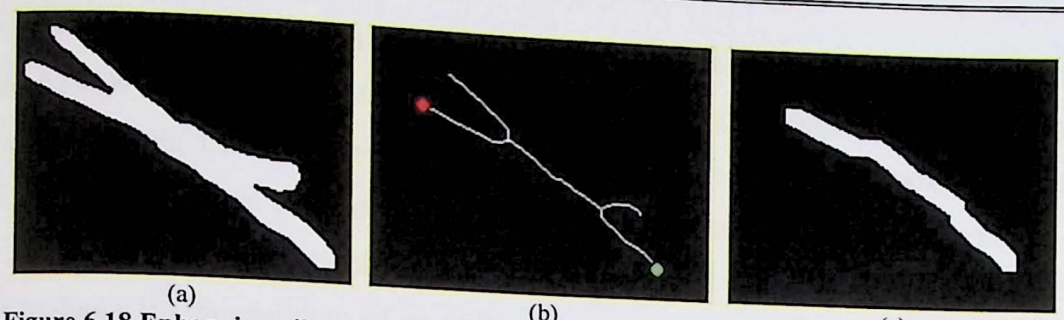


Figure 6.17 Disconnecting upper jointed ribs mask.
(a) Three connected ribs. The minor axis length (green line) of the red ellipse is greater than 30 pixels; (b) Result of applying binary morphology closing operation to fill the gap between the ribs; (c) Result of applying XOR operation between (a) and (b); (d) Connected ribs with the cut line (red line) to disconnect them; (e) Result of cut the connected ribs to separate them

Enhancing the rib segmentation

A common problem especially in the lower ribs is the distorted shape of the ribs. This happens when some unwanted objects or part of the rib reflection fuse with the rib as shown in Figure 6.18(a). In addition to giving inaccurate information about the location and size of the rib, it may also affect the final counting when the rib appears bigger than the expected size for a single rib.

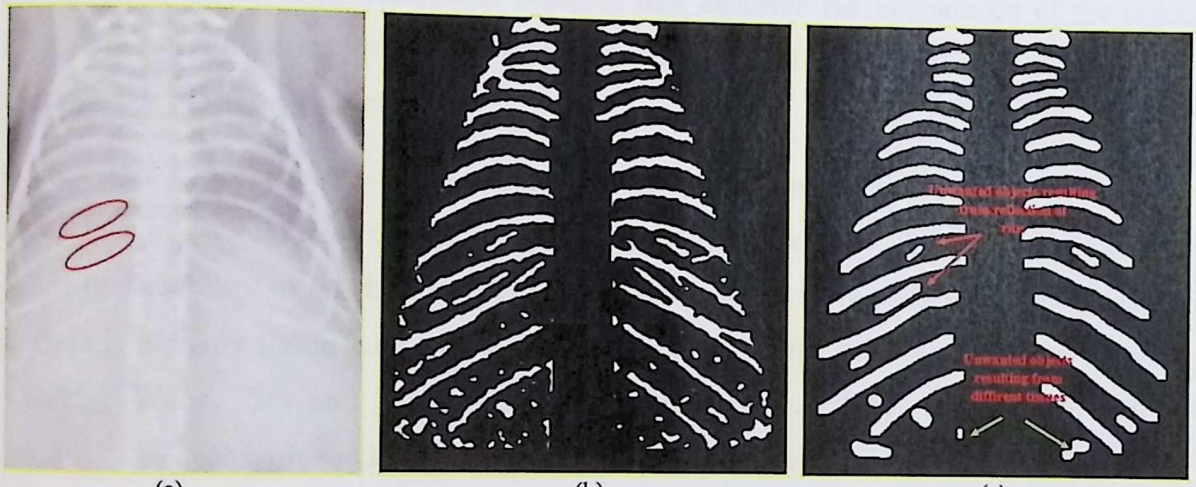
Trying to solve such cases by using traditional binary morphological operations such as opening and closing usually gives undesired results and leads to lose parts of the rib itself. We propose a solution that can approximate the real rib structure and remove most of non-rib objects fused to the rib based on shortest distance between the two ends of the rib. The advantage of using graph based to extract the real path of the rib was presented in the literature by Li et al. (Li et al., 2015). In the first step of the solution, the rib is thinned to remove all pixels but keep one pixel width object as shown in Figure 6.18(b). Following the thinning, the two far right and left points of the thinning object are determined as shown by red and green points in Figure 6.18(b). To get rid of any branches, we found the shortest path between the red and green points by calculating the geodesic distance between the two extreme points. The geodesic distance between two pixels in a binary object is the minimum length of the path joining those pixels and included in the binary object (Soille, 2004). Finally, the found path is taken as approximation for the shape of the rib after thickening it as shown in Figure 6.18(c).



(a) (b) (c)

Figure 6.18 Enhancing rib mask segmentation.
(a) Rib mask shape with connected unwanted objects; (b) Thinning rib mask to a single pixel width mask and the two extreme right (green) and left (red) points; (c) The final rib mask as the thicken path between the two extreme points

Figures 6.19(a) shows the original rib area with highlighting for the ribs reflection (red ellipses). Figures 6.19(b) and (c) show the rib area before and after applying the rib shape smoothing and enhancements respectively.



(a) (b) (c)

Figure 6.19 Enhancements and smoothing of rib area.
(a) Mouse rib area which shows some reflection of ribs (red ellipses); (b) Isolated rib masks and non-rib objects in the rib area before applying enhancements and smoothing; (c) Rib area after applying enhancements and smoothing operations. It shows existence of many non-rib objects which result from reflection of ribs and different tissues in the rib area

6.3.5 Two-level Classification Scheme to Remove non-Rib Objects

As we can see from Figure 6.14B, the rib area still has several non-rib binary objects. There are two sources of the non-binary objects at this stage. The first source are mainly coming from the reflections of the ribs on the other side as shown in the red arrow in the Figure 6.19(c). Such objects are usually have the same texture features as the real ribs. The other non-rib objects include the background tissues of the lungs and vessels in the homogenous area between intermediate ribs and below the final ribs. Optimally, those types of non-rib objects have a different texture features from the real ribs.

To deal with this non-rib objects removal problem, one of thoughts is to exploit a well-known properties of the ribs (i.e. size and orientation) to define some rules to filter out any non-rib

objects. However, such approach may faces several challenges. One of the challenges is that there can be many different rules to cover the diversity of non-rib objects shapes and properties from various images. Besides, it would not be enough to use only the known properties from a limited number of images that describe the shape of objects to eliminate all possible non-rib objects. Therefore, it can be concluded that using such kind of conventional approaches is not feasible to deal with the problem of non-rib objects removal.

The other approach to tackle this problem is to use a machine learning approach. Such approach can offer several advantages. Once we train a model with features that extracted from a sufficient number of samples, the created model can be general to cope with different types of non-rib objects. In addition, on the contrast to conventional approach, the trainable approach could be used with a high dimension features that describe the appearance of the objects. Finally, we can train different models with different types of features to deal effectively with different types of non-rib objects.

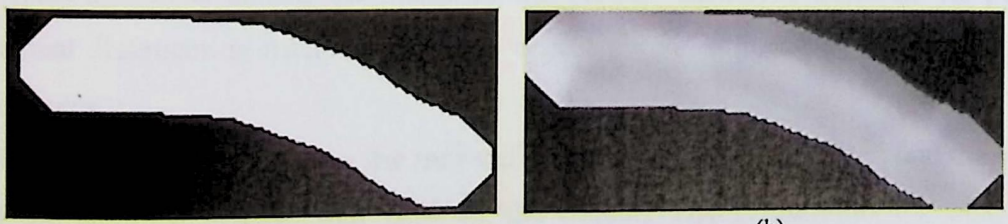
For our method, we propose to use two-level SVM classifier (Section 2.2.3) with a different features type in each level to classify the objects to rib and non-rib. We select SVM classifier because its suitability to work with high dimension feature space. It has also shown to work successfully in many fields such as image classification and medical image recognition ((Li et al., 2015), (Wang, Zhu and Liang, 2001) ,and (Hoi et al., 2006)). The idea of two-level classifiers is to deal with earlier mentioned two types of non-rib objects effectively. The first level should deal with any non-rib objects exploiting their different texture features from the ribs. However, although this level should exclude many non-rib objects, it usually cannot distinguish the non-rib objects that similar to the rib in terms of texture features (bony objects). This justifies the need for the second level that exploit the geometrical features of the objects to remove remaining non-rib objects. By observing many images with different degree of difficulty, we can conclude that the majority of challenging non-rib objects are existing in the lower part of rib area. This is because the possibility of non-rib objects formation in the lower rib area is much higher than the other parts. So, for effectivity and efficiency purposes, we build the classifiers for lower rib area only. As we mentioned before, the middle point of the rib area was considered to separate it to lower and upper rib area. We hereby argue that this cut is not very sensitive and should not affect the accuracy of the classifiers based on our observation and understanding of the images. This is because the changing in the rib features are smooth and the lower ribs of upper area and upper ribs of lower area share same common features.

In the following subsections, we will describe each level of the classifier scheme along with selected features and how to obtain those features.

Level 1- SVM classifier

In this level, we select to train the SVM classifier using the LBP features (Section 2.2.3) which extracted from the texture of the object and the small region around it. The LBP is selected due to its ability to distinguish the change in contrast between the rib and its background. The procedure of LBP feature extraction for each object in the image is as follows:

1. To include the texture of the rib and its surrounding area, we apply binary morphological operation to thicken the rib mask as shown in Figure 6.20(a). The mask is then superimposed on the original image to obtain the texture of the rib object. We then take the smallest bounding rectangle shape around the masked area, and then maintain the intensity values of the pixels inside the mask, but set the values of the pixels outside the mask area to zero. Figure 6.20(b) illustrates the rectangle area after this step.
2. We apply the LBP transformation of the image in the bounding rectangle into an LBP image and then extracted the normalized 59–bin histogram of uniform LBP from the LBP image of the rectangle. The normalized histogram is important for making this texture feature invariant of the sizes of the rectangles.
3. Since the dark area, i.e. zero intensity is imposed, we have decided to remove the bin that represents the feature generated from the dark area from the LBP histogram.



(a) (b)

Figure 6.20 Uniform LBP features extraction from rib object.
(a) The binary mask of rib object; (b) Texture of rib object obtained from mouse rib area

Normally, within a rib, there should be some gradual changes in intensity values of pixels as we can see from Figures 6.20(b). Such changes in intensity would either not occur or occur differently in the non-rib objects (see Section 2.2.4 for more details). Such intensity changes could be exploited to train a classifier using LBP feature. We propose to train one SVM-LBP classifier for both right and left sides using the uniform LBP features extracted from 700 ribs and 700 non-rib objects were collected from the lower right and lower left rib area. Figure 6.21(b) shows the resulting image of one example after removing non-rib objects from the rib

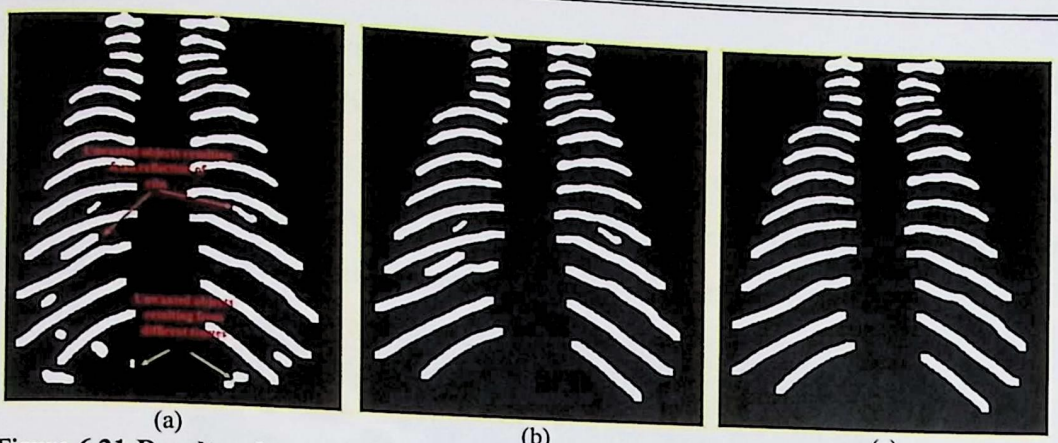
area of Figure 6.21(a) using the SVM-LBP classifier. As we can see, many of "non-bone" objects were removed at this stage.

Level 2 - SVM classifier by using geometrical-based features

As we mentioned in Section 6.5.3, most of non-bone objects could be removed by applying the SVM-LBP classifier based on their texture features. However, some other "bony" objects are usually not removed because they have similar texture features to the ribs. Those objects represent part of the ribs but for the ventral side and usually appear between the ribs as a consequence of 2D X-ray images. As we are not interested in those objects, we will refer them as bony non-rib objects. To remove the bony non-rib objects, we decide to use some intrinsic rib geometrical features to train another SVM classifier. The effectiveness of geometrical feature to filter non-rib objects was shown in the literature ((Li et al., 2015) and (Al Okashi et al., 2014)). We propose to use the following geometrical features:

1. Horizontal distance from spine. In normal cases, all the ribs should be jointed to the spine or very close to it, so that the horizontal distance between the spine boundaries and the end points of the ribs close to the border should be very narrow. We calculate an actual horizontal distance in terms of pixels between every object in the image and the relevant boundaries of the spine.
2. Vertical distance between ribs. The vertical gap between any two consecutive ribs should not be very small or very big. To approximate the vertical distance between two objects in the image, we calculate the distance between every opposite points in both objects and then taking the mean of all vertical distances between the points. The vertical distances is measured for the object above, object below and then sum both distances.
3. Shape parameters including the major diameter, the minor diameter, and the orientation of the smallest ellipse that best fit the object. The orientation for each object is calculated as the angle between the x-axis and the axis of the major diameter of the best fit ellipse.
4. Area. Calculated as the number of pixels within the boundary of the object.

One of the important features that contributes in rib recognition is the orientation. Because this feature value is dependent on the side of the rib, we decide to train two classifier, one for each side. Each SVM-Geometrical classifier was trained with feature extracted from 400 ribs and 400 non-rib objects. The final result of applying SVM-Geometrical classifier on a rib area of Figure 6.21(b) is as shown in Figure 6.21(c).



(a) (b) (c)

Figure 6.21 Results of applying the two-level classifier scheme to remove non-rib objects from rib area.

(a) Rib area before applying the two-level classification scheme. It shows existence of many non-rib objects which result from reflection of ribs and different tissues in the rib area; (b) Resulting rib area after applying SVM-LBP classifier. Most of non-bony objects were removed by exploiting their different texture features from the rib; (c) Resulting rib area after applying SVM-Geometrical classifier on (b). All the remaining non-rib objects were removed by exploiting their different geometrical features from the rib

6.3.6 Reliable Rib Count

In normal cases when there are no joint ribs, we can consider number of ribs is equal to the number of binary objects in the image. We consider two rib counts: one for the ribs on the left and the other for the ribs on the right.

