

**A Machine Learning and Computer Assisted Methodology
for Diagnosing Chronic Lower Back Pain on Lumbar
Spine Magnetic Resonance Images**

By

Ala Sami Al-Kafri

A thesis submitted in partial fulfillment of the requirements of Liverpool John Moores
University for the degree of Doctor of Philosophy

July 2019

Acknowledgments

First and foremost, all praise is due to our God (s.w.t) for giving me the strength and ability to complete this journey throughout my PhD.

I would like to express my sincere gratitude to my supervisor, Dr.Sud Sudirman for the support and advice he offered to me during my PhD which gave me the ability to successfully face any challenges. Dr.Sud gave me all I needed to achieve work of such a high standard. Special thanks to Prof.Abir Hussain and Prof Dhiya Al Jumeily for their continuous advice and support since day one of my PhD journey. I would like to express my thankfulness to my external supervisor Dr. Mohammed Al Jumeily for his inspired guidance and providing me with all the medical advice required for my research.

I would like to express my appreciations to Dr.Wasfi Al Rashdan and Dr.Mohammed Bashtawi (Irbid Speciality Hospital), Dr.Paul Fergus and everyone who supported me in completing this thesis. Thanks to all the staff members in the Computer Science Department at LJMU for the admin and technical advice and support. Thanks to Tricia Waterson, Elizabeth Hoare, Carol Oliver and from the IT department special thanks to Neil Rowe, Paul Cartwright, Warren Anacoura and Ian Fitzpatrick.

Finally, my heartfelt thanks to my parents, brothers, and sisters for their unlimited support and encouragement. I would love to thank my wife, my daughters (Maya and Yara) and my son (Karim) for their patience, tolerance, and support during this long journey in my PhD.

Abstract

Chronic Lower Back Pain (CLBP) is one of the major types of pain that affects many people around the world. It is estimated that 28.1% of US adults suffer from this illness and 2.5 million of the UK population experience this type of pain every day. Most CLBP cases do not happen overnight and it is usually developed from a less serious but acute variant of lower back pain. An acute type of lower back pain can develop into a chronic one if the underlying cause is serious and left untreated. The longer a person is disabled by back pain, the less chance he or she returns to work and the more health care cost he or she will require. It is therefore important to identify the cause of back pains as early as possible in order to improve the chance of patient rehabilitation. The speediness of early diagnosis can depend on many factors including referral time from a general practitioner to the hospital, waiting time for a specialist appointment, time for a Magnetic Resonance Imaging (MRI) scan and time for the analysis result to come out. Currently diagnosing the lower back pain is done by visual observation and analysis of the lumbar spine MRI images by radiologists and clinicians and this process could take up much of their time and effort. This, therefore, rationalizes the need for a new method to increase the efficiency and effectiveness of the imaging diagnostic process.

This thesis details a novel methodology to automatically aid clinicians in performing diagnosis of CLBP on lumbar spine MRI images. The methodology is based on the current accepted medical practice of manual inspection of the MRI scans of the patient's lumbar spine as advised by several practitioners in this field. The main methodology is divided into three sub-methods the first sub-method is disc herniation detection using disc segmentation and centroid distance function. While the second sub-method is lumbar spinal stenosis detection via segmentation of area between anterior and posterior (AAP) Elements. Whereas, the last sub-method is the use of deep learning to perform semantic segmentation to identify regions in the MRI images that

are relevant to the diagnosis process. The method then performs boundary delineation between these regions, identifies key points along the boundaries and measures distances between these points that can be used as an indication to the health of the lumbar spine.

Due to a limitation in the size and suitability of the currently existing open-access lumbar spine dataset necessary to train and test any good classification algorithms, a dataset consisting of 48,345 MRI slices from a complete clinical lumbar MRI study of 515 symptomatic back pain patients from several specialty hospitals around the world has been created. Each MRI study is annotated by expert radiologists with notes regarding the observed characteristics, condition of the lumbar spine, or presence of diseases. The ground-truth dataset containing manually labelled segmented images has also been developed. To complement this ground-truth dataset, a novel method of constructing and evaluating the suitability of ground truth data for lumbar spine MRI image segmentation has been developed.

A subset of the dataset, which includes the data for 101 patients, is used in a set of experiments that have been conducted using a variety of algorithms to conclude with using SegNet as the image segmentation algorithm. The network consists of VGG16 layers pre-trained using a subset of non-medical images from the ImageNet database and fine-tuned using the training portion of the ground-truth dataset. The results of these experiments show the accurate delineation of important boundaries of regions in lumbar spine MRI. The experiments also show very close agreement between the expert radiologists' notes on the condition of a lumbar spine and the conclusion of the system about the lumbar spine in the majority of cases.

Table of Contents

Acknowledgments.....	ii
Abstract.....	iii
Table of Contents.....	v
List of Figures	viii
List of Tables	x
Thesis Acronyms	xi
List of Publications	xiii
Chapter 1 . Introduction.....	1
1.1 Research Problems.....	3
1.2 Research Aims and Objectives	4
1.2.1 Research Aims.....	4
1.2.2 Research Objectives.....	4
1.3 Research Contributions.....	5
1.4 Research Goal	6
1.5 Thesis Structure	7
Chapter 2 Background Information	9
2.1 Lower Back Pain	9
2.2 Lumbar Spine Anatomy.....	11
2.3 Causes of lower back pain.....	14
2.3.1 Spinal Canal Stenosis.....	15
2.3.2 Disc Herniation.....	16
2.4 MRI Imaging	17
2.5 Diagnosis of the Lower Back Pain	18
2.5.1 The current method of diagnosing the spinal canal stenosis and disc herniation.	18
2.6 Chapter Summary	21
Chapter 3 Literature Review	22
3.1 Medical Image Analysis.....	22
3.2 Medical Images in the Computer-Aided Diagnosing System.....	24
3.3 Computer-Aided Diagnosing System for Lumbar Spine Stenosis	26
3.4 Machine Learning.....	28
3.4.1 Artificial neural networks.....	30
3.4.2 Convolutional Neural Network	32
3.4.3 Support Vector Machine	32
3.4.4 Decision Tree.....	36
3.4.5 Boosting	36

3.4.6	K-Nearest Neighbour	38
3.5	Chapter Summary	39
Chapter 4	The Dataset Developed for this Research.....	41
4.1	The Available Open-Access Dataset.....	41
4.2	Lumbar Spine MRI (LSMRI) Dataset.....	44
4.3	Dataset Used in the First Two Experiments.....	49
4.4	Summary	52
Chapter 5	Disc Herniation Detection using Disc Segmentation and Centroid Distance Function.....	53
5.1	Basic Edge Detection Algorithms to detect disc boundaries	56
5.2	Image Labels and Ground Truth.....	57
5.3	Classifier Model Selection.....	60
5.3.1	Classifier Training Result	63
5.3.2	Automatic Segmentation Result	64
5.4	Detecting the Disc Herniation in Segmented Lumbar Spine MR Image using Centroid Distance Function	66
5.5	Chapter Summary	71
Chapter 6	Lumbar Spinal Stenosis Detection via Segmentation of Area between Anterior and Posterior (AAP) Elements.....	72
6.1	MRI Image Cropping	73
6.2	Deep Neural Network Architecture	75
6.3	Training	77
6.4	Discussion and Analysis.....	82
6.5	Two Regions Boundary Delineation through Semantic Segmentation.....	84
6.6	Locating the main points in the delineated boundaries	89
6.7	Chapter Summary	98
Chapter 7	Semantic Segmentation using Convolutional Neural Networks	99
7.1	SegNet Training Setup.....	101
7.1.1	Image Labels and Ground Truth.....	102
7.1.2	Ground Truth Quality Metrics.....	105
7.1.3	Compositing T1- and T2-weighted MRI images	107
7.2	Development of Ground Truth Data for Automatic Lumbar Spine MRI Image Segmentation 109	
7.2.1	Analysis of image labelling results	109
7.2.2	Individual Region Analysis.....	110
7.2.3	Confidence and consistency metrics.....	112
7.3	Boundary Delineation through Semantic Segmentation	114

7.3.1	Segmentation Results and Analysis	114
7.4	Chapter Summary	118
Chapter 8 Conclusion and Future Work.....		120
8.1	Conclusion.....	120
8.2	Research Contributions.....	122
8.3	Summary and Future Research.....	125
References		127

List of Figures

Figure 2.1: Anatomy of the Spine [25]	11
Figure 2.2: Lumbar Spine that contains the vertebrae from L1 to L5 [26].	12
Figure 2.3: A T2-weighted mid-sagittal view of lumbar spine MRI.	13
Figure 2.4: Axial View of the Lumbar Spine.	14
Figure 2.5: The three important boundaries for stenosis detection between AAP and 1) IVD, 2) Facet Joints, and 3) LF to measure a) the anteroposterior diameter of the spinal canal and b) the left and right width of the foramen	16
Figure 2.6: Axial view of (A) non-herniated and (B) herniated Disc [35].	17
Figure 2.7: The selected slice for the disc L3-L4.	19
Figure 2.8: Illustrate a point at the bottom centre of the disc L3-L4.	19
Figure 2.9: Shows the lowest point on the spinal canal for the disc L3-L4.	20
Figure 2.10: Illustrate the line between the lowest point on the spinal canal and the point at the bottom centre for the disc L3-L4.	20
Figure 3.1: illustrations of biological neuron parts.	30
Figure 3.2: Network of artificial neuron or perceptron.	31
Figure 3.3: Convolutional deep neural network layers	32
Figure 3.4: Diabetic and non-diabetic data separated linearly by hyper-plane.	33
Figure 3.5: SVMC linearly separable set of two classes.	34
Figure 3.6: Two and three-dimensional illustration of data.	35
Figure 3.7: K-nearest neighbour algorithm (KNN) example.	39
Figure 4.1: Illustrates an example for two different patients MRI (A) Four slices of images for the level 3 disc. (B) Five slices of images for the level 3 disc.	47
Figure 4.2 Some examples of the discarded MRI scans. Image (a) is T1-weighted and (b) T2-weighted MRI scans of the same part of the lumbar spine of a patient. There is a significant difference in the position of the same organ in both images resulting in a large number of mismatch pixels (purple regions) in the resulting (c) composite image. Image (d) and (e) are example cases where multiple regions are fused together making manual segmentation unreliable. And (f) is an example of a scan that contains unexplained imaging distortion or artefact.	49
Figure 4.3: Example of the radiologist report for one of the patients	51
Figure 5.1: Machine Learning Framework.	54
Figure 5.2: Shows the methodology followed to label the lumbar spine disc using axial view MRI based on pixels coordinate and grey level features.	55
Figure 5.3: The results of Canny edge detection using a combination of two threshold values and two sigma value where a) low threshold – low Gaussian width, b) low threshold – high Gaussian width, c) high threshold – low Gaussian width and d) high threshold and high Gaussian width.	56
Figure 5.4: Dual-view of MRI images showing the mid-sagittal view (left) and the corresponding axial-view slice (right) at the cut line	58
Figure 5.5: The selected MRI Images of the Three Lumbar Spine Discs.	58
Figure 5.6: Discs MRI and the developed Mask for each disc.	59
Figure 5.7: The classifiers for the two experiments and the accuracy for each of them.	64
Figure 5.8: The Original and Segmented Images by Weighted KNN.	65
Figure 5.9: The Original and Segmented Images by SVM Fine Gaussian.	66
Figure 5.10: Axial view of (A) non-herniated and (B) herniated discs.	67
Figure 5.11: Illustration of the radial distances calculated from (A) non-herniated and (B) herniated disc.	67
Figure 5.12: The distance between the centroid and the disc boundary (d) in the angle (θ).	68
Figure 5.13: Plot of the Centroid Distance Function of a non-herniated disc.	69
Figure 5.14: Plot of the Centroid Distance Function of a herniated disc. The red region marked the area of the disc where the herniation occurs.	69
Figure 5.15: Plot of the Centroid Distance Function of a non-herniated disc.	70
Figure 5.16: Plot of the Centroid Distance Function of a herniated disc.	70

Figure 6.1: Methodology used for segmenting the AAP area in the lumbar spine MRI images. _____	73
Figure 6.2: An example of MRI and ground-truth image pair used for training. _____	74
Figure 6.3: The process flow of image segmentation using patch-based pixel classification method. _____	74
Figure 6.4: The three important boundaries for stenosis detection between AAP and 1) IVD, 2) Facet Joints, and 3) LF to measure a) the anteroposterior diameter of the spinal canal and b) the left and right width of the foramen _____	82
Figure 6.5: The result of superimposing the edges of the segmentation result on to the ground truth (right), and their corresponding original MRI images (left). The top images are taken from D3, middle from D4, and bottom from D5. _____	83
Figure 6.6: Visual evidence of overfitting of SegNet in this experiment. The figure shows prediction on (left) a training image and (right) on a test image. The bright pixels outline the edges of the segmented image. _____	84
Figure 6.7: The process flow of SegNet's encoder-decoder architecture. _____	86
Figure 6.8 Example of the reconstructed images using SegNet using full-size MRI images _____	88
Figure 6.9: The four labelled Regions of Interest namely 1) Intervertebral Disc (IVD), 2) Posterior Element (PE), 3) Thecal Sac (TS) and 4) the Area between Anterior and Posterior (AAP) vertebrae elements. Delineation of boundaries between these regions is used to measure A) the anteroposterior diameter of the spinal canal width. _____	89
Figure 6.10: Axial View of the Lumbar Spine. _____	90
Figure 6.11: Sample of the reconstructed image for the AAP for L3-L4, L4-L5, and L5-S1 _____	90
Figure 6.12: Important points locations _____	92
Figure 6.13: The distribution of the differences between the manually computed spinal canal width and the width measured using Fine Gaussian SVM _____	95
Figure 6.14: The distribution of the differences between the manually computed spinal canal width and the width measured using the SegNet _____	96
Figure 6.15: (A) – (F), Illustrate the System Diagnosed Result in % for L3-L4, L4-L5 and L5-S1 Using SegNet and Fine Gaussian SVM. _____	98
Figure 7.1 Shows the methodology followed in the segmentation using the convolutional neural network _____	100
Figure 7.2 The four labelled RoIs namely 1) Intervertebral Disc, 2) Posterior Element, 3) Thecal Sac and 4) the AAP. _____	103
Figure 7.3 Dual-view of MRI images showing the mid-sagittal view (left) and the corresponding axial-view slice (right) at the cut line. _____	104
Figure 7.4: Visualizing vote counts that each pixel in and around each RoI receives as a heat map, where the brighter pixels have more votes than darker pixels. The RoIs are a) IVD, b) PE, c) TS and d) the AAP _____	110
Figure 7.5: The plot of a) Pixel Accuracy and b) Mean Accuracy of the SegNet semantic segmentation results at different training data percentages. _____	115
Figure 7.6: The plot of a) Frequency-Weighted IoU and b) Mean IoU of the SegNet semantic segmentation results at different training data percentages. _____	116
Figure 7.7: The plot of the class accuracy of the semantic segmentation results using a) SegNet-FS and b) SegNet-TL at different training data percentages. _____	117
Figure 7.8: The plot of class IoU of the semantic segmentation results using a) SegNet-FS and b) SegNet-TL at different training data percentages. _____	117

List of Tables

Table 4.1: Shows all the available dataset in the SpineWeb web page with its details.	44
Table 4.2: MRI Scanning Parameters.	45
Table 4.3: Data Categorization for 101 Patients.	50
Table 4.4: Sample of the provided data	51
Table 5.1: Shows sample of the exported data including the pixel position, Intensity, and Disc which represented by (1) or Non-Disc which represented by (0).	60
Table 5.2 The Models trained in the two experiments	61
Table 5.3: The first experiment result based on the evaluation of 6 models. The Models trained using one patient data.	62
Table 5.4: The second experiment result based on the evaluation of 6 models. The Models trained using seven patients' data.	62
Table 6.1: The Architecture of PALMSNet	76
Table 6.2 Performance using All Intervertebral Discs	79
Table 6.3 Performance when trained and tested using D3 only	80
Table 6.4 Performance when trained and tested using D4 only	80
Table 6.5 Performance when trained and tested using D5 only	81
Table 6.6 PALMSNet Performance when tested on D5	82
Table 6.7: The classifier's performance using all intervertebral discs.	91
Table 6.8: Sample of the calculated distances using the Fine Gaussian SVM to reconstruct the images are compared with the manually calculated distances for the three lumbar spine disc L3-L4, L4-L5, and L5-S1 respectively	93
Table 6.9: sample of the calculated distances using SegNet to reconstruct the images are compared with the manually calculated distances for the three lumbar spine disc L3-L4, L4-L5, and L5-S1 respectively	94
Table 7.1 Intersection-over-union values of different vote-threshold groups.	109
Table 7.2: Confidence and consistency values of the resulting ground truth data.	113
Table 7.3: The performance of the best semantic segmentation network (SegNet-TL trained using 80% PTD)	118
Table 8.1: The research contributions and how it has been achieved.	124

Thesis Acronyms

AAP	Area between Anterior and Posterior
a_p	Pixel Accuracy
a_m	Mean Accuracy
CAD	Computer-Aided Diagnosis
CDF	Centroid Distance Function
CLBP	Chronic Lower Back Pain
CNN	Convolutional neural network
CSI	Challenges on Spine Imaging
CT	Computed Tomography
DNN	Deep Neural Network
GBDTs	Gradient-Boosted Decision Trees
GBM	Gradient Boosting
iou_c	The Intersection-Over-Union
IVD	Intervertebral Discs
IVUS	Intravascular Ultrasound
KDD	Knowledge discovery in databases
KNN	K-Nearest Neighbour Algorithm
LSMRI	Lumbar Spine MRI
LSS	Lumbar Spinal Stenosis

MA	Media Adventitia
MFAAS	Multi-Feature and Adaptive Spectral Segmentation
MICCAI	Medical Image Computing and Computer-Assisted Intervention
ML	Machine learning
MLP	Multilayer Perceptron
MRI	Magnetic Resonance Image
MRM	Magnetic Resonance Myelography
PALMSNet	Patch-based Axial-view Lumbar-spine MRI
PE	Posterior Element
PTD	Percentage of Training Data
RBF	Radial Basis Kernel Function
ReLU	Rectified Linear Unit
RoIs	Regions of Interest
SMAD	Symmetric Mean Absolute Surface Distance
SOM	Self Organised Map
SVM	Support vector machines
TS	Thecal Sac
UK	United Kingdom

List of Publications

- [1] Al-Kafri, Ala S., Sud Sudirman, Abir Hussain, Dhiya Al-Jumeily, Friska Natalia, Hira Meidia, Nunik Afriliana, Wasfi Al-Rashdan, Mohammad Bashtawi, and Mohammed Al-Jumaily. "Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection through Semantic Segmentation using Deep Neural Networks." *IEEE Access* 7 (2019): 43487-43501.
- [2] Al Kafri, Ala S., Sud Sudirman, Abir J. Hussain, Dhiya Al-Jumeily, Paul Fergus, Friska Natalia, Hira Meidia et al. "Segmentation of lumbar spine MRI images for stenosis detection using patch-based pixel classification neural network." In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1-8. IEEE, 2018.
- [3] Natalia, Friska, Hira Meidia, Nunik Afriliana, Ala S. Al-Kafri, Sud Sudirman, Andrew Simpson, Ali Sophian, Mohammed Al-Jumaily, Wasfi Al-Rashdan, and Mohammad Bashtawi. "Development of ground truth data for automatic lumbar spine MRI image segmentation." In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 1449-1454. IEEE, 2018.
- [4] Khalaf, Mohammed, Abir Jaafar Hussain, Dhiya Al-Jumeily, Thar Baker, Robert Keight, Paulo Lisboa, Paul Fergus, and Ala S. Al Kafri. "A Data Science Methodology Based on Machine Learning Algorithms for Flood Severity Prediction." In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1-8. IEEE, 2018.
- [5] A. S. Al Kafri, S. Sudirman, A. J. Hussain, P. Fergus, D. Al-Jumeily, H. Al Smadi, M. Khalaf, M. Al-Jumaily, W. Al-Rashdan, and M. Bashtawi, "Detecting the Disc Herniation in Segmented Lumbar Spine MR Image Using Centroid Distance Function," *2017 10th Int. Conf. Dev. eSystems Eng.*, pp. 9–13, 2017.
- [6] A. S. Al Kafri, S. Sudirman, A. J. Hussain, P. Fergus, D. Al-jumeily, H. Al Smadi, M. Khalaf, M. Al-jumaily, W. Al-rashdan, M. Bashtawi, and J. Mustafina, "Lumbar Spine Discs Labeling Using Axial View MRI Based on the Pixels Coordinate and Gray Level Features," vol. 10363, pp. 107–116, 2017.
- [7] M. Khalaf, A. J. Hussain, D. Al-jumeily, R. Keight, A. S. Al Kafri, and I. O. Idowu, "A Performance Evaluation of Systematic Analysis for Combining Multi-class Models for Sickle Cell Disorder Data Sets," vol. 10362, pp. 115–121, 2017.
- [8] A. S. Al Kafri, S. Sudirman, A. J. Hussain, P. Fergus, D. Al-Jumeily, M. Al-Jumaily, and H. Al-Askar, "A Framework on a Computer Assisted and Systematic Methodology for Detection of Chronic Lower Back Pain Using Artificial Intelligence and Computer Graphics Technologies," in *International Conference on Intelligent Computing*, 2016, pp. 843–854.
- [9] Natalia, Friska, Hira Meidia, Nunik Afriliana, Ala Al-Kafri, and Sud Sudirman. "Methodology to Determine Important-Points Location for Automated Lumbar Spine Stenosis Diagnosis Procedure." In *Proceedings of the 2019 International Conference on Intelligent Medicine and Health*, pp. 53-57. ACM, 2019.

Chapter 1 . Introduction

Lower back is pain considered as the second most common illness after the common cold. More than half of the world population are affected by lower back pain once in their lives [1]. In the UK, the figure is higher where about sixty to eighty percent of its population will suffer from back pain once in their lives [2]. As a result back pain costs the UK government £1.3 million per day [3]. One of the main techniques for diagnosing the cause of lower back pain is through MRI examination by a radiologist. It has been reported that the number of MRI examination in 2014 has increased in the UK by 11.3% to reach 2.61 million compared with the number of imaging tests in 2013 [4]. However, the number of radiologists in the UK is insufficient for the clinical demand made by the radiology services as there is an increase of 26% for the MRI requests in comparison to only 3% increase in the consultant radiology workforce [5]. The gap between the increasing rate of the radiology services compared to the number of radiologists is very wide which justifies the need for automating the diagnosing process which usually includes two steps that start with labelling the intervertebral discs area then diagnosing the disc abnormality. Lumbar spine MRI images can also be used to diagnose different kinds of diseases such as Lumbar Spinal Stenosis (LSS) which is one of the leading causes of chronic lower back pain. It is a narrowing of the lumbar spinal canal caused by inflammation of bone or soft tissues, which in turn produces pressure on spinal nerve roots. Patients will experience symptoms ranging from radicular pain to atypical leg pain to neurogenic claudication [6]. Chronic lower back pain is suffered by millions of people around the world and MRI is one of the main techniques used for diagnosing patients with back pain. The lumbar spine MRI can be taken of the patients in different positions. However, the most common position is while the patient is in the supine position where that can affect the load and pressure on the disc. Subsequently, there have been a number of studies that suggest

improvements in the way the scans are carried out which include taking the scans while the patient is in an upright position [7] and using a Lumbar Pillow [8]. Other studies propose to use different radiology techniques such as Radiographic Myelography [9] to detect the most difficult cases of LSS. While we acknowledge that these methods can serve as better alternatives to using supine MRIs in some cases, we also argue that supine MRIs remain, in the majority of cases, the most practical way to study the lumbar spine. Our argument lies on a simple fact that the equipment to carry out supine MRI is much more ubiquitous in hospitals than the others. Furthermore, scans taken in a standing or sitting up position are most likely affected by patients' movement due to discomfort and can result in bad quality images. Because our study relies on gathering and analysing a large amount of data, we, therefore, focus on using supine MRI scans rather than other types of scans. For the purpose of brevity, we refer to them simply as MRI in the remainder of the thesis.

In this research, a method is proposed to aid clinicians in performing lumbar spinal stenosis detection through semantic segmentation and delineation of important boundaries in axial-view MRI images. We will first provide a concise review of related methodologies in the literature before presenting the rationale of our proposed method. There is a wide range of algorithms for computer-aided medical diagnosis depending on the type of disease they are targeting. These algorithms are often specific to a certain type of medical images, such as CT scan, X-Ray, intravascular ultrasound, or MRI. This is the case because certain types of images can capture features of certain diseases better than others. For example, the detection of lumen and media adventitia (MA) borders is the key procedure to determine anomalies inside coronary arteries. Production, costs, and insurance are usually matters that involve economics [10] [11]. To achieve the purpose of this research, utilisation has been made of the current advances in image analysis and machine-learning technologies to develop computer-assisted diagnostic software. Clinical studies have shown that there are strong correlations between clinical

outcome and the morphological characteristics of the patient's lumbar spine as captured in the MRI images [12]. Using a combination of different image analysis and machine learning techniques, computers can analyse the images and identify regions of interests for clinicians to focus on. The techniques could also be used to provide tools for specialists to help them come to the correct diagnosis quickly. As with many other medical imaging processes, the aforementioned approach will apply image segmentation as one of the earliest yet important steps. To ensure the validity of this step, the process must be grounded in sound medical evidence and reasoning as well as proven methodology.

1.1 Research Problems

We deduce from analysing the different relevant algorithms in the literature to date that boundary delineation through image segmentation is a very popular yet effective method to medical image analysis. Nonetheless, we have identified three main problems in the current state-of-the-art research in this field.

1. The first one is the size of the freely available dataset. The most comprehensive database of lumbar-spine-related medical images is hosted by SpineWeb [18], however, it only contains a relatively small sample size and incomplete datasets taken from between 8 and 125 patients. To address this problem, we work together with one specialty hospital in Jordan and several physicians and radiologists around the world to gather a significant number of relevant MRI scans complete with medical annotation to develop our dataset. We made this dataset freely available to the research community.
2. The second problem is the lack of ground truth data and the means to assess its quality. Since the ground truth data is tied to the dataset, the argument of needing a suitably large ground truth dataset also applies. Furthermore, because the task of developing ground truth data by manually labelling the MRI images is a laborious one, it becomes prone to errors. In other words, the data can be inaccurate and inconsistent. We address

the problem of how to measure accuracy and variability in ground truth data by developing two new metrics which are derived from the existing machine learning metrics.

3. Despite the rapid advances in machine learning techniques, to the best of our knowledge, there is only a limited study on their application to lumbar spine image segmentation. In this research, we show how the new advances machine learning methods can be used to perform the segmentation for the important region of lumbar spine MRI which can subsequently be used for disc herniation and lumbar spinal stenosis detection.

1.2 Research Aims and Objectives

1.2.1 Research Aims

The aim of this research is to develop a computer-assisted and systematic methodology for detection and prediction of potential sources of chronic lower back pain.

1.2.2 Research Objectives

To achieve the above-stated aim, the following research objectives are set:

1. To produce a comprehensive literature review that catalogues previous achievements in the related area of research including image processing, computer vision, artificial intelligence, computer graphics, and physics simulation that would provide valuable knowledge, validation criteria and information of previous work relating to this research field.
2. To understand the reasons and causes of CLBP and how this disease affects the lumbar spine. In addition, a clear understanding of the process carried out by the radiologist and orthopaedist when diagnosing CLBP. This can be achieved by interviewing several

radiologists and orthopaedists and understand their ways of analysing related medical data.

3. To gather a sufficient number of data that is required for this research which includes the lumbar spine MRI images and related radiologist reports.
4. To define methods to do the following: A) Select the MRI image with the best view for the lumbar spine disc that can help in diagnosing and detecting the CLBP. B) Separate the different important regions in the lumbar spine MRI image to help in help in diagnosing patients with chronic lower back pain. C) Locate the important points in the lumbar spine MRI image to determine and measure the relevant features that can be used to detect the cause of the CLBP.
5. To implement a working prototype of the proposed methodology and evaluate its performance. The evaluation will be done by comparing its results with the reports provided by the radiologist for the MRI images used in this research

1.3 Research Contributions

1. This research presented a method to aid clinicians in performing CLBP detection through three sub-methods: the first one is Disc Herniation Detection using Disc Segmentation and Centroid Distance Function, the second sub-method is Lumbar Spinal Stenosis Detection via Segmentation of Area between Anterior and Posterior (AAP) Elements, the third sub-method is Semantic Segmentation using Convolutional Neural Networks. The mentioned sub-methods will be used to delineate the important boundaries in MRI images using different types of machine learning algorithms. The delineated images will be used to allocate specific points which have been used to calculate the distance for the spinal canal width as an indicator for the spinal canal stenosis.

2. Due to the limitation in size and suitability of the currently existing open-access lumbar spine dataset, there are a limited number of images in all the available datasets which range from 30 to 242 images and the majority of them are CT images. Moreover, the currently available datasets are not supported by radiologist reports to help the researchers in analysing those images. In addition, the process for receiving the dataset is very complicated. For all these reasons, it has been decided to develop our own dataset. Our dataset contains clinical lumbar spine MRI study of 515 patients with symptomatic back pains. From this dataset, a ground truth label image dataset is developed. It can be used to train and test an image segmentation model.
3. The development of two novel metrics to assess the accuracy and variability of a ground truth dataset, namely the confidence and consistency metrics, respectively. These metrics are derived from the widely used intersection-over-union metric to measure the accuracy of image segmentation algorithms.

1.4 Research Goal

The main goal of this research is to provide all researchers “who aim to aid the clinicians in diagnosing patients using different types of images” the base needed to start their research by providing them with the up to date techniques and methods used in this area. The work done in this research can help in developing methodologies to diagnose patients with chronic lower back pain and many other types of diseases like detecting lung cancer or brain tumor. Moreover, the data used in this research is now freely available for all researchers to speed up the research process and benefit in diagnosing other diseases.

1.5 Thesis Structure

The thesis is divided into 8 chapters, each part covering a specific area of the research work.

The remainder of this thesis is organised as follows:

- **Background Information (Chapter 2):** background on lower back pain and the economic cost of this disease has been provided. In addition, the spine anatomy including a detailed explanation about the lumbar spine and MRI have been also discussed.
- **Literature review (Chapter 3):** a literature review on medical image analysis and methods that currently used medical images. The different types of machine learning have been also discussed.
- **Dataset (Chapter 4):** the dataset has been introduced in which all our experiments are performed in the following two chapters. We introduce parts of our image analysis and explain them in more detail, along with explaining the publicly available dataset.
- **Disc Herniation Detection using Disc Segmentation and Centroid Distance Function (Chapter 5):** This chapter discusses in detail a method for detecting the disc herniation using the centroid distance function that uses the disc boundary from the segmented MRI images. This chapter will discuss the selected classifiers and their training results in addition to the experiments' results.
- **Lumbar Spinal Stenosis Detection via Segmentation of Area between Anterior and Posterior (AAP) Elements (Chapter 6):** A method for presenting an automated delineation for our area of interest (AAP) that are used to help in diagnosing patients with spinal canal stenosis is explained. Furthermore, we have discussed an automatic calculation of the spinal canal width. This chapter also includes the technique used for cropping the images and the architecture for the deep neural network.

- **Semantic Segmentation using Convolutional Neural Networks (Chapter 7):** The Convolutional neural network is used to segment four important regions in the lumbar spine MRI image which includes the area of interest (AAP) and three other areas which are the intervertebral disc, posterior element and thecal sac which can be used for diagnosing more diseases. The development of ground Truth data for automatic lumbar Spine MRI Image segmentation will be also discussed in detail in this chapter.
- **Conclusion and Future Work (Chapter 8):** we conclude the thesis, listing our aims, objectives, contributions and discussing potential future work.

Chapter 2 Background Information

Back pain is one of the major musculoskeletal pain problems which affects many people and it is considered as one of the main causes of disability all over the world [13]. The Pain Community Centre [14] indicated that in the United Kingdom (UK), 2.5 million people have back pain every day of the year. The survey also found that back problems are the leading cause of disability with nearly 119 million days per year lost. The survey also found that one in eight unemployed people give back pain as the reason for unemployment. Statistically, an individual who has been off sick with back pain for a month has a 20% chance of still being off work a year later [15]. The percentage of people who return to see their general practitioner (GP) with back pain within 3 months is more than 29% [16].

2.1 Lower Back Pain

According to which part of the spine the pain originates from, back pain can be classified into two categories, upper and lower. Upper back pain is not as common as lower back pain. However, although it is not a very common spinal disorder, upper back pain can cause significant discomfort and pain when it occurs. The most common causes of upper back pain are muscular irritation and joint dysfunction. While most patients will have a resolution of their back pain with simple measures such as using simple analgesia and exercise, a small proportion of them develop a more chronic condition, referred to as Chronic Lower Back Pain or (CLBP) [2][17]. CLBP is a debilitating illness that affects the health, social life, and employment of millions of people around the world. In the UK, the cost of treating patients with CLBP is estimated to be around £500 million annually to the National Health Service (NHS) [18]. This is on top of other economic costs resulting from the loss of productivity and other informal care which is estimated to reach around £10,668 million [19].

It is important to emphasise that CLBP is a chronic illness. While millions of people experience some sort of back pain at one point in their lives, only a small percentage of them experience what is termed acute lower back pain. This form of back pain is often non-specific and may not necessarily be caused by a serious condition. However, a small percentage of people with acute lower back pain can develop CLBP if the underlying cause is serious and left untreated. A study [15] discovered that the longer a person is disabled by back pain, the less chance that he or she returns to work and the more health care cost he or she will require. It is therefore important for early identification of back pain cause to happen, in order to improve the chance of patient rehabilitation.

The speediness of early diagnosis can depend on many factors including referral time from GP to the hospital, waiting time for a specialist appointment, time for an MRI scan and time for the analysis result to come out. Due to the heavy demand for radiologists and clinicians, the whole process could take weeks to complete.

As a short case study, NHS England currently put a target for all imaging and radio diagnostics to be within 6 to 13 weeks [20]. However, specialist shortages have caused many patients to wait longer. A report by NHS England [20] reveals that there is a significant number of cases where diagnostic wait time can be extended to more than 13 weeks. This problem is expected to get worse since the number of scans is historically on an uptrend. A report [21] by the Office of National Statistics shows that there is a 12.3% average annual growth in demand for MRI scans since 1995. This rationalises the need for a new method to increase the efficiency and effectiveness of the imaging diagnostic process.

To have more understanding about this disease detailed explanations and descriptions of the lumbar spine anatomy and the MRI will be discussed in the next sections.

2.2 Lumbar Spine Anatomy

The lumbar spine is the lower back area in the spinal column which contains five vertebrae labelled L1 to L5 [22][23]. Figure 2.1 shows the diagram of the relative position of each part of the spinal column and Figure 2.2 describes the lumbar spine and its parts which are the area affected by Chronic Lower Back Pain (CLBP) [24].

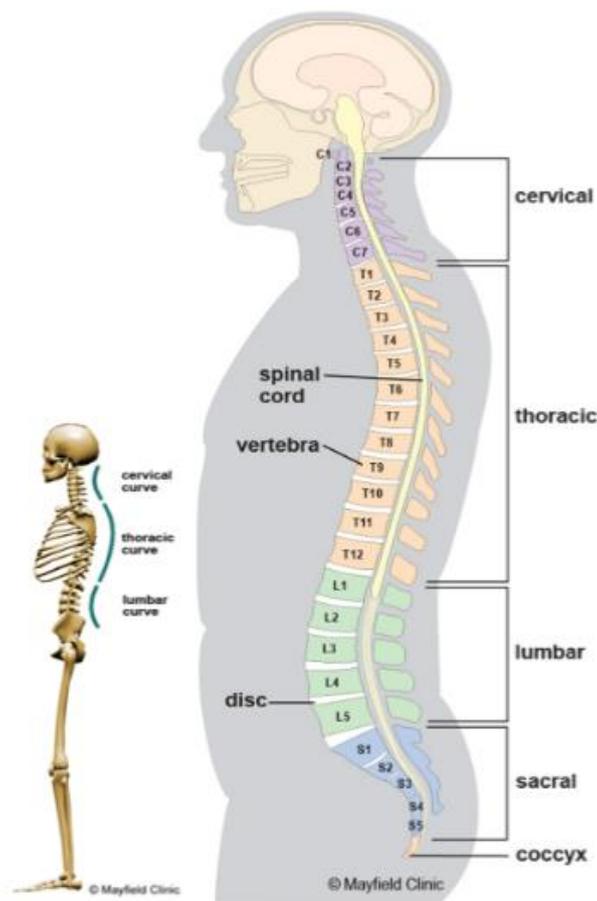


Figure 2.1: Anatomy of the Spine [25]

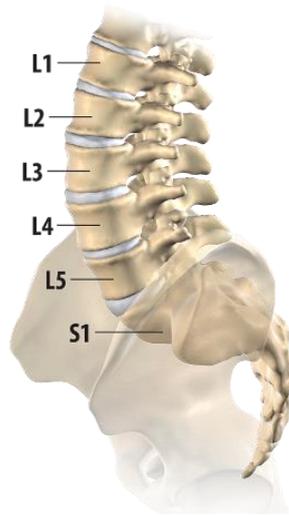


Figure 2.2: Lumbar Spine that contains the vertebrae from L1 to L5 [26].

The lumbar spine is made up of five lumbar vertebrae and each vertebra consists of an anterior (body) arch and posterior (rear) arch. In a T2-weighted mid-sagittal view of the lumbar spine, as shown in Figure 2.3, the two arches are separated by a long white opening. The part of this opening that is visible in this mid-sagittal cut is the Thecal Sac (TS) which contains Cerebrospinal Fluid (CSF), the same type of fluid that resides inside the brain. The back of the opening, which borders with the anterior of the posterior arch, is covered with Ligamentum Flavum (LF). The figure shows the five vertebrae of the lumbar spine labelled L1 to L5. Adjacent vertebrae are separated by an Intervertebral Disc (IVD) which is denoted by the two vertebrae that sandwich the disc, e.g., L1-L2, L2-L3, L3-L4, L4-L5 and L5-S1, the latter being the large triangular-shaped bone at the bottom of the spine, called the sacrum.

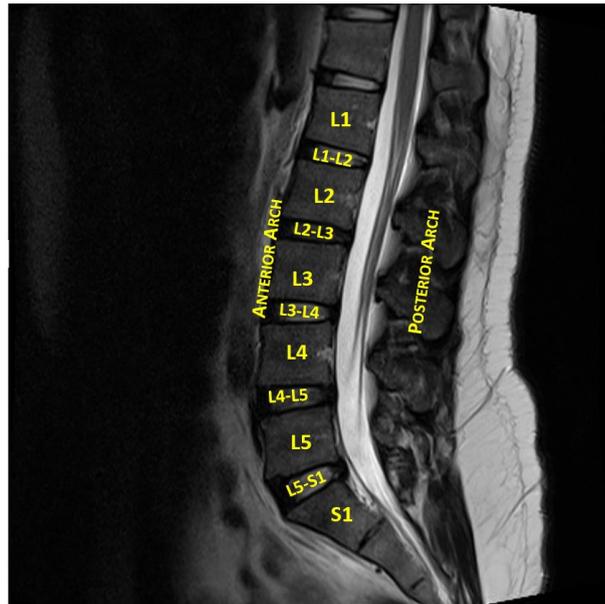


Figure 2.3: A T2-weighted mid-sagittal view of lumbar spine MRI.

The sagittal view of lumbar spine MRI is the easiest to understand and traditionally is considered the best view that can be used to detect certain types of pain-inducing conditions. This can be seen from the methods used by some of the earliest computer-assisted diagnosis of disc herniation [1], [27], [28]. On the other hand, the axial view of MRI is more difficult to read because it has much more information about the lumbar spine at each vertebrae segment. As a result, in this view, the clinicians could identify the exact location of any tissue inflammations, facet cysts, intervertebral disc herniation, stenosis, or fractures. This view, as illustrated in Figure 2.4, can show more information on the various tissues surrounding the vertebra, the IVD and the posterior element (PE) of the vertebral body. Also in this view, we can see clearly the area between the anterior and posterior element of the vertebral body which contains the Thecal Sac and the nerve roots in both lateral recesses. This area extends from the cervical spine down to the lumbar spine. For the lack of a suitable medical terminology, in this thesis, we will refer to this area as AAP, which is short for Area between Anterior and Posterior vertebrae elements.

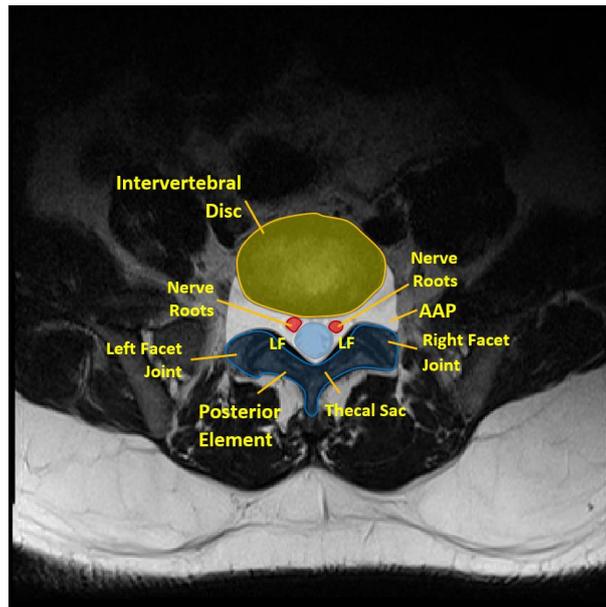


Figure 2.4: Axial View of the Lumbar Spine.

2.3 Causes of lower back pain

When a patient experiences lower back pain, there may be many possible causes for it, including [29], [30]:

- An irritation to either the large nerve roots in the lower back that go to the legs or the smaller nerves supplying the lower back.
- Straining of the large paired lower back muscles.
- Damage to the bones, ligaments or joints in the lumbar spine.
- Degeneration or injury of one or more of the intervertebral discs.
- Lumbar spinal stenosis.
- Disc Herniation.

Lumbar spine stenosis and disc herniation are widespread medical problems as they can occur when carrying heavy items or having incorrect movements like twisting the lower back and

many other reasons which cause heavy stress on the disc [31], [32]. In addition, the lumbar spinal stenosis and disc herniation affect a wide range of ages mostly between 30 and 60 years old which means a huge percentage of the population may be affected by the disc herniation in their life. That was the main reason for choosing these two diseases in our research, in this section, two of these reasons will be discussed starting by the spinal canal stenosis followed by the disc herniation.

2.3.1 Spinal Canal Stenosis

Lumbar spinal stenosis is a narrowing of the spinal canal caused by inflammation of bone or soft tissues, which in turn produces pressure on spinal nerve roots. This pressure could result in Chronic Lower Back Pain (CLBP) with symptoms ranging from radicular pain to atypical leg pain to neurogenic claudication [19]. The area between the anterior arch and the posterior arch of the vertebral body, that we discussed previously, extends from the cervical spine down to the lumbar spine. For the lack of a better word, in this research, we will refer to this area as AAP, which is short for Area between Anterior and Posterior. The width of AAP varies depending on the location in the spine where the measurement is taken. Furthermore, there have been a number of studies [20]–[22], which record the measurements of spinal canal widths in a population of different countries and the results suggest that width of AAP is also affected by the patient's ethnicity.

Lumbar spine stenosis can manifest as the narrowing of any parts of the AAP and depending on where it occurs can be classified as either central stenosis or lateral stenosis. This phenomenon is associated as one of the main causes of CLBP because of both the spinal canal and the lateral part of the AAP, called foramen, house many nerve roots from different parts of the spine to other parts of the body. An abnormal compression of either of them would exert pressure on these roots and creates a sensation of pain.

The stenosis could occur in any part of the AAP and could be caused by different types of defect such as posterior or posterolateral disc herniation, osteoarthritic thickening of the posterolateral vertebral body, or hypertrophy of LF. In all of these cases, clinicians will measure three distances in the AAP, namely the anteroposterior diameter of the spinal canal and the left and right width of the foramen. This process starts by manual delineation of the boundaries between AAP and the IVD, between AAP and the left and right facet joints, and between the AAP and LF. The three distances and the three boundaries are illustrated in Figure 2.5.

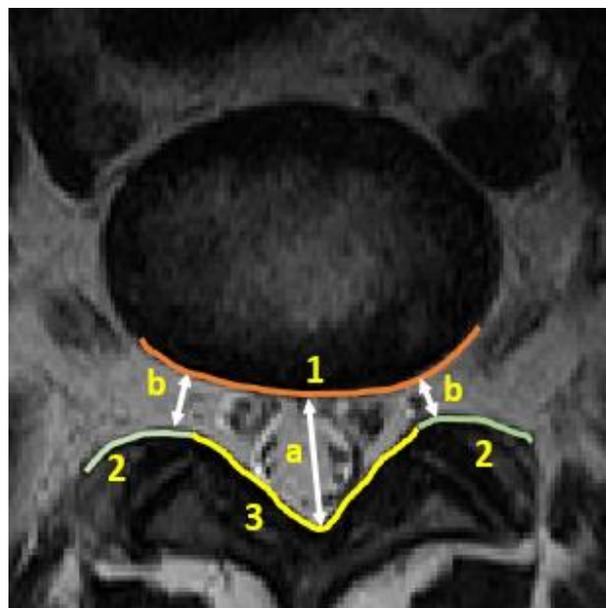
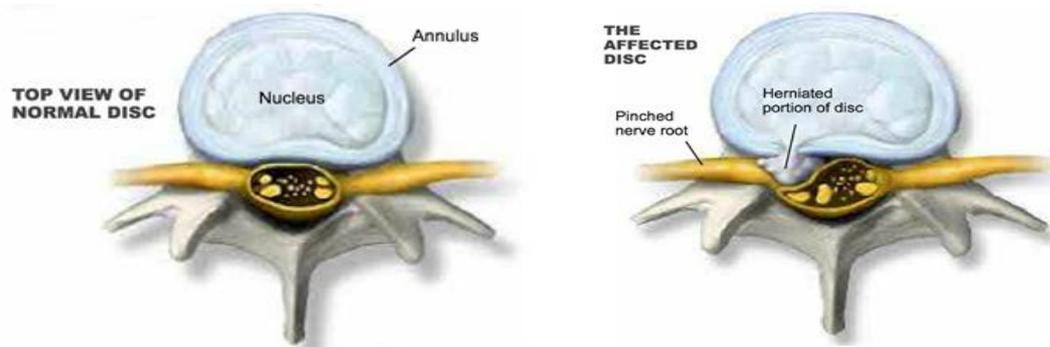


Figure 2.5: The three important boundaries for stenosis detection between AAP and 1) IVD, 2) Facet Joints, and 3) LF to measure a) the anteroposterior diameter of the spinal canal and b) the left and right width of the foramen

2.3.2 Disc Herniation

Disc herniation is one of the main possible causes of lower back pain. It affects the spinal column by bulging out the nucleus pulposus to the outer ring through the annulus fibrosus [33] causing the pain by applying pressure to the nerve root during most of the body's movements [34]. Figure 2.6 (A) illustrates a non-herniated disc seen from an axial view

whereas Figure 2.6 (B) illustrates a herniated disc by highlighting the herniated area which produces pressure to the nerve root.



A. Axial View Model Showing a non-herniated Disc.

B. Axial View Model Showing the Herniated Area on the Disc.

Figure 2.6: Axial view of (A) non-herniated and (B) herniated Disc [35].

2.4 MRI Imaging

Magnetic Resonance Image (MRI) is mainly used to diagnose patients with CLBP or those with symptoms consistent with radiculopathy or spinal column stenosis [36]. MRI is the preferred method of medical scans for detecting the causes of back pains. MRI images can be used to visualise lumbar spine, slice by slice, in three view-planes namely sagittal (side), axial (top-down) and coronal (frontal) – typically only the first two are used in lumbar spine MRI.

Sagittal view of lumbar spine MRI is the easiest to understand and traditionally is considered the best view that can be used to detect certain types of pain-including conditions. This can be seen from the methods used by some of the earliest computer assisted diagnosis of disc herniation [28], [37], [38]. On the other hand, the axial view of MRI is much more difficult to read because it has much more information about the lumbar spine at each vertebrae segment.

As a result, in this view, the clinicians could pinpoint the exact location of any tissue inflammations, facet cysts, disc herniation, stenosis, and fractures.

Based on the timing of radiofrequency pulse sequences used, MRI images can be one of two types namely T1- and T2-weighted MRI [39]. The same tissue could have markedly different intensity levels when imaged in each type. One example is Cerebrospinal Fluid (CSF), a clear and colourless body fluid that can be found inside our brain and spinal cord. Identification of CSF tissue in MRI images has significant relevance to our case since it surrounds the many nerve roots inside the spinal cord. Further and more detailed information on MRI and its uses as medical imaging technology are explained in chapter 4.

2.5 Diagnosis of the Lower Back Pain

The diagnosis of back pains is usually carried out by a radiologist or an orthopedist through manual inspection and observation of the patient's spine MR images, because of the nature of different causes of back pains and the way each of them is diagnosed and treated.

2.5.1 The current method of diagnosing the spinal canal stenosis and disc herniation.

Currently diagnosing the lower back pain is done by visual observation and analysis of the lumbar spine MR images and this process could take up much of a radiologist's time and effort. To understand this process, several interviews have been done with experienced clinicians in this field to summarise these steps. The current visual observation approach by the radiologist for diagnosing the spinal canal stenosis and disc herniation is explained in the following steps.

1. Selecting the best axial view for the appropriate disc. Normally, each disc has a range of slices between 3 and 5 images and the best axial view slice is the image that shows all the required details in a way that gives the radiologist the ability to take a diagnosing

decision in the current disc for a particular disease. An example of the selected image for the disc L3-L4 is shown in Figure 2.7.

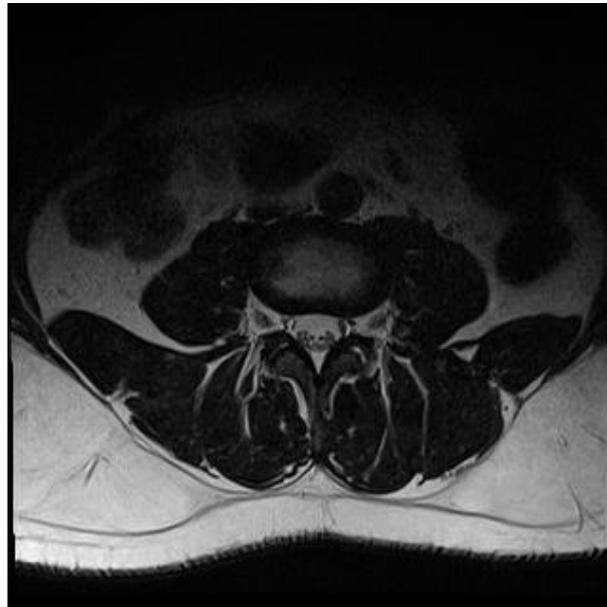


Figure 2.7: The selected slice for the disc L3-L4.

2. The radiologist or the orthopaedist then will find a point at the bottom centre of the disc as in Figure 2.8.

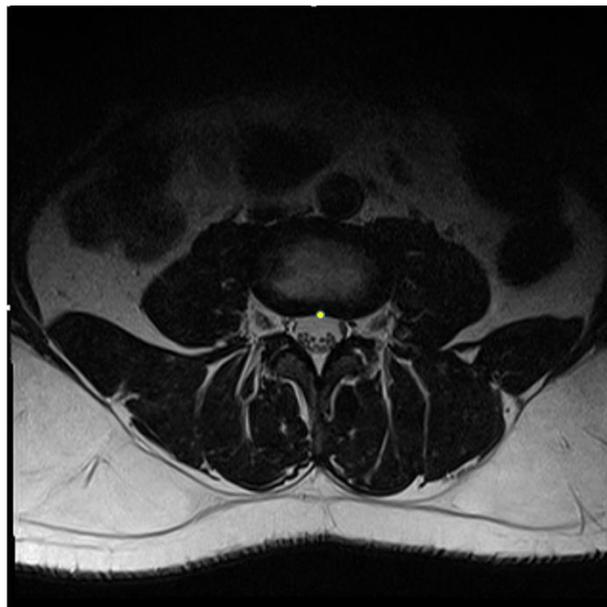


Figure 2.8: Illustrate a point at the bottom centre of the disc L3-L4.

3. On this step, they need to allocate the lowest point in the spinal canal as shown in Figure 2.9.

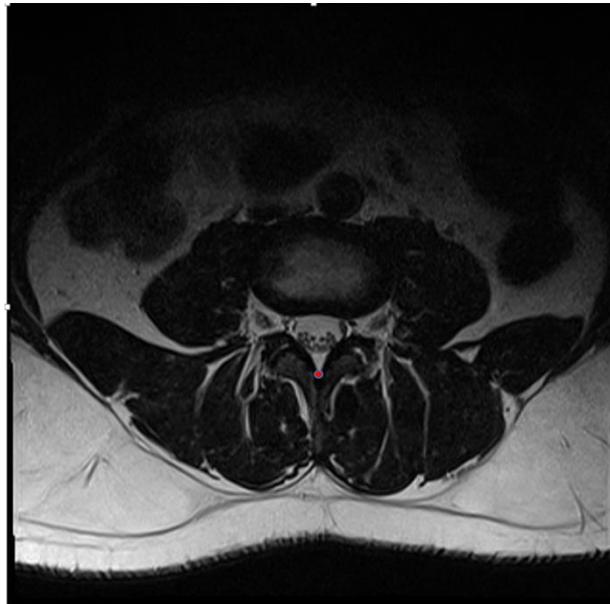


Figure 2.9: Shows the lowest point on the spinal canal for the disc L3-L4.

4. Find the distance between the point at the centre of the disc and the lowest point of the spinal canal as displayed in Figure 2.10.

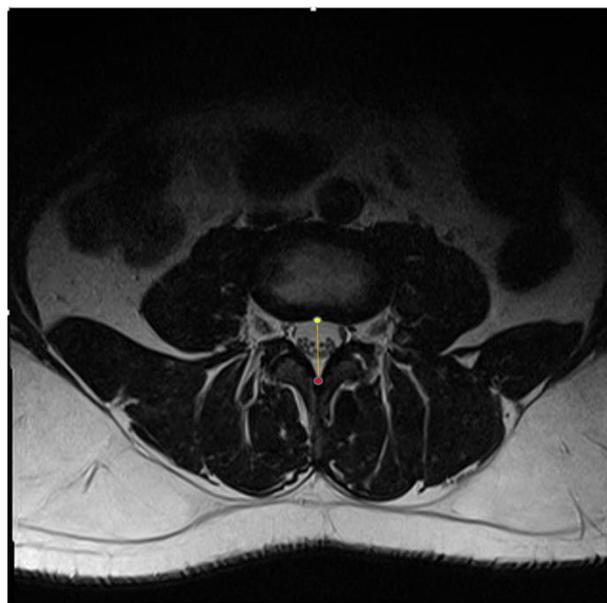


Figure 2.10: Illustrate the line between the lowest point on the spinal canal and the point at the bottom centre for the disc L3-L4.

5. Compare the distance with the standard distance for the normal lumbar spine and then decide if spinal canal stenosis is there or not.

2.6 Chapter Summary

While diagnosis accuracy is most paramount in medicine, we argue that it is also important to improve the effectiveness of the early diagnosis process to prevent benign cases from becoming more acute and chronic. A diagnosis of lumbar spine stenosis and disc herniation is performed through inspection of a Magnetic Resonance Imaging (MRI) scan of the patient's lumbar spine. The current diagnosing process of the lower back pain could take up much of the physician's time and effort. Moreover, it can increase the probability of misdiagnosis. The speed of early diagnosis can depend on many factors including referral time from the general practitioner to the hospital, waiting time for a specialist appointment, time for an MRI scan and time for the analysis result to come out. Due to the heavy demand for radiologists and clinicians, the whole process could take weeks to complete. Developing a computer-aided diagnosing system will help the clinicians to efficiently diagnose this kind of disease. At the same time, there is a pressing need for this type of application to help radiologists and orthopedists in their tasks. In our method, however, we argue that it is best to use the axial-view slice of the MRI because it has much more information about the lumbar spine at each vertebrae segment than a sagittal-view slice. In the following chapter, the literature review for the current methodologies that use the medical images in the computer-aided diagnosing is discussed.

Chapter 3 Literature Review

Clinical studies have indicated that morphological characteristics of lumbar discs and signal intensity of the patient's MR image have a strong relationship with clinical outcome [34]. To this end, Computer Vision and Artificial Intelligence algorithms can be utilised to exploit these facts by analysing the MR images, calculating appropriate image features, and classifying them to decide if any particular regions in the image belong to problematic areas. The literature review in this research is build using the up to date published research which includes the medical images, computer-aided diagnosing system, and machine learning. The research started by analysing most of the available research and only the significant research is referenced in this thises.

3.1 Medical Image Analysis

Image features are as a set of important information derived from an image or a subset of an image that can uniquely describe the image contents. This information is extremely important in computer vision as it can be used to label or mark specific locations of the image and can be used in comparing different images. There are two types of image features namely global and local features.

Local features are computed at different locations in the image using only the small support area around the location point. As such, local features describe only the image in the context of that small subset and nothing else. That means even when the other parts of the image undergo changes, as long as the support area remains the same, local features would more likely not be affected. This is one of the strong points of local features over global features because they are robust to occlusion. Examples of local features are corners, edges, and texture descriptors. On the other hand, global features are derived from the entire image which results in their ability to generalise the entire image into one single feature vector. One example of

global features are image code, which is a compressed form of the image using an appropriate coding technique that preserves the high-level information of the image contents [40]. Alternatively, global features could be constructed from a collection of local features such as shape descriptors, contours descriptors and/or texture descriptors.