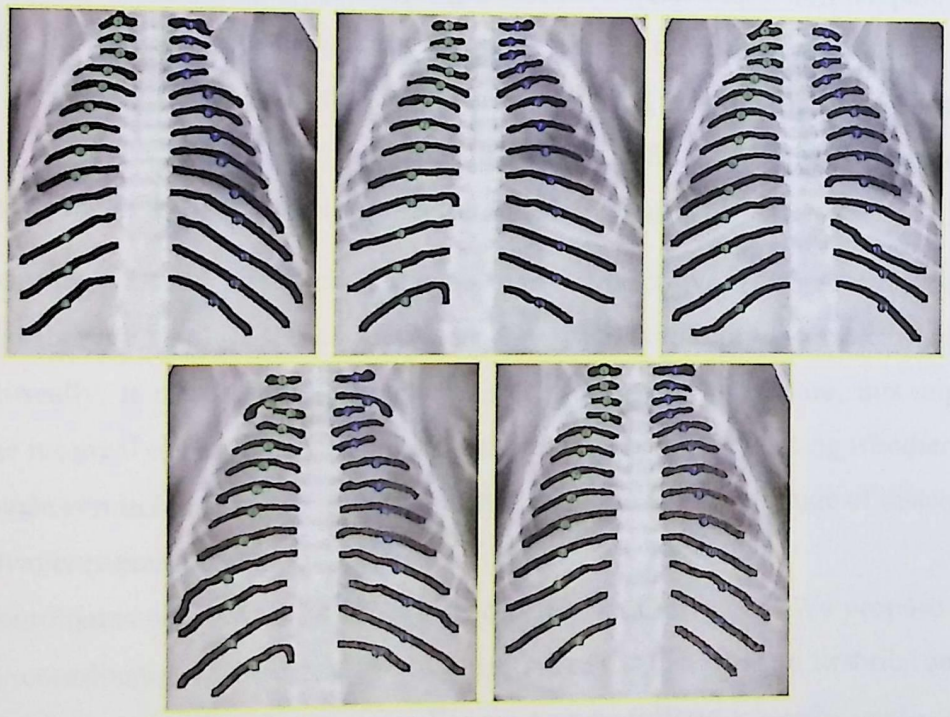
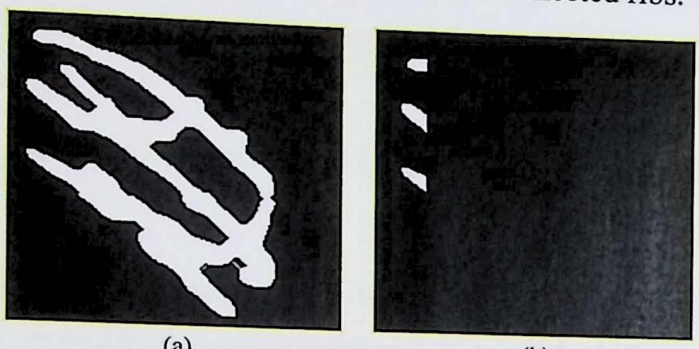


Figure 6.22 Several examples of detected and labelled rib mask superimposed on original rib area images.

To illustrate the result of rib detection and counting, we superimposed the detected ribs and the counting indices for the ribs against the original segmented chest area for the five rib area images as shown in Figure 6.22.

Sometimes, two or three ribs are interconnected to form a big binary object as shown in Figure 6.23(a).We observe that this problem does not usually affect the inner part of the ribs close to the spine. Therefore, to deal with such cases, we propose to cut a small region (e.g. 10 pixels wide) closer to the spine as shown in Figure 6.23(b). If there are jointed ribs, the number of generated objects should represent the actual number of connected ribs.



(a) (b)

Figure 6.23 Rib count from connected ribs mask.
(a) Three connected ribs mask; (b) Result of cutting the first 10 pixels closer to spine to enable counting of the ribs in (a)

Although the classifiers explained in Section 6.3.5 are designed to produce accurate ribs, some errors such as removing a real rib or keeping a non-rib objects might still happen. To increase the reliability of the final counting, some tests are needed after final detection of all the ribs. Those tests are built upon the fact that any extra or missing ribs to either side should be either from upper or lower rib area (i.e. not from the middle). Some location, distance and count information are then exploited to compose the following tests:

1. Maximum vertical distance between two consecutive ribs. A suitable range for acceptable vertical distance between any two consecutive ribs can be estimated empirically. If any unusual distance is detected by the procedure, this might indicate either removal of real ribs or existence of non-rib objects depending whether the vertical distance is much too big or much too small than the expected range of distance between the two consecutive ribs.
2. Y-coordinates of the last and first ribs on right and left sides. We propose to compare the y-coordinates of the close end to the spine of the last ribs and first ribs on both sides. This test could lead to the two unreliable cases as follow:
    - If the y-coordinates of both upper and lower pairs of ribs are equal or very close yet the count of ribs is different for each side, then this may indicate a possibility of unremoved non-rib object between two ribs.

- If the y-coordinates of either the upper pair or lower pair of ribs are highly different yet the count of ribs is equal for each side, then this may indicates a possibility of unremoved non-rib object and removed rib on one side.

3. Minimum and maximum bounds of the number of ribs. From the domain knowledge, it is known that the number of ribs for each side should be in a range between 12 and 14. The number of ribs beyond this range is biologically impossible and hence must be due to miscount.

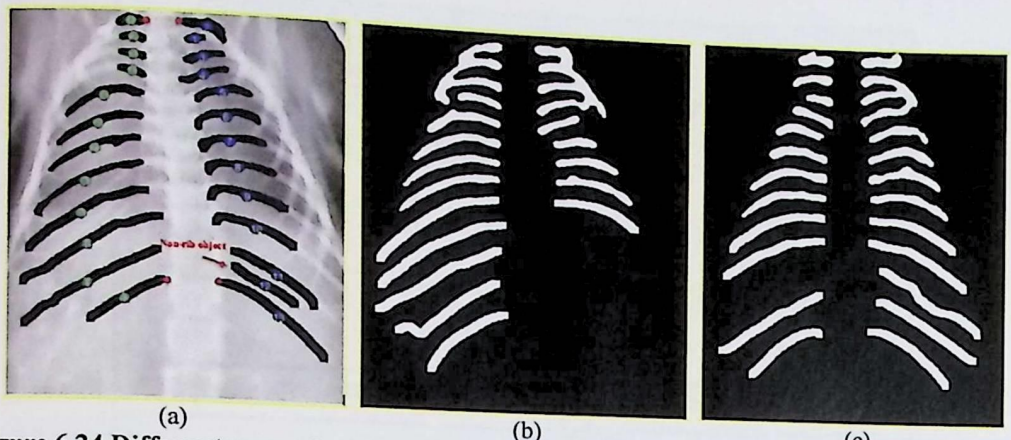
Based on those tests, and the early ROI reliability check (Section 6.3.2), there are three possible categories.

- The image passes the test and so labelled as reliable counting.
- There is a degree of doubt of possible existence of non-rib objects or possible removal of real ribs and so the image is labelled as unsure which needs a further manual examination.
- There is an obvious evidence that the image could not be segmented reliably (insufficient quality) and hence it is labelled as rejected (Section 6.3.2).

It is important to highlight that in the second category the ribs are segmented and counted (but it needs to be manually verified) whereas no output is produced from images in the third category.

An examples of some unreliable rib area cases are given in Figure 6.24. Figure 6.24(a) shows an example where the number of ribs in this image should be 12 ribs on each side. As we can see, our proposed solution succeeded to define all the ribs in the right and left sides. However, there still one non-rib object on the right side as indicated by red arrow. By applying the proposed checks, we found that the y-coordinates of the consecutive ribs are very close. Also we found that one of the object has small upper and lower vertical distances. This give an indication of the existence of a non-rib object and so we label this image as unsure.

In some cases, several ribs are connected in a way that makes them unrecognizable as a rib by the classifier, and so they are removed as a non-rib object. One example of such cases is shown in the Figure 6.24(b) where the left rib count is correctly done as 13 ribs while the right side is wrongly counted as 9 ribs. Such case is automatically labelled as unsure. The final example given in Figure 6.24(c) shows an equal or very close y-coordinates of all most upper and lower ribs while the count of both side is different (i.e. missing rib).



(a) (b) (c)

Figure 6.24 Different examples of unreliable rib count.
(a) Unremoved non-rib object as indicated by the red arrow. This object results from the reflection of the rib and its features identical to the rib feature and so it is not removed as a non-rib object; (b) Right rib count much fewer than the expected range; (c) Removed rib which leads to high vertical distance between the two ribs above and below it and so the image is labelled as unreliable rib count

6.4 Experiment Results and Analysis

We have tested our method on a dataset of 6023 full body dorsoventral view X-ray images of mutant mice that were previously manually annotated by a domain expert. Those 6023 images include 5948 images with normal counts of ribs (i.e. 13 ribs each side). The other 75 images involve missing or extra ribs either on one or both sides. After applying the reliability measures (Section 6.3.2) and reliable counting (Section 6.3.6), we categorise the resulting images for both normal and abnormal counts into three categories as shown in Table 6.1.

Automatic Metric Used	Manual annotator	Labelled as "reliable count"	Labelled as "unsure count"	Rejected images (can't be reliably segmented)
Normal Images (Ground truth)	5948	4906	1003	39
Abnormal Images (Ground truth)	75	57	16	2
Total	6023	4963	1019	41

Table 6.1 Different categories of rib count (reliable count, unsure count and rejected images).

The first category should contain any image was not excluded based on reliability measures and also passed the reliable counting successfully. The images that were counted with low reliability put in the second category. The final category composed of the images that failed reliability check and thus not counted. The rationale behind this division is to minimize the images that need to be manually annotated to only group 2 (verifying the count) and group 3 (counting from scratch). However, not all the images in the first category are correctly counted. Figure 6.25 presents the confusion matrix which shows sensitivity and specificity results for the reliable counting category when we take the given manual counting results as the ground truth.

	Actual	
	Abnormal rib count	Normal rib count
Prediction	True Positive = 41	False Positive = 1172
	False Negative = 16	True Negative = 3734
	Sensitivity = TP/(TP + FN) = 71.92%	Specificity = TN/(FP + TN) = 76.11%
	Type II error = 28.08%	Type I error = 23.89%
Accuracy = $\frac{TP+TN}{Total\ population}$ = 76.06%		

Figure 6.25 Confusion matrix for the reliable rib count category.

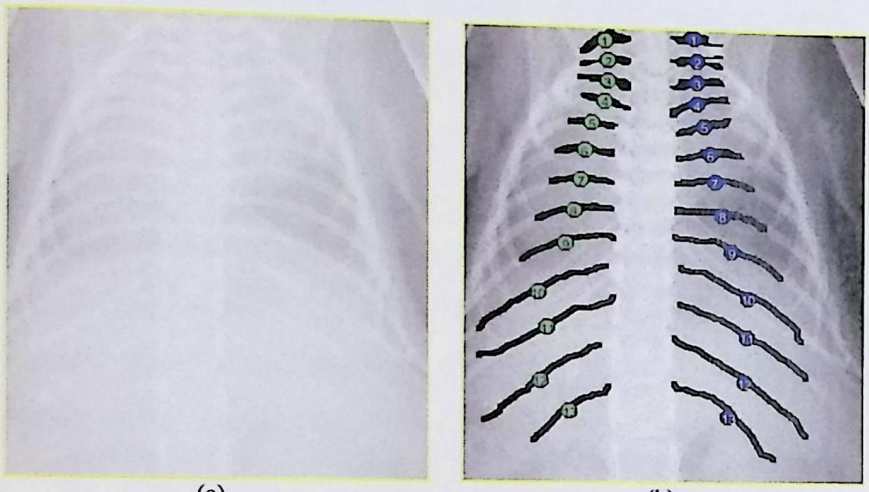
As we can see from the above figure, both sensitivity and specificity are above 70%. The specificity is higher than the sensitivity. The proposed method works therefore better for normal mice than mutant mice. Although the accuracy in recognizing the positive and negative cases for those images categorised as a reliable count is reasonable, it still not as good as we like it to be. To further understand the obtained results, we need to address three issues. Firstly, the ultimate effectiveness of rib count should be judged in terms of its ability to classify the mutant line (7 mice) as normal or abnormal. As we explained in Section 3.4.1, the mutant line is said to be "hit" if at least 5 out of 7 (60%) rib count are abnormal. To investigate along this direction, we examined 8 different mutant lines that were previously annotated by an expert. Table 6.2 shows how many images in each line are counted as abnormal (positive) or normal (negative) by the manual annotator and the proposed automatic solution.

Gene	Gender	Positive cases (Manual)	Negative cases (Manual)	Hit/ Normal	Positive cases (Automatic)	Negative cases (Automatic)	Hit/ Normal
Oas1g/Oas1g	Female	3	4	Normal	4	3	Normal
Oas1g/Oas1g	Male	3	4	Normal	4	3	Normal
Kdm7a/Kdm7a	Female	6	1	Hit	5	2	Hit
Kdm7a/Kdm7a	Male	1	6	Normal	3	4	Normal
Setd5/+	Female	3	4	Normal	5	2	Hit
Setd5/+	Male	1	6	Normal	3	4	Normal
Orc3/+	Female	0	7	Normal	3	4	Normal
Orc3/+	Male	0	7	Normal	3	4	Normal

Table 6.2 Manual and automatic phenotyping of 8 mutant lines for mouse rib count.

As we can see from the above table, even the automatic solution didn't not recognize all the cases in the line, but eventually it can define one hit line out of one (Kdm7a/Kdm7a- Female) and 6 normal lines out of 7, indicating good performance of the proposed method in predicting normal lines and hit lines.

The second issue to debate about, is the inaccuracy involved in the ground truth for the rib count. In other words, the ground truth is not as solid as expected. Figure 6.26 shows one of examples which labelled as abnormal counting in the ground truth while it looks perfectly normal and our system defined it as normal counting. Following several meetings and discussion sessions with the expert annotator, we can conclude that there is a level of uncertainty in the counting process especially regarding upper and lower ribs (see Section 6.5.1).



(a) (b)

Figure 6.26: Example of uncertainty in the ground truth.
(a) Rib area of mouse X-ray image which shows normal rib count (13,13) but labelled as abnormal in the GT; (b) Result of normal rib count by the proposed solution which shows normal rib count

The other example of problem with the ground truth can be seen from Figure 6.8(c). Even the mention image was rejected, such cases of corrupted ribs may also affect the ground truth due to biological reasons or inaccurate mouse position when x-rayed. If such cases are not rejected, it will be infeasible to obtain any accurate counting from them.

The final issue to be addressed is whether or not the levels of accuracy for the reliable category images are higher than the levels of accuracy for the unsure category of images. If not, the so called reliable category would make no sense at all, indicating that the idea of excluding unreliable images is fundamentally flawed. Therefore, we conducted a test on unsure category of images and calculated the sensitivity, specificity and accuracy for the unsure count group as shown by confusion matrix in Figure 6.27. As we can see from Figure 6.27, the accuracy of rib count from the unsure count category is significantly lower on both sensitivity and specificity than those for the reliable count category as stated in Figure 6.25. This confirms the effectivity of our proposed tests to classify the results to reliable and unsure count.