Image analysis and comparison are performed by means of classifying its features. This is done by comparing the features from the test image in question with those from training data. A brute force approach for comparing two sets of image features would compare every feature in one set to every feature in the other and keep track of the "best so far" match. This results in a heavy computational complexity in the order of $O(N^2)$ where N is the number of features in each image. A number of algorithms have been proposed to improve the computational complexity, including the popular kd-tree technique [41]. This technique uses the exact nearest neighbour search and works very well for low dimensional data but quickly loses its effectiveness as dimensionality increases. The popularity of the kd-tree technique has seen a number of proposals to further improve the algorithm including [42], [43]. The success of a more recent matching technique called Fast Library for Approximate Nearest Neighbour (FLANN) [44] is another example how the computational complexity of image feature comparison can be further reduced to allow near real-time execution.

The uniqueness of each proposed algorithm in this category often lies in the choice of features and matching algorithms as well as a novel application of existing methods to a new or untested application. Below we review a number of algorithms that were proposed to identify regions in the MRI that are responsible for CLBP.

A method for curvature scale space image in shape similarity retrieval is developed in [45]. The paper used the curvature scale space CSS images to solve the problem in the shape similarity retrieval in image databases. The arch-shape contours in the CSS images are used to

represent the flexion of the shape in the image. The shape is represented by having the maximum of this contour. The CSS representation provides a method which can be easily used for the process of the shape similarity retrieval.

Ma and Manjunath [46] have provided an implementation of NeTra which is an image retrieval system by using many different features including the image colour, information of the location for the segmented region of the image, texture and the shape. An algorithm for image automatic segmentation is developed that gives the ability to search for a specific region on the image. They are segmenting the images into different regions after which a calculation for the image attribute is done. Moreover, there are many other processes performed to increase the performance of the retrieval process which represents the colour, colour indexing, texture and calculating the shape features. The curvature, centroid, and complex functions are used by the author's experiment as a contour type derived from the boundary coordinate.

Saleh and Eswaran in [47] developed an automatic diagnosing system for diabetic retinopathy. The system has a user-friendly interface which gives the ability to grade the severity of the diabetic retinopathy after locating and examining the first indication for the disease. The authors use the centroid distance function and depend on the shape to find the exact location of the optic disc.

3.2 Medical Images in the Computer-Aided Diagnosing System

Computer-Aided Diagnosis (CAD) systems have been researched and developed to improve the effectiveness and efficiency of many medical procedures. As an example, when locating regions of interest in medical images for radiologists to focus on, or to provide a "second opinion" before a physician makes their final decision. It is important to note that fully automated diagnosis systems are not expected to replace, in the foreseeable future, experts performing medical diagnoses. Instead, a CAD system is expected to complement the

physicians' ability rather than replace it [48]. The development of a CAD system involves modelling and embedding medical knowledge and thus incorporate a design of artificial intelligence through machine learning.

Visual observation and analysis of MR images could take up much of a physician's time and effort. As a result, physicians would opt to use a Computer-Aided Diagnosis (CAD) to help with this task. There are a number of CAD systems that can be used for various clinical purposes ranging from a CAD system for detecting colonic polyps and breast cancer in mammography, to another for detecting prostate cancer using MR images [27]. Despite the availability of these systems, physicians still have to overcome a number of technical challenges due to the wide range of imaging characteristics and resolutions [49] as well as due to the limitation of the algorithms employed to highlight areas of interest.

Segmenting the area of interest and allocating the border for the segmented area is one of the main required tasks to help in automising the diagnostic process. There are many different methods that have been developed for segmentation and disease detection. One of the most recent frameworks for detecting these borders was proposed by Gao et. al. [10]. In [50], the authors then apply this framework by using an Artificial Neural Network (ANN) method as the feature learning algorithm for detecting the borders. In this method, two types of imaging information, including spatial-neighbouring features, were used as the input data to one ANN that has two sparse auto-encoders as a classifier. Another ANN was used as an optimizer. The paper presented a comparison between this method's performance and the manual drawing approach on 461 IVUS images and reported a highly accurate result. A variant of the method using three types of features has also been proposed in [51] using 538 IVUS images.

Intravascular ultrasound images are not, however, the only possible source of information for detecting vascular anomalies. Recently, a regression segmentation framework to delineate the

boundaries of bi-ventricle from cardiac MRI was proposed [52]. In this framework, a regression model has been trained automatically on a deep belief network by using extracted DAISY feature [53] as input and using automatically generated boundary points as labels. The method was reported to yield high performance when tested on 2,900 images taken from 145 clinical subjects.

MRI scans and Neural Networks have also been used to diagnose other illnesses. A recent study [54] shows that a Convolutional Neural Network (CNN) can be used to reliably (with a reported accuracy of 98.8% when data augmentation and stochastic pooling are applied) identify Multiple Sclerosis in brain MRI scans. The experiment was conducted on images from an open-source dataset [55] (containing 676 Multiple Sclerosis slices from 38 subjects) and another [56] (containing 681 healthy slices from 26 subjects) as a control dataset

3.3 Computer-Aided Diagnosing System for Lumbar Spine Stenosis

The previous methods are some of the most recent examples of how image segmentation can be used for border detection and delineation in medical images. There is a wide range of algorithms for computer-aided medical diagnosis depending on the type of disease they are targeting. These algorithms are often specific to a certain type of medical images, such as computed tomography (CT) scan, X-Ray, intravascular ultrasound, or MRI. This is the case because certain types of images can capture features of certain diseases better than others. For example, the detection of lumen and media adventitia (MA) borders is the key procedure to determine anomalies inside coronary arteries. This procedure is most often carried out using intravascular ultrasound (IVUS) images rather than any other types of images because it has been well recognized as one powerful imaging techniques to evaluate the stenosis inside the coronary arteries.

Jiang et. al. proposed a visualization and quantitative analysis framework using the image segmentation technique to derive the features that are extracted from patients' MR images, which were found to have a close relationship with the Lumbar Disc Herniation score [57]. The features are the distribution of the protruded disc, the ratio between the protruded part and the dural sacs, and its relative signal intensity.

Alomari et. al. [27], [28] proposed a probabilistic model for automatic herniation detection work by combining the appearance and shape features of the lumbar intervertebral discs. The technique models the shape depending on both the T1-weighted and T2-weighted co-registered sagittal views for building a 2D feature image. The disc shape feature is modelled using Active Shape Model algorithm while the appearance is modelled using the normalised pixel intensity. These feature-pairs are then classified using Gibbs-based classifier. The paper reported that 91% accuracy is achieved in detecting the herniation.

A vertebrae detection and labelling algorithm of lumbar MR images is proposed in [49]. The paper firstly converts the 2D MR images to 3D before using them as an input to the detection algorithm. This detection algorithm is a combination of two detectors namely Deformable Part Model (DPM) [58] and inference using dynamic programming on chain [59]. After the spines were detected in the 3D images a graphical model of the spine layout is built and bounded boxes for all vertebrae in the images are labelled. The algorithm is evaluated on a set of 291 lumbar spine test images with a variable number of vertebrae visible and is reported to achieve 84.1% and 86.9% correct identification rate for overall vertebrae and lumbar vertebrae respectively.

A computational method to diagnose Lumbar Spinal Stenosis (LSS), which is a medical condition in which the spinal canal narrows and compresses the spinal cord, from a patient's Magnetic Resonance Myelography (MRM) and MRI was proposed. An image segmentation

process is first carried out as a pre-processing step to identify the affected dural sac area in the input images. It then produces the relevant image features based on the inter and intra-context information of the segments and uses them to detect the presence of LSS [1].

A more recent and relevant technique to our work is an unsupervised neural foramina boundary delineation framework by He et al. [60]. This framework uses the Multi-Feature and Adaptive Spectral Segmentation (MFASS) algorithm to automate the delineation process of neural foramina in the mid-sagittal view of a lumbar spine. MFASS utilizes a combination of region and edge features to generate spectral features that can be used to separate neural foramina and its surroundings. The separation process is controlled by adjusting the separation threshold, which is optimally and automatically estimated for each individual image. The framework is tested on 280 neural foramina MR images from 56 clinical subjects. The results are compared with manual boundary delineation performed by experienced physicians and analyzed using two metrics namely the Dice metric (a variant of the intersection-over-union metric) to measure segmentation overlap, and the Symmetric Mean Absolute Surface Distance (SMAD) to evaluate the delineation accuracy [61]. The paper reported a high consistency with manual delineation results (Dice: 90.58% \pm 2.79%; SMAD: 0.5657 \pm 0.1544 mm).

3.4 Machine Learning

Machine learning is often confused with data mining and knowledge discovery in databases (KDD), which share a similar methodology. Before the modern machine learning techniques, performing simple tasks was developed by giving the computer instructions or controlling the computer system by defining a complete task algorithm and then programming the algorithm into the computer which could consume much time and effort for computer specialists.

Machine learning algorithms are used to optimize the performance criterion using example data or past experience. Models are defined based on some parameters, and the learning process is

the execution of a computer program to optimize the parameters of the model using training data or past experience [62]. The model is used to make predictions in the future or descriptive to gain knowledge from data or both. There are many examples of machine learning problems. This thesis will discuss the classification problem in detail as this study is focusing on this specific concept. In classification, objects need to be categorized into fixed categories by taking input and labelling them as belonging to a given class. The decision being modelled is to assign labels to new unlabelled data. This can be thought of as a discrimination problem, modelling the differences or similarities between groups [63].

In regression, the aim is to predict a real value using historical data. For instance, ML can be used to predict how much it will rain tomorrow, or predict how much a house will be worth a couple of months in the future; other examples including the prediction of time series data like the price of a stock over time, the decision being modelled is what value to predict for new unpredicted data.

Another ML problem where the data is not labelled but can be divided into groups based on similarity and other measures of natural structure in the data is defined as clustering. A clustering or cluster analysis is the procedure of grouping a set of objects to discover that objects belonging to the same group (known cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). The Self Organized Map (SOM) network is a strong network and it can be implemented for feature selection, visualisation, and clustering tasks

The main goal of the data processing job is to select appropriate methods to a wide range of challenging real-world data-sets. Experienced data scientists cannot judge which algorithm will perform best before trying them, hence data need to be studied before using them in ML processing.

3.4.1 Artificial neural networks

Biologically, a simple neuron has four basic components: dendrites, soma (cell body), axon, and synapses as shown in Figure 3.1. Dendrites are assumed as input channels which receive inputs (signals) from the synapses of other neurons and send them to the soma. The soma processes the signals over time to transform the processed value as output and send it to other neurons through the axon and the synapses [64].

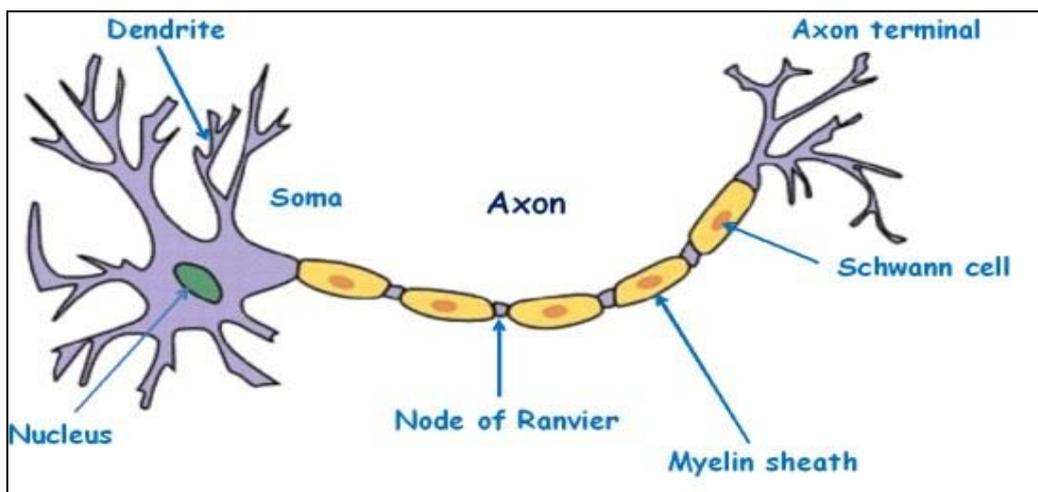


Figure 3.1: illustrations of biological neuron parts.

The ANNs try to replicate the basic component of the human brain. Right up to the present moment researchers and scientists continue evolving various methods and models of neural networks to reach a higher accuracy rate in solving different problems [65].

ANNs consist of a number of basic units called neurons. The artificial unit is often called an artificial neuron, perceptron, node, unit or processing element (PE). It is designed in a method to stimulate the four natural neuron functions. As shown in Figure 3.2 the unit or processing element receives inputs from other units or external inputs x_i . These inputs are connected with other processing elements by weights W_{ij} which is the weight connection from node j to node i . Each input is associated with different synaptic weights. The inputs are multiplied by the

connection weights and the products are summed to obtain the net input. Then they are fed through a transform function, thus generating the neuron's output or unit activation, Y at the output layer.

$$\text{SUM} = \sum w_{ij}x_i \quad (3.1)$$

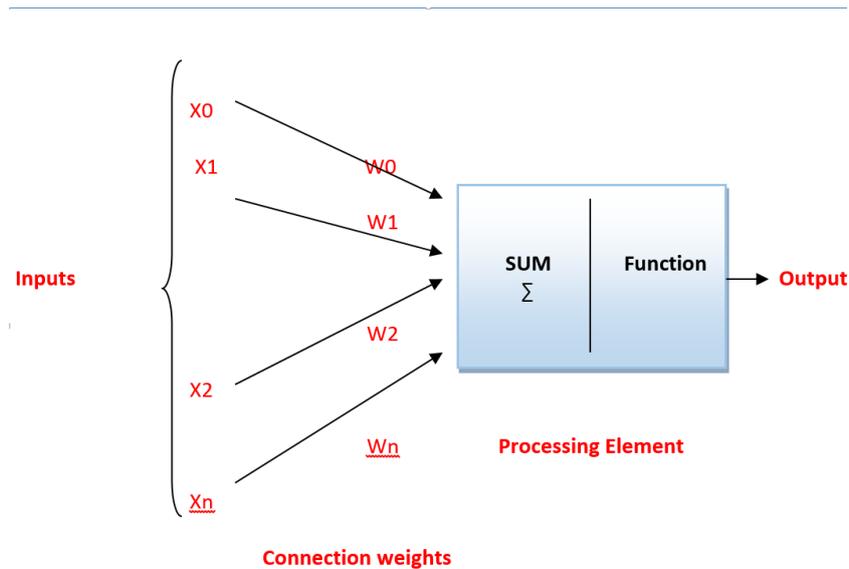


Figure 3.2: Network of artificial neuron or perceptron.

It is important to understand that the input nodes are used to provide the data to the neural network while the hidden nodes and the output units are considered as processing elements [64]. According to A.C.Knowles 2005 [66] there are two main types of NN architectures; feedforward networks and feedback or recurrent networks. The neural network is named feedforward NN when it is organized into layers which are input layers, hidden layer/s , output layer and the connections between consecutive layers have the same direction [67]. These NNs can include any number of neurons per layer, as well as having any number of hidden layers [64].

3.4.2 Convolutional Neural Network

A convolutional neural network (CNN) is considered as one of the most commonly used deep neural networks for deep learning images. The convolution is the mathematical linear operation between matrixes and this is the reason for naming this algorithm. CNN has multi-layers counting the convolutional, non-linearity, pooling, and fully connected layer in addition to the input and output layer [68]. In the convolutional layer, the input images will go through a set of convolutional filters where certain features from the image will be activated. Nonlinear downsampling and the number of the parameter will be reduced to simplify the output in the pooling layer. Another layer called rectified linear unit (ReLU) is implemented to map the negative values to zero and maintain positive values to have quicker and more efficient training as seen in Figure 3.3 [69] [70].

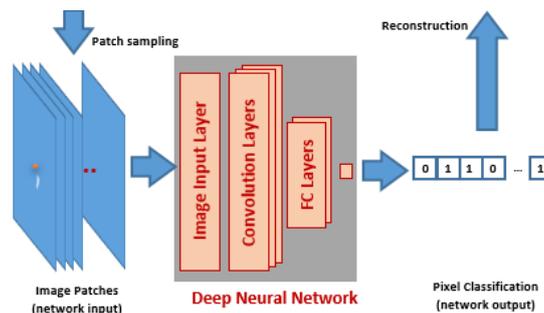


Figure 3.3: Convolutional deep neural network layers

3.4.3 Support Vector Machine

Support vector machines (SVM) is one of the supervised machine learning methods with the ability to analyse the datasets that can be used for the regression and classification role [166], minimising the misclassification done using SVM through a training phase, which is known as the maximum margin point [71].

The SVM model, on the other hand, is one of the dichotomous, kernel-based learning methods in which the Open Virtualization Format method extends its functionality to multi-class

classification. The MLP model uses a backpropagation algorithm to adjust the weights and determine the set of weights and bias values with the goal of minimising the error rate. In contrast, the SVM model in this study uses a Gaussian radial basis kernel function (RBF) to map the data into high dimensional space, where it is easier to create a linear decision boundary in the features space. The decision boundary, also called hyper-plane, should maximise the margin between the classes for optimal diagnosis.

For instance, using LSVM to predict the health state, diabetic versus non-diabetic for a particular patient, based on genetic data by giving a set of data samples with a two-class problem where the classes are linearly separable.

$$D = \{(x_1, y_1), \dots, (x_m, y_m)\} \quad (3.2)$$

For the training data a set of input vectors x_i with each one having a number of component *features*. These input vectors are paired with corresponding *labels* y_i , and there are m such pairs ($i= 1, 2, \dots, m$) where

$$\left\{ \begin{array}{l} x \in R^n \\ y \in \{-1, +1\} \end{array} \right\}, \text{ n is the data dimension} \quad (3.3)$$

As shown in Figure 3.4 diabetic cases would be labelled $y_i = +1$, non-diabetic $y_i = -1$, and the matching x_i are input vectors encoding the variables data derived from each patient. Separated by hyper-plane and the feature vector (Support vectors) represents the data points subset of the training data that lie closest to the hyper-plane or decision line [72].

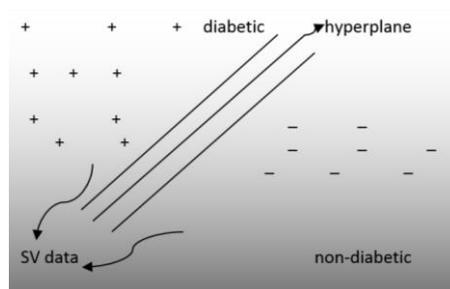


Figure 3.4: Diabetic and non-diabetic data separated linearly by hyper-plane.

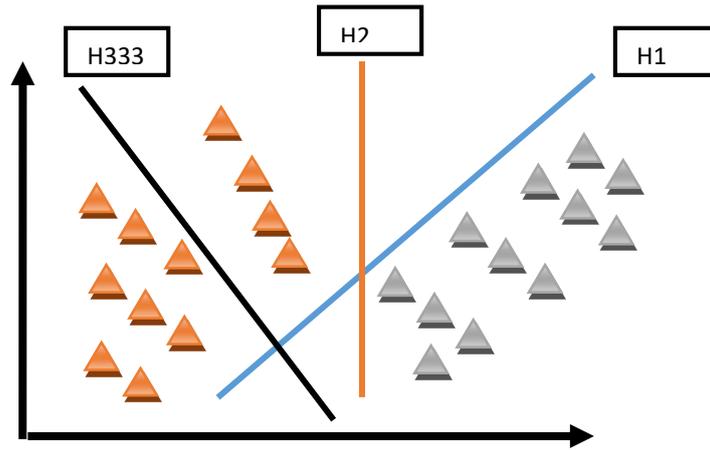


Figure 3.5: SVMC linearly separable set of two classes.

Classifying datasets is considered a common procedure in machine learning algorithms such as SVM. In this model, a dataset is shown as a p -dimensional vector in order to create a model that is able to separate such points with a $(p - 1)$ -dimensional hyper-plane. A number of hyper-planes can be applied to separate the datasets into group sets as shown in Figure 3.5. The main function of applying hyper-plane is to deal with the largest margin, or separation, between the two sets or more multi-class labels. In this regard, we select the hyper-plane to maximize the distance from the nearest data point on each side.

Figure 3.5 illustrates an example of a group of instances, with using optimal separating hyper-plane 1 in the purpose of maximum margin. The hyper-plane usually needs to be drawn in the midway point between the two margins. The SVM models require learning where the optimal hyper-plane should be fitted. The main aims behind maximizing the margin are to minimize the probability between points of different classes, that are unclassified or unseen, that may drop on the wrong side [73]. The first hyper-plane (H1) work is the best one among others, which is able to separate them with the maximum margin. The second hyper-plane (H2) is capable of separating points, but with a small margin. The last hyper-plane (H3) isn't able to separate the data sample.

Given the optimal hyper-plane function $\mathbf{h}(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + \mathbf{b}$, for any new point \mathbf{z} , the class can be predicted as

$$\hat{y} = \text{sign}(\mathbf{w}^T \cdot \mathbf{z} + \mathbf{b}) \quad (3.4)$$

Where the sign function returns +1 if its argument is positive, and -1 if its argument is negative. Non-linear SVM can separate datasets of the two classes that may not be linearly separable in input space. It may not be possible to find a directed hyper-plane separating the two classes of data [74], [75]. For this reason, an alternative representation of the data will be the key in such classification by performing mapping the data points into an input space with different dimensionality, called feature space. A dataset which is not separable in input space can always be separated in a space of high enough dimensionality. For example, Figure 3.6 shows a dataset of the two classes in a 2-dimensional space, mapping to a third dimension, such that the +1 labelled points are moved forward and the -1 labelled are moved backward, thus two classes become separable. The idea is to map the original d-dimensional points \mathbf{x}_i in the input space to points $\Phi(\mathbf{x}_i)$ in a high-dimensional feature space via some nonlinear transformation Φ (mapping function for each point). The linear SVM method can be used, where the points $\Phi(\mathbf{x}_i)$ might be linearly separable in the feature space.

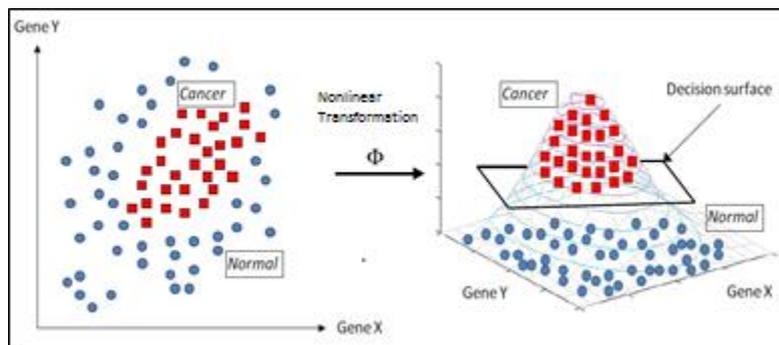


Figure 3.6: Two and three-dimensional illustration of data.

3.4.4 Decision Tree

The method followed in the decision tree was based on a logical or binary operation to implement an automatic computing procedure. The terminology associated with the decision tree is as follows:

- **Root Node:** this node represents the entire sample of data, which gets divided into two or more similar sets [76].
- **Splitting:** this node represents a process of dividing a node into two or more sub-nodes.
- **Decision Node:** this node shows when a sub-node separates into more sub-nodes.
- **Terminal Node:** This node doesn't split [77].
- **Pruning:** this node represents the process of removing sub-nodes from the Decision Node. It is the opposed process of Splitting.
- **Sub-Tree:** this is a whole sub-section of the whole tree [78].
- **Parent and Child Node:** a parent node is a node which is divided into sub-nodes which are called the child node of the parent node [79].

Using the Tree-based Model VS Linear Model

- A. The linear model works best if a linear model estimates the relationship between dependent & independent variables.
- B. If there is a complex relationship between dependent & independent variables, then a tree model works best.
- C. For simplicity and ease of interpretation by people, the decision tree is the best method to use.

3.4.5 Boosting

Boosting combines weak learner to form strong rule. In order to find a weak rule, we apply base learning (ML) algorithms with a different distribution. Every time the algorithm is applied

it generates a new weak prediction rule. This process is repetitive. After many repetitions, the boosting method combines weak rules together into a single prediction rule [80].

The best ways to choose the right distribution are:

1. The base learner takes all the distributions and assigns equal weight to each observation.
2. If there is any prediction error caused by a first base learning algorithm, then we focus on observations that are having prediction error. Therefore, we apply the next base learning algorithm.
3. Repeat step 2 until the limit of the base learning algorithm reaches the maximum accuracy level.

In conclusion, this algorithm combines the outputs from the weak learner and creates strong learner which improves the prediction power of the model.

There are many boosting algorithms, which convey an additional boost to the model's accuracy. The two most commonly used algorithms are, Gradient Boosting (GBM) and XGboost.

Extreme Gradient Boosting 'XGBoost'

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. It was engineered for the efficiency of computation time and memory resources.

The main goals of using XGBoost are execution speed and model performance [80].

- A. Execution Speed: XGBoost is fast in comparison to other implementations of gradient boosting.
- B. Model Performance: XGBoost controls structured datasets on classification and regression predictive modelling problems.

There are advantages and disadvantages for using this algorithm, and some of the features are:

- 1) Sparse aware implementation with automatic handling of missing data values.
- 2) Block structure to handle parallelization of tree construction.
- 3) Training so that you can boost an already fitted model on new data.

However, the price and the low number of algorithms are considered as a major drawback of XGBoost.

Gradient Boosting and Random Forests are both ensemble learning methods that predict by joining the outputs from individual trees. They are different in the way the trees are constructed and the way the output results are combined. Although it may appear that Gradient-boosted decision trees (GBDTs) are better than random forests, they are prone to over fitting. Random Forests train each tree independently using random samples of the data. This way aids in making the model more vigorous than a single decision tree and unlikely to over-fit on the training data. Whereas GBDTs build trees one at a time, the new tree helps to rectify the errors made by the previously trained tree. Every additional tree adds more expressiveness to the model. Training GBDTs generally takes longer because the trees are built successively. Benchmark results have shown GBDTs are better learners than Random Forests.

3.4.6 K-Nearest Neighbour

K-Nearest Neighbour Algorithm (KNN) is one of the most used algorithms. It's used in many areas including machine learning and pattern recognition [81]. The closest training example in the feature space is the method followed for all the objects' classification. The classifying process is done depending on the plurality of its neighbours. Meanwhile, the K-Nearest Neighbour method hangs on the theorem in principle. However, the KNN method in the category decision relates to a small group of adjacent samples. Figure 3.7 illustrates the instances with the blue and orange signs and the query with a yellow hexagon [82]. The main

procedure in this algorithm is divided into two parts starting by finding the nearest K instance followed by taking the majority vote of its neighbours to classify the dataset.

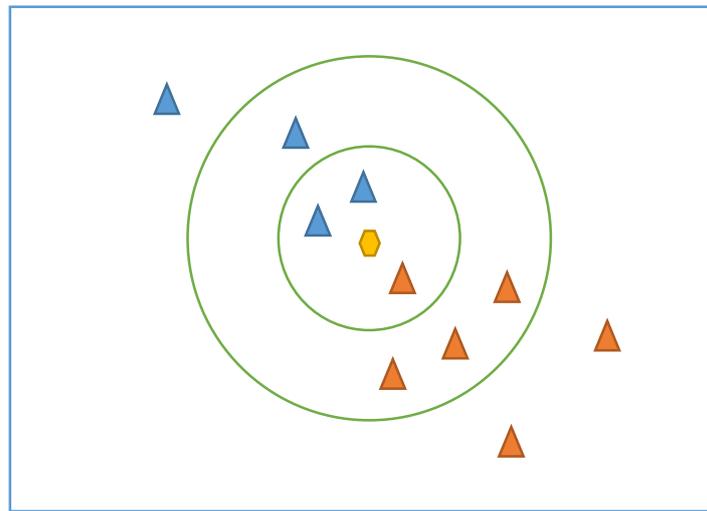


Figure 3.7: K-nearest neighbour algorithm (KNN) example.

The test sample in Figure 3.7 is yellow hexagon, which is classified into two classes, blue and orange triangle. The hexagon object indicates a new data point and we want to predict the class of it. If $K=3$, the algorithm will look at the 3 nearest neighbours to this new data point (inner circle) where we have two blue class objects and one orange object. Due to the majority of the blue class objects being near to the hexagon, it will be classified into blue class B. On the other hand if $K=7$, the algorithm will predict the new data point into the orange class as now there are more of the orange class in the outer circle as compared to the blue class. [83].