		Actual	
		Abnormal rib count	Normal rib count
Prediction		True Positive = 9	False Positive = 666
		False Negative = 7	True Negative = 337
		Sensitivity = TP/(TP + FN) = 56.2%	Specificity = TN/(FP + TN) = 33.6%
		Type II error = 43.8%	Type I error = 66.4%
Accuracy = $\frac{TP+TN}{Total\ population}$ = 33.9%			

Figure 6.27: Confusion matrix for the unsure rib count category.

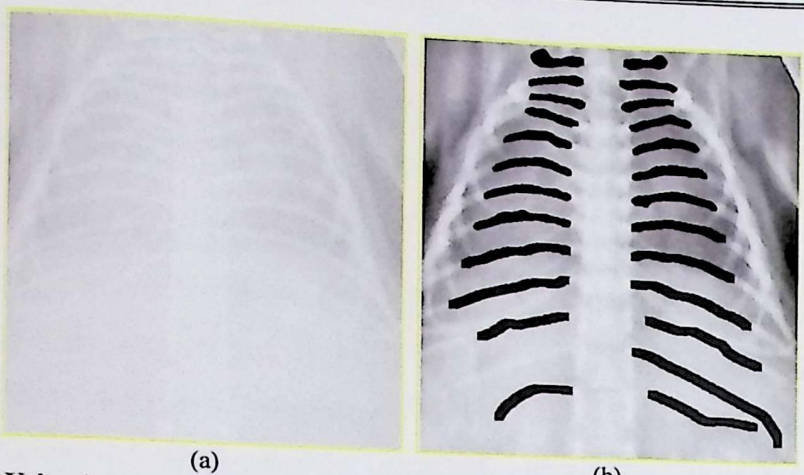
To form a more solid picture about the ability of proposed method for rib detection, we collected a random sample of 100 images which contains 2598 ribs to evaluate them in terms of number of correctly detected ribs and missing ribs as shown in Table 6.3. These 100 images include different degrees of difficulties in the rib area.

	Correctly Detected ribs	Missing ribs
Proposed method	99.2%	0.8%

Table 6.3 Accuracy of rib detection. The proposed method succeeded to detect most of the ribs in the tested images.

The results above indicated that our proposed method has an ability to detect the most of the ribs in the tested images. However, there is still small percentages of missing ribs in the segmented images.

It can be argued that even for the images automatically labelled as unsure (i.e. to be manually checked), the fact that the proposed solution could detect most of the ribs in the images could make the manual verification easier. For those images, the annotator does not need to look at the original images (Figure 6.28(a)) but he can use the segmented images by the system (Figure 6.28(b)). This should make the manual check much easier without compromising the accuracy due to the high accuracy of rib detection.



(a) (b)

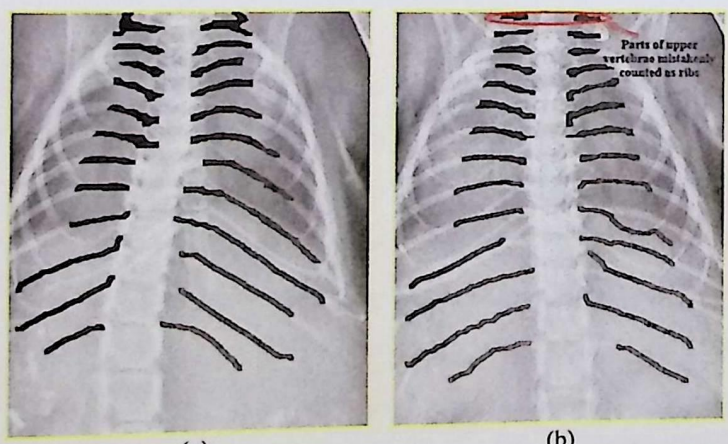
Figure 6.28 Using images from unsure rib count in manual rib count.
(a) Rib area of mouse X-ray image before rib detection process which shows very low contrast and so difficult to perform rib count manually; (b) Result of rib detection of (a) which categorised as unsure rib count because of missing rib which leads to increasing the vertical distance between the two ribs. Even the image is classified as unsure rib count, the manual annotator could use it as a guide for manual count which makes the rib count easier for him

6.5 Discussions

This section presents a thorough discussion on three different but related issues: 1) the main reasons of inaccuracies, 2) parameter sensitivity of two important parameters that used in the solution, and 3) the possibility of using different view for rib count.

6.5.1 Main Reasons of Inaccuracies

In this section, we discuss the main reasons that might lead to inaccurate rib count for our method. The first reason is related to the ROI segmentation. The ROI segmentation affects all the later stages, so it is considered as an essential stage in our proposed method. This segmentation is said to be good enough if the segmented ROI area includes all the ribs.



(a) (b)

Figure 6.29 Examples of inaccurate rib count because of inaccurate upper cut.
(a) Undetected upper ribs because of inaccurate start of the rib area; (b) False upper rib resulting from part of cervical vertebrae which mistakenly considered as ribs

However, from a discussion with an expert, it is established that the upper start of the ROI is very difficult to determine even for a trained annotator. This is because the small sizes of the

upper ribs, which may look very similar to the processes of the cervical vertebrae. Inaccurate start of the rib area (Section 6.3.1) could either lead to undetected ribs (Figure 6.29(a)) or labelling processes of cervical vertebrae as ribs (Figure 6.29(b)).

The other reason of non-rib object resulting in the final rib count is the rib shape criteria i.e. the project of a rib on the other side of rib cage looks like a rib. The ribs are curved bones (C shape) that extend from the vertebrae at dorso side of the body to the sternum bone at ventral side. Because of this shape, the 2D dorsoventral X-ray images show the dorso side as well as the reflection of the ventral side as shown in Figure 6.24(a). Even our method was designed to remove such cases as a non-rib objects, the geometrical features of some of them are very similar to the real rib.

Another problem that may lead to missing some ribs is the very low contrast of the lower ribs. This fact makes the final mask of the lower ribs in some images is relatively small and may not cover their texture in the input image entirely. Such situation might increase the probability that the SVM classifiers to consider them as a non-rib objects. In most of these cases, the final mask of the rib was automatically corrected. We can argue that our solution was succeeding to deal with many difficult cases but few other cases were highly challenging.

6.5.2 Parameters Sensitivity

Some of steps in our proposed solution are dependent on some fixed parameters or thresholds. Even we try to minimise the number of parameters, in some cases it is still unavoidable. Here we discuss the parameter sensitivity for some values that may have a big impact on the final result as follow.

The cut off frequency for the BHPF

The value of cut off frequency for BHPF (see Section 2.2.1) controls the attenuation of the frequencies in the frequency domain. Setting this parameter to a relatively big value means more distant frequencies from the centre of the frequency domain will be attenuated and the vice versa. For the rib area, decreasing this parameter to a small value will produce a strong and complete objects of the ribs (sharp edges objects) but at the price of producing many non-rib objects which complicate rib masking at later stages. Increasing the cut off frequency to a relatively high value will diminishes some non-rib objects but at the same time makes the ribs fade out and so prone to be removed in the later stages. To explain our selection for this parameter, we firstly make a comparison among the outputs of the BHPF of the same image but with different values for the cut off frequency as shown in Figure 6.30. To balance between

strong rib edges and many non-rib objects associated with low threshold values (Figures 6.30(a) and (b)) and fewer non-rib objects and the less contrast on rib objects associated with high threshold values (Figures 6.30(d), (e) and (f)), we found that the possible values for the threshold should be around 50 (Figure 6.30(c)).

To make more sensible decision about the threshold selection, we test different values of cut off frequency around (50) on a sample of 100 images while fixing other parameter values and then find the accuracy of rib count. The chart in Figure 6.31 below shows the accuracy of rib count for 7 different values (25, 35, 40, 45, 50, 55, and 60). The accuracy measured confirmed that the value 50 gave the best results in terms of counting which is up to 70%.

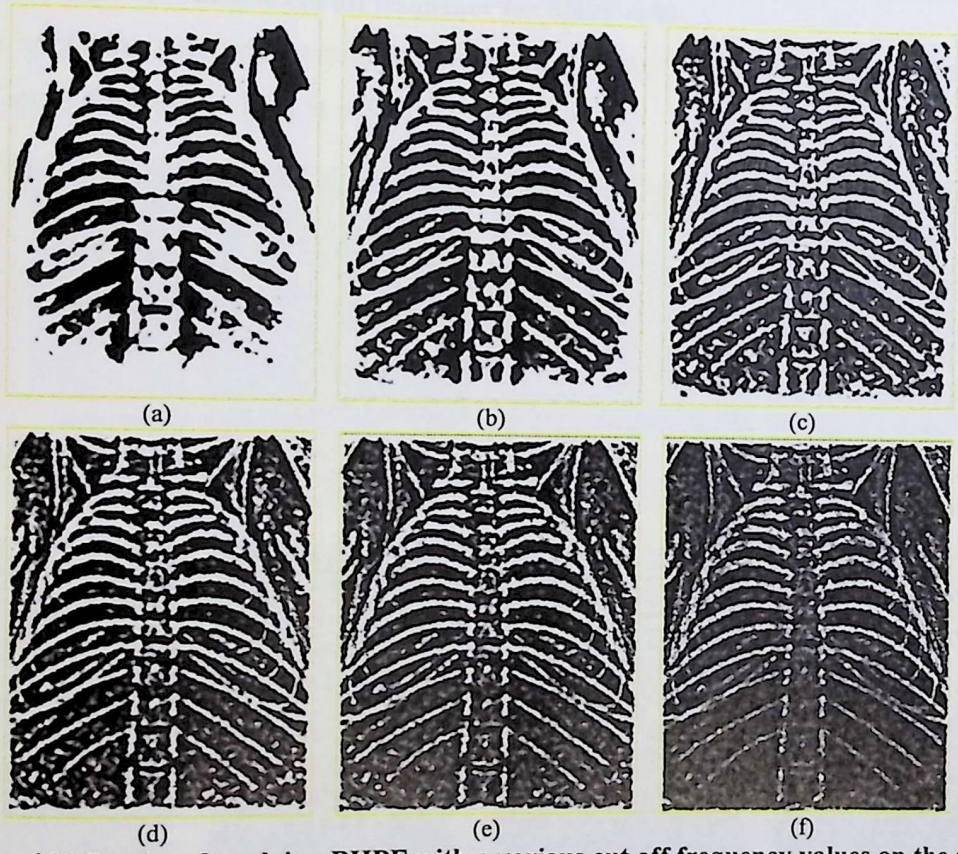


Figure 6.30 Results of applying BHPF with a various cut off frequency values on the same rib area image

(a) Small cut off frequency value (10) which leads to strong rib edges and many non-rib objects; (b) Small cut off frequency value (25) which leads to strong rib edges and many non-rib objects; (c) Cut off frequency value (50) which leads to good isolation of ribs and decreases non-rib objects; (d) High cut off frequency value (75) which leads to weak rib edges and less non-rib objects;(e) High cut off frequency value (100) which leads to very weak rib edges and less non-rib objects; (f) High cut off frequency value (150) which leads to very weak rib edges and less non-rib objects

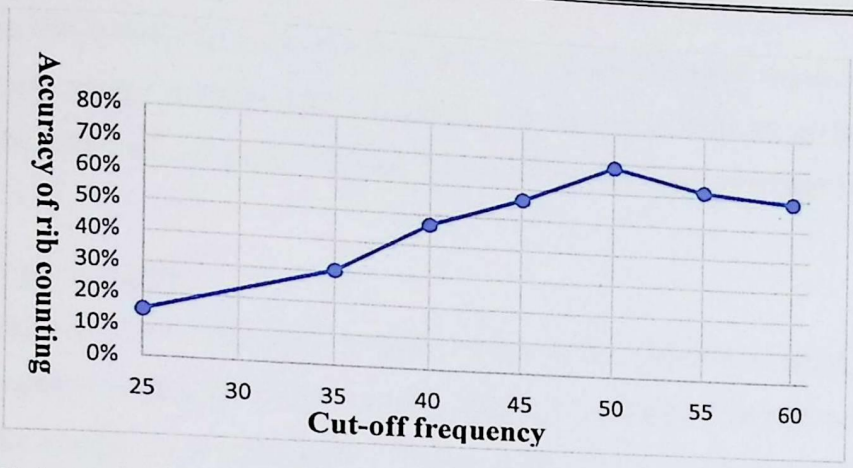
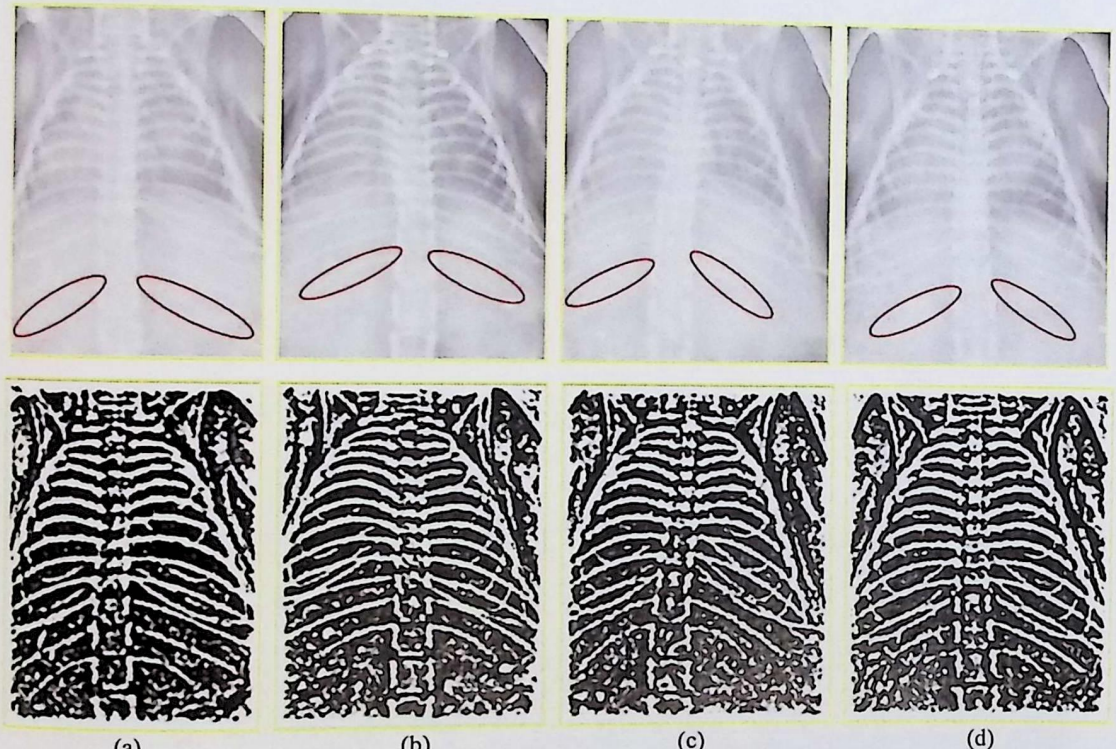


Figure 6.31 Accuracy figures of rib count resulting from using different cut off frequency values (25, 35, 40, 45, 50, 55, and 60) for BHPF. The Figures shows that the best accuracy obtained when the value of cut off frequency for BHPF is 50.