3.5 Chapter Summary

As a summary of the literature. It is noticeable that there is only a limited study on the machine learning techniques and their applications to lumbar spine image segmentation. This limitation was a motivation for this research to go further and improve the currently available methods. To improve the currently available methods it is important to have a dataset that contains enough number of images in addition to radiologist reports. The next chapter will include a discussion about the available dataset and the type of images used in each one of the datasets.

A comparison between those datasets and the available features will be used to state the reason to start developing our own dataset which contains a lot of properties and features to help in achieving the required result.

Chapter 4 The Dataset Developed for this Research

Accessing the publicly available dataset was a big challenge for many reasons such as the complicated process to receive the data, the very small amount of available data, the non-availability of medical reports for the images, and the type of available images. All these issues required us to build our own dataset. In this chapter, the dataset used to illustrate the work in this thesis will be described. The dataset consists of a scan of 575 patients with lower back pain with at least the images for the last three discs for each patient's lumbar spine. The publicly available dataset provided from the *SpineWeb* website is described in section 4.1 along with the Lumbar Spine MRI dataset which is used in our thesis.

4.1 The Available Open-Access Dataset

Dataset 1: The dataset published in the cross-modality spinal images for spine workshop contains MRI and CT images for the 30 patients' lumbar spine [84]. However, only the data for 20 patients are available. As we are only interested in the MR images and after analysing this dataset we have noticed that the T1 images are only available for this dataset. Meanwhile, the radiologist reports for this dataset are not available.

Dataset 2: A set of 10 patients with 18 trauma CT scan images with their labels are provided by the Department of Radiological Sciences, University of California, Irvine, School of Medicine. The manual label provided for these images is for the vertebrae only. However, there are no medical reports provided with this dataset [85].

Dataset 3: This training dataset contains 242 CT scan images for 125 patients as some of the patients have a repeated scan for a long diagnosis and follow-up procedure. Manual annotations have been provided for the vertebra centroid. Spine data information for the patients is provided which contains the scan ID, number of scans for each patient, gender and age [84]. The Department of Radiology at the University of Washington provided the dataset for the

Computational challenges on spine imaging workshop CSI 2014. This dataset does not contain the radiologist report to help in the diagnosing process.

Dataset 4: A set of spinal canal CT scan images provided by Melih S. Aslan [86]. The dataset contains CT images for 5 patients with ground truth for all the images. However, the manual segmentation is not provided if the full scan is not fully provided for the vertebrae. The dataset has no medical report for the diagnosis prediction stage.

Dataset 5: A binary mask manual segmentation and the CT scan images are provided by [87] for 10 patients. Moreover, for each one of those patients lumbar spine vertebrae images are provided. This dataset has used landmark-based shape representation in the 3D view. This dataset does not provide the medical reports for those patients to be used in the diagnosing process.

Dataset 6: A dataset of four patients' lumbar spine images has been used in [88]. These images are provided in a CT scan image with a reference manual segmentation for each vertebra from Level 1 to Level 5. The manual segmentation for each one of the vertebrae is provided as a binary mask in the meta-image format used. The overall images have been released during the Challenge on Spine Imaging CSI 2016. The medical report for these images is not provided with the dataset

Dataset 7 : In the International Conference on Medical Image Computing and Computer-Assisted Intervention "MICCAI" a dataset of 30 pairs (antero-posterior and lateral) of vertebral fracture assessment VFA images has been released for the challenge "Automatic vertebral fracture analysis and identification from vertebral fracture assessment VFA by dual-energy x-ray absorptiometry DXA" [84]. The dataset contains 10 pairs with a compression fracture, 10 pairs for normal vertebrae and another 10 with another type of deformation. Moreover, these

images do not contain medical reports to help in the diagnosis process for any other research issues.

Dataset 8: A dataset to help in understanding the name, location and the different positions of the vertebrae is provided by [89] which also includes the 3D for the vertebra centre location. The ground truth is developed from the manually annotated vertebra 3D orientation. However, there are no medical reports with this dataset.

Dataset 9: The CT images of the spine for 10 patients are provided by the trauma centre at the Department of Radiological Sciences, University of California, Irvine, School of Medicine. The data is divided into two groups as five of the patients have a vertebral compression fracture and the rest of the patients are an adult group with an age range from 20 – 35 years old. The images are in mhd/raw format. The data has been used in [90] for the vertebra segmentation as a part of the spinal column analysis. Similar to the previous datasets, the medical reports are not provided with this dataset. Table 4.1 shows all the available dataset with its details. However, some of this dataset is not accessible and no longer available to the public.

Table 4.1: Shows all the available dataset in the SpineWeb web page with its details.

Dataset #	Image Type	Number of Images	Number Of Patients	Type of availability	Comments
1	CT and MRI	30	N/A	The online form needs to be filled to obtain the data. The ground truth is not provided.	
2	CT	N/A	10	Segmentation challenge of the CSI 2014 Workshop. Published Paper is used for the citation Detection of Vertebral Body Fractures Based on Cortical Shell Unwrapping, ODC Public Domain Dedication and License (PDDL) (http://opendatacommons.org/licenses/pddl/1.0/).	The segmentation is also provided.
3	CT	242	125	localization and identification challenge of the CSI 2014 Workshop http://research.microsoft.com/spine/	
4	CT	N/A	5	Published Paper Clinically desired segmentation method for vertebral bodies 18 (10 female 8 male)	
5	CT	10	10	Published Paper. The code and image databases are publicly available through http://lit.fe.uni-lj.si .	
6	CT	N/A	30	CSI 2016 challenge + Published	
7	X-Ray	N/A	30	Testing Set of the MICCAI 2015 Challenge	
8	MRI + CT	N/A	20	Published Paper	
9	CT	N/A	10	Published Paper	

4.2 Lumbar Spine MRI (LSMRI) Dataset

All procedures performed in this research are both in accordance with the ethical standards of the United Kingdom and comply with the 1964 Helsinki declaration and its later amendments. The approval was granted by the Medical Ethical Committee of Hospitals which provided this data. The data was collected between September 2015 and July 2016 from patients who attended the hospital who reported relevant pains. Written formal consent was obtained from each patient prior to the data collection. A personal-data cleaning process was applied by the hospitals to the collected data to remove any information that can be used to relate the data to any patient such as the patient's name and contact details. We assign each patient's data with a unique identification number and only refer to particular data with its assigned number for subsequent processes. This allows the data to be accessed anonymously as stipulated by the ethical committee's condition of approval.

Our original dataset contains a clinical MRI study, or a set of scans, of 575 patients with symptomatic back pains. The MRI scanning parameters used in the scans can vary depending on the sequence and view plane types. The values of the most important parameters are summarized in Table 4.2

Table 4.2: MRI Scanning Parameters.

View Plane Types	Sagittal		Axial	
Sequence Types	T1-weighted	T2-weighted	T1-weighted	T2-weighted
Number of Echoes (ETL)	3	15 to 18	3	9 to 16
Repetition Time (ms)	330 to 926	3190 to 4000	385 to 953	1900 to 5000
Echo Time (ms)	9.2 to 12.0	67.0 to 96.0	11.0	84.0 to 96.0
Slice Thickness (mm)	3.0 to 4.0	3.0 to 5.0	4.0	3.0 to 5.0
Spacing Between Slices (mm)	3.3 to 4.8	3.3 to 6.5	4.4 to 4.4	3.3 to 6.5
Field of View (mm)	280	280	220	220
Matrix (Freq. x Phase)	100%	100%	100%	100%
Imaging Frequency (MHz)	63.6765 to 63.6828	63.6765 to 63.6828	63.6801 to 63.6828	63.6801 to 63.6828
Number of Phase Encoding Steps	288 to 540	408 to 544	295 to 336	272 to 360
Flip Angle	150	150	150	150

During the data selection stage, we had to remove 60 of the studies that do not meet the requirements set by the proposed methodology described in this thesis. The requirements that need to be satisfied for each MRI study are as follows:

1. The study needs to include at least the last three lumbar vertebrae and their adjacent posterior elements, the last three intervertebral discs (IVD), and the topmost sacral bones.
2. The study has to contain both T1-weighted and T2-weighted scans. An image registration algorithm should be able to align both scans within a reasonable search space and time period. This means the patients have not made any significant movement during the scanning procedure to make the two scans completely different.

3. The study needs to have both sagittal and axial view scans and their corresponding cross-view information. This should allow us to see the direction and position of the image plane of an axial view slice on the sagittal view and vice versa in a DICOM viewer application.
4. The study needs to have at least one axial view slice close to the centre of each intervertebral disc.
5. The study should be of reasonably good quality in terms of focus, sharpness, and distortion.
6. The study should not contain destroyed or fused lumbar spine elements to make manual region labelling difficult or impossible.
7. The study should be from patients with a minimum age of 17 to ensure common physiology of the lumbar spine.

At the end of the data selection process, we have an MRI study of 515 patients. To help us maintain consistency in data numbering for all processes, we opted to keep the ID number assignment for each patient data the same instead of reassigning them every time an item of data is removed from the dataset. This approach will also be adopted in the future should we be required to remove more data. If this need arises, we will simply remove the data from the dataset but keep the unused numbers intact in the record.

Each study of the remaining 515 patients in our dataset is annotated by expert radiologists with notes regarding the observed characteristics, condition of the lumbar spine, or the presence of diseases, these include bone marrow disease, endplate degeneration, IVD bulges, *Thecal Sac* compressing, central or foraminal stenosis, annular tears, scoliosis, endplate defects (Modic type), facet joint and *Ligamentum Flavum* hypertrophy, and spondylolisthesis.

Each patient's data can have one or more MRI studies associated with it. Each study contains *slices*, i.e., individual images taken from either sagittal or axial view, of the lowest three vertebrae and the lowest three IVDs. The axial view slices are mainly taken from the last three IVDs – including the one between the last vertebra and the sacrum. The orientation of the slices

of the last IVD is made to follow the spine curve whereas those of the other IVDs are usually made in blocks – i.e., parallel to each other. There are four to five slices per IVD and they begin from the top of the IVD towards its bottom. Many of the top and bottom slices cut through the vertebrae leaving between one and three slices that cut the IVD cleanly and show purely the image of that IVD. In most cases, the total number of slices in axial view ranges from 12 to 15. However, in some cases, there may be up to 20 slices because the study contains slices of more than the last three vertebrae. Figure 4.1 illustrates an example for two different patients' MRI which contain in the first case four slices of images for the disc while containing five slices for the second case. The scans in sagittal view also vary but all contain at least the last seven vertebrae and the sacrum. While the number of vertebrae varies, each scan always includes the first two sacral links.

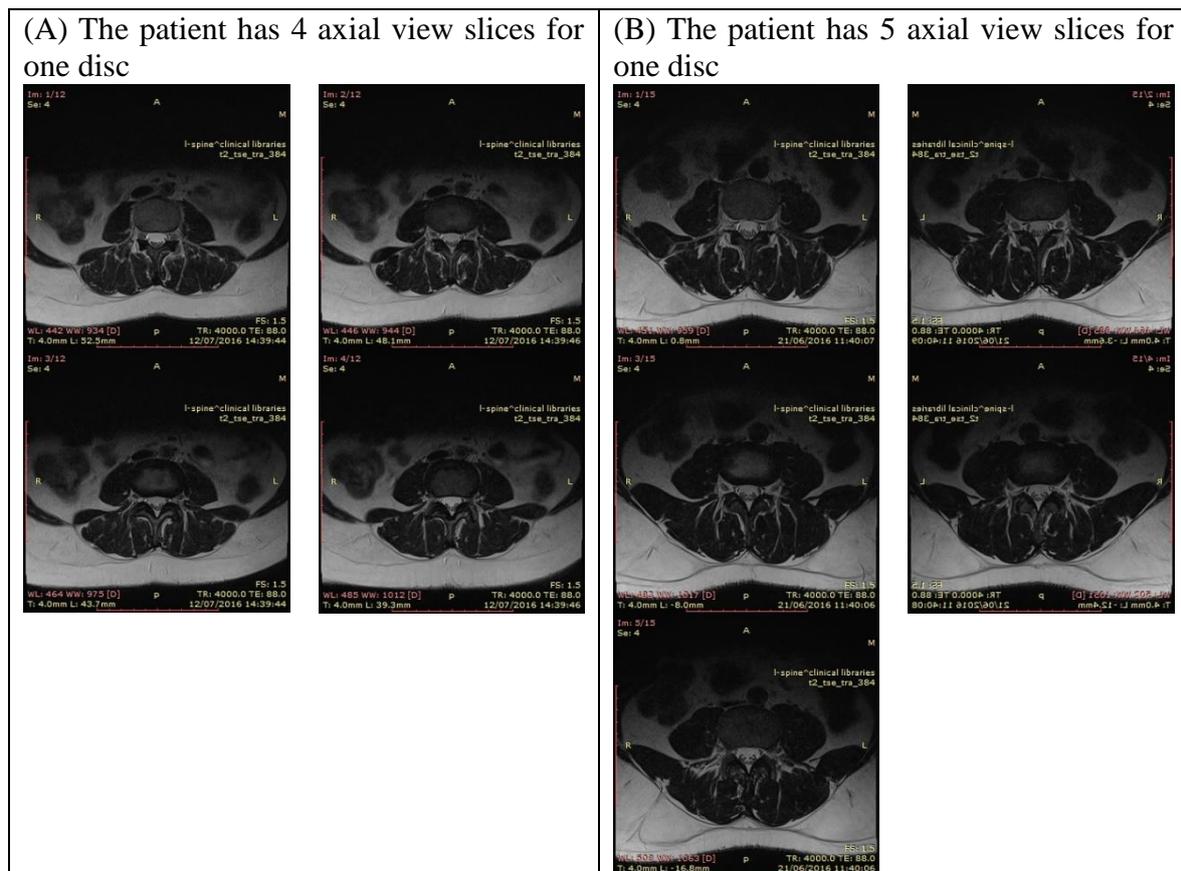


Figure 4.1: Illustrates an example for two different patients MRI (A) Four slices of images for the level 3 disc. (B) Five slices of images for the level 3 disc.

There are a total 48,345 MRI slices in our dataset. The majority of the slices have an image resolution of 320x320 pixels, however, there are slices from three studies with 320x310 pixel resolution. The pixels in all slices have 12-bit per pixel precision which is higher than the standard 8-bit greyscale images. Specifically for all axial-view slices, the slice thickness is uniformly 4 mm with the centre-to-centre distance between adjacent slices to be 4.4 mm. The horizontal and vertical pixel spacing is 0.6875 mm uniformly across all axial-view slices.

The majority of the MRI studies were taken with the patient in Head-First-Supine position while the rest were taken with the patient in Feet-First-Supine position. Each study can last between 15 and 45 minutes and a patient may have one or more studies associated with them taken at a different time or a few days apart. Because of the requirement of the method we are employing, we only select studies that contain both T1- and T2-weighted images in both sagittal and axial views. The difference in acquisition time between T1- and T2-weighted scans ranges from 1 to 9 minutes. The long-time difference could suggest that corresponding slices may not necessarily align and may require an application of an image registration algorithm to align them. As before, due to the requirement of the method we are employing, we removed the studies where the difference in T1- and T2-weighted scans causes the image registration process to fail to provide our reader with a better picture on the unsuitability of the images that we discarded, we show several examples of them in Figure 4.2.

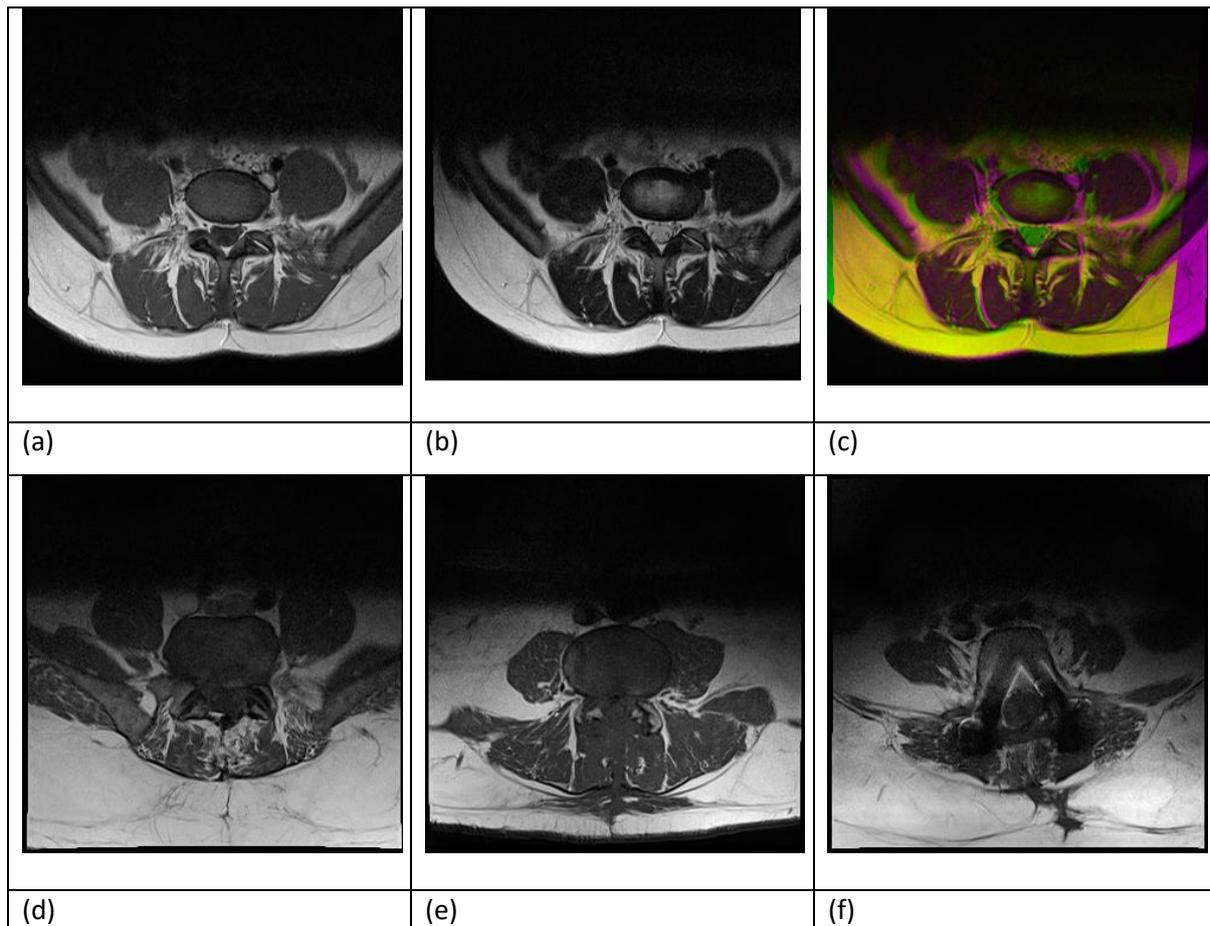


Figure 4.2 Some examples of the discarded MRI scans. Image (a) is T1-weighted and (b) T2-weighted MRI scans of the same part of the lumbar spine of a patient. There is a significant difference in the position of the same organ in both images resulting in a large number of mismatch pixels (purple regions) in the resulting (c) composite image. Image (d) and (e) are example cases where multiple regions are fused together making manual segmentation unreliable. And (f) is an example of a scan that contains unexplained imaging distortion or artefact.

4.3 Dataset Used in the First Two Experiments

The dataset to be used to develop this methodology is a subset of the LSMRI described in this chapter; the subset we are going to use contains 101 patients out of 515. The distribution for age and gender is shown in table 4.3. We have produced a dataset from the anonymized axial view MR images. The data consists of the digitised MRI and the radiologist report for each one of these MR image sets, in addition to the patients' details (*Case Number, Patient Age, and*

Gender). The LSMRI dataset is observed by our external partner including the MRI and here we expect no less than 3 axial view MR images for each patient.

Before the data is used, it has been analysed to make sure that all the data used will have the necessary information and match our requirements to run the experiment. Then a filtering and cleaning process is done by removing all the data which does not fulfill the conditions discussed in section 4.2. As a result of this process, only the data for 515 patients remain as acceptable data and 101 patients have been selected for this experiment which has been categorized as in Table 4.3 whereas a sample of the data in terms of the patients’ age and gender is shown in Table 4.4. The radiologist report for each MRI scan will be used by the orthopedist to have more information about the reason for the patient’s pain. The radiologist reports are adopted as ground truth labels within this study, where each case reflects the observation of a single patient. Figure 4.3 shows an example of the radiologist report for one patient.

Table 4.3: Data Categorization for 101 Patients.

Patients Ages	Numbers of Patients
21 - 30	39
31 - 50	53
50>	9
	Total of 101

Patients Gender	Number of Patients
Male	70
Female	31
	Total of 101

Table 4.4: Sample of the provided data

Patient DOB	MRI date	Patient Age	Gender
06/09/1979	22/07/2015	36	M
16/12/1985	03/01/2016	30	M
13/07/1979	26/12/2013	34	M
01/01/1975	18/11/2015	40	F
17/06/1978	21/01/2015	37	F
08/12/1961	22/07/2015	54	M
05/02/1956	24/11/2015	59	F
10/10/1993	02/06/2015	22	M
25/01/1974	28/07/2015	41	F
26/07/1944	19/01/2016	71	F

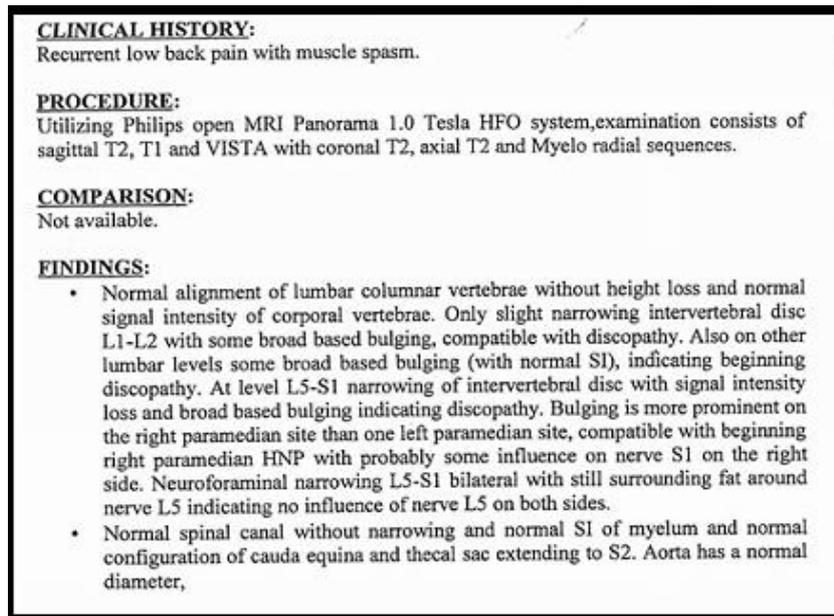


Figure 4.3: Example of the radiologist report for one of the patients

4.4 Summary

In summary, different types of spinal column dataset are available. However, none of the available data are suitable for our research either because the number of patients included was very low or because the medical reports for the MRI images were not provided and that made the use of these data not possible for the disease prediction stage. The dataset used in our research covers a wide range of cases fulfilling all the approval requirements in addition to having the medical reports for each one of these cases. Meanwhile, the dataset used which contains the medical MRI images, radiologist reports, and ground truth data is available for the public in [91]–[93] respectively. The dataset used in our research was also used in developing the labels and the ground truth data which are going to be explained in the next chapter.

Chapter 5 Disc Herniation Detection using Disc Segmentation and Centroid Distance Function

The methodology used is derived from the main machine learning framework as illustrated in Figure 5.1 and begins with the measurement of a real-world process, in this case, the MR images of the lumbar spine. Measurement is followed by the transcription of MRI into a usable form where the collection can take place. Such collected data is then pre-processed according to broadly distinct phases: cleaning and evaluation, followed by feature engineering. The cleaning and validation phases are potentially iterative, ensuring that subsequent analytical stages are not contaminated by systematic errors or the effect of unintended artefacts. Upon the output of a clean set of data, feature engineering is undertaken in our method to provide a problem representation suitable for input to the machine learning elements in subsequent analysis. The primary analytical phase follows feature derivation, comprising model building and model evaluation. The model building process involves the choice and specification of model architectures, in addition to the optimisation of the free parameters of the models using appropriate learning algorithms. The data used during this learning process is referred to as the training set (in-sample data). Following the modelling process, the generalisation performance of such models is evaluated, a procedure that aims to establish the degree to which the previous process of model development succeeded in producing a general model, capable of explaining future unseen data. The portion of data used to test generalisation performance is referred to as the test set (out of sample data), comprising data not used to influence the training procedure when forming the models.

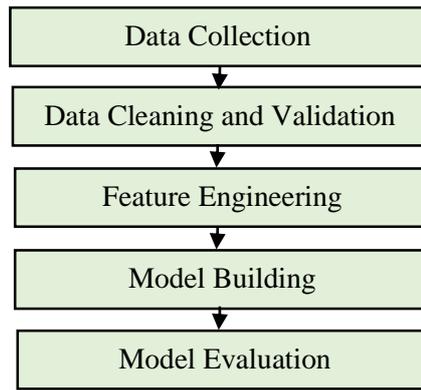


Figure 5.1: Machine Learning Framework.

The methodology in this thesis has been divided into three sub-methods, the first method is for segmenting the disc area in the lumbar spine MRI to diagnose the disc herniation. The second method is done by segmenting the AAP area in the lumbar spine MRI but using a cropped and full-size images which are going to be explained in chapter 6. Whereas, the final method is done by segmenting five different regions of the lumbar spine MRI using the full image size as explained in chapter 7. The decision of moving from one experiment to another is to improve the result of the segmented area which can lead to better diagnosing for the targeted diseases. Figure 5.2 shows the methodology followed to label the lumbar spine disc using axial view MRI based on pixels coordinate and grey level features.

In order to detect the disc herniation, we need to find the disc boundary; in the next section we will explain how to detect this boundary using a traditional edge detection algorithm

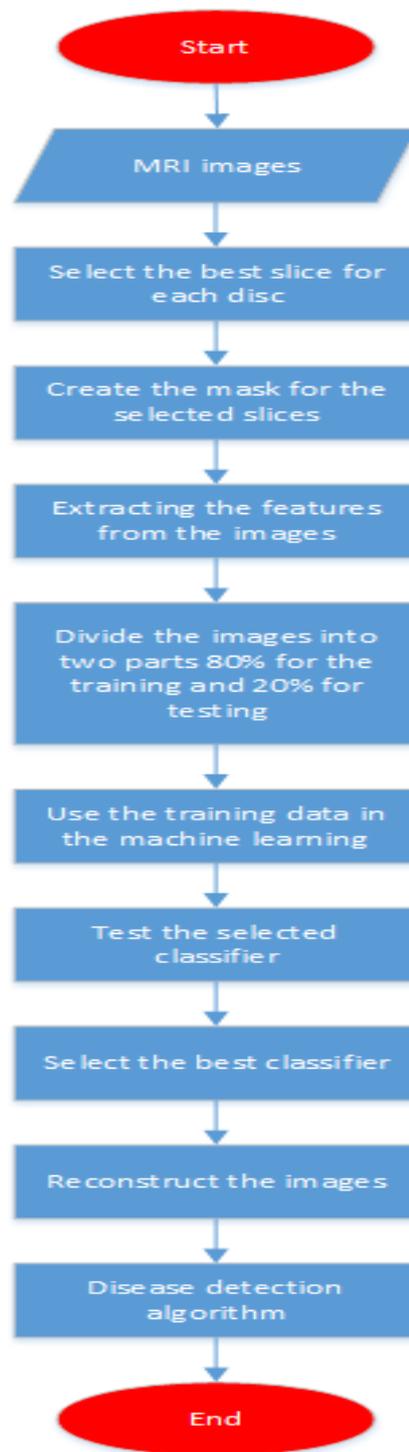


Figure 5.2: Shows the methodology followed to label the lumbar spine disc using axial view MRI based on pixels coordinate and grey level features.

5.1 Basic Edge Detection Algorithms to detect disc boundaries

A hastily done observation for locating the disc boundary in addition to the three important boundaries illustrated in Figure 2.5 in chapter 2 would suggest that the boundaries could be located by applying an edge detection algorithm. However, it should be noted that lumbar spine MRI images can have a varying wide spectrum of edge strengths while the strength of the disc and the other three edges are somewhat convoluted in the middle. This makes detection of those edges alone a challenge. To illustrate this claim, four results of an application of the Canny edge detector using a combination of two threshold values and two Gaussian sigma values to capture different levels of edge strength are shown in Figure 5.3. The figure shows that the edges cannot be reliably located without losing accuracy or having too much noise.