We further apply the selected parameter value on 4 images with different contrast levels as shown in Figure 6.32 to see how it affects the outcomes.



(a) (b) (c) (d)

Figure 6.32 Results of applying BHPF with the selected cut off frequency value (50) on four different mouse rib area images. Each column repressents the mouse rib area image before and after applying BHPF.

(a) The result of applying BHPF shows a good isolation of the ribs; (b) The result of applying BHPF shows a good isolation of the ribs;(d) The result a good isolation of the ribs;(c) The result of applying BHPF shows a good isolation of most of ribs except for the lower left rib. This is because the of applying BHPF shows a good isolation of most of ribs except for the lower left rib. This is because the lower left rib in the original image has very low contrast and so it is difficult to be recognized

We can see that the selected value can maintain the whole rib shapes for most of the cases, especially the last two ribs encircled by the red ellipses. However, the most challenging case is shown in column (d) in Figure 6.32 where the last two ribs have an extremely low contrast from the background, especially for the left rib. After applying the BHPF with the selected

parameter value, the whole rib shapes for the two ribs cannot be maintained. The surrounding areas contain too many background non-rib objects. Such case is considered as very challenging to be detected correctly because of the rib in the original image is extremely low contrast.

Number of tiles for CLAHE

CLAHE is an important tool that used in this chapter to increase the contrast of the rib area. However, this method is dependent on a very important parameter that can control the degree of contrast of the image. This parameter is the number of tiles as described in Section 2.2.1. Figures 6.33 (b), (c), (d), (e), and (f) show results of enhancing the same rib area image in Figure 6.33(a) with the CLAHE by using four different values for the number of tiles.

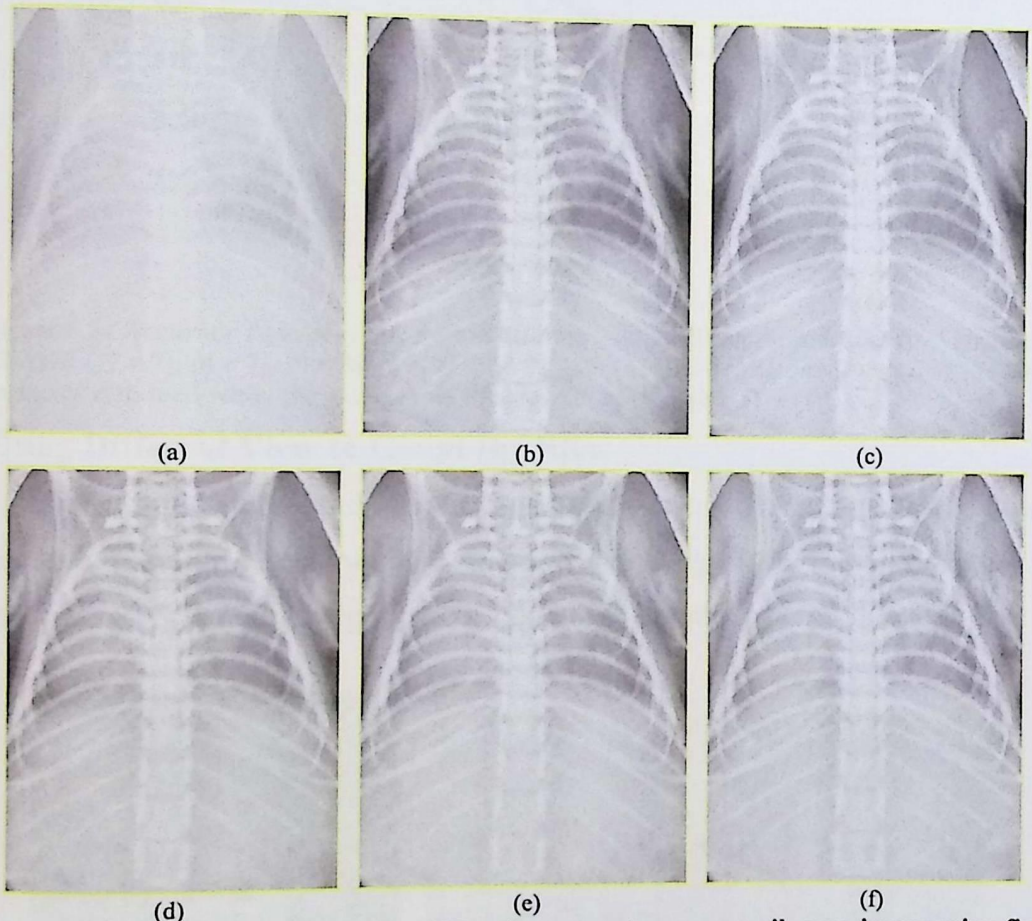


Figure 6.33 Results of applying CLAHE enhancements on same rib area image using five various values for number of tiles. All the result images show better contrast than the original image.
(a) Rib area input image before applying CLAHE;(b) Rib area input image after applying CLAHE with number of tiles as [4 × 4]; (c) Rib area input image after applying CLAHE with number of tiles as [6 × 6]; (d) Rib area input image after applying CLAHE with number of tiles as [8 × 8]; (e) Rib area input image after applying CLAHE with number of tiles as [9 × 9]; (f) Rib area input image after applying CLAHE with number of tiles as [10 × 10]

However, as we explained in Section 2.2.1, increasing the number of tiles should increase the contrast but it will also produce more background noises in the homogenous area. As we can

see from the above figure, it will be difficult to draw a solid conclusion by naive eye on which value gives best contrast with minimum generated noises. To make clearer picture about the suitable value for number of tiles, we change the number of tiles while fixing other parameters and check the accuracy of rib count on the sample of 100 images. Figure 6.34 shows the accuracy for the values ([7 × 7], [8 × 7], [8 × 8], [8 × 9], [9 × 9], [10 × 10]). It has been found that the number of tiles around [8 × 8], (64) tiles gave the best accuracy (up to 70%) because it makes a good balance between the degree of the enhancement and the generated noises.

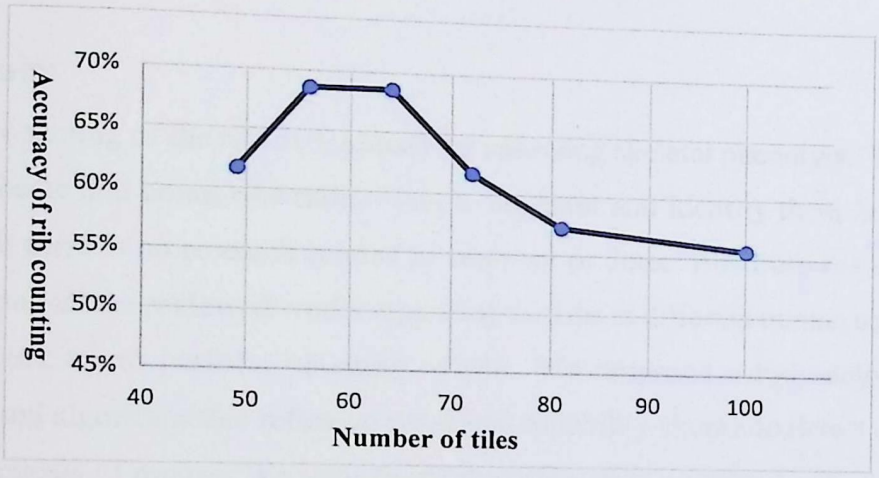
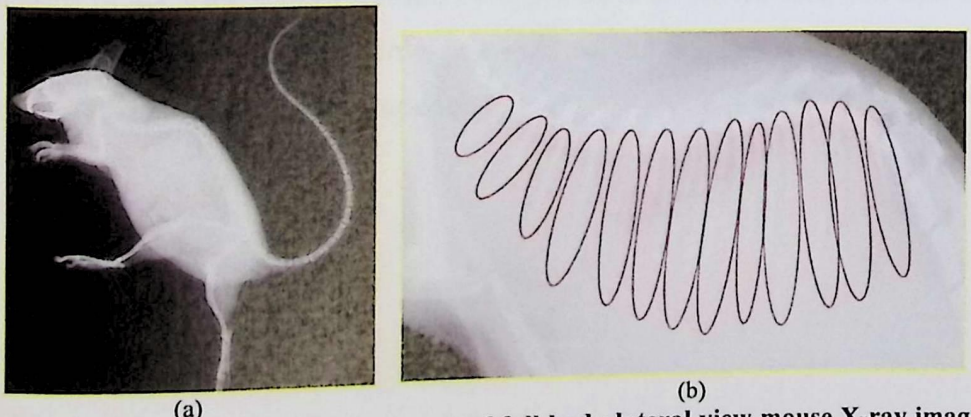


Figure 6.34 Accuracy figures of rib count resulting using different number of tile values for CLAHE ([7 × 7], [8 × 7], [8 × 8], [8 × 9], [9 × 9], [10 × 10]). The Figures shows that the best accuracy obtained when the number of tiles for CLAHE is [8 × 8].

6.5.3 Using Different View to Count the Ribs

As we mentioned in Section 6.2, the dorsoventral view is considered as the best view for rib detection and counting. In here, we will assess if the using of the full body lateral view as shown in Figure 6.35(a) could help in rib detection and counting.



(a) (b)

Figure 6.35 Assessing mouse X-ray image of full body lateral view mouse X-ray image for rib count.
(a) Mouse X-ray image of full-body lateral view; (b) Zoom-in chest area for full body lateral image where each rib is encircled by red ellipse

The lateral view can show all the ribs as shown by red ellipses in zoom in rib area in Figure 6.35(b). However, this view is considered as not very helpful view for rib count for the following reasons. First, for the first 3 or 4 ribs, only the ribs on the close side are shown, whereas the ribs on the distant side are completely obscured. This will affect correct counting of ribs on both sides. Furthermore, for the rest of the ribs, there are a big interference between the ribs from both sides. So it is not clear which rib belongs to which side of the spine. All these reasons justify our choice of constraining the task of rib count to the dorsoventral view images.

6.6 Summary

Detection and counting of the ribs is essential for assessing skeletal phenotype. The aim of this study was to locate and count ribs rather than to segment and identify them in their entirety. We argued that there is no research related to segment or detect ribs from the mouse images. Moreover, among all the reviewed works regarding the ribs in different human medical images, no previous work try to perform counting of ribs. We proposed a fully-automated pattern recognition based algorithm that relies on two-level reliability checks to detect and counts the ribs in X-ray images of mouse. We investigate the difficulties presented with detection of 3D objects in 2D X-ray images. We demonstrated that the mothed does not only reduce the number of images that need to be checked by manual annotators by 82% but also it makes the manual counting easier by producing accurate rib detection (99.2% accuracy). The evaluation on 100 images showed the ability of proposed method to detect majority of ribs in mouse X-ray images with a minimum amount of missing ribs. We also showed the ability of our solution to overcome a variety of challenges presented by the segmentation of the X-ray images. However, some limitations in accurate determination of upper start of the rib area and loss of the lower ribs due to very low contrast still remain not entirely solved. Further research for more accurate and robust methods is still very much needed.

Chapter 7: Work in Progress: Other Automated Measurements of Mouse Body from X-ray Images

Besides the main objectives of our study presented in Chapters 4, 5 and 6, there are still some other automatic annotations of mouse body that can be of a potential interest for the biomedical community. This chapter presents several automatic methods to estimate the additional measurements that are still at the *work in progress stage*. The automated measurements include additional head dimension measurements and pelvis segmentation that could be both related to certain changes due to genotyping. Investigating these measurements helps to establish any significant link between phenotyping changes and the genes responsible for the changes.

Initial evaluations for our proposed methods have been conducted on a data set of 100 images. Automatically obtained measurements and segmentation are subjectively rated as either 'good', 'acceptable' or 'poor'. However, the results for the three head measurements showed high percentage for the 'good' level which ranges from 95% to 100%. For the pelvis segmentation, the accuracy was up to 85%.

The rest of this chapter is organized as follow. Sections 7.1 presents a brief discussion of the three proposed methods to automate the estimation of three measurements in the mouse head. The segmentation of the pelvis from the full body mouse is presented in Section 7.2. Section 7.3 reports and analyses the initial evaluation results.

7.1 The Proposed Methods for the Additional Measurements of the Mouse Head

This section presents the proposed methods to automate three different measurements in the mouse head X-ray image. Those methods benefit from some of pre-processing operations and ideas that were developed in Chapter 4, such as the automatic alignment and the mask preparation and using of intensity based approach. In addition, some of landmarks points obtained in Chapter 4 will also be used in the proposed solutions.

7.1.1 Brain Case Perimeter Estimation

One of important measurements is the brain case perimeter which refers to the border of the brain. The perimeter measurement provides valuable information about the circumference, the area, and the roundness of the brain (information provided by Sanger Institute). The proposed method to estimate the brain case perimeter is mainly depending on the binary skull mask, L_1

and L_2 landmarks (see Chapter 4). The brain case perimeter could be obtained by trimming the skull mask using the location of length landmarks and then using Canny edge operator (Canny, 1986) to extract the border of the mask. Figure 7.1 shows the main stages of estimating the brain case perimeter.

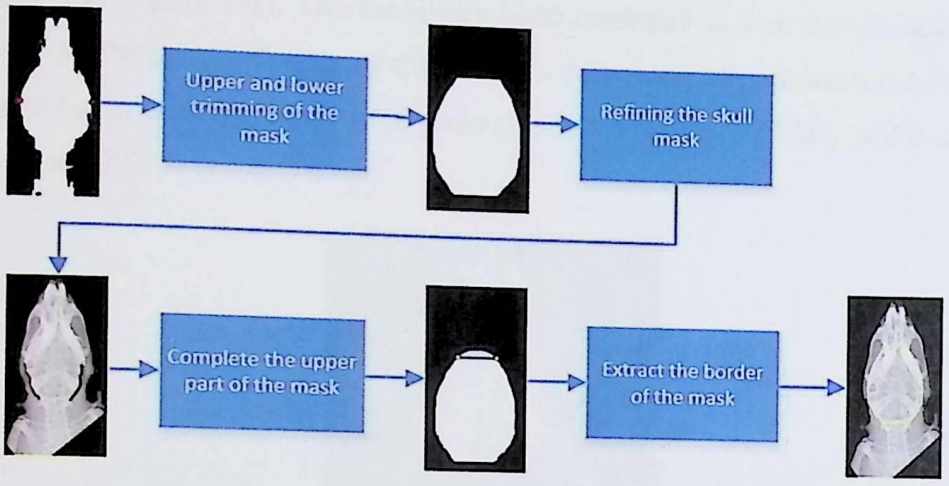
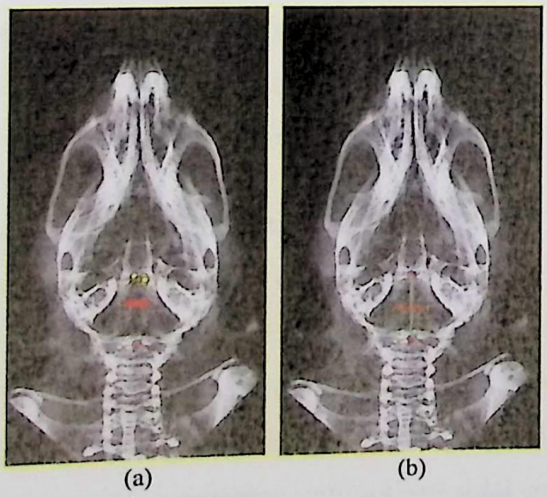


Figure 7.1: Main stages of the automatic brain case perimeter estimation. The process starts by trimming the upper and lower area of the skull mask. After that the edge of the prepared mask is obtained and superimposed on the input skull image (yellow border).

7.1.2 Lambdoid Sutures Measure Estimation

The second measurement is lambdoid sutures which is a dense tissue that connects between the occipital bone and parietal bone of the skull (information provided by Sanger Institute). To estimate the lambdoid sutures length, the lower landmark for the lambdoid sutures is considered as the lower length L_1 (Chapter 4). For the upper landmark detection, we use the same idea of intensity based as explained in Chapter 4. Figure 7.2 shows the lower landmark and the setting of the intensity based (a) and the final estimated length (b).



(a) (b)

Figure 7.2 Lambdoid suture length measure.
(a) The initial line (red) line and the updated locations of the lines' points (yellow points); (b) The upper and lower landmarks of the lambdoid suture (red points) and the distance between them (green line)

7.1.3 Maxillary Width Estimation

The maxillary width measure is manually estimated as the distance between the two maxillary bones on both sides of the mouse's head. To locate the maxillary bone, we start from a reference point (blue point in Figure 7.3) and then annotate points on the edges of the nasal bone (red and green points in Figure 7.3). The maxillary bone landmark is then determined as the first sudden change of the x-coordinates of the nasal bone points. The distance between the two maxillary bone landmarks (green and red points) is taken as the maxillary width as shown by blue line in Figure 7.3.



Figure 7.3 Automatic maxillary width estimation. The green and red points represent the edge of the nasal bone. The maxillary bone landmark is determined as the first sudden change of the x-coordinates of the nasal bone points. The maxillary width is the distance between the two maxillary bone landmarks (blue line).

7.2 Pelvis Segmentation

In this section an automatic solution for pelvis segmentation based on structural and orientation properties of the mouse pelvis in the X-ray images is proposed. In this proposal, we are interested in the segmentation of the pelvis girdles. The main steps of the proposed solution are shown in Figure 7.4. The solution consists of following steps:

1. Initial pelvis ROI segmentation from the full body mouse aligned image exploiting global landmarks (see Section 5.3.2 for details).
2. Obtain the pelvis mask using grey scale morphology operations (Section 2.2.2), Gabor filters (Aach, Kaup and Mester, 1995) with the set of angles between (85°) and (95°) and Otsu method (Section 2.2.2). Gabor filters are bandpass filters which obtained by modulating a Gaussian envelope function with a sinusoidal wave and they are employed for feature extraction, edge detection and texture analysis in image processing (Aach,

Kaup and Mester, 1995). The Gabor filter is chosen with (85° and 95°) because in the aligned pelvis image the two girdles of the pelvis are most likely to be vertical.

3. Increase the contrast of pelvis image using Gamma transformation (Section 2.2.1) and grey scale morphology enhancement (Section 2.2.2).
4. Obtain the pelvic bones using a bit-wise multiplication between the initial pelvis mask obtained from (2) and the enhanced version of the pelvis obtained from (3).

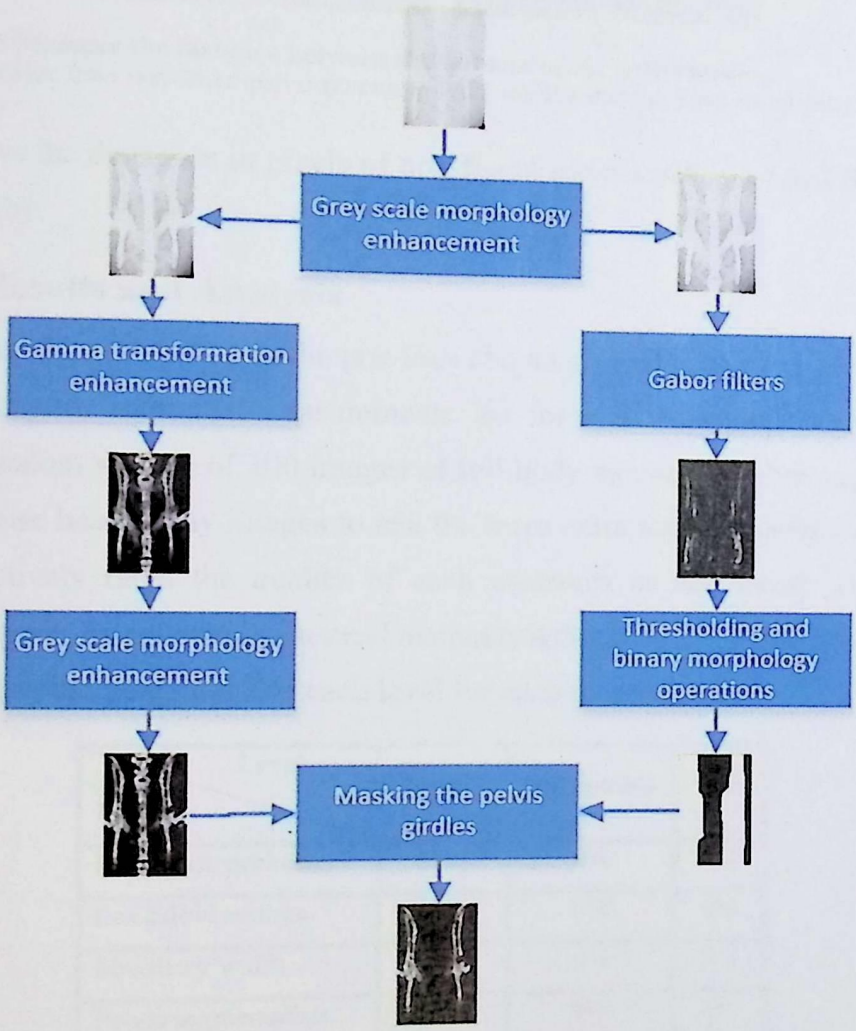
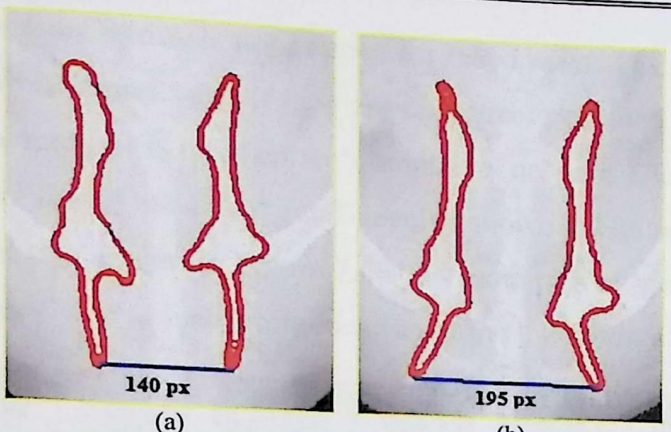


Figure 7.4 Pelvis segmentation process from mouse X-ray image. The process starts by preparing pelvis mask by applying grey scale morphology enhancements and Gabor filter on the pelvis ROI image. The pelvis mask is then used to mask the pelvis object from an enhanced pelvis image.

One of the possible phenotyping that needed to measure from the pelvis is the flaring in the ischia, the base bone of the pelvis (information is provided by Sanger Institute). Following the pelvis segmentation, this measure can be estimated automatically by calculating the distance between the lowest points of each pelvis girdle as shown in Figure 7.5.



(a) (b)

Figure 7.5 Measure the distance between the ischium of the pelvis girdles.
(a) The distance from non-flared pelvis (normal pelvis); (b) The distance from flared pelvis (abnormal pelvis)

The figure shows the distances in pixels of non-flared normal pelvis (a) and flared one at the ischium pelvis (b).

7.3 Initial Results and Analysis

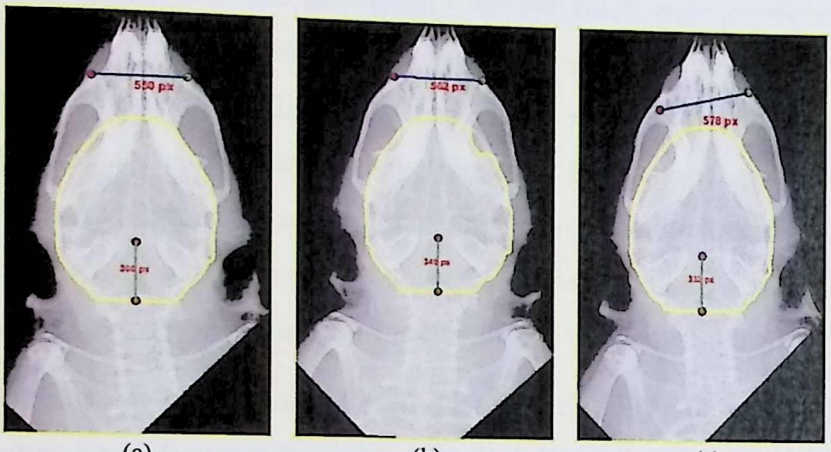
Unlike the evaluations presented in the previous chapters, we do not have a complete ground truth for validating the estimated measurements. So, for evaluation purposes, we applied our methods on a random sample of 100 images of full body mouse for pelvis segmentation and another 100 mouse head X-ray images to test the three extra measurements in the head area. We then subjectively rated the quality of each measures or annotation result as 'good', 'acceptable' or 'poor' by referring to several manually annotated images provided by an expert. Table below shows the percentage of each level for each measurement.

Level / Measure	Good	Acceptable	Poor
Brain case perimeter	96%	4%	0%
Lambdoid sutures	100%	0%	0%
Maxillary width	95%	0%	5%
Pelvis segmentation	85%	9%	6%

Table 7.1 Percentage of accuracy for proposed methods to estimate different measurements in the head area and the pelvis segmentation.

For the head measurements, different accuracy levels are shown in Figure 7.6. For the brain case perimeter, the result is considered as 'good' if the line covers the whole supposed perimeter without including any unwanted or missing wanted area as shown in Figure 7.6(a) and (c). Any small bias of the obtained boundary to inside or outside of the brain case is classified as 'acceptable' level as shown in Figure 7.6(b). Such cases classified as 'acceptable' because it seems that small curve to inside or outside of the brain case perimeter should not affect the overall result of the perimeter estimation. The case is classified as 'poor' only if the

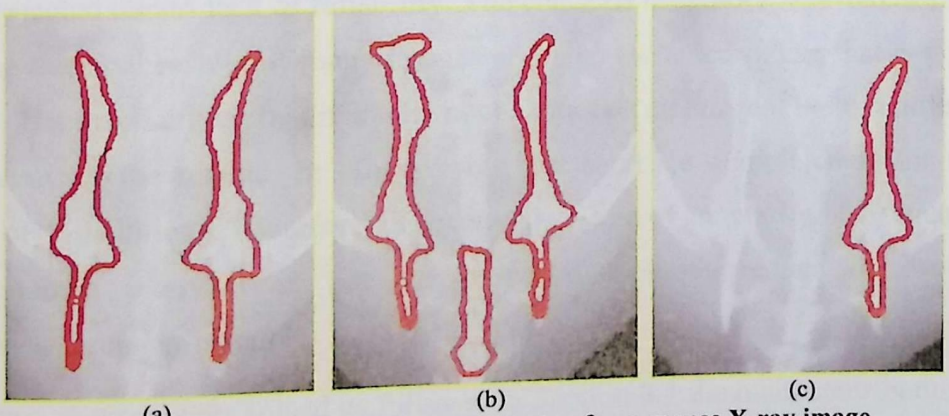
part of the border is missing which is not happen for our dataset. The other two measures in the head depend on two landmarks and if they are on correct positions, the measure is said to be 'good' otherwise the measure is rated either 'acceptable' or 'poor'. For all the tested images, the lambdoid sutures are shown to be in a 'good' level as shown in Figure 7.6. For the maxillary width measure, the first two images in Figure 7.6 show examples of 'good' measure while the third one is rated as 'poor' due to inaccurate location of the left landmark.



(a) (b) (c)

Figure 7.6 Accuracy of extra measurements in head area.
(a) Good level estimation for brain case perimeter; (b) Acceptable level estimation for brain case perimeter; (c) Poor level estimation for maxillary width measure

The rating for the pelvis segmentation depends on the accuracy of the pelvis girdle segmentation which is done subjectively. The result is said to be 'good' if all or almost all of the pelvis girdles are segmented without extra unwanted objects as shown in Figure 7.7(a). In some cases, the whole required part of the pelvis is well segmented but with some small extra unwanted objects stick to the pelvis or some unwanted object segmented also. Such cases are rated as 'acceptable' quality as shown in Figure 7.7(b). The final 'poor' level happens when one or both girdles are missing either completely or partially as shown in Figure 7.7(c).



(a) (b) (c)

Figure 7.7 Different examples of pelvis segmentation from mouse X-ray image.
(a) Good level pelvis segmentation; (b) Acceptable level pelvis segmentation; (c)Poor level pelvis segmentation

Chapter 8: Conclusion and Future works

The X-ray imaging is one of the oldest and yet still very effective medical imaging technique (Haidekker, 2013). In genetic research, the X-ray imaging of mice was standardized as one element of the phenotyping pipeline (Koscielny et al., 2014). This thesis has mainly focused on examining X-ray images of the genetically altered mice and consequently developing solutions to automate the manual annotations for the images with the ultimate goal to detect phenotyping abnormalities and candidate the responsible genes for such changes. The need of fully automatic solutions was motivated by the plenty of genetic projects, the huge number of produced X-ray images from those projects and plenty of genotype lines needs to be investigated. Such detections have a high value in terms of identifying the genetic reasons behind some acute diseases and abnormalities that affect the human.

This thesis presented a set of automated methods towards fully automatic phenotyping. Our experimental studies have demonstrated the effectiveness and efficiency of these methods in detecting and quantifying changes in a mouse body skeleton in a consistent manner, and linking the changes to candidate genes that are responsible for some changes in specific parts of mouse body skeleton. Due to the challenges associated with the automatic phenotyping highlighted in Chapter 1, the novelty of thesis was centred around self-assessment algorithms and hybridization of efficient intensity based techniques and robust texture-based trainable ones.

To increase the reliability of automatic phenotyping results, we have considered different reliability measures and embedded them into the proposals to enable the detection of any segmentation problems at an earlier stage and recover from them if possible; otherwise, the image was rejected due to lack of reliability. The check of reliability at different stages helped in enhancing the final results of many images and also excluded others that might affect the final results. The thesis argues that it is wise not to annotate an image if there is little confidence in the accuracy of the results. In other words, we sacrifice completeness for the sake of correctness of the solutions. Therefore, the eventual system of annotation still requires a certain degree of human intervention, but the proposed methods reduce the workload and inconsistency as much as possible

The rest of this chapter is organised as follows. In Section 8.1 the main contributions achieved by this thesis are summarised. The impact of the research results and the transferability of them

to the human X-ray image analysis are discussed in Section 8.2. Finally, Section 8.3 outlines the future research work and directions.