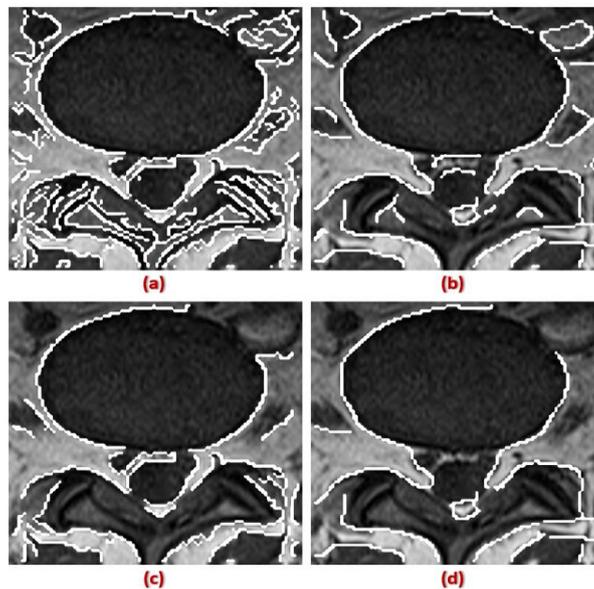


Figure 5.3: The results of Canny edge detection using a combination of two threshold values and two sigma value where a) low threshold – low Gaussian width, b) low threshold – high Gaussian width, c) high threshold – low Gaussian width and d) high threshold and high Gaussian width.

Therefore we conclude each detection alone cannot be used to detect the boundary of the disc reliably. In order to do that we need first to segment the MRI images ie. separate the different region of the MRI images before we locate the boundary of that region

The methodology we take to address the problem of disc segmentation is characterised by the use of a supervised machine learning framework, which we used to analyse the information exported from the axial view MR images, representing both disc and non-disc area. The machine learning method has been identified as a promising direction in our literature review, offering advantages over manual forms of analysis by human actors and conventional statistical methodologies. The computer-aided system has widened the scope of human grounded analysis, providing a shift in capability in disease diagnosis. However, the analytical forms permissible under the constraints of statistical validity, that is those accompanied by mathematical proof, restricts the class of addressable problems to those that can be represented using existing mathematical theorems. Machine learning, in contrast, allows for the analysis of arbitrary problem domains, while minimising the need for human grounded assumptions.

5.2 Image Labels and Ground Truth

Using machine learning required us to have ground truth data and this process starts from the manual labelling process of the disc area which is carried out on each T2-weighted axial-view MRI slice of the last three IVDs. The steps to develop the ground truth data are detailed below:

1. We use a dual-view setup on a DICOM/MRI viewer illustrated in Figure 5.4. On the mid-sagittal (left) view, we observe the cut line (yellow line) of the corresponding axial-view slice on the right. The best axial-view slice of a disc, defined as one which cuts closest to the half-height of the disc, is then selected.

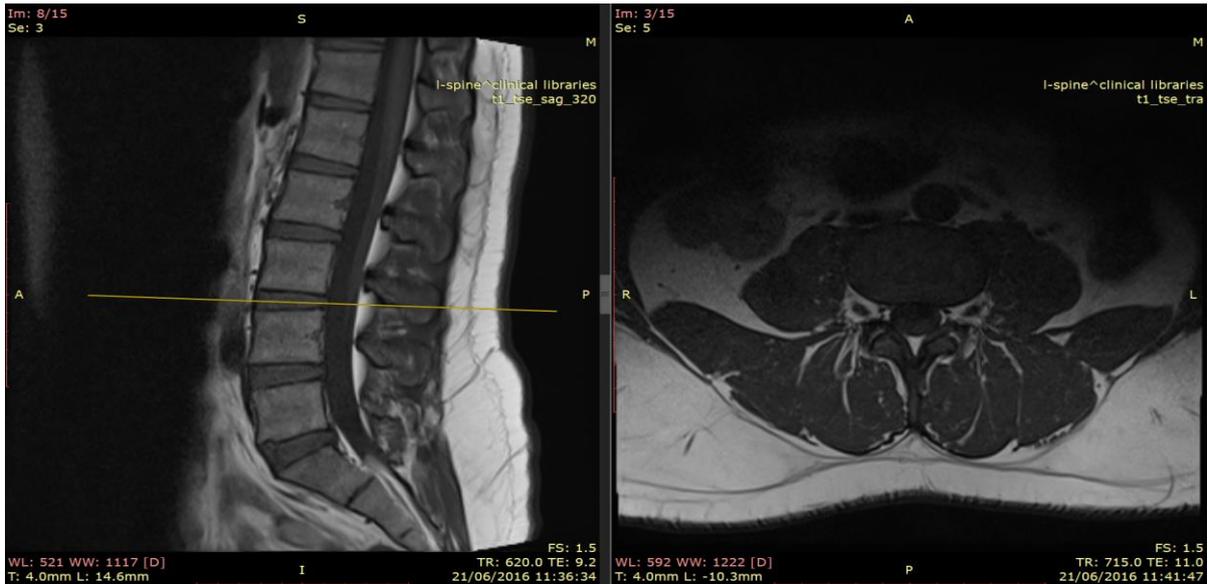


Figure 5.4: Dual-view of MRI images showing the mid-sagittal view (left) and the corresponding axial-view slice (right) at the cut line

2. We then extract three best T2-weighted MRI slices, one from each of the lowest three IVDs, i.e., L3-L4, L4-L5, and L5-S1, from each patient’s MRI study. The extracted images are uncompressed. For brevity, we will refer to these selected slices as the images from now on. Figure 5.5 shows some examples of the selected images for each disc.



(A) Middle Image of the Disc L3-L4



(B) Middle Image of the Disc L4-L5



(C) Middle Image of the Disc L5-S1

Figure 5.5: The selected MRI Images of the Three Lumbar Spine Discs.

3. The disc area in each image is then manually labelled. We call these images mask images or label images. An example of such mask images with the corresponding MRI can be seen in Figure 5.6.

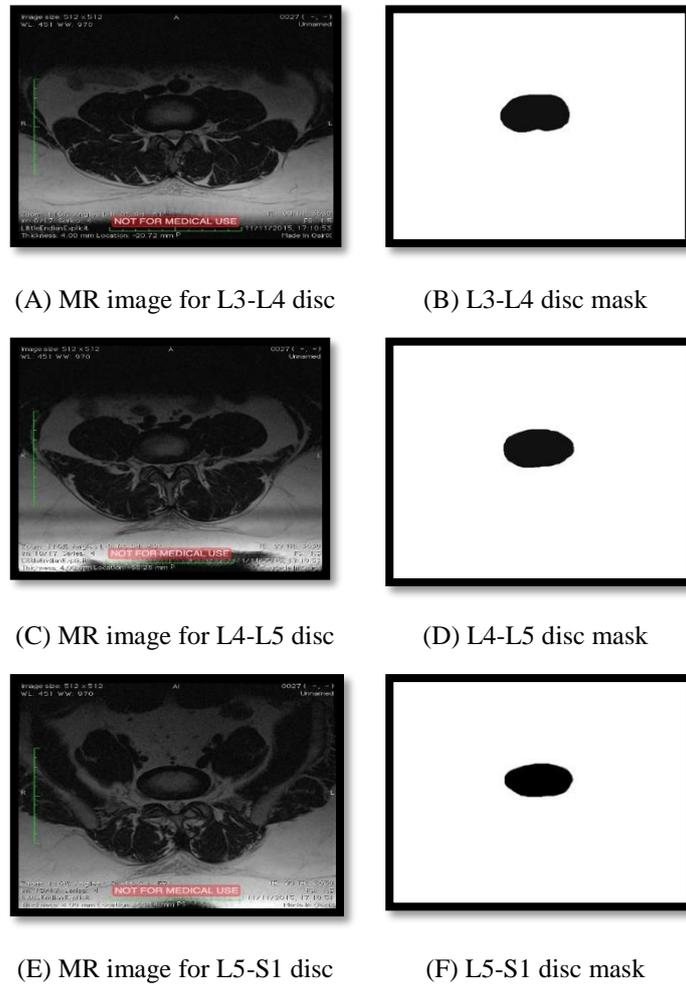


Figure 5.6: Discs MRI and the developed Mask for each disc.

The intensity level for the disc on the mask is used to differentiate between the disc area and the non-disc area. The disc area on the mask is filled in black colour with intensity level 0 and the remaining area of the mask image has the white colour with intensity level 255. The feature vector is the x-, y- coordinate, and intensity of each pixel. Table 5.1 shows a sample of the exported data and the intensity in addition to the pixel values.

Table 5.1: Shows sample of the exported data including the pixel position, Intensity, and Disc which represented by (1) or Non-Disc which represented by (0).

X coordinate	Y coordinate	Pixel Intensity	Disc/Non-Disc area Class 1/Class 0
288	292	43	1
375	323	7	0
486	544	69	0
491	83	86	0
238	92	30	1
288	266	85	1

There are a number of algorithms that could be used in the knowledge-based/artificial intelligence system. Our method is to experiment with a six of classifiers and perform the training and classification process. The best classifier will be chosen based on the achieved accuracy. To illustrate the training and classification process, the training will use the truth data which have been developed using contrast weighted MR images as discussed earlier. Contrast weighted images are used to emphasise different types of tissues within the same MR images. The trained system will then be able to produce labelled images of the affected areas if a disc herniation is detected in the input MR images.

5.3 Classifier Model Selection

To evaluate the capability of our classifiers, we conducted empirical simulations using the extracted features from the MRI. The models used in our experiment are listed in Table 5.2 while the experiment results are shown in Table 5.3 and Table 5.4.

Table 5.2 The Models trained in the two experiments

The Trained Model
Medium Tree [94] [95]
Coarse KNN [96] [97]
Weighted KNN [96] [97]
Medium Gaussian SVM [98]
Fine Gaussian SVM [98]
Ensemble RUSBoosted Trees [99] [100]

Three features are extracted from the selected MR images and the masks produced for 101 patients using Matlab as defined earlier in this chapter. Two different experiments were carried out using two data samples. We tried to use the maximum possible amount of data for the training and after many trials, it has been realized that the algorithm will run for around two days and then the machine will hang up due to limited resources in the machine used for the training process. That is why we have a limitation on the amount of chosen data. First, the classifiers were trained using 7,500 randomly chosen feature data from the three-disc images of one patient. The limit of the data is due to the maximum number of practically possible data that can be used at one time. Table 5.3 shows the list of classifiers and training results. The table shows that the weighted KNN model produces the highest accuracy compared to the other classifiers. Figure 5.8 shows the original and segmented images using the weighted KNN classifier. In the second experiment, the classifiers were trained using 7,000 randomly chosen feature data from seven patients. The number of feature data is drawn equally from each patient, i.e., one thousand points from each patient equally in all three discs. Table 5.4 shows the performance of the classifiers. In this experiment, SVM Fine Gaussian produced the best accuracy from the entire classifier set. An example of pixel classifier using SVM Fine Gaussian is shown in Figure 5.9. The trials presented in this work were conducted using a common dataset partitioning scheme. We reserved 80% of the data for model training and the remaining 20%.

Table 5.3: The first experiment result based on the evaluation of 6 models. The Models trained using one patient data.

Model	Accuracy	True Negative Rate (%)	True Positive Rate (%)	False Positive Rate (%)	False Negative Rate (%)	AUC
Medium Tree [94] [95]	98.3	99	93	1	7	0.98
Coarse KNN [96] [97]	98.6	99	93	1	7	1
Weighted KNN [96] [97]	98.9	99	94	1	6	0.99
Medium Gaussian SVM [98]	98.7	99	93	1	7	1
Fine Gaussian SVM [98]	98.8	99	94	1	6	0.99
Ensemble RUSBoosted Trees [99] [100]	98.0	98	98	2	2	1
Key: SVM Support Vector Machine KNN K-Nearest Neighbour AUC Area Under Curve PS Prediction Speed			Hint: True positive – correctly classified disc area True negative – correctly classified non-disc area False positive – incorrectly classified non-disc area False negative – incorrectly classified disc area Note: The selected classifier is highlighted in the table			

Table 5.4: The second experiment result based on the evaluation of 6 models. The Models trained using seven patients' data.

Model	Accuracy	True Negative Rate (%)	True Positive Rate (%)	False Positive Rate (%)	False Negative Rate (%)	AUC
Medium Tree [94] [95]	89.0	92	70	8	30	0.87
Fine Gaussian SVM [98]	91.1	96	62	4	38	0.88
Medium Gaussian SVM [98]	90.8	95	64	5	36	0.95
Coarse KNN [96] [97]	90.7	97	3	44	56	0.95
Weighted KNN [96] [97]	90.8	95	66	5	34	0.94
Ensemble RUSBoosted Trees [99] [100]	84.1	83	93	17	7	0.94
Key: SVM Support Vector Machine KNN K-Nearest Neighbour AUC Area Under Curve PS Prediction Speed			Hint: True positive – correctly classified disc area True negative – correctly classified non-disc area False positive – incorrectly classified non-disc area False negative – incorrectly classified disc area Note: The selected classifier is highlighted in the table			

The results of our experiment are divided into two parts the first part is the classifier training result, whereas the second part is the reconstructed image of predicted data from the classifier which got the highest accuracy; these two parts will be discussed later in this chapter.

5.3.1 Classifier Training Result

The results from our experimental procedure are presented and organised for each respective classifier for the first experiment in Table 5.3 when the data source was one patient's details with 7,500 pixels. Whereas, in the second experiment in Table 5.4 the data source was seven patients' details with 7,000 pixels. We then proceeded to present our evaluation of the classifiers according to the classifier accuracy. We have experimented with all the available classifiers in Matlab app and there were 23 classifiers. Nevertheless, only the best 6 classifiers with the highest accuracy have been discussed. Figure 5.7 shows the classifiers for the two experiments and the achieved accuracy for each of them. In our experiments, we concentrate on the disc area; that's why a positive value has been given when the classifier predicts a disc area while a negative value is given when the non-disc area has been segmented. However, if the classifier has correctly segmented the area, either disc or non-disc, a true value is given for this case and a false value is given for the incorrect prediction.

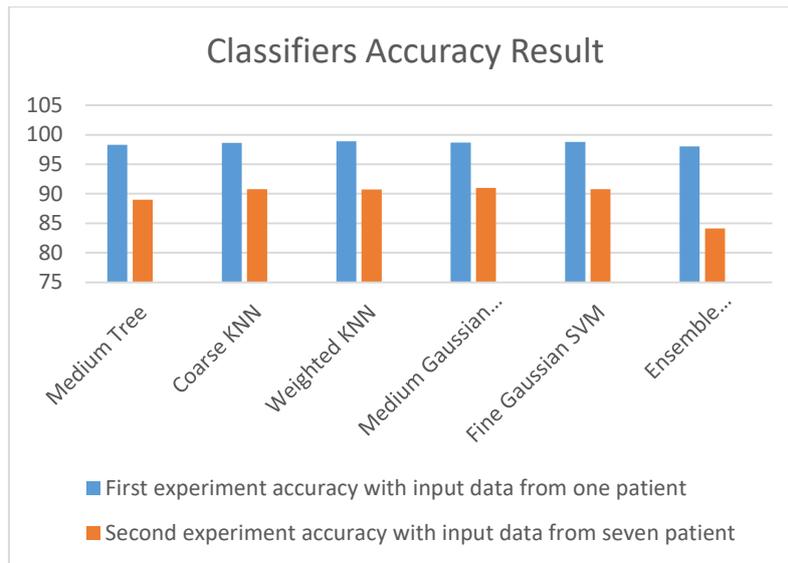


Figure 5.7: The classifiers for the two experiments and the accuracy for each of them.

In our experiments, we concentrate on classifiers' accuracy. The best accuracy for the first experiment was for the Weighted KNN classifier with accuracy = 98.9. Whereas, for the second experiment SVM Fine Gaussian classifier got accuracy to 91.1. However, Weighted KNN is the selected classifier in the first experiment because of the high achieved accuracy. For the second experiment, SVM Fine Gaussian classifier has been selected for the same reason. The selected classifiers will be used in subsequent steps and analysis. Regarding the other performance measures like the sensitivity and specificity, we have done an investigation and we found that the quality of the segmented images is more closely aligned to the value of the accuracy compared to the other performance measures.

5.3.2 Automatic Segmentation Result

In this section, we have reconstructed the images depending on the result of the classifier as discussed in the previous section. From the first experiment, we took the result of the Weighted KNN classifier for each x and y coordinate as an input to Matlab code which converts this matrix to image as shown in Figure 5.8. On the other hand, the result of the SVM Fine Gaussian classifier has been selected in the second experiment and used as an input with x and y

coordinate to the Matlab code to reconstruct the output image. Figure 5.9 shows the segmented image for SVM Fine Gaussian classifier.

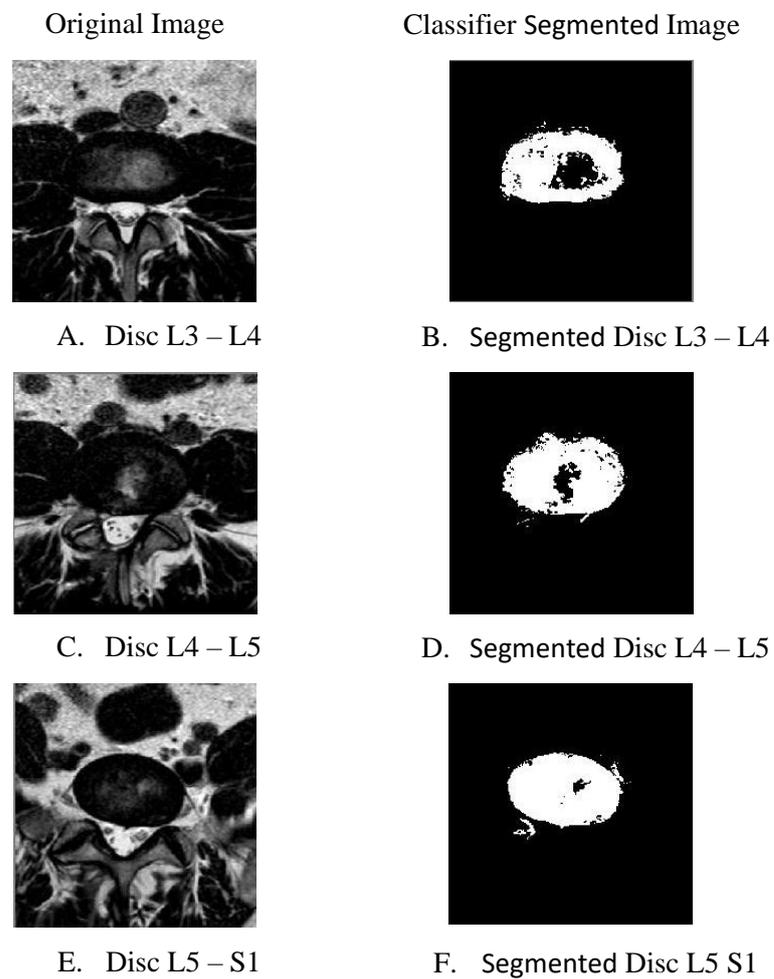


Figure 5.8: The Original and Segmented Images by Weighted KNN.

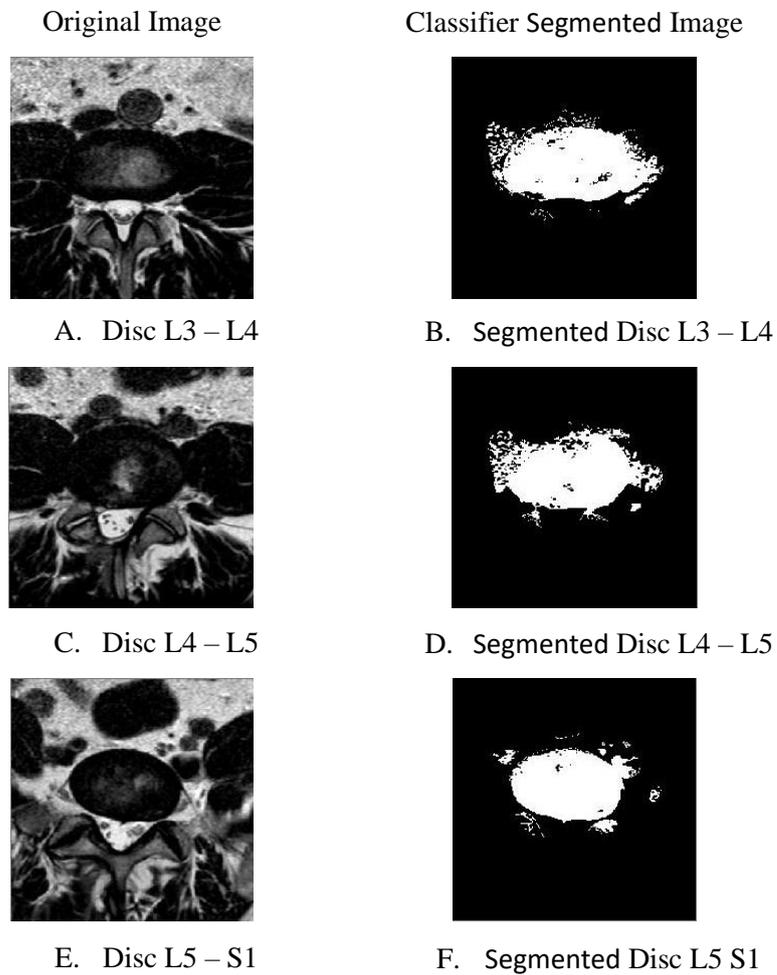


Figure 5.9: The Original and Segmented Images by SVM Fine Gaussian.

5.4 Detecting the Disc Herniation in Segmented Lumbar Spine MR Image using Centroid Distance Function

The main goal of this work is the automatic detection of disc herniation in lumbar spine MRI. The process involves automatic labelling of the disc area, taken from an axial view, using machine learning and artificial intelligence techniques and using the resulting segmented images to generate shape features for the discs.

As a proof of concept, we will use a manually-created mask image marking the area of the disc in the MRI. Figure 5.10 shows the masked regions for a non-herniated and herniated disc.

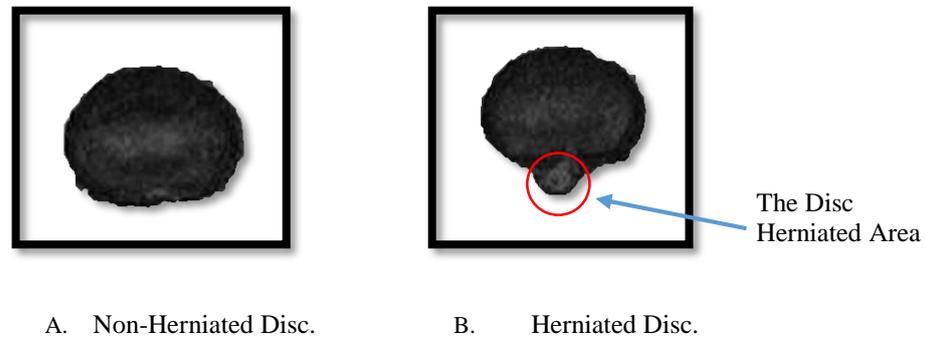


Figure 5.10: Axial view of (A) non-herniated and (B) herniated discs.

Our observation on a number of non-herniated and herniated discs yields a conclusion that if we measure the distance from the centre to the edge points of the disc taken over 360° and plot the distance values against the angle, we will get a distinctly different shaped curve in each case. The illustration for the process to calculate these distances is shown in Figure 5.11. This shape feature is known as the Centroid Distance Function (CDF) [101].

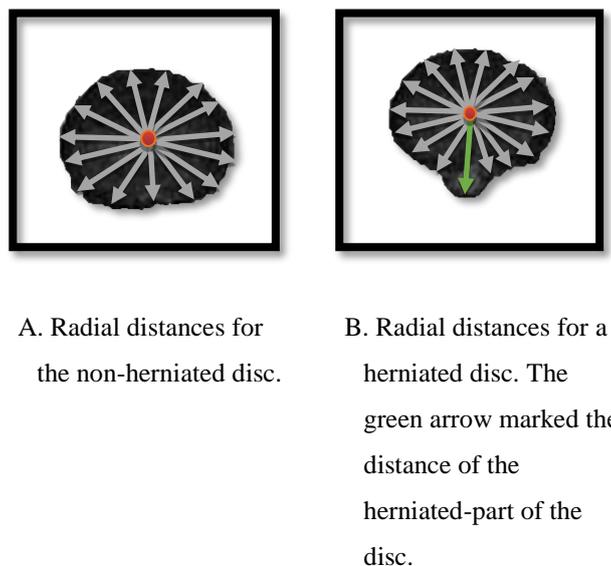


Figure 5.11: Illustration of the radial distances calculated from (A) non-herniated and (B) herniated disc.

The CDF is one of the shape signature functions [102] which is normally used as a tool for the preprocessing level for other algorithms [103] such as calculating the distance between boundary points from the centroid [104]. The position of the centroid is static relative to the shape. The average of all points inside the shape will be calculated to find the centroid. The region information is included in the centroid distance function which records the local and global feature. Moreover, the performance of using centroid distance function is very high based on [105]. To calculate CDF, the coordinate of the centroid of each disc needs to be calculated. This is followed by a boundary or edge detection process to calculate the coordinates of the disc's boundaries. The distance between each of the boundary points and the disc centroid as well as the angle are calculated as illustrated in Figure 5.12.

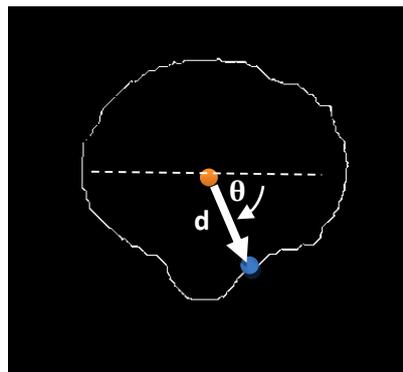


Figure 5.12: The distance between the centroid and the disc boundary (d) in the angle (θ).

The results are then plotted as a graph for visual inspection as shown in Figure.5.13 and Figure.5.14. Moreover, Figure. 5.13 shows the change of the distance for the non-herniated disc whereas Figure. 5.14 shows that for a herniated disc. We can clearly see how the distance is increased sharply in comparison to the distances before and after the highlighted area.

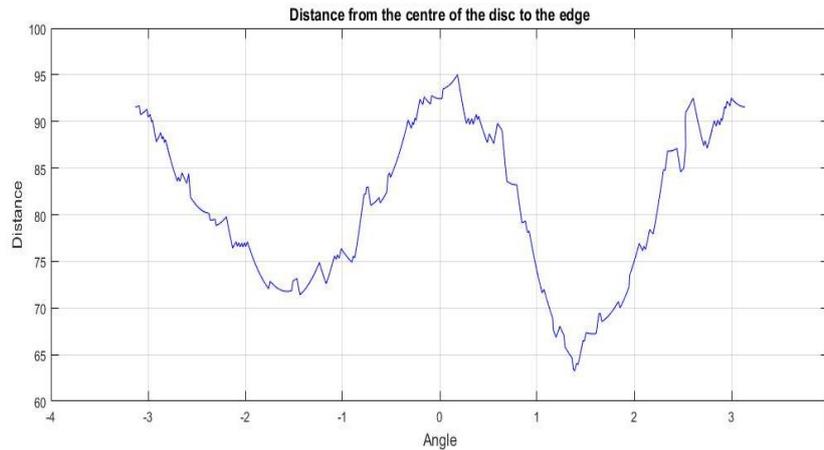


Figure 5.13: Plot of the Centroid Distance Function of a non-herniated disc.



Figure 5.14: Plot of the Centroid Distance Function of a herniated disc. The red region marked the area of the disc where the herniation occurs.

Although this finding has provided a useful indicator for herniation detection, the process requires a clear segmentation of disc area to produce reliable edge points from which the radial distances were measured. At this point, this is provided by manually segmenting the MRI to separate the disc and the non-disc areas.

However, manual segmentation of each patient's MRI would require an additional burden on the radiologists or orthopaedists. Therefore, this process needs to be automated by the computer by using machine learning and artificial intelligence. Unfortunately, the achieved segmentation of the disc area using the experiments discussed earlier is not accurate enough to apply the

centroid distance function methodology for detecting the disc herniation. Figure 5.15 shows the plot of the centroid distance function after using the automated segmentation of the disc area for a normal disc while Figure 5.16 shows the plot of the centroid distance function for an abnormal disc. As seen from the two plots there is no difference between the two figures for the normal and abnormal disc which required us to think widely about a new method in diagnosing the lumbar spine diseases. In the next chapter, we will discuss the method of segmenting the AAP in the axial view MRI and how this can lead to diagnosing the lumbar spine diseases.