8.1 The Main Contributions of the Thesis

The work described in this thesis was mainly focused on automated and reliably annotations of different organs from the X-ray images of the mouse body. The automatic solutions add value in terms of speeding up the process and reducing the intra- and inter-observer variation problems. The solutions proposed in this thesis presented a set of findings and made contributions that are summarized as follows.

1- Developing reliable automatic solution to locate landmarks to estimate the brain width and length from X-ray images of mice head. The proposed solutions are based on
   a. Combining intensity-based techniques and machine learning techniques based on RF classification model to improve the overall robustness without compromising the efficiency factor required for the high-throughput analysis.
   b. Automatic self-assessment components to control the flow of the solutions e.g. deciding on when to exclude images with poor quality or unreliable skull modelling and when to use machine learning over efficient intensity-based techniques.
   c. Proposing a RF landmark detection model based on combining two types of texture features (LBP and HOG) at the score level (i.e. combining the output of the decision trees of each feature) to enhance the overall accuracy of landmarks detection.
2- Identifying a list of candidate genes associated with abnormal brain size. The proposed landmarks detection system succeeded to define a list of genes that might be responsible for changes in the brain size. Some of those genes were already confirmed to cause such abnormalities.
3- Proposing an automatic intensity-based and machine learning based solutions for reliable spine segmentation from the whole body of the mouse X-ray images. A list of index and angle based metrics were used to quantify the curvature of the segmented mouse spine. The obtained metrics were found effective to identify cases of abnormal spines.
4- Proposing a rib segmentation and counting method. The automatic solution to segment and count the ribs from the full body mouse X-ray images is based on two-level

classification scheme to filter out non-rib objects depending on texture and geometrical features such as the horizontal distance from the spine and the orientation of the rib.

5- Proposing solutions to deal with the problem of aligning and initial cropping of different parts of mouse body and head from the X-ray images. It was demonstrated that these solutions are both simple and effective.

6- Proposing a solution to segment the pelvis from the full body mouse X-ray image. The solution exploited the known angles of the pelvis girdles in an aligned image.

7- As a work in progress, investigating a solution to locate different landmarks in the head of the mouse to estimate important measures such as the maxillary width and lambdoid sutures. In addition to those landmark based measurements, a brain case perimeter of the mouse was annotated with high accuracy.

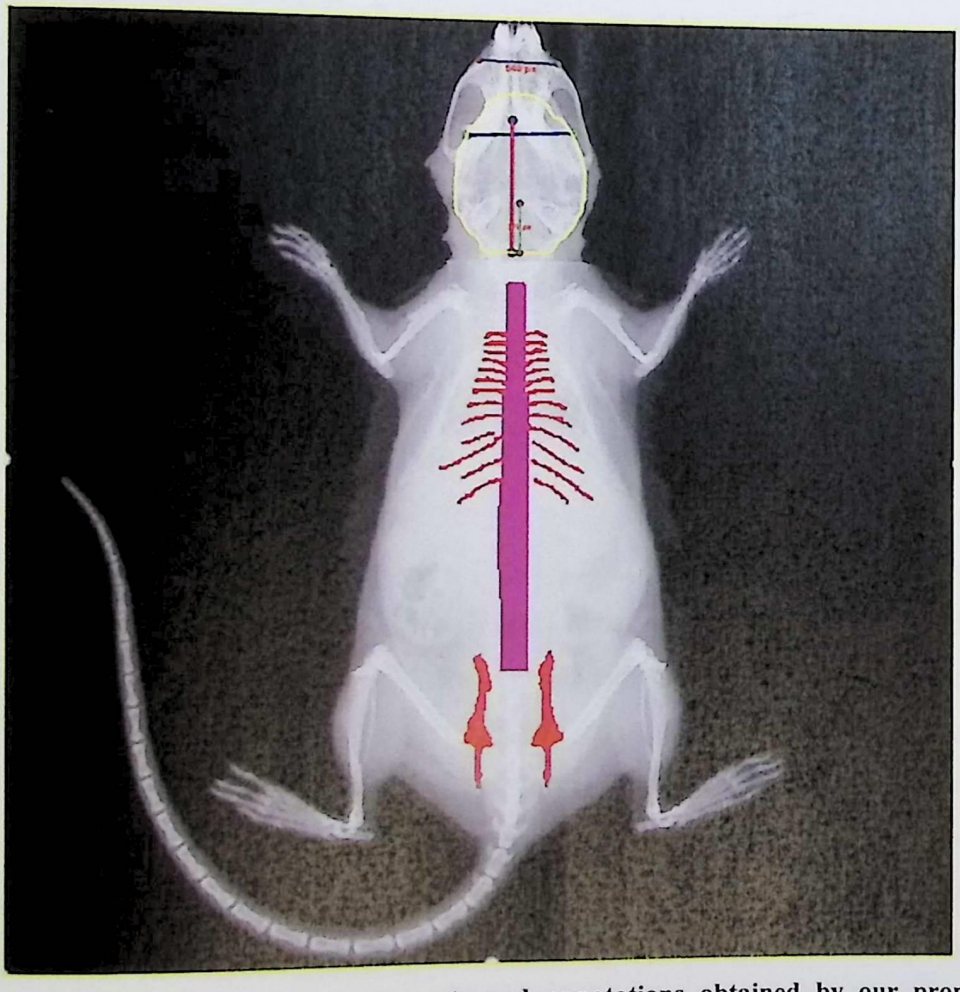


Figure 8.1 The automatic measurements and annotations obtained by our proposed solutions superimposed on a full body mouse X-ray image. Those annotations include different measurements and landmarks annotation in the head of the mouse. Also, the spine and pelvis segmentation. Finally, it includes the rib detection and count.

Behind most contributions listed above are some common approaches of problem solving, such as rotating images for alignment, using masks to segment ROIs, locating initial positions of

important landmarks, refining landmarks for ideal eventual positions, and measuring the annotation variables. This sequential process is augmented with the assessment of reliability of the processed images and measurement results. Figure 8.1 shows all the key landmarks and segments obtained from the different proposed methods in the thesis.

8.2 Impacts of the Research and its Applicability to the Human Images and other Mouse Data

All the works described in this thesis was designated to segment and annotate different organs in the X-ray images of the mouse body, and extract some useful information from them. The direct and indirect benefits of the work achieved by the thesis as well as the impacts of our work can therefore be summarised into the following points:

1- Assisting the identification of novel genes related to specific diseases (e.g. Down's syndrome and Scoliosis) or abnormalities. Even any defined gene is obtained from manipulation of the mouse data set, it should be highly related to the human because of the big similarity between the human and the mouse genetic backgrounds.

2- Providing the biology researchers with consistent and accurate software tools that can segment, annotate and estimate the measures from different organs in the mouse's body.

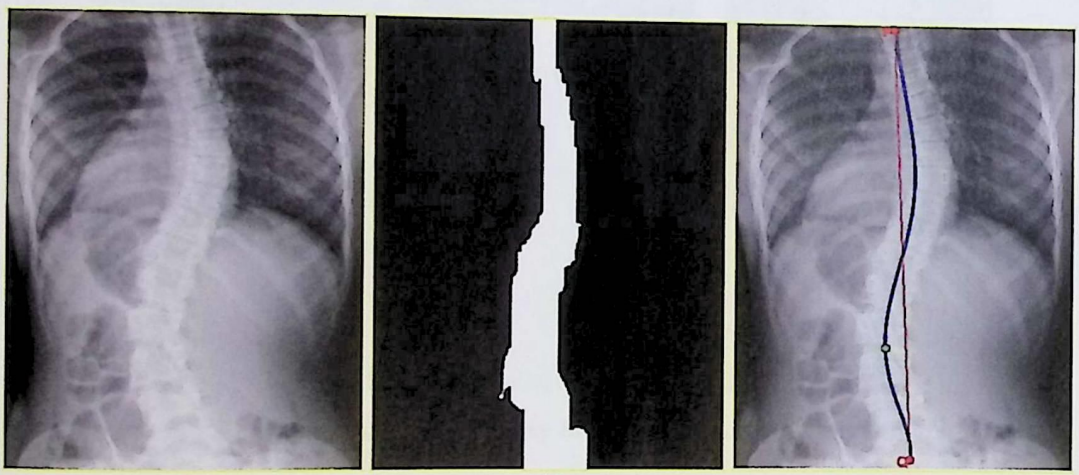
3- Most of the Developed solutions can be applied on the corresponding organs in human medical images for assisting doctors in accurate diagnoses. As a matter of interest, we investigated the initial applicability of two of the proposed methods in this thesis to the medical human X-ray images after some adjustments. Results of applying the spine curvature estimation method on a scoliosis X-ray image of humans can be seen in Figure 8.2. We believe that the spine curvature estimation method could be a good solution for low quality human X-ray images especially where the vertebrae boarders are very difficult to be identified. Also, the main steps of the proposed rib segmentation method were applied on the human chest X-ray image and the result is shown in Figure 8.3. The result indicates that the rib segmentation method is indeed applicable to a human chest X-ray image. With some adjustment, the ribs can not only be detected and counted, but also measured in terms of their lengths. Since we currently have no access to a sufficiently number of human X-ray images, those two available images are used only as a proof of concept.

4- There are different scientific institutions which are interested in studying the relationship between genes and diseases by using animal models. One of the well-known centres for mouse genetics is MRC Harwell (MRC Harwell, 2017). They

provided us with a limited data set of 19 images of mouse X-ray images to conduct an initial investigation of using our proposed solutions to estimate the different measurements in the head area. The initial results were promising in terms of the applicability of our solutions (with some modifications) to a different data set. Figure 8.4 shows samples of annotated images from MRC Harwell data set.

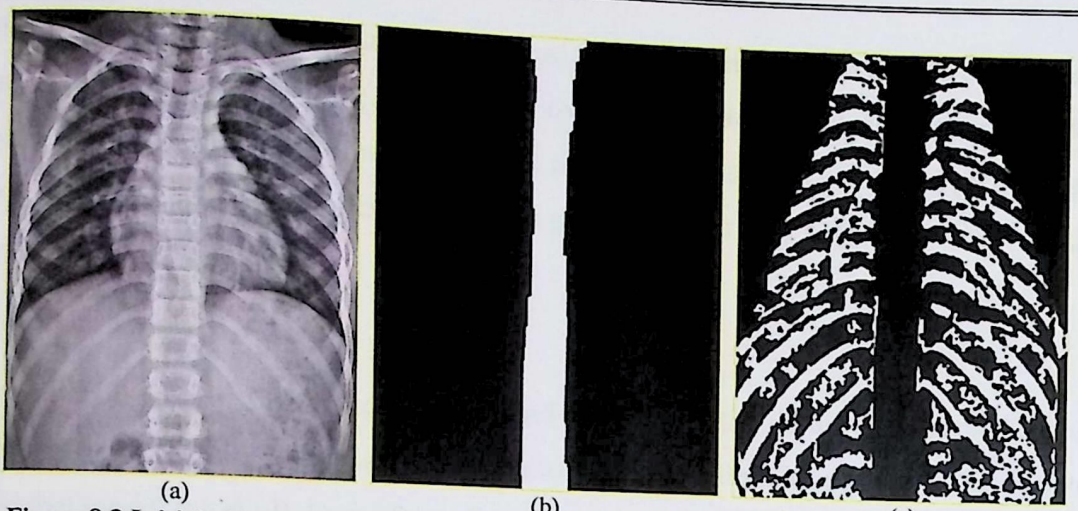
5- In addition to applicability on the corresponding parts on human medical images (highlighted in point 3) and different animal data sets (highlighted in point 4), the algorithms and thoughts developed in this research can be generalised to solve other problems in different medical/biomedical imaging such as CT scan.

6- Useful knowledge and research skills are transferrable to the home institutions. All the obtained experience in biomedical image processing could bring great advantages to the home institutions in terms of knowledge transfer, which will benefit teaching by down-streaming the knowledge, and research by building research excellence centre(s) in the home institutions.



(a) (b) (c)

Figure 8.2 Human spine curvature estimation from X-ray image.
(a) Chest X-ray image, (b) Segmented spine mask from (a); (c) Obtained spine curvature using our proposed method for mouse spine curvature estimation



(a) (b) (c)

Figure 8.3 Initial rib segmentation from human X-ray image.
(a) Human X-ray chest image; (b) Segmented spine mask from (a); (c) Initial segmentation of ribs from (a)

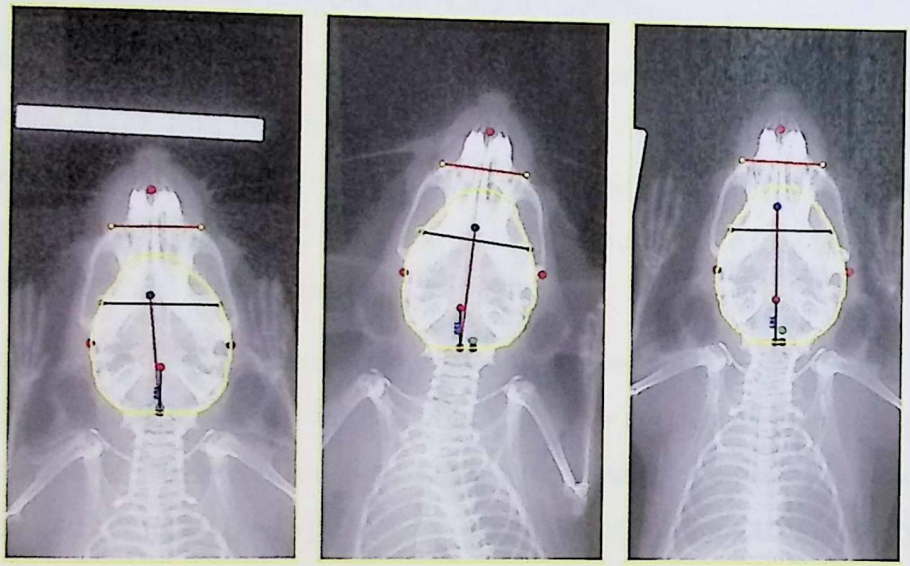


Figure 8.4 Samples of annotated head X-ray images with different measurements (images are provided by MRC Harwell).

8.3 Future Work

The aim of the thesis is to automate the manual annotation and produce extra measurements on different parts of the mouse body to identify responsible phenotyping genes with a higher reliability and accuracy. Achieving such a big goal needs to expand the scope of our reported work further. As a part of bigger research team, our work should end by producing a list of genes which seems to cause a different phenotyping in the mouse's body. The job of other researchers in the team is to conduct more experiments to confirm the genotype to phenotype outcomes in the mouse. Moreover, other researchers will then explore the feasibility of translating such findings to humans.

In general, the future work should be dedicated to improvements and further development of the current proposed methods to increase the overall accuracy of the annotation. In addition, the work should continue to extract important measurements from annotated parts yielding to produce a list of genes from phenotyping. In particular, we will explore the following areas that still need extra work.

Head measurements

All the measurements estimated in the skull and the brain of the mouse depend on the detection of landmarks. The initial search areas estimation of the landmarks rely on accurate detection of circle that model the skull. One of possible improvements to enhance the reliability measures of the circle's position is to use machine learning approach to predict whether the circle is in an acceptable position and size (see Section 4.3.3 for details).

The results of machine learning based approach for landmarks detection in Chapter 4 were shown to change depending on the type of feature used for training. Our intended accuracy enhancement of landmarks detection involves investigating different features that can describe the target area in a better way. Also, one of the possible thoughts is to fuse between the different features to enhance the overall accuracy.

In addition to investigating different features, we will investigate to the effectiveness of deep learning. Deep learning is a potential solution for end to end learning i.e. without a need to extract measurements/features. However, this requires large number of samples from both types (normal and abnormal). Typical CNN network of sufficient hidden layers requires hundreds of thousands of images. This in not currently available especially the number of abnormal images. For example, we found only handful of genes associated with brain size abnormality i.e. only tens of abnormal images which is not sufficient. However, when our algorithm is put in practice we can potentially collect the output and reuse it as increment learning.

Another piece of research is to exploit the RF approach to detect some important landmarks for the three additional measurements presented in Section 7.1. After that, we will apply the refined proposed methods on the whole available data set of mouse head X-ray images, and evaluate the effectiveness of the annotations in defining the normal ranges of the measurements and hence highlighting any possible abnormalities.

Spine curvature estimation

In our proposed solution, we use the spine mask to determine the search areas for obtaining accurate spine borders. Our current solution relies on the use of heuristically determined parameters and thresholds, which may limit the potential use of our solution in other domains of application. One possible direction for future investigation is the estimation of the initial search areas without using a spine mask.

In our current work, the main goal of segmenting the spine is to estimate its curvature. However, the estimated part may not cover all the spine from the pelvic vertebrae to the cervical vertebrae. The suggestion for future work is to make more accurate segmentation for all the spine by combining some ideas already tested in different chapters of this thesis. This complete segmentation would help the work in vertebrae segmentation (i.e. pelvic vertebrae). For the lumbar vertebrae segmentation, a potential area of enhancing the proposed solution presented in Section 5.5.1 is to use more sophisticated features that describe the appearance/shape of vertebra such as ASM. There are some important phenotyping interests regarding vertebrae shapes and vertebrae fusion (Sanger, 2013).

Rib segmentation

The main point to be addressed in the rib segmentation is to develop a reliable automatic method to accurately determine the start of the rib area from the upper side. This point is the trickiest point because of the similarity between the upper rib and the processes of the cervical vertebrae in mouse X-ray image.

The rib segmentation work in this thesis focused to rib detection and counting rather than accurate segmentation of the whole ribs, and automatic measurements of rib lengths. The future work will involve using the detected rib masks to segment rib borders more accurately, making those further quantifications possible. Furthermore, using more sophisticated features to train a more effective classifier to distinguish between rib and non-rib objects will be considered in the future work.

Pelvis segmentation

As we mentioned, the pelvis segmentation method is still a work in progress. The next goal is to complete this method and enhance the segmentation results. This results should then exploited to extract different measurements to assess any abnormalities in the pelvis shapes that may be the results of certain responsible genes.

Finally, we can conclude that if humans can be trained with experience and knowledge to detect the right landmarks for making different measurements, a machine-based system may also be trained to locate the landmarks more precisely. As a future work, it is interesting to investigate the effectiveness of trainable methods such as the RF and deep learning in detecting different landmarks and annotation different parts in the biomedical images.

References

Aach, T., Kaup, A. and Mester, R., 1995. On texture analysis: Local energy transforms versus quadrature filters. *Signal Processing*, 45(2), pp.173–181.

Adissu, H.A., Estabel, J., Sunter, D., Tuck, E., Hooks, Y., Carragher, D.M., Clarke, K., Karp, N.A., Project, S.M.G., Newbigging, S., Jones, N., Morikawa, L., White, J.K. and McKerlie, C., 2014. Histopathology reveals correlative and unique phenotypes in a high-throughput mouse phenotyping screen. *Disease Models & Mechanisms*, 7(5), pp.515–524.

Al-Helo, S., Alomari, R.S., Chaudhary, V. and Al-Zoubi, M.B., 2011. Segmentation of lumbar vertebrae from clinical CT using active shape models and GVF-snake. In: *Conference proceedings :Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. pp.8033–8036.

Ali, A.A., Dale, A.M., Badea, A. and Johnson, G.A., 2005. Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain. *NeuroImage*, 27(2), pp.425–435.

Anderson, S.M., 2007. Spinal Curves and Scoliosis. *Radiologic Technology*, 79(1), pp.44–65.

Aubert, B., Vazquez, C., Cresson, T., Parent, S. and De Guise, J., 2016. Automatic spine and pelvis detection in frontal X-rays using deep neural networks for patch displacement learning. In: *Proceedings - International Symposium on Biomedical Imaging*. pp.1426–1429.

Badea, A., Nicholls, P.J., Johnson, G.A. and Wetsel, W.C., 2007. Neuroanatomical phenotypes in the Reeler mouse. *NeuroImage*, 34(4), pp.1363–1374.

Baiker, M., Milles, J., Vossepoel, A.M., Que, I., Kaijzel, E.L., Lowik, C.W.G.M., Reiber, J.H.C., Dijkstra, J. and Lelieveldt, B.P.F., 2007. Fully automated whole-body registration in mice using an articulated skeleton atlas. In: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings*. pp.728–731.

Baple, E.L., Maroofian, R., Chioza, B.A., Izadi, M., Cross, H.E., Al-Turki, S., Barwick, K., Skrzypiec, A., Pawlak, R., Wagner, K., Coblentz, R., Zainy, T., Patton, M.A., Mansour, S., Rich, P., Qualmann, B., Hurles, M.E., Kessels, M.M. and Crosby, A.H., 2014. Mutations in KPTN cause macrocephaly, neurodevelopmental delay, and seizures. *American Journal of*

Human Genetics, 94(1), pp.87–94.

Becker, L., Mito, T., Takashima, S. and Onodera, K., 1991. Growth and development of the brain in Down syndrome. *Progress in clinical and biological research*, 373, pp.133–52.

Birkfellner, W., Figl, M. and Hummel, J., 2014. *Applied Medical Image Processing*. 2nd ed. CRC press.

Bland, M. and Altman, D., 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, 327(8476), pp.307–310.

Bogue, M., 2003. Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *Journal of Applied Physiology*, 95(3), pp.1335–1337.

Breiman, L., 1999. Random forest. *Machine Learning*, 45(5), pp.1–35.

Brown, S.D.M. and Moore, M.W., 2012. The International Mouse Phenotyping Consortium: Past and future perspectives on mouse phenotyping. *Mammalian Genome*, 23(9–10), pp.632–640.

Canny, J., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp.679–698.

Chen, Y.-L., Chen, W.-J. and Chiou, W.-K., 2007. An alternative method for measuring scoliosis curvature. *Orthopedics*, 30(10), pp.828–831.

Ching, Y.P., Qi, Z. and Wang, J.H., 2000. Cloning of three novel neuronal Cdk5 activator binding proteins. *Gene*, 242(1–2), pp.285–294.

Chu, C., Chen, C., Wang, C.W., Huang, C.T., Li, C.H., Nolte, L.P. and Zheng, G., 2014. Fully Automatic Cephalometric X-Ray Landmark Detection Using Random Forest Regression and Sparse Shape Composition. In: *Proc. ISBI Int. Symp. Biomed. Imag. 2014, Automat. Cephalometric X-Ray Landmark Detection Challenge*. pp.9–16.

Cobb, J., 1948. Outlline for the Study of Scoliosis. *American Academy of Orthopaedic Surgeons*, 5, p.261.

Collins, F.S., Morgan, M. and Patrinos, A., 2003. The Human Genome Project: Lessons from large-scale biology. *Science*, 300(5617), pp.286–290.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. and Walters, L.R., 1998. New goals for the U.S. Human Genome Project: 1998-2003. *Science*, 282(5389), pp.682–689.

Cong, L., Guo, W. and Li, Q., 2016. Segmentation of ribs in digital chest radiographs. *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, 9788(2258), pp.2–7.

Cortes, C. and Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 20(3), pp.273–297.

Cover, T. and Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp.21–27.

Cui, W., Yu, T., Ren, L. and Shao, H., 2012. Ribs segmentation based on image fusion and wavelet de-noising. In: *2012 5th International Conference on BioMedical Engineering and Informatics*. pp.362–366.

Dalal, N. and Triggs, B., 2005. Histograms of oriented gradients for human detection. In: *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*. pp.886–893.

Diab, K.M., Sevastik, J.A., Hedlund, R. and Suliman, I.A., 1995. Accuracy and applicability of measurement of the scoliotic angle at the frontal plane by Cobb's method, by Ferguson's method and by a new method. *European Spine Journal*, 4(5), pp.291–295.

Ferguson, A.B., 1930. The Study and Treatment of Scoliosis. *Southern Medical Journal:*, 23(2), pp.116–120.

Gall, J., Yao, A., Razavi, N., Van Gool, L. and Lempitsky, V., 2011. Hough forests for object detection, tracking, and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11), pp.2188–2202.

Gonzalez, R.C. and Woods, R.E., 2017. *Digital Image Processing*. 4th ed. Pearson International Edition.

Green, E., 2015. *National Human Genome Research Institute*. [online] Available at: <https://www.genome.gov/> [Accessed 1 Sep. 2018].

Greenspan, A., Pugh, J.W., Norman, A. and Norman, R.S., 1978. Scoliotic index: a comparative evaluation of methods for the measurement of scoliosis. *Bulletin of the Hospital for Joint Diseases*, 39(2), pp.117–25.

Griffiths, A., Wessler, S., Lewontin, R., Gelbart, W., Suzuki, D. and Miller, J., 2005. *An Introduction to Genetic Analysis*. 8th ed. W.H.Freeman & Co Ltd.

Grüneberg, H., 1955. Genetical studies on the skeleton of the mouse xv. Relations between

major and minor variants. *Journal of Genetics*, 53(3), pp.515–533.

Guyon, I. and Elisseeff, A., 2006. An Introduction to Feature Extraction. *Feature Extraction: Springer*, pp.1–25.

Haidekker, M., 2013. *Medical Imaging Technology*. 1st ed.

Haralick, R., Shanmugan, K. and Dinstein, I., 1973. *Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics*, Available at: <http://dceanalysis.bigr.nl/Haralick73-Textural features for image classification.pdf>.

Hastie, T., Tibshirani, R. and Friedman, J., 2008. *The Elements of Statistical Learning*. 2nd ed. *Springer 2001*, Springer.

Hoi, S.C.H., Jin, R., Zhu, J. and Lyu, M.R., 2006. Batch mode active learning and its application to medical image classification. In: *ICML: International Conference on Machine Learning*. pp.417–424.

Hough, P.V.C., 1962. Method and means for recognizing complex patterns. *US Patent 3,069,654*, 21, pp.225–231.

Ibragimov, B., Likar, B., Pernuš, F. and Vrtovec, T., 2014. Automatic Cephalometric X-Ray Landmark Detection by Applying Game Theory and Random Forests. *Machine Learning*.

Innes, A., Ciesielski, V., Mamutil, J. and John, S., 2002. Landmark Detection for Cephalometric Radiology Images using Pulse Coupled Neural Networks. In: *Int. Conf. on Artificial Intelligence*. pp.391–396.

Jackson, A.P., Eastwood, H., Bell, S.M., Adu, J., Toomes, C., Carr, I.M., Roberts, E., Hampshire, D.J., Crow, Y.J., Mighell, A.J., Karbani, G., Jafri, H., Rashid, Y., Mueller, R.F., Markham, A.F. and Woods, C.G., 2002. Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet*, 71(1), pp.136–142.

Janumala, T. and Ramesh, K.B., 2015. A Review of Different Analysing Methods to Detect Spinal Deformity. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(8), pp.352–355.

Jimenez-Sanchez, G., Childs, B. and Valle, D., 2001. Human disease genes. *Nature*, 409(6822), pp.853–855.

Johnson, D.R., 1976. The interfrontal bone and mutant genes in the mouse. *J. Anat*, 121(3), pp.507–513.

Kaindl, A.M., Passemard, S., Kumar, P., Kraemer, N., Issa, L., Zwirner, A., Gerard, B., Verloes, A., Mani, S. and Gressens, P., 2010. Many roads lead to primary autosomal recessive microcephaly. *Progress in Neurobiology*, 90(3), pp.363–383.

Kapur, J.N., Sahoo, P.K. and Wong, A.K.C., 1985. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3), pp.273–285.

Kawakami, M. and Yamamura, K., 2008. Cranial bone morphometric study among mouse strains. *BMC Evolutionary Biology*, 8(1), p.73.

Khmelinskii, A., Baiker, M., Kok, P., De Swart, J., Reiber, J.H.C., De Jong, M. and Lelieveldt, B.P.F., 2011. Atlas-based articulated skeleton segmentation of??sPECT mouse data. In: *Proceedings - International Symposium on Biomedical Imaging*. pp.437–440.

Klinder, T., Lorenz, C., von Berg, J., Dries, S.P.M., Bülow, T. and Ostermann, J., 2007. Automated model-based rib cage segmentation and labeling in CT images. *Medical image computing and computer-assisted intervention : MICCAI*, 10(2), pp.195–202.

Koscielny, G., Yaikhom, G., Iyer, V., Meehan, T.F., Morgan, H., Atienza-Herrero, J., Blake, A., Chen, C.K., Easty, R., Di Fenza, A., Fiegel, T., Grifiths, M., Horne, A., Karp, N.A., Kurbatova, N., Mason, J.C., Matthews, P., Oakley, D.J., Qazi, A., Regnart, J., Retha, A., Santos, L.A., Sneddon, D.J., Warren, J., Westerberg, H., Wilson, R.J., Melvin, D.G., Smedley, D., Brown, S.D.M., Flicek, P., Skarnes, W.C., Mallon, A.M. and Parkinson, H., 2014. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Research*, 42(D1).

Kundu, R., Chakrabarti, A. and Lenka, P.K., 2012. Cobb Angle Measurement of Scoliosis with Reduced Variability. *MedImage*, 1211(5355), pp.1–6.

Lam, G.C., Hill, D.L., Le, L.H., Raso, J. V. and Lou, E.H., 2008. Vertebral rotation measurement: A summary and comparison of common radiographic and CT methods. *Scoliosis*, 3(1).

Lee, J., Jomier, J., Aylward, S., Tyszka, M., Moy, S., Lauder, J. and Styner, M., 2009. Evaluation of Atlas based Mouse Brain Segmentation. In: *Proceedings of SPIE*. pp.725943–725949.

Lee, J. and Reeves, A.P., 2010. Segmentation of individual ribs from low-dose chest CT. In: *SPIE*. p.76243J.

Lee, J.S., Wang, J.W., Wu, H.H. and Yuan, M.Z., 2012. A nonparametric-based rib suppression method for chest radiographs. In: *Computers and Mathematics with Applications*. pp.1390–1399.

Li, H., Li, J., Pan, S., Guo, Q., Liu, J. and Kang, Y., 2010. Automatic rib positioning method in CT images. In: *2010 4th International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2010*. pp.1–4.

Li, X., Luo, S. and Hu, Q., 2015. An automatic rib segmentation method on X-ray radiographs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8935, pp.128–139.