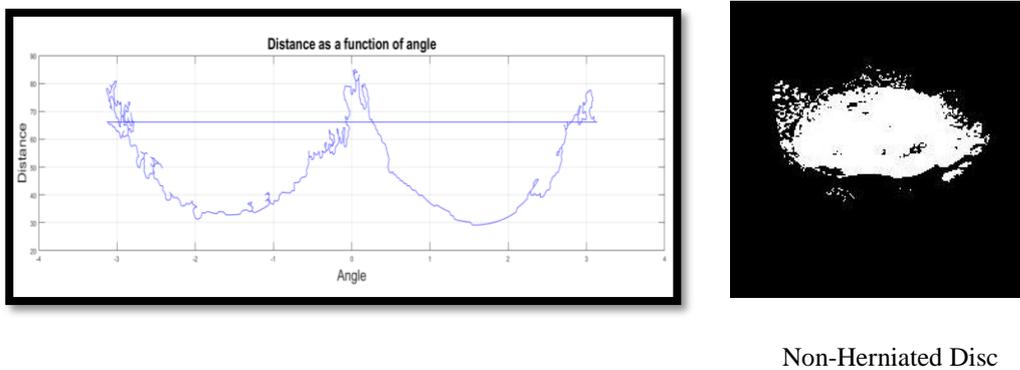


Figure 5.15: Plot of the Centroid Distance Function of a non-herniated disc.

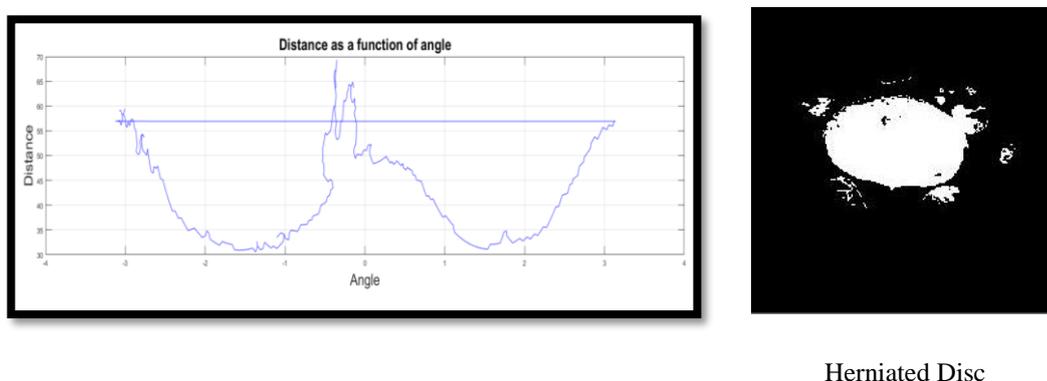


Figure 5.16: Plot of the Centroid Distance Function of a herniated disc.

5.5 Chapter Summary

The methodology of detecting the lumbar spine disc herniation using the centroid distance function works in theory but in practice, it has a problem as the segmented disc from the lumbar spine MRI images were not accurate enough to be used as an input image for the centroid distance function. However, because of the complexity of segmenting the disc, we developed a new methodology to segment the area AAP which will allow us to use another factor in the segmented area that will help in diagnosing different diseases from the lower back pain. The next chapter will discuss the method of segmenting the lumbar spine MRI images for stenosis detection.

Chapter 6 Lumbar Spinal Stenosis Detection via Segmentation of Area between Anterior and Posterior (AAP) Elements

Lumbar spinal stenosis is illustrated as a narrowing of the AAP that produces pressure on the spinal nerve canal or roots. An abnormal compression of either of them would exert pressure and create a sensation of pain. Traditional methods to image segmentation include clustering techniques based on pixel values such as k-means clustering, histogram-based clustering, and probabilistic-based model. However, current state-of-the-art image segmentation techniques are dominated by different types of Deep Neural Network to achieve semantic segmentation. There are two main methods of semantic segmentation namely patch-based pixel classification [106]–[108] using convolutional and fully connected layers, and whole-image segmentation [109]–[111] using fully convolutional and convolution-transpose layers. Figure 6.1 illustrates the methodology used for segmenting the AAP area in the lumbar spine MRI images. Some of the processes used in segmenting the AAP are used in segmenting the lumbar spine disc which is explained in detail in chapter 5. However, the other steps will be explained in detail in this chapter.

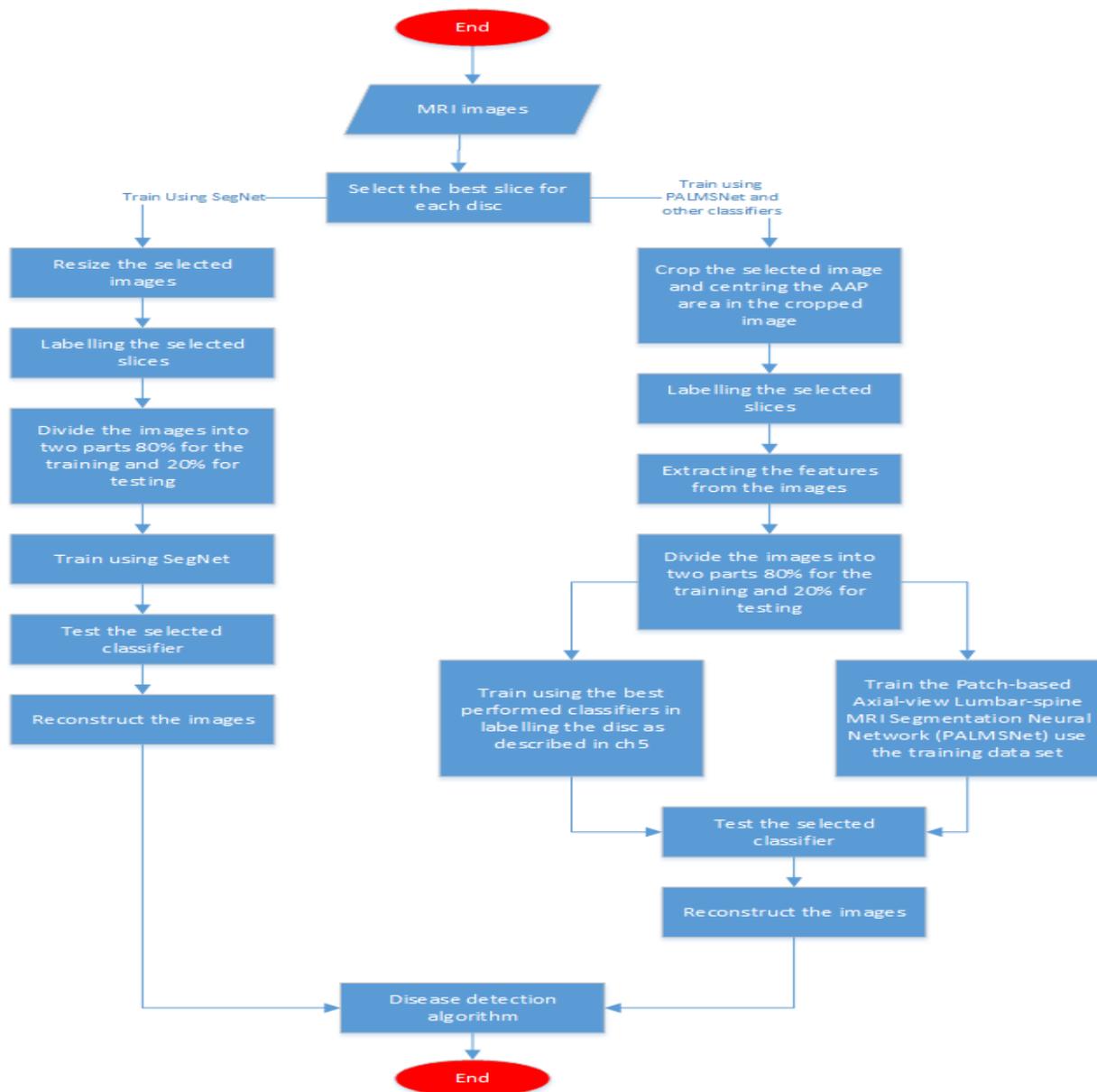


Figure 6.1: Methodology used for segmenting the AAP area in the lumbar spine MRI images.

6.1 MRI Image Cropping

In the previous chapter, we have used a full-size image to segment the lumbar spine disc while for segmenting the AAP the images are cropped to provide better focus on the areas around the IVD, AAP, and posterior element, which can help in labelling the ground truth data. Ground truth images were created with the assistance of physicians in this field by creating a binary

image marking the area of interest. An example of the MRI and ground-truth (label) image pair is shown in Figure 6.2.



Figure 6.2: An example of MRI and ground-truth image pair used for training.

Furthermore, we exploit the fact that all axial views of lumbar spine MRI images are always taken from the same orientation and similar scale to avoid having to augment the image patches synthetically.

In this section, we utilise a patch-based classification method to image segmentation to separate the AAP region from other tissues in lumbar spine MRI. For each pixel, p , in the input image, a patch centred at p is considered. The patch dimension, w , is set to an odd number to enforce symmetry. The network will classify each pixel based on the information contained in the patch into one of two possible classes namely AAP (0) and not-AAP (1). The process flow of image segmentation using this architecture is illustrated in Figure.6.3

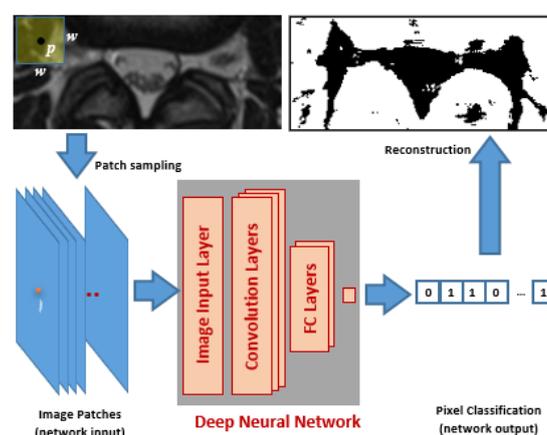


Figure 6.3: The process flow of image segmentation using patch-based pixel classification method.

6.2 Deep Neural Network Architecture

Our neural network shares a similar architecture to other patch-based deep neural networks. It includes an Image Input Layer, a cascade of Convolution Layers that will produce multiscale classification features and a Fully Connected Layer for classification of these features.

The Image Input Layer contains a 2D array of nodes that act as a receptor to each pixel in the patch. The dimension of this layer is identical to the dimension of the patch. A data normalisation step is applied to the pixel values by subtracting the mean pixel value before generating the output of this layer.

The output of the Image Input Layer is fed to a cascade of Convolution Layers. In our network, each convolution layer has a fixed number of equal-sized kernels. The kernel size chosen is relatively small since the input of the network is already a small subset of the entire image. Each kernel corresponds to each classification feature trained within the layer. The output of each of these layers is calculated as the dot product of the input and the kernel. As the signal is passed across these layers, it gets down-sampled hence the output of each layer covers an increasing support area of the input patch. Each layer, however, is set to have an increased number of kernels compared to the last. This decision is made to allow more features to be trained as the support area of the convolution layer increases.

To further speed up the training process and reduce the sensitivity of the network to its initial weight values, we opt to apply batch normalisation [112] [113] to the convolution layer output. This method works by subtracting the output of the previous convolution layer by the batch mean and dividing the result by the batch standard deviation. The increase in training speed and stability are achieved since the batch normalisation process reduces the variability of the input signal into two trainable parameters namely, the batch means and batch standard deviation.

As with other deep neural network architectures, we opt to use the Rectified Linear Unit (ReLU) activation function [114]. This decision is based on the function’s advantages over others, such as Sigmoid function, that include faster training speed due to a reduced likelihood of vanishing gradient. ReLU activation function also has been proven to train networks with sparser, and hence considered better, weights representation.

We set the convolution stride to one and apply an appropriate amount of pixel padding to its input to ensure the length of input and output of each convolution layer remains the same. A downsampling process is applied at the end of each convolution layer by taking the maximum of the activation signals inside a 2x2 non-overlapping region to produce a signal that has half the dimension. In total, our network uses three convolution layers to perform the feature weight training.

After the end of the last convolution layer, we use two Fully Connected Layers to learn all high-level combinations of the features learned by the earlier layers. Since the network will classify the central pixel in each patch into one of the two possible classes, the last layer will have two neurons. For classification, our network uses the softmax classifier. This classifier produces the probability of the input to belong to each class and can produce a more representative loss value when evaluating the network [115].

Table 6.1: The Architecture of PALMSNet

Layer	Type	Number of Neurons	Kernel Size	Number of Kernels
0	Image Input	25x25		
1	Convolution	25x25	3x3	16
2	Batch Normalisation	25x25		
3	ReLU	25x25		
4	Max Pooling	25x25		
5	Convolution	12x12	3x3	32
6	Batch Normalisation	12x12		
7	ReLU	12x12		
8	Max Pooling	12x12		

9	Convolution	6x6	3x3	64
10	Batch Normalisation	6x6		
11	ReLU	6x6		
12	Max Pooling	6x6		
13	Fully Connected	2304		
14	Fully Connected	2		
15	Soft Max	1		

For ease of reference, throughout the remainder of this thesis, we will refer to this network as PALMSNet which is an acronym for Patch-based Axial-view Lumbar-spine MRI Segmentation Neural Network. The overall architecture of PALMSNet is summarised in Table 6.1

6.3 Training

The training data consists of 400,680 image patches of size 25x25 pixels. They are taken from T2-weighted axial lumbar spine MRI images of the last three IVDs (D3-D5) of seven patients. The network is trained using the popular Stochastic Gradient Descent with Momentum algorithm [116] to update the network weights and biases. The algorithm works by taking small steps in the direction of the negative gradient of the loss function to minimize the error function. The size of the step is modulated by a learning rate parameter that is set to 0.001.

The training is performed on a personal computer in Windows 10 with i7-7700 CPU @ 3.60GHz, 64 GB RAM, and two NVIDIA Titan X GPUs. The training process takes 2,102 seconds and 9.5×10^4 iterations to complete.

Once trained, the network is used to classify pixels from test images. The test data consists of 171,720 image patches of size 25x25 pixels extracted from T2-weighted axial lumbar spine MRI images of D3-D5 of three patients. They are different patients from the seven whose MRI images are used for training. We carried out four different experiments to measure the performance of PALMSNet. The first experiment uses the images from all three IVDs for

training and testing whereas the other three use only those from each disc separately. This experimental set-up is used based on our hypothesis made during observation and analysis of the MRI images, that separation of the disc images during training will improve the classification results.

We will use four performance metrics that are common in other semantic segmentation evaluations [110]. They are pixel accuracy (a_p), mean accuracy (a_m), mean intersection over union (iou_m) and frequency weighted intersection over union (iou_{fw}). The formulas to calculate these metrics are shown in Eq. [6.1] to [6.4].

$$a_p = \frac{\sum_i n_{ii}}{\sum_i t_i} \quad [6.1]$$

$$a_m = \frac{\sum_i \frac{n_{ii}}{t_i}}{n_{cl}} \quad [6.2]$$

$$iou_m = \frac{\sum_i \frac{n_{ii}}{t_i + \sum_i n_{ii} - n_{ii}}}{n_{cl}} \quad [6.3]$$

$$iou_{fw} = \frac{\sum_i \frac{t_i \cdot n_{ii}}{t_i + \sum_i n_{ii} - n_{ii}}}{\sum_i t_i} \quad [6.4]$$

Where n_{ij} is the number of pixels of class i predicted to belong to class j , and n_{cl} is the number of classes and $t_i = \sum_j n_{ij}$ is the total number of pixels of class i .

For comparison purposes, we perform another image segmentation using SegNet [111], a fully convolutional neural net and using six other, but more conventional, image segmentation algorithms. Four of these six algorithms are variants of the decision tree (complex/medium [117]–[119], and ensemble [120], [121]). A further four are based on the modified k-nearest neighbour (medium, weighted, coarse and cubic [97], [122]–[124]) and the other two are based on modified SVM (fine and medium Gaussian [125]). The features used in these six

conventional algorithms are the raw pixel values and their corresponding locations. Pixel locations have been used in conjunction with pixel values to provide better spatial coherence to the segmentation result [126].

Table 6.2 Performance using All Intervertebral Discs

Methods	a_p	a_m	iou_m	iou_{fw}
Fine Gaussian SVM	0.78	0.77	0.62	0.64
Medium Gaussian SVM	0.79	0.78	0.64	0.66
Weighted k-nearest neighbour	0.77	0.77	0.62	0.64
Coarse k-nearest neighbour	0.79	0.79	0.64	0.66
Medium Tree	0.75	0.75	0.59	0.61
Boosted Tree Ensemble	0.76	0.76	0.60	0.62
SegNet	0.75	0.67	0.53	0.58
PALMSNet	0.86	0.85	0.75	0.76

The results of the segmentation as measured using the four performance metrics are shown in Table 6.2 to 6.5. The results shown in Table 6.2 to 6.4 indicate that PALMSNet performs significantly better than the other algorithms. The other DNN-based algorithm, however, has a similar level of pixel accuracy to the more conventional methods but performs slightly worse using the other three metrics. It is also important to note from Table 6.3 and 6.4 that the result of training and testing the network using only D3 or D4 data separately yielded significantly better results than training and testing using all of the discs together.

Table 6.3 Performance when trained and tested using D3 only

Methods	a_p	a_m	iou_m	iou_{fw}
Fine Gaussian SVM	0.81	0.77	0.62	0.70
Medium Gaussian SVM	0.79	0.78	0.60	0.68
Weighted k-nearest neighbour	0.79	0.78	0.61	0.69
Coarse k-nearest neighbour	0.80	0.78	0.61	0.69
Medium Tree	0.76	0.75	0.56	0.64
Boosted Tree	0.71	0.74	0.52	0.59
Ensemble				
SegNet	0.79	0.73	0.58	0.68
PALMSNet	0.93	0.92	0.85	0.87

Table 6.4 Performance when trained and tested using D4 only

Methods	a_p	a_m	iou_m	iou_{fw}
Fine Gaussian SVM	0.82	0.78	0.63	0.71
Medium Gaussian SVM	0.81	0.78	0.62	0.70
Weighted k-nearest neighbour	0.80	0.78	0.61	0.69
Coarse k-nearest neighbour	0.81	0.78	0.63	0.70
Medium Tree	0.77	0.75	0.57	0.65
Boosted Tree	0.79	0.77	0.59	0.67
Ensemble				
SegNet	0.73	0.69	0.53	0.57
PALMSNet	0.92	0.91	0.84	0.85

The last experiment results when the network is trained and tested using D5 only are shown in Table 6.5. The results also indicate the superiority of PALMSNet over the other methods. However, the difference between them is less pronounced than the previous three results. For example, only 0.02% difference in mean accuracy between the best of the methods and PALMSNet in Table 6.5 compared to 0.06% in Table 6.2.

Table 6.5 Performance when trained and tested using D5 only

Methods	a_p	a_m	iou_m	iou_{fw}
Fine Gaussian SVM	0.68	0.71	0.48	0.55
Medium Gaussian SVM	0.68	0.71	0.49	0.56
Weighted k-nearest neighbour	0.67	0.69	0.47	0.54
Coarse k-nearest neighbour	0.69	0.72	0.49	0.56
Medium Tree	0.65	0.65	0.44	0.52
Boosted Tree Ensemble	0.64	0.68	0.45	0.51
SegNet	0.60	0.62	0.43	0.43
PALMSNet	0.74	0.74	0.59	0.59

The results provide a mixed degree of validation to our initial hypothesis. While the decision to separate the disc yields much better results for D3 and D4, the same cannot be said for D5. As can be seen in Table 6.6 the network trained using data from all discs still yields better performance on D5 test data than the network that is trained using only D5 data.

Table 6.6 PALMSNet Performance when tested on D5

Methods	a_p	a_m	iou_m	iou_{fw}
PALMSNet trained using all discs	0.80	0.80	0.66	0.66
PALMSNet trained using D5 only	0.74	0.74	0.59	0.59

6.4 Discussion and Analysis

The experiment results presented in the previous section clearly show that PALMSNet can produce more accurate segmentation than the other techniques. However, while the segmentation metrics that are used provide us with a measure of how well each algorithm is in segmenting the MRI images, they do not provide a good visual indicator on how each of the algorithms fares in providing accurate delineation of the MRI images. This is because if we were to use the result of this segmentation to assist detection of lumbar spinal stenosis, the next step will rely heavily on the accuracy of the edges found around the areas shown in Figure 6.4 rather than the entire segmented image.

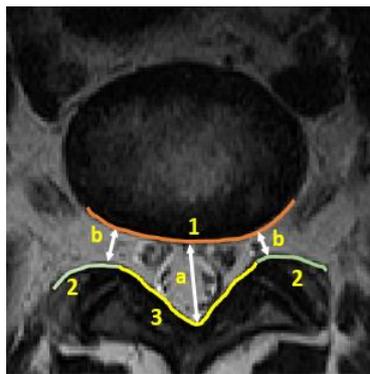


Figure 6.4: The three important boundaries for stenosis detection between AAP and 1) IVD, 2) Facet Joints, and 3) LF to measure a) the anteroposterior diameter of the spinal canal and b) the left and right width of the foramen

To provide better visual cues on the suitability of the network for this purpose, we show the result of superimposing the resulting edges with the ground truth. In the interest of conciseness of the discussion, we only show the result using PALMSNet trained on all discs. These are shown in Figure 6.5.

The images illustrated in Figure 6.5 confirm our finding in the previous section that the results are markedly worse when applied to D5 compared to the other two discs.

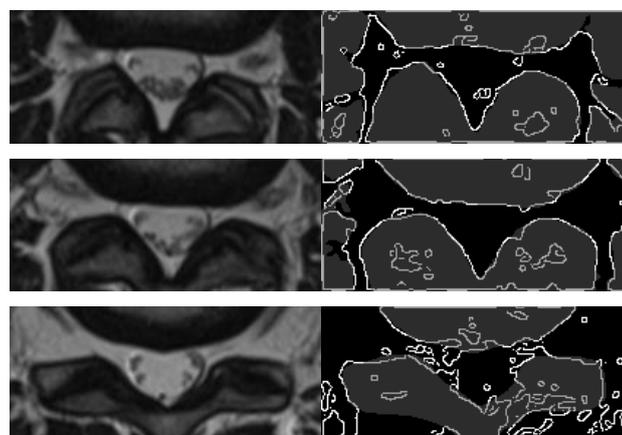


Figure 6.5: The result of superimposing the edges of the segmentation result on to the ground truth (right), and their corresponding original MRI images (left). The top images are taken from D3, middle from D4, and bottom from D5.

The simulation result indicated that PALMSNet produces significantly better segmentation results than SegNet, another deep network architecture. Further investigation into the low metric score for SegNet reveals that this network suffers greatly from overfitting to the training dataset which may be attributed to the high number of layers that the network has. The network overfitting is evidenced by the significantly high accuracy of the prediction on the training data (0.95) compared to that on the test data (0.75). As illustrated in Figure 6.6. A major challenge of using PALMSNet, in addition to the requirement to crop the image, is that we still couldn't train a large number of images due to the limitation of the resources in terms of the machine

capabilities and the time required to train using this algorithm. For this reason, in the next section, we will use the convolutional neural network to train the data.



Figure 6.6: Visual evidence of overfitting of SegNet in this experiment. The figure shows prediction on (left) a training image and (right) on a test image. The bright pixels outline the edges of the segmented image.

6.5 Two Regions Boundary Delineation through Semantic Segmentation

In this section, we evaluate the applicability of deep learning using CNN in performing semantic segmentation on our dataset. Prior to the advent of deep learning in computer vision, image segmentation is traditionally performed using clustering techniques such as k-means clustering [127], nearest-neighbour [128], support vector machine [129], and Texton Forest [130]. With the recent increase in popularity of deep learning in image classification, so is its use in performing pixel-wise classification. This gives rise to a special type of image segmentation, namely semantic segmentation [131], which surpasses other methods by a large margin in terms of accuracy and efficiency.

Image segmentation can be generally formulated as a discrete transformation function $f: n \rightarrow l$ for each pixel n where $n \in \mathbb{R}^3$ for colour images or $n \in \mathbb{R}$ for monochrome images to a label $l_n \in C$, where C is a set of all available classes. However, image segmentation techniques do not often use a singular value of a pixel to decide its label. Instead, the techniques consider a set of pixels surrounding and including the pixel n to take into account local pixel distribution and local texture properties. The result is a partition of the image into several parts, each of which has common low-level features such as colour distributions, texture properties, or smoothness

of boundary, etc. Semantic segmentation differs from traditional image segmentation because whereas traditional image segmentation is carried out without any attempt at understanding what these different parts represent, semantic segmentation, on the other hand, performs the partition into parts that share high-level meanings. Although semantic segmentation can be performed by classifying each pixel (i.e., pixel-wise classification), the process considers groups of pixels that belong to a semantically identical entity as one class regardless of their low-level properties. Semantic segmentation solves many key problems in the field of computer vision, paving the way towards complete scene understanding. Numerous applications such as self-driving vehicles, human-computer interaction, and medical image analysis benefit greatly from the advances in high-level image segmentation.

There are a number of proposed solutions to semantic segmentation including Fully Convolutional Networks [131], SegNet [111], DeepLab [132], and RefineNet [133]. We will focus on one of the most popular algorithms, namely SegNet, in carrying out the semantic segmentation of our dataset.

SegNet consists of a series of convolutional layers arranged in an encoder-decoder architecture. The architecture of SegNet's encoder network is topologically identical to the first 13 convolutional layers of the VGG16 network [134]. The input image is passed on to the first layer of the encoder that performs convolution with a trainable filter bank to produce the first set of feature maps. These feature maps are then batch normalised [112] before an element-wise Rectified Linear Unit (ReLU) function [114] is applied to them. The resulting signal is then applied to a max-pooling function with a 2x2 window and stride 2, non-overlapping window, before it is sub-sampled by a factor of 2. The result is then fed to the next set of convolution layer, batch-normalisation layer, ReLU layer, max pooling, and sub-sampling layer, and so on, until the end of the encoder section. The purpose of the encoder layers is to capture image features at varying resolutions. It is also intended to achieve translation

invariance over the various size of spatial image context at the cost of increasingly lossy, or inaccurate, boundary detail in its final output signal.

The process to restore this detail starts by up-sampling that output signal using max unpooling layer in the decoder. The decoder will need the memorised indices of the corresponding max-pooling process that was carried out previously in the encoder, to create sparse feature maps from the lower resolution input signal. These sparse feature maps are then convolved with a trainable decoder filter bank to produce dense feature maps. Similar to the step in the encoder, a batch normalisation step is applied to the dense feature maps before applying the ReLU function. The result is then fed to the next set of max unpooling and up-sampling layer, convolution layer, batch-normalisation layer, and ReLU layer and so on until the end of the decoder section. The role of the decoder is to map the low resolution, sub-sampled, feature maps to full resolution and dense feature maps, which then are pixel-wise classified using SoftMax function [115]. The signal propagation in the first encoder and the last decoder of SegNet is illustrated in Figure 6.7.

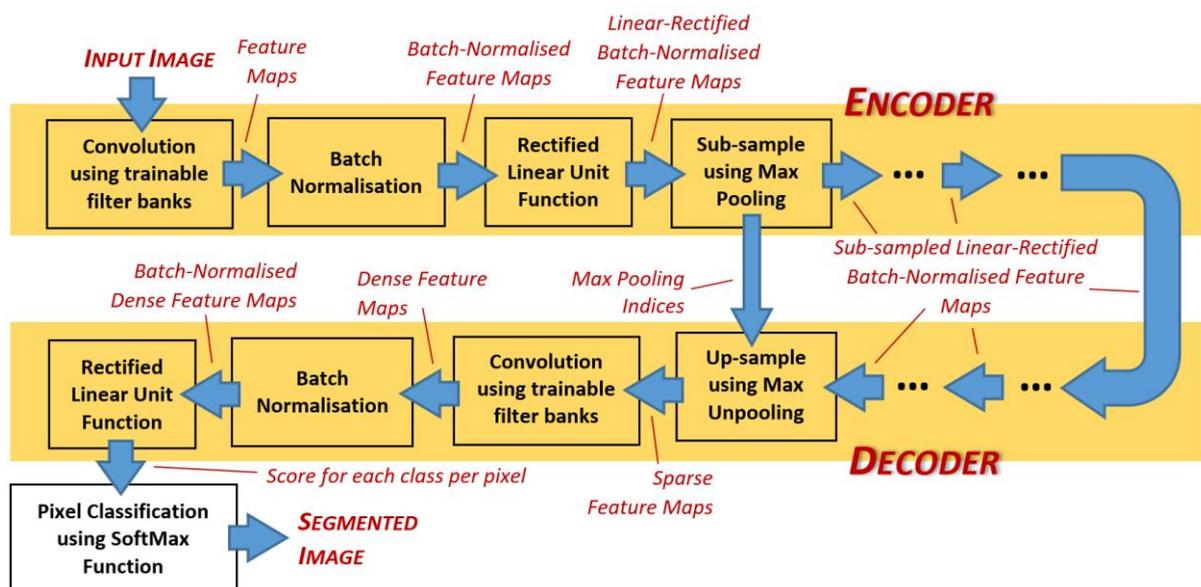


Figure 6.7: The process flow of SegNet's encoder-decoder architecture.