Li, X., Luo, S., Hu, Q., Li, J. and Wang, D., 2015. Rib suppression in chest radiographs for lung nodule enhancement. In: *2015 IEEE International Conference on Information and Automation, ICIA 2015 - In conjunction with 2015 IEEE International Conference on Automation and Logistics*. pp.50–55.

Lindner, C. and Cootes, T.F., 2015. Fully automatic cephalometric evaluation using Random Forest regression-voting. *Isbi*, pp.1–8.

Lindner, C., Thiagarajah, S., Wilkinson, J.M., Consortium, T., Wallis, G.A. and Cootes, T.F., 2013. Fully Automatic Segmentation of the Proximal Femur Using Random Forest Regression Voting. *Medical Image Analysis*, 32(8), pp.1462–1472.

Lonstein, J.E., 1999. Congenital spine deformities: Scoliosis, kyphosis, and lordosis. *Orthopedic Clinics of North America*, 30(3), pp.387–405.

Lootus, M., Kadir, T. and Zisserman, A., 2014. Vertebrae detection and labelling in lumbar MR images. *Lecture Notes in Computational Vision and Biomechanics*, 17, pp.219–230.

Löscher, W., 2011. Critical review of current animal models of seizures and epilepsy used in the discovery and development of new antiepileptic drugs. *Seizure : the journal of the British Epilepsy Association*, 20(5), pp.359–68.

Matheron, G., 1975. Random sets and integral geometry. *Computer Vision, Graphics, and Image Processing*, Wiley(3), pp.283–305.

Michalak, K., Devenish, J., Linggard, R., Parker, K. and Emelyanova, I., 2002. Automated Measurement of Brain Dimensions. *Seventh International Conference on Control, Automation, Robotics and Vision*, pp.525–529.

Miloro, M., Borba, A.M., Ribeiro-Junior, O., Naclério-Homem, M.G. and Jungner, M., 2014. Is there consistency in cephalometric landmark identification amongst oral and maxillofacial surgeons? *International Journal of Oral and Maxillofacial Surgery*, 43(4), pp.445–453.

Mirzaalian, H. and Hamarneh, G., 2014. Automatic Globally-Optimal Pictorial Structures with Random Decision Forest Based Likelihoods For Cephalometric X-Ray Landmark Detection. In: *Automatic Cephalometric X-Ray Landmark Detection Challenge 2014, in conjunction with IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*. pp.1–12.

Moura, D.C., Correia, M. V., Barbosa, J.G., Reis, A.M., Laranjeira, M. and Gomes, E., 2007. Automatic vertebra detection in X-ray images. In: *Proceedings of the International Symposium CompIMAGE 2006 - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*. pp.307–312.

MRC Harwell, 2017. *MRC Harwell : An International Centre for Mouse Genetics*. [online] Available at: <http://www.har.mrc.ac.uk/> [Accessed 10 Feb. 2017].

Nanni, L., Lumini, A. and Brahnam, S., 2012. Survey on LBP based texture descriptors for image classification. *Expert Systems with Applications*, 39(3), pp.3634–3641.

Nie, J. and Shen, D., 2013. Automated segmentation of mouse brain images using multi-atlas multi-ROI deformation and label fusion. *Neuroinformatics*, 11(1), pp.35–45.

Nixon, M.S. and Aguado, A.S., 2008. *Feature Extraction and Image Processing*. 3rd ed. *Academic Press*.

Ojala, T., Pietikainen, M. and Harwood, D., 1994. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: *Proceedings of 12th International Conference on Pattern Recognition*. pp.582–585.

Ojala, T., Pietikainen, M. and Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp.971–987.

Al Okashi, O., Du, H., Selway, J., Lelliott, C., Maguire, S., Melvin, D., Langlands, K. and Al-assam, H., 2014. Automatic Ribs Segmentation and Counting From Mouse X-ray Images. In: *Medical Image Understanding and Analysis (MIUA)*. pp.143–148.

Ólafsdóttir, H., Darvann, T.A., Hermann, N. V., Oubel, E., Ersbøll, B.K., Frangi, A.F., Larsen, P., Perlyn, C.A., Morriss-kay, G.M. and Kreiborg, S., 2007. Computational mouse atlases and their application to automatic assessment of craniofacial dysmorphology caused by the

Crouzon mutation Fgfr2C342Y. *Journal of Anatomy*, 211(1), pp.37–52.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62–66.

Ouellet, J. and Odent, T., 2013. *Animal models for scoliosis research: State of the art, current concepts and future perspective applications. European Spine Journal*, .

Pattison, et al., 2000. A Fifth Locus for Primary Autosomal Recessive Microcephaly Maps to Chromosome 1q31. *The American Journal of Human Genetics*, 67, pp.1578–1580.

Peng, Z., Zhong, J., Wee, W. and Lee, J.-H., 2005. Automated Vertebra Detection and Segmentation from the Whole Spine MR Images. In: *Conference proceedings :Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. pp.2527–2530.

Pitas, I. and Venetsanopoulos, A.N., 1992. Order Statistics in Digital Image Processing. *Proceedings of the IEEE*, 80(12), pp.1893–1921.

Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B. and Zuiderveld, K., 1987. ADAPTIVE HISTOGRAM EQUALIZATION AND ITS VARIATIONS. *Computer vision, graphics, and image processing*, 39(3), pp.355–368.

Powers, D.M.W., 2011. Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, [online] 2(1), pp.37–63. Available at: <http://www.bioinfopublication.org/files/articles/2_1_1_JMLT.pdf>.

Rajiah, P., 2015. *Idiopathic Scoliosis Imaging*. [online] Available at: <http://emedicine.medscape.com/article/413157-overview> [Accessed 4 Sep. 2017].

Randen, T. and Husøy, J.H., 1999. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), pp.291–310.

Rangayyan, R.M., 2004. *Biomedical Image Analysis*. 1st ed. CRC Press.

Redwine, J.M., Kosofsky, B., Jacobs, R.E., Games, D., Reilly, J.F., Morrison, J.H., Young, W.G. and Bloom, F.E., 2003. Dentate gyrus volume is reduced before onset of plaque formation in PDAPP mice: a magnetic resonance microscopy and stereologic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3),

pp.1381–1386.

Reeves, R.H., Irving, N.G., Moran, T.H., Wohn, A., Kitt, C., Sisodia, S.S., Schmidt, C., Bronson, R.T. and Davisson, M.T., 1995. A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genetics*, 11(2), pp.177–184.

Reyes-Aldasoro, C.C. and Bhalerao, A., 2006. The Bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognition*, 39(5), pp.812–826.

Ribeiro, E.A., Nogueira-Barbosa, M.H., Rangayyan, R.M. and Azevedo-Marques, P.M., 2010. Detection of vertebral plateaus in lateral lumbar spinal X-ray images with Gabor filters. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*. pp.4052–4055.

Rimol, L.M., Agartz, I., Djurovic, S., Brown, A.A., Roddey, J.C., Kahler, A.K., Mattingsdal, M., Athanasiu, L., Joyner, A.H., Schork, N.J., Halgren, E., Sundet, K., Melle, I., Dale, A.M. and Andreassen, O.A., 2010. Sex-dependent association of common variants of microcephaly genes with brain structure. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1), pp.384–388.

Roby, K.F., Taylor, C.C., Sweetwood, J.P., Cheng, Y., Pace, J.L., Tawfik, O., Persons, D.L., Smith, P.G. and Terranova, P.F., 2000. Development of a syngeneic mouse model for events related to ovarian cancer. *Carcinogenesis*, 21(4), pp.585–591.

Rogers, S. and Girolami, M., 2011. *A first course in machine learning*. 2nd ed. *Finance*. CRC press.

Samuvel, B., Thomas, V., Mini, M.G. and Renjith Kumar, J., 2012. A mask based segmentation algorithm for automatic measurement of Cobb angle from scoliosis x-ray image. In: *Proceedings - 2012 International Conference on Advances in Computing and Communications, ICACC 2012*. pp.110–113.

Sanger, 2013. *Wellcome Sanger Institute*. [online] Available at: <http://www.sanger.ac.uk/> [Accessed 9 Jan. 2018].

Sardjono, T.A., Wilkinson, M.H.F., Veldhuizen, A.G., Van Ooijen, P.M.A., Purnama, K.E. and Verkerke, G.J., 2013. Automatic cobb angle determination from radiographic images. *Spine*, 38(20), pp.E1256–E1262.

Sharief, A.A., Badea, A., Dale, A.M. and Johnson, G.A., 2008. Automated segmentation of the actively stained mouse brain using multi-spectral MR microscopy. *NeuroImage*, 39(1),

pp.136–145.

Simon, M.M., Greenaway, S., White, J.K., Fuchs, H., Gailus-Durner, V., Wells, S., Sorg, T., Wong, K., Bedu, E., Cartwright, E.J., Dacquin, R., Djebali, S., Estabel, J., Graw, J., Ingham, N.J., Jackson, I.J., Lengeling, A., Mandillo, S., Marve, J., Meziane, H., Preitner, F., Puk, O., Roux, M., Adams, D.J., Atkins, S., Ayadi, A., Becker, L., Blake, A., Brooker, D., Cater, H., Champy, M.F., Combe, R., Danecek, P., Di Fenza, A., Gates, H., Gerdin, A.K., Golini, E., Hancock, J.M., Hans, W., Hölter, S.M., Hough, T., Jurdic, P., Keane, T.M., Morgan, H., Müller, W., Neff, F., Nicholson, G., Pasche, B., Roberson, L.A., Rozman, J., Sanderson, M., Santos, L., Selloum, M., Shannon, C., Southwel, A., Tocchini-Valentini, G.P., Vancollie, V.E., Westerberg, H., Wurst, W., Zi, M., Yalcin, B., Ramirez-Solis, R., Steel, K.P., Mallon, A.M., De Angelis, M.H., Herault, Y. and Brown, S.D.M., 2013. A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biology*, 14(7), p.R82.

Soille, P., 2004. *Morphological Image Analysis. Morphological Image Analysis Principles and Applications*.

Song, Y., Cai, W., Zhou, Y. and Feng, D.D., 2013. Feature-based image patch approximation for lung tissue classification. *IEEE Transactions on Medical Imaging*, 32(4), pp.797–808.

Spencer, H.T., Gold, M.E. and Hresko, M.T., 2014. Abnormal rib count in scoliosis surgery: impact on the reporting of spinal fusion levels. *Journal of Children's Orthopaedics*, 8(6), pp.497–503.

Staal, J., van Ginneken, B. and Viergever, M.A., 2007. Automatic rib segmentation and labeling in computed tomography scans using a general framework for detection, recognition and segmentation of objects in volumetric data. *Medical Image Analysis*, 11(1), pp.35–46.

Stolojescu-Crisan, C. and Holban, S., 2013. A Comparison of X-Ray Image Segmentation Techniques. *Advances in Electrical and Computer Engineering*, 13(3), pp.85–92.

Stone, M., 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, 36(2), pp.111–147.

Tan, C.A., Topper, S., Ward Melver, C., Stein, J., Reeder, A., Arndt, K. and Das, S., 2014. The first case of CDK5RAP2-related primary microcephaly in a non-consanguineous patient identified by next generation sequencing. *Brain and Development*, 36(4), pp.351–355.

Tang, F.H., Chan, L.W.C., Lau, H.P., Tsui, P.Y. and Cheung, C.W., 2008. Computer-generated index for evaluation of idiopathic scoliosis in digital chest images: A comparison with digital

measurement. *Journal of Digital Imaging*, 21(SUPPL. 1).

Tomasi, C. and Manduchi, R., 1998. Bilateral filtering for gray and color images. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. pp.839–846.

Vandaele, R., Maree, R., Jodogne, S. and Geurts, P., 2015. Automatic Cephalometric X-Ray Landmark Detection Challenge 2014: A machine learning tree-based approach. In 14th International Congress for Stereology and Image AnalysisLiège.

Vasamsetti, S., Sardana, V., Kumar, P. and Sardana, H.K., 2017. Feature extraction through Wavelet decomposition for automatic detection of landmarks on Cephalogram Images. *bioRxiv*, 125286.

Vrtovec, T., Pernuš, F. and Likar, B., 2009. A review of methods for quantitative evaluation of spinal curvature. *European Spine Journal*, 18(5), pp.593–607.

Wang, P., Krishnan, S.M., Kugean, C. and Tjoa, M.P., 2001. Classification of endoscopic images based on texture and neural network. In: *Annual Reports of the Research Reactor Institute, Kyoto University*. pp.3691–3695.

Wang, S., Zhu, W. and Liang, Z., 2001. Shape deformation: SVM regression and application to medical image segmentation. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2, pp.209–216.

Wattanasirichaigoon, D., Prasad, C., Schneider, G., Evans, J.A. and Korf, B.R., 2003. Rib defects in patterns of multiple malformations: A retrospective review and phenotypic analysis of 47 cases. *American Journal of Medical Genetics*, 122A(1), pp.63–69.

White, J.K., Gerdin, A., Karp, N. a, Ryder, E., Buljan, M., Bussell, J.N., Salisbury, J., Clare, S., Ingham, N.J., Podrini, C., Houghton, R., Estabel, J., Bottomley, J.R., Melvin, D.G., Sunter, D., Adams, N.C., Tannahill, D., Logan, D.W., Macarthur, D.G., Flint, J., Mahajan, V.B., Tsang, S.H., Smyth, I., Watt, F.M., Skarnes, W.C., Dougan, G., Adams, D.J., Ramirez-Solis, R., Bradley, A. and Steel, K.P., 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell*, 154(2), pp.452–64.

Willekens, I., Buls, N., Lahoutte, T., Baeyens, L., Vanhove, C., Caveliers, V., Deklerck, R., Bossuyt, A. and De Mey, J., 2010. Evaluation of the radiation dose in micro-CT with optimization of the scan protocol. *Contrast Media and Molecular Imaging*, 5(4), pp.201–207.

Witten, I.H., Frank, E. and Hall, M. a., 2011. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition. *Annals of Physics*, 54(2), p.664.

Wu, H., Bailey, C., Rasoulinejad, P. and Li, S., 2017. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp.127–135.

Zadeh, L. a., 1953. Theory of Filtering. *Journal of the Society for Industrial and Applied Mathematics*, 1(1), pp.35–51.

Zhang, J., Li, H., Lv, L. and Zhang, Y., 2017. Computer-Aided Cobb Measurement Based on Automatic Detection of Vertebral Slopes Using Deep Neural Network. *International Journal of Biomedical Imaging*, 2017.

Zhang, J., Lou, E., Le, L.H., Hill, D.L., Raso, J. V. and Wang, Y., 2009. Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. *Journal of Digital Imaging*, 22(5), pp.463–472.

Zhang, J., Peng, Q., Li, Q., Jahanshad, N., Hou, Z., Jiang, M., Masuda, N., Langbehn, D.R., Miller, M.I., Mori, S., Ross, C.A. and Duan, W., 2010. Longitudinal characterization of brain atrophy of a Huntington's disease mouse model by automated morphological analyses of magnetic resonance images. *NeuroImage*, 49(3), pp.2340–2351.

Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization. In: *Graphics Gems*. pp.474–485.