2D convolution of an image is computed from the sum of element-wise multiplication of image pixels and filter coefficients of a certain size. The VGG16 architecture is an improvement over AlexNet [135], an earlier deep network architecture, by replacing large filter size with multiple smaller-sized, 3x3 pixel, filters. This method increases the depth of the networks while at the same time lowering the number of trainable parameters. For example, a convolution of a single-channel image with a 5x5 filter can be replaced by two 3x3 filters with stride 1. The first method is implemented as one convolution layer with 25 parameters whereas the latter method is implemented as two convolution layers with 18 parameters. Note that the batch-normalisation and ReLU function activation still have to be applied at the end of each convolution while max pooling is only applied at the end of the last ReLU layer.

The training of SegNet adjusts the value of filter coefficients in all convolution layers so as to minimise the loss function between the resulting predicted segmentation and the ground truth. In this section, we compare the result of the segmentation of our dataset using two versions of SegNet. The first SegNet has pre-trained VGG16 coefficients which had been trained using more than a million images from the ImageNet database [136]. Since the type and number of classes in the ImageNet database do not match the type and number of classes in our case, we replaced the last classification layer and retrained the SegNet with our dataset. While the second SegNet has an identical architecture as SegNet – TL. However, the initial value of the weight and bias of the convolution layer were set using uniform distribution random number generator between -1 and 1. This allows to train the SegNet from scratch and as a result, we refer to this network as SegNet – FS.

When training both SegNets, we experimented using 80:20 ratios of training and testing images, we used the first number (i.e., training data percentage) to represent the ratio and referred to the parameter as Percentage of Training Data (PTD).

The network is trained using the popular Stochastic Gradient Descent with Momentum algorithm [116] to update the network weights and biases. The algorithm works by taking small steps in the direction of the negative gradient of the loss function, which is set as the cross-entropy of probability distribution of each class, in order to minimize the loss function. The size of the step is modulated by a learning rate parameter that is set to 0.001. The training is performed in small batches of 40 images each, for a maximum of 100 epochs. Due to the input size requirement of VGG16 network, the input images need to be resized to 360x480 beforehand. Figure 6.8 shows an example of the reconstructed images using SegNet using full-size MRI images. This reconstructed images using this algorithm have been used in addition to the reconstructed images using Fine Gaussian SVM to allocate the important points and measure the required distances to help in diagnosing the lumbar spine stenosis.

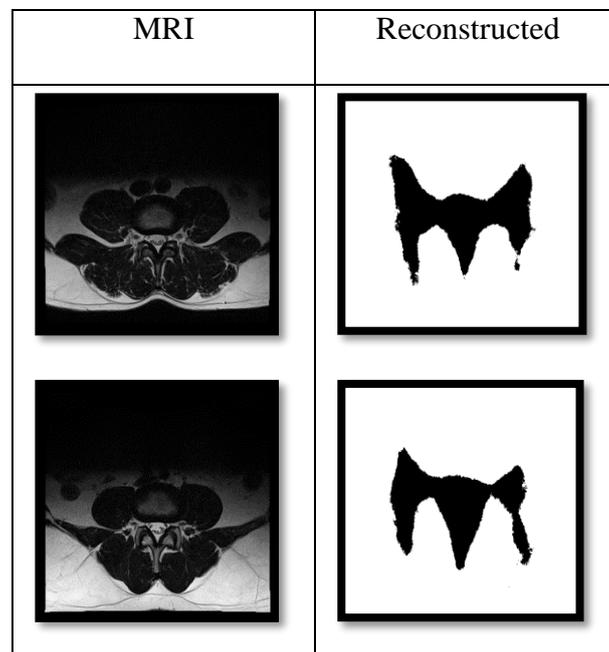


Figure 6.8 Example of the reconstructed images using SegNet using full-size MRI images

6.6 Locating the main points in the delineated boundaries

The manual diagnosing process for patients with spinal canal stenosis is done by calculating the spinal canal width which is labelled “A” as illustrated in Figure 6.9. The same distance will be calculated automatically using the reconstructed images from fine gaussian SVM and SegNet. The calculated distance will be compared with an approved distance for a healthy lumbar spine. The agreed anteroposterior diameter of the spinal canal width for the normal spine shouldn't be less than 15 mm [137]. The patient's MRI images will be used as an input to find the border of the AAP area as mentioned in Figure 6.10. An image will be reconstructed for the delineated area from the axial view MRI image for the lumbar spine, as shown in Figure 6.11. which illustrates the reconstructed images for 12 axial views MRI images for disc L3-L4, L4-L5, and L5-S1, which then will be used for further processes starting by calculating the important distance as is going to be explained later in this section.

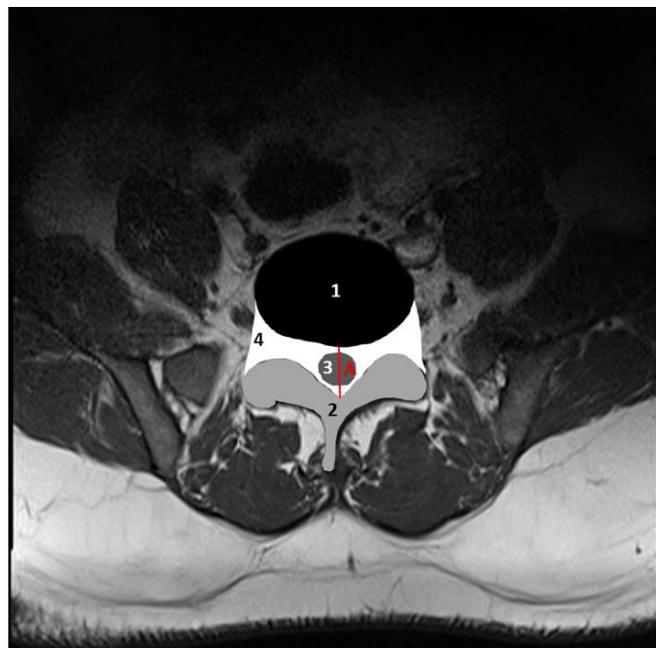


Figure 6.9: The four labelled Regions of Interest namely 1) Intervertebral Disc (IVD), 2) Posterior Element (PE), 3) Thecal Sac (TS) and 4) the Area between Anterior and Posterior (AAP) vertebrae elements. Delineation of boundaries between these regions is used to measure A) the anteroposterior diameter of the spinal canal width.

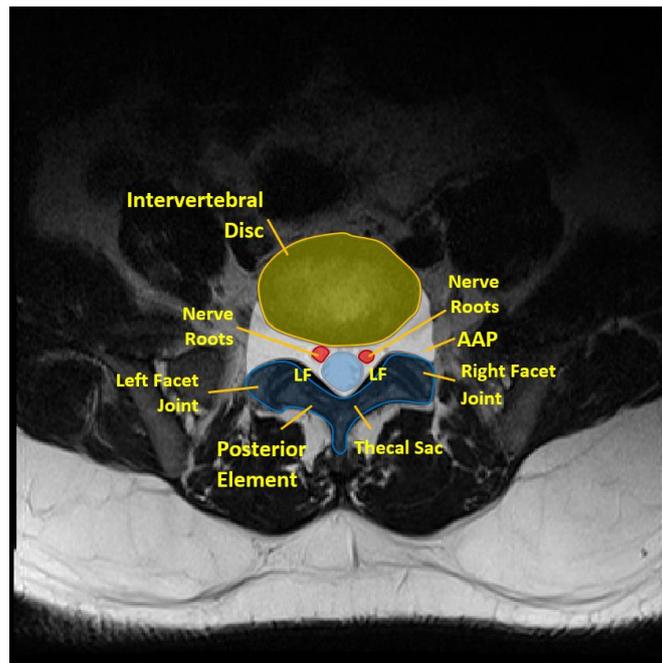


Figure 6.10: Axial View of the Lumbar Spine.

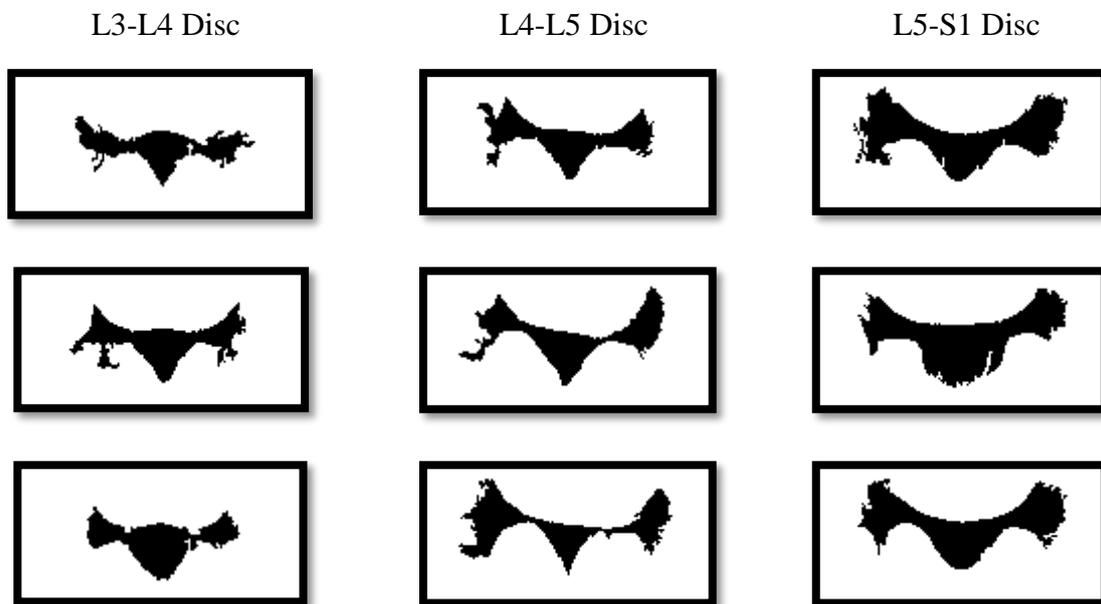


Figure 6.11: Sample of the reconstructed image for the AAP for L3-L4, L4-L5, and L5-S1

The results of the segmentation as measured using the four performance metrics are shown in Table 6.7. The results shown in Table 6.7 include all the data used in this research. Full-size images are used in these calculations for SegNet, whereas cropped images are used for the rest of the classifiers. Determining the important-points location for the automated lumbar spine diagnosis procedure has been done using the reconstructed images from Fine Gaussian SVM as the best performance classifier using cropped images and the SegNet has been used to complete the experiment in using a full-size image which does not require the radiologist or orthopaedist to crop the image to make sure that the AAP is in the middle of the image.

Table 6.7: The classifier's performance using all intervertebral discs.

Methods	Pixel Accuracy a_p	Cropped Image a_p	Mean Accuracy a_m	Cropped Image a_m	Mean Intersection over Union iou_m	Cropped Image iou_m	Frequency weighted Intersection over Union iou_{fw}	Cropped Image iou_{fw}
Fine Gaussian SVM		0.92		0.89		0.81		0.85
Medium Gaussian SVM		0.90		0.85		0.76		0.81
Cubic k-nearest neighbour		0.91		0.88		0.80		0.84
Medium k-nearest neighbour		0.91		0.88		0.80		0.84
Weighted k-nearest neighbour		0.91		0.88		0.80		0.84
Coarse k-nearest neighbour		0.91		0.86		0.79		0.83
Complex Tree		0.89		0.86		0.76		0.81
Medium Tree		0.87		0.81		0.71		0.77
Boosted Tree Ensemble		0.90		0.87		0.78		0.82
SegNet	0.94		0.90		0.81		0.90	
PALMSNet		0.70		0.53		0.41		0.53

Determining the important-points location for the automated lumbar spine diagnosis procedure will help to calculate the width of the spinal canal which is going to be used as the main factor for diagnosing the patients with spinal canal stenosis. As seen in Figure 6.12 the process to find the important points will be by dividing the image into two parts from the middle then find the

points A and B which indicate the shortest distance between all the points in the green line and all the points in the yellow line, The same procedure will be done to find the points C and D then to find Y as the lowest point between B and D. To calculate Z point we find the midpoint between A and C which can help to determine the X point as nearest point in the disc border to the point Z. Determining all these values will help to calculate the width of the spinal canal as a distance between X and Y to be compared with the spinal canal width for the healthy lumbar spine; as seen in Table 6.8 a sample of the calculated distances using the Fine Gaussian SVM to reconstruct the images is compared with the manually calculated distances for the three lumbar spine disc L3-L4, L4-L5, and L5-S1 respectively. Meanwhile, Table 6.9 shows the same calculations using SegNet reconstructed images.

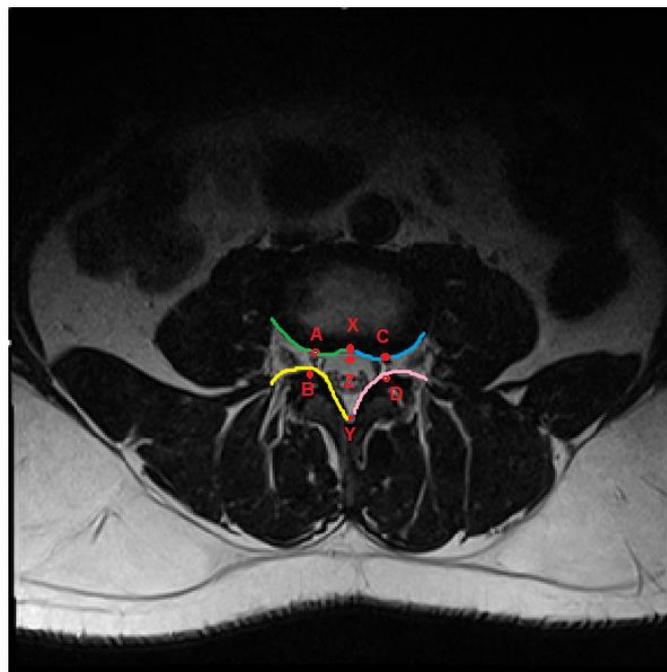


Figure 6.12: Important points locations

Table 6.8: Sample of the calculated distances using the Fine Gaussian SVM to reconstruct the images are compared with the manually calculated distances for the three lumbar spine disc L3-L4, L4-L5, and L5-S1 respectively

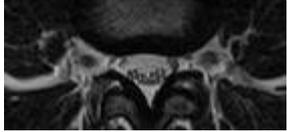
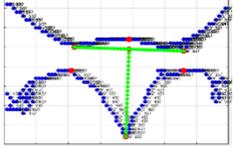
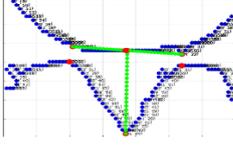
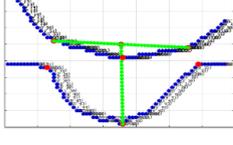
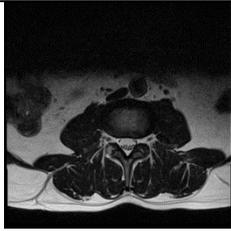
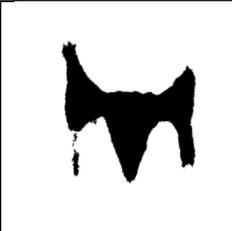
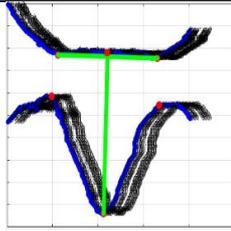
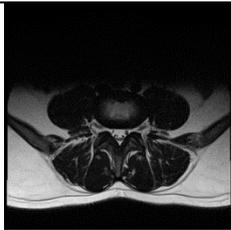
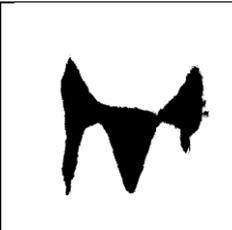
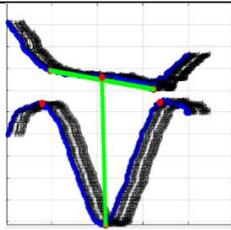
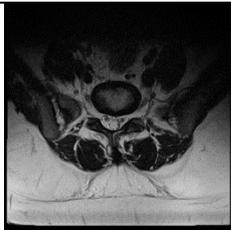
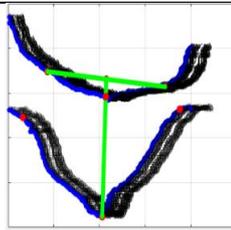
Original Cropped MRI Image	Reconstructed Image using Fine Gaussian SVM Classifier	Using Fine Gaussian SVM	Manual Calculate Distance for the Spinal Canal Width in MM	Auto Calculated Distance for the Spinal Canal Width in MM using Fine Gaussian SVM	Difference between the Manual Distance and Fine Gaussian SVM Calculated Distance
Disc L3 – L4					
			17.104	17.201	0.097
Disc L4 – L5					
			13.547	15.510	1.963
Disc L5 – S1					
			13.041	13	0.041

Table 6.9: sample of the calculated distances using SegNet to reconstruct the images are compared with the manually calculated distances for the three lumbar spine disc L3-L4, L4-L5, and L5-S1 respectively

Original MRI Images used in SegNet	Reconstructed Images Using SegNet	The located points using SegNet	Manual Calculate Distance for the Spinal Canal Width in MM	Auto Calculated Distance for the Spinal Canal Width in MM using SegNet	Difference between the Manual Distance and SegNet Calculated Distance
Disc L3 – L4					
			15.906	19.902	3.996
Disc L4-L5					
			13.547	17.255	3.708
Disc L5 – S1					
			13.041	17.127	4.086

A comparison between the automatic and manual calculated result is illustrated in Figure. 6.13 and Figure. 6.14 for the anteroposterior diameter of the spinal canal width. It has been found that 53% of the calculated distances using Fine Gaussian SVM were less than 1 mm different and 34% were less than 2 mm different. On the other hand, the calculated percentage for SegNet shows 21% only with distances less than 1 mm and 26% for distances less than 2 mm which indicate a better result in using Fine Gaussian SVM. That's related to using cropped images in

Fine Gaussian SVM and full-size images in SegNet which aim to fully automatize the diagnosing process. In addition, the percentage for a difference between 2-3 mm and 3-4 mm was 8% and 5% for Fine Gaussian SVM. Whereas, it increased to 21% and 16% for SegNet.

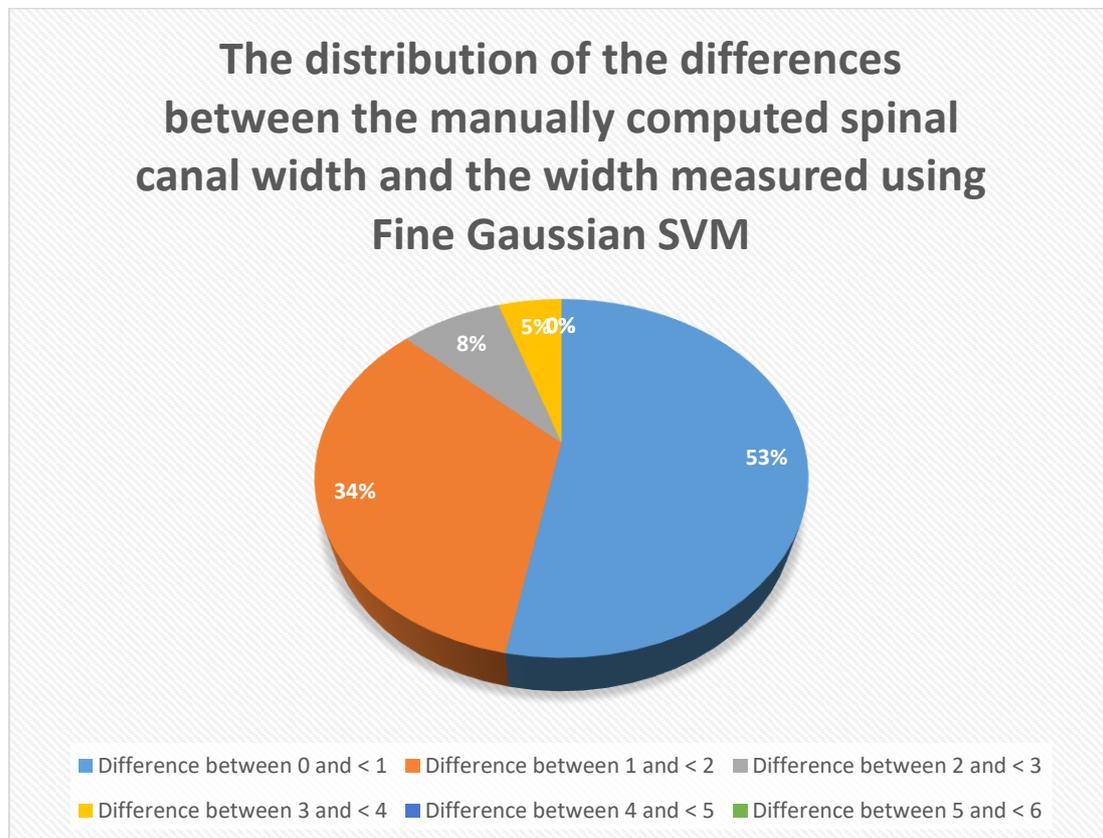


Figure 6.13: The distribution of the differences between the manually computed spinal canal width and the width measured using Fine Gaussian SVM

The distribution of the differences between the manually computed spinal canal width and the width measured using the SegNet

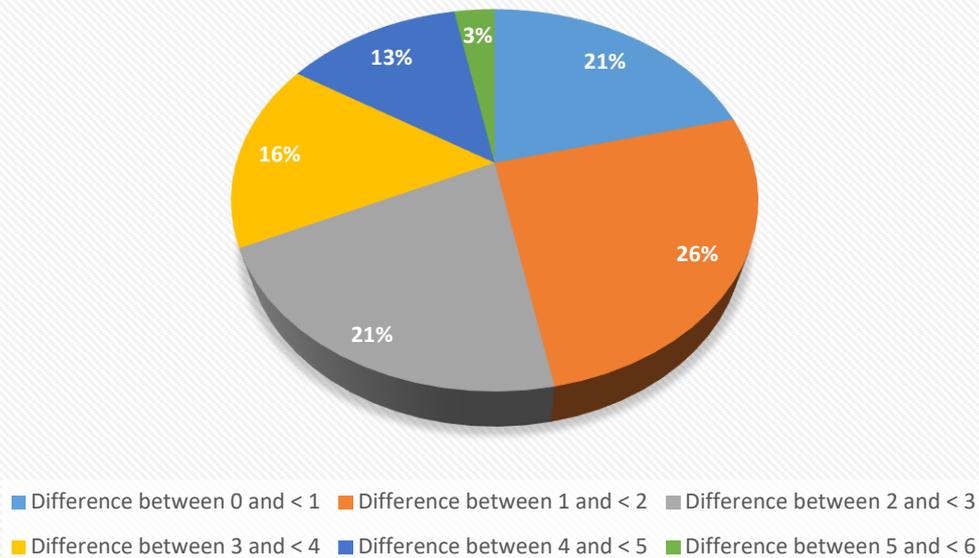


Figure 6.14: The distribution of the differences between the manually computed spinal canal width and the width measured using the SegNet

The calculated distances from Fine Gaussian SVM and SegNet are used as an indicator for a patient with spinal canal stenosis. A comparison has been done between the auto-calculated distance for the spinal canal width in MM (scw) and the approved distance for the healthy patients as follows:

For *L3-L4* and *L4-L5*

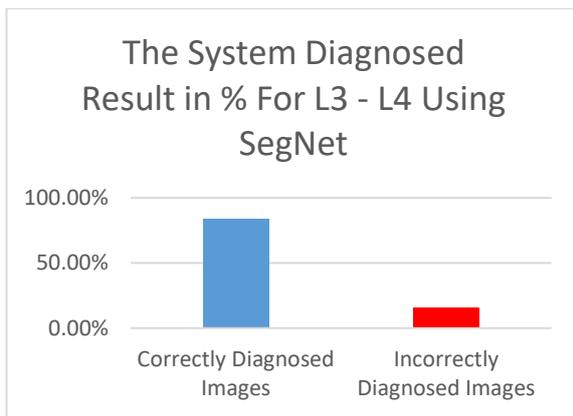
If $scw \leq 15$ patients will have a probability of spinal canal stenosis.

For *L5-S1*

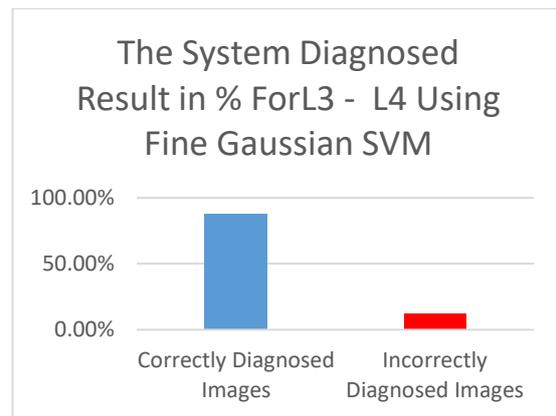
If $scw < 15$ patients will have a probability of spinal canal stenosis.

The results we achieved are shown below in Figure 6.15 where we can find that 84% of the images we correctly diagnosed using SegNet for disc L3-L4 as illustrated in Figure 6.15. (A), while the Fine Gaussian SVM reached 88% for the same disc as shown in Figure 6.15. (B).

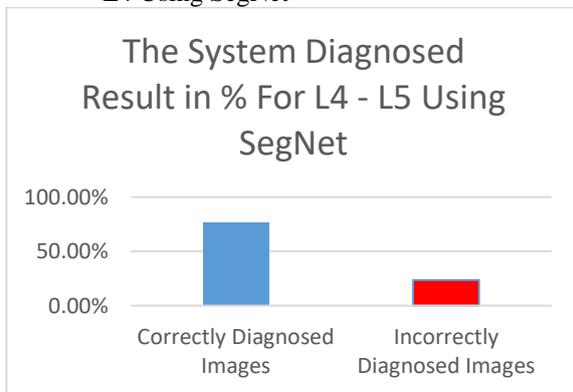
Meanwhile, Figure 6.15.(C) and (D), show that the success in the diagnosing for disc L4-L5 slightly decreased for both classifiers because the accuracy of the segmentation of the AAP area was lower than the other discs, which reduced the accuracy for the diagnosing to reach 76.5% and 82.4 for SegNet and Fine Gaussian SVM respectively. However, the diagnosing accuracy is increased in disc L5-S1 to 90% for both classifiers as demonstrated in Figure 6.15.(E) and (F).



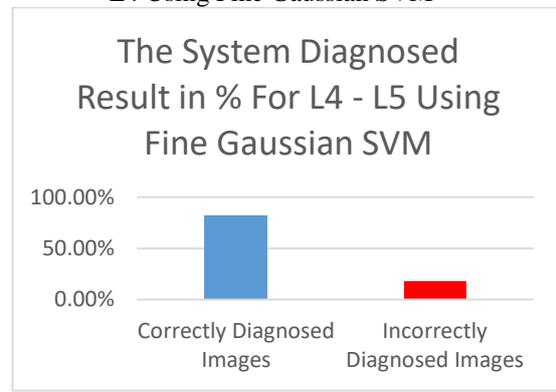
(A) The System Diagnosed Result in % For L3 - L4 Using SegNet



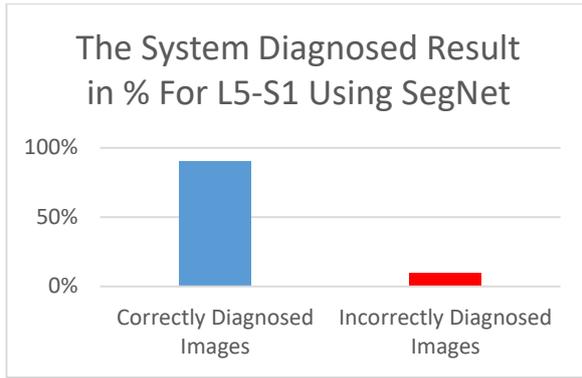
(B) The System Diagnosed Result in % For L3 - L4 Using Fine Gaussian SVM



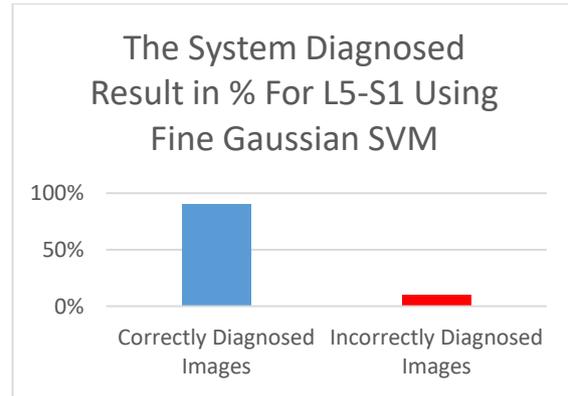
(C) The System Diagnosed Result in % For L4 - L5 Using SegNet



(D) The System Diagnosed Result in % For L4 - L5 Using Fine Gaussian SVM



(E) The System Diagnosed Result in % For L5-S1 Using SegNet



(F) The System Diagnosed Result in % For L5-S1 Using Fine Gaussian SVM

Figure 6.15: (A) – (F), Illustrate the System Diagnosed Result in % for L3-L4, L4-L5 and L5-S1 Using SegNet and Fine Gaussian SVM.

6.7 Chapter Summary

In this chapter, the lumbar spinal stenosis has been detected via segmentation of the area between anterior and posterior (AAP). Segmenting the AAP area helped us to determine the important-points location for automated lumbar spine stenosis diagnosis procedure. The determined points are used to measure the width of the spinal canal and compare it with the spinal canal width for the normal lumbar spine. Segmenting the AAP will produce an output image having only two regions that can be used to detect a limited number of diseases. However, in the next chapter, the number of segmented regions has been expanded to segment four different important regions which are: Intervertebral Disc, Posterior Element, Thecal Sac and the AAP in the lumbar spine MRI images. Having all these regions segmented can help in diagnosing more diseases other than the lumbar spine stenosis such as disc herniation and disc degeneration.

Chapter 7 Semantic Segmentation using Convolutional Neural Networks

As I have discussed in chapter 6 labelling the lumbar spine MRI image can help in diagnosing spinal canal stenosis. However, there are other parts in the image that can be also segmented and help the clinician in diagnosing more diseases such as disc herniation and disc degeneration. In this chapter, the methodology and the result of using the Semantic Segmentation using Convolutional Neural Networks for labelling the Intervertebral Disc, Posterior Element, Thecal Sac, and AAP will be discussed. Figure 7.1 illustrates the methodology followed in the segmentation using the convolutional neural network which includes some similar steps as the previous two methodologies such as selecting the best slice and feature extraction. Whereas, the other steps will be discussed in detail in this chapter. In this method, the labelling has been done for 515 MRI image using different experts and the results of all experts are analysed.

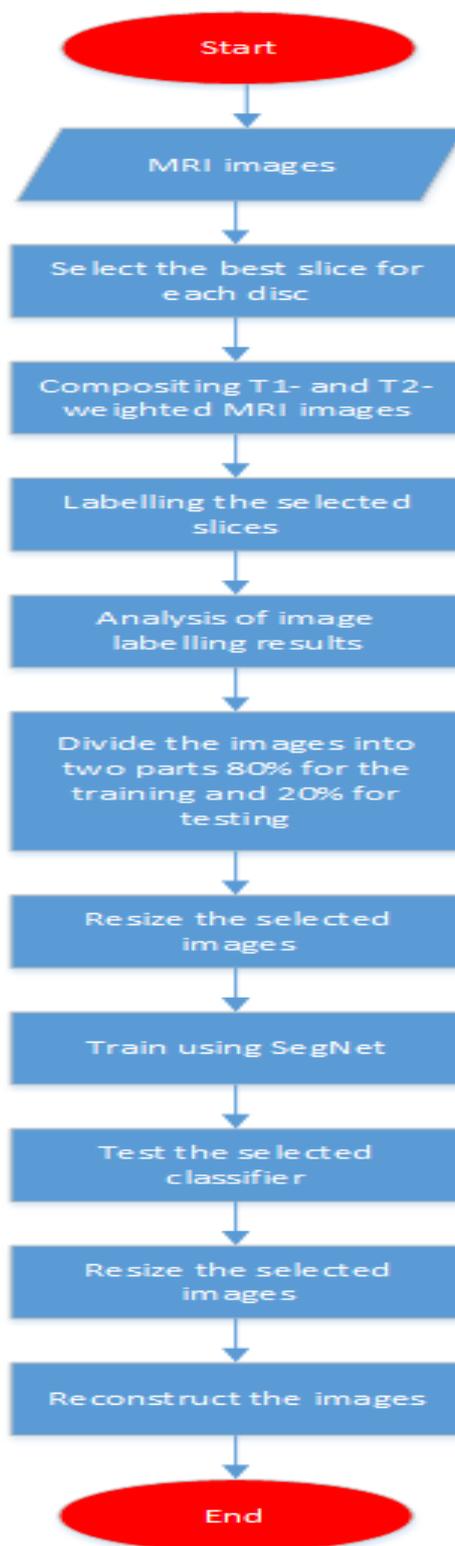


Figure 7.1 Shows the methodology followed in the segmentation using the convolutional neural network

7.1 SegNet Training Setup

When training both SegNets, we experimented using different ratios of training and testing images, namely 20:80, 40:60, 50:50, 60:40, 70:30, 75:25, and 80:20. For convenience, we use the first number (i.e., training data percentage) to represent the ratio and refer to the parameter as a Percentage of Training Data (PTD). By evaluating the output of a network when it is trained using varying values of PTD we will be able to evaluate how well increasing training data ratio improves the network's segmentation performance.

The network is trained using the popular Stochastic Gradient Descent with Momentum algorithm [116] to update the network weights and biases. The algorithm works by taking small steps in the direction of the negative gradient of the loss function, which is set as the cross-entropy of probability distribution of each class, in order to minimize the loss function. The size of the step is modulated by a learning rate parameter that is set to 0.001. The training is performed in small batches of 40 images each, for a maximum of 100 epochs. Due to the input size requirement of VGG16 network, the input images need to be resized to 360x480 beforehand.

Additionally, due to class population imbalances, we use class weighting to balance the importance of the classes. To make sure that small classes, such as the TS and AAP, are not under-represented in our training data we set the class weighting to be inversely proportional to the class population.

The training is performed using MATLAB on a personal computer in Windows 10 with i7-7700 CPU @ 3.60GHz, 64 GB RAM, and two NVIDIA Titan X GPUs. The time taken to complete the maximum number of training epochs is, as expected, linearly dependent on the number of training images used. On average it takes about 18.6 seconds per image to complete

the training and it is also worth noting that there is very little difference between the average training times of each SegNet network.

7.1.1 Image Labels and Ground Truth

The labelling achieved by using Canny edge detector, using a combination of two threshold values and two Gaussian sigma values, leads to our conclusion that the three boundaries cannot be detected merely using edge detection alone. They should be located more reliably by first performing segmentation on the MRI image to separate the AAP and other regions from the rest of the lumbar spine areas before applying the edge detection algorithm. However, medical image segmentation itself possesses its own unique challenges. One of the major difficulties in medical image segmentation is the high variability in medical images which is caused by the variability in human anatomy itself, the severity of the illness, the effect of age and gender, and also the intrinsic factors of the equipment such as calibration and sensitivity. To overcome these difficulties, we use a deep neural network to perform the segmentation because of the technique's widely acknowledged ability to take into account these variabilities [138]. The ground truth used for training and testing of the deep learning algorithm consists of labelled axial-view MRI slices of the IVDs [139]. It is important to note that we do not use the slices of all lumbar IVDs, but instead, we use the slice of the last three only. The rationale of this was provided in our previous explanations.

The labels in the ground truth mark several regions of interest (RoIs). Since lumbar spinal stenosis occurs inside the AAP, we only focus on parts of the MRI which contain and are around this region. Subsequently, we decided to have four RoIs which are a) the IVD, b) the PE, c) the TS, and d) the AAP itself. Any other pixels that do not belong to any one of the above four regions are labelled as e) other. The labelling of these four regions is illustrated in Figure 7.2.



Figure 7.2 The four labelled RoIs namely 1) Intervertebral Disc, 2) Posterior Element, 3) Thecal Sac and 4) the AAP.

The labelling process is carried out using the T1-weighted axial-view MRI slice of the last three IVDs. The task of manually labelling the four areas on each MRI slice is a laborious one. On average, five to ten minutes are spent to label each slice. This provides a significant challenge for us if we were to use the highly valuable expert's time to perform it. As an alternative, we opted to use the expert's time to label several MRI images as examples and use them to guide non-experts when labelling the dataset. We also use the examples that the expert has created to select the best results from the output. The steps to develop the ground truth data are detailed below:

1. We use a dual-view set-up on a DICOM/MRI viewer to select the images as illustrated in Figure 7.3. On the mid-sagittal (left) view, we observe the cut line (yellow line) of the corresponding axial-view slice on the right. The best axial-view slice of a disk, defined as one which cuts closest to the half-height of the disc, is then selected.

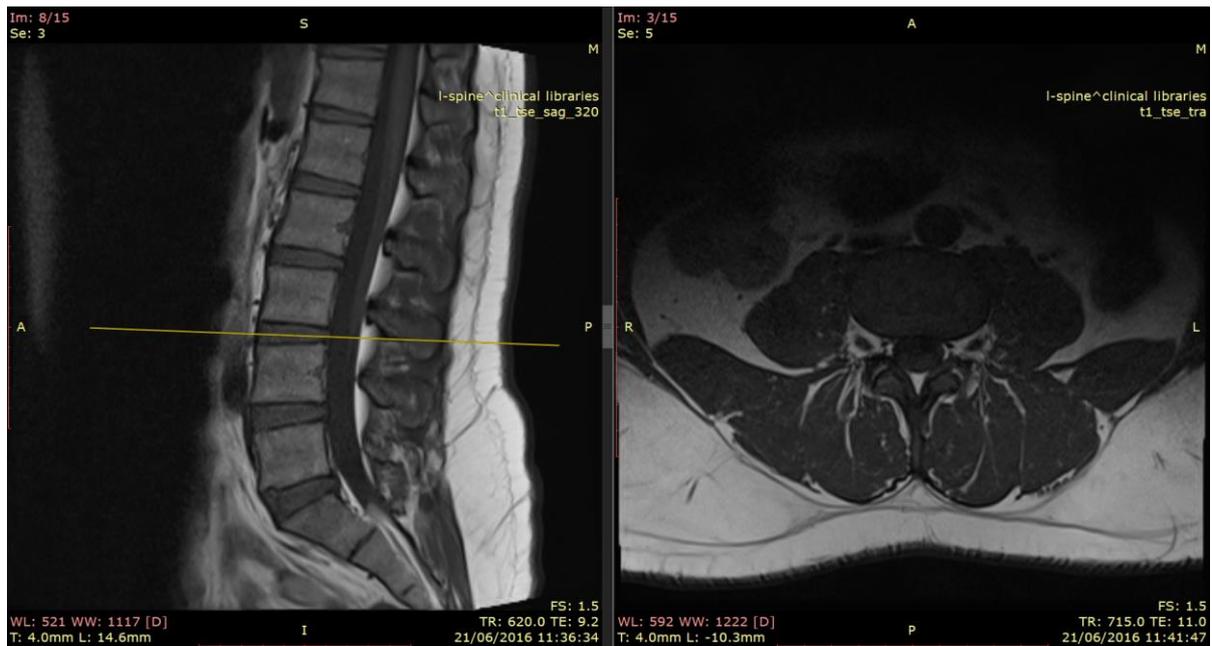


Figure 7.3 Dual-view of MRI images showing the mid-sagittal view (left) and the corresponding axial-view slice (right) at the cut line.

2. We then extract the three best T1-weighted MRI slices, one from each of the lowest three IVD, i.e., L3-L4, L4-L5, and L5-S1, from each patient's MRI study. For brevity, we will refer to these selected slices as the images from now on.
3. We asked our expert to label the four RoIs on ten patients (30 slices in total) as examples. The ten selected images are chosen in such a way to provide as best representation to the rest of the dataset as possible.
4. We use these example images to train several undergraduate students, as project participants, to perform the same task. The training runs twice with a test at the end to assess their suitability. In the end, fourteen participants were recruited.
5. We split the database into two groups and assigned seven participants to label all images in each dataset group independently. Their results are regularly checked by the expert to ensure

a good level of consistency and accuracy. In the end, we have seven batches of labelled images per group.

6. From each group, the five best batches were selected by the expert to be used for subsequent stages.

7.1.2 Ground Truth Quality Metrics

Compared to other topics in computer vision, little formal or analytical work has been published to guide the creation of ground truth data. There is some guidance [140], [141] provided by the machine learning community for measuring the quality of ground truth data used for training and test datasets, but this tends to revolve only around the size of the dataset [142]. To address this issue, we propose a novel method to assess ground truth quality, not from the size of the dataset but through calculating its confidence and consistency levels to measure its accuracy and variability, respectively. The detail of this method has been presented in a limited form in our previous publication [143]. In the interest of completeness, in this section, we will describe the algorithm again and provide an updated result when the metrics are calculated using the final MRI dataset.

We define the confidence level of ground truth data as a sureness measure that all labelled regions contain all the pixels that should be in that class and nothing less. On the other hand, we define consistency level as to how varied the ground truth data is across its sources. To measure the confidence level of the resulting labelled images we use a variant of the Jaccard Index, which is also known as the *intersection-over-union* metric [131]. The intersection-over-union (iou_c) of class c is calculated as the ratio between the number of correctly predicted pixels (intersection) and the sum (union) of the number of correctly and incorrectly predicted pixels.

$$iou_c = \frac{m_{cc}}{t_c + m_c - m_{cc}} \quad [7.1]$$

Where m_{cc} is the number of pixels of class c correctly predicted to belong to class c , and t_c is the total number of pixels of class c – according to the ground truth, and m_c is the total number of pixels predicted to belong to class c . An ideal classifier would classify correctly all pixels that belong to class c as that class (i.e., $m_{cc} = t_c$) and only those pixels (i.e., $m_{cc} = m_c$), resulting in $iou_c = 1$. Since in this case we do not have and are still developing the ground truth, therefore by definition, the value of both m_{cc} and t_c cannot be determined. Hence, we will develop an alternative intersection-over-union metric, denoted as iou'_{cv} to estimate iou_c , which will be used as a measure of confidence of the ground truth.

Consider a set C , defined as $C = \{1, 2, 3, 4\}$, of the four classes or RoIs. A pixel, n , can be labelled by a participant p , where $p \in \{1, 2, 3, 4, 5\}$, as l_{np} where $l_{np} \in C$. We define a vote count, k_{nc} , as the number of votes from all five participants that assign class c to pixel n , where $c \in C$.

$$k_{nc} = \sum_p [l_{np} = c] \quad [7.2]$$

Where $[z]$ is the Iverson Bracket notation of logic proposition z , i.e., $[z] = 1$ if z is true or 0 otherwise. The vote count has values in the range of $0 \leq k_{nc} \leq 5$, e.g., $k_{n2} = 0$ means the pixel n receives zero vote that assigns class 2 to it and $k_{n3} = 5$ means the pixel n receives all five votes that assign class 3 to it.

Next, we define the intersection of c -labelled regions, denoted as s_{cv} , as the normalised number of pixels that receive at least v number of votes that assign class c . Here, we refer v as the vote-threshold.

$$s_{cv} = \frac{1}{n} \sum_n [k_{nc} \geq v] \quad [7.3]$$

Note that s_{c1} is the normalised number of pixels that receive at least one vote that assigns class c . This represents the union of all pixels receiving a non-zero number of votes for that class. Therefore, these pixels will serve as the union, or denominator, in our alternative intersection-over-union iou'_{cv} calculation. Another important fact to consider is that for $\forall c$, the following composite inequality applies:

$$s_{c1} \geq s_{c2} \geq s_{c3} \geq s_{c4} \geq s_{c5} \quad [7.4]$$

Based on the above argument, we define our alternative intersection-over-union metric iou'_{cv} of class c and vote threshold v as,

$$iou'_{cv} = \frac{s_{cv}}{s_{c1}} \quad [7.5]$$

Substituting the equation to the above inequalities we have the following relationship:

$$1 \geq iou'_{c2} \geq iou'_{c3} \geq iou'_{c4} \geq iou'_{c5} \quad [7.6]$$

Hence, the closer the value of iou'_{cv} is to unity for all vote thresholds the better in-agreement the five participants are in labelling the region of class c .

7.1.3 Compositing T1- and T2-weighted MRI images

The labelling of the MRI slices used T1-weighted images because it provides us with the ability to identify and locate the TS region. However, the information in its corresponding T2-weighted image is also as important as that contained in the T1-weighted image. The combination of the two will provide better and richer discriminating features when carrying out the segmentation process.

Despite the fact that both T1-weighted and T2-weighted images have identical dimensions, not all pixels at the same location in both images correspond to the same point in an organ or tissue.

We have observed a wide gap, typically between 1 and 9 minutes, between the time data recorded on the T1-weighted and T2-weighted scans. The big time difference suggests that corresponding T1-weighted and T2-weighted MRI slices may not necessarily align.

The process to align the two images begins by fixing one of the images and transforming the other to match the first image. We set the T1-weighted image as the fixed image because we used them when constructing our image labels. A set of affine transforms, i.e., a linear combination of translation, rotation, scaling, and shearing, are then applied to the T2-weighted image to produce transformed images and calculate the difference between them and the fixed image. The whole process is known as image registration, which is essentially an error-minimisation problem over a search-space. In our experiment, we set the minimum and maximum limit to the radius of this search-space to $1.5e-6$ and $13e-3$, respectively with a growth factor of 1.05. To avoid a long or an indefinite search time, we limit the number of searches to 300 iterations. It is also expected that both modalities are affected by both high-frequency noise and low-frequency inhomogeneity field. To counter the latter, a parametric bias field estimation is applied before being corrected using PABIC method [144]. This method also employs a search optimisation algorithm called (1+1)-Evolutionary Strategy that locally adjusts the search direction and step size while at the same time providing a mechanism to step out of non-optimal local minima.

In the event that the registration process fails to converge, we perform a manual inspection on the images. It is very likely that it is a result of large discrepancies between the two modalities which may be caused by significant movement by the patient during the MRI scan. If that is the case, we will treat the images as unusable and remove the pair from our dataset.

Once the image registration is completed, a composite 3-channel image is created from the T1-weighted and T2-weighted slices. The first channel is the original T1-weighted image whereas

the second channel is constructed from the aligned T2-image. The third channel can be either a linear or non-linear combination of the two. We have experimented with different functions, including Manhattan distance, Euclidean distance, Mahalanobis distance, and polynomials and settled with the simplest one which is the Manhattan distance of the images. The registration process may also produce a set of pixels locations where T1-weighted pixels have no correspondence with any of the registered T2-weighted pixels. To accommodate the classification of these pixels, we create a new class of pixels in addition to the five classes we already have. We refer to this new set of pixels as ‘Unregistered’. Next, we use these images for training and testing a Convolutional Neural Network (CNN) to perform automatic semantic segmentation on them.

7.2 Development of Ground Truth Data for Automatic Lumbar Spine MRI Image Segmentation

7.2.1 Analysis of image labelling results

We calculate the values of iou'_{cv} for the four regions of interest in 1545 (three IVDs in each study of 515 patients) MRI images in our dataset and show the results in Table 7.1. We will next discuss and analyse the result of each individual region.

Table 7.1 Intersection-over-union values of different vote-threshold groups.

Regions (label/class)	iou'_{c2}	iou'_{c3}	iou'_{c4}	iou'_{c5}
Intervertebral Disc (1)	0.96	0.93	0.91	0.87
Posterior Element (2)	0.89	0.82	0.76	0.66
Thecal Sac (3)	0.87	0.81	0.74	0.66
The AAP (4)	0.66	0.48	0.34	0.21

The MRI scans in our dataset show varying degrees of severity. Some scans, for example, show strong evidence of intervertebral disc collapse while some others show a non-existent gap

between the intervertebral disc and posterolateral arch. The condition in these cases is so severe that we could not reliably label the images with a high degree of confidence.

7.2.2 Individual Region Analysis

The analysis of the labelling results is done by correlating the calculated iou'_{cv} for the four regions of interest with the heat map of the vote each pixel has as shown in Figure 7.4. In this figure, pixels with the most votes are rendered the brightest and vice versa. High contrast between a region and its surrounding suggests stronger vote confidence whereas a blurry boundary suggests a presence of ambiguities. The images are cropped tightly to the size of each region of interest to maximise the visual clarity of the heat map.

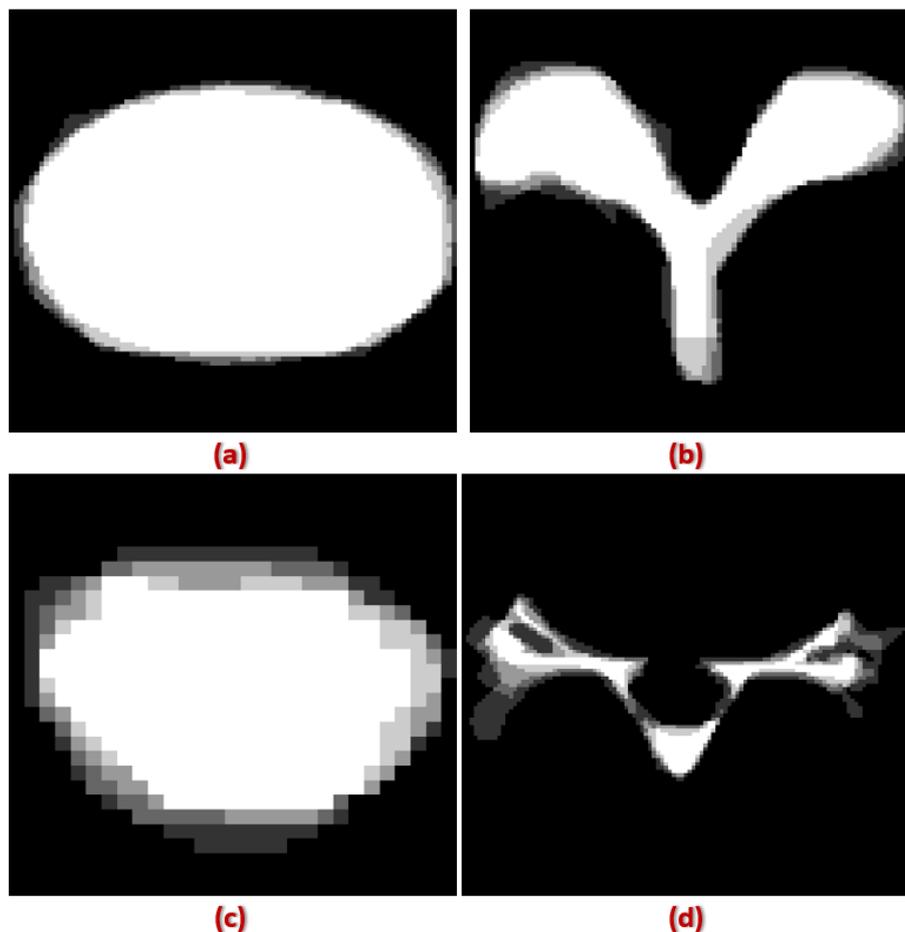


Figure 7.4: Visualizing vote counts that each pixel in and around each RoI receives as a heat map, where the brighter pixels have more votes than darker pixels. The RoIs are a) IVD, b) PE, c) TS and d) the AAP

The IVD region is by far the easiest region to the label. The region has more consistent visual characteristics across all patients which manifest as a narrow range of pixel grey level values, smoother texture, as well as high contrast to the surrounding tissues. This fact is reflected by the high iou'_{cv} values the region has compared to the other three as shown in Table 7.1 The heat map image of the IVD region shown in Figure 7.4 (a) shows a high contrast between the region and the surrounding area indicating a high level of confidence in the data for this region.

The PE region has a unique shape, similar to the letter Y in many cases, as exemplified in Figure 7.4 (b). The region has a relatively wide range of pixel grey level values and more varied texture compared to IVD. Furthermore, it has a lower contrast to the surrounding tissue than the latter, making it more difficult to determine its edges correctly especially towards its lower end. This fact is reflected in its lower iou'_{cv} values than the IVD region.

In a healthy patient, the TS region will appear distinct to its surrounding in T1-weighted MRI and has a round shape as depicted in Figure 7.4 (c). However, when central spinal stenosis occurs, the spinal canal may be squashed between the IVD and the PE. This, in turn, could make accurate identification of its edges more difficult. Furthermore, because of its small size, the ratio between its edges and inner pixels is large hence lowering its iou'_{cv} values.

By far, the hardest region to label is the AAP region. One of the reasons for this is because it does not strictly represent any part of human tissues like the other three, but instead it represents a large osseous opening in the lumbar spine structure [145]. Its shape can vary significantly depending on many factors such as the location of the slice, the patient's posture as the MRI is performed, as well as the presence of illness or defects.

Usually, the top and bottom boundaries of the AAP with the IVD and the PE respectively, are pronounced. However, in cases where central or lateral stenosis occurs, these boundaries narrow and become unclear. The AAP region also contains many spinal nerves, spinal arteries,

and veins whose locations vary significantly between patients and depending on which part of the lumbar spine the AAP is. This results in inconsistent and different pixel values and texture in MRI scans. Furthermore, judging from the example segmentation provided by our clinician experts, there seems to be a very loose definition of where the region should end on each side of the vertebrae. All these reasons result in poor values of iou'_{cv} across all vote thresholds as shown in Table 7.1

7.2.3 Confidence and consistency metrics

At the end of the experiment, we have five groups of labelled images for each class. We need to choose one from each class that best represents the groups and to be selected as the final ground truth data. From them, we will calculate the confidence metric and consistency metric, denoted as x_c and y_c respectively.

Each of the five groups is developed from its corresponding vote threshold. There is a question of which vote threshold value is to be chosen as the *selected* vote-threshold, denoted as v_t , that will provide the best and most appropriate ground truth. The most liberal option would be to pick $v_t = 1$ as it provides the highest iou'_{cv} value. However, this presents one problem, that it is very likely that pixels which have only one vote for one class also have at least one vote for other classes, hence we need to reconcile the different votes before assigning those pixels a label. Another alternative is to pick the most conservative group, i.e., choosing only those pixels that have all the votes, i.e., $v_t = 5$. One of the issues with this is the lowest iou'_{cv} that the group has compared to all other options, hence it provides the least confidence level.

One intuitive, yet arguably the best, answer to the question is to pick the majority vote. In our case, since there are in total five participants we will select $v_t = 3$ as our chosen vote threshold. This is a compromise solution that avoids both the problem of having the lowest iou'_{cv} value and having to perform vote reconciliation. Therefore, a pixel n will be assigned a final

classification c when $g_{nc3} = 1$. Our generalised solution is, therefore, to set the confidence metric x_c as the alternative intersection-over-union iou'_{cv} at the selected vote-threshold v_t .

$$x_c = iou'_{cv_t} \quad [7.7]$$

We have previously defined consistency level as to how varied the ground truth data is across its sources. One way to measure this is by calculating the rate of change of iou'_{cv} along the vote threshold dimension. Using the mean first derivative of the class's confidence level at the chosen vote-threshold v_t , we calculate the consistency metric y_c as:

$$y_c = 1 + 2 \times \frac{iou'_{cv_{t+1}} - iou'_{cv_{t-1}}}{v_{t+1} - v_{t-1}} \quad [7.8]$$

Note that the value of y_c ranges between 0 and 1, where low value suggests low consistency and high variability between the labellers, and vice versa. The final values of confidence and consistency values of the ground truth data are presented in Table 7.2.

Table 7.2: Confidence and consistency values of the resulting ground truth data.

Regions (label/class)	x_c	y_c
Intervertebral Disc (1)	0.93	0.95
Posterior Element (2)	0.82	0.87
Thecal Sac (3)	0.81	0.87
The AAP (4)	0.48	0.68

7.3 Boundary Delineation through Semantic Segmentation

7.3.1 Segmentation Results and Analysis

In this experiment, we will use a number of performance metrics to measure how well both SegNet networks perform semantic segmentation on our dataset. These metrics include general as well as class-specific metrics. To assess how the segmentation process performs on a specific class, we calculate class-specific metrics, namely class accuracy and class intersection-over-union. The class accuracy of class c , denoted as a_c , is calculated as:

$$a_c = \frac{m_{cc}}{t_c} \quad [7.9]$$

The definitions m_{cc} and t_c were given previously when we defined the class intersection-over-union (iou_c) in Eq. 7.1 in this chapter.

$$iou_c = \frac{m_{cc}}{t_c + m_c - m_{cc}} \quad [7.10]$$

General metrics that we use to assess the overall performance of the segmentation process are pixel accuracy (a_p), mean accuracy (a_m), mean intersection-over-union (iou_m) and frequency-weighted intersection-over-union (iou_{fw}). The formula used to calculate these metrics are provided as follows:

$$a_p = \frac{\sum_c m_{cc}}{\sum_c t_c} \quad [7.11]$$

$$a_m = \frac{\sum_c a_c}{m_{cl}} \quad [7.12]$$

$$iou_m = \frac{\sum_c iou_c}{m_{cl}} \quad [7.13]$$

$$iou_{fw} = \frac{\sum_c (t_c \times iou_c)}{\sum_c t_c} \quad [7.14]$$

The pixel accuracy and mean accuracy results at different PTD values are shown in Figure 6.5. The results show that SegNet-FS produces very low accuracies at the low end of the PTD scale and increases in performance as the PTD increases. On the other hand, SegNet-TL produces more accurate segmentation even at low PTD and is significantly more insensitive to the values of PTD used. The figure also shows significant differences between the two accuracy metrics. The figure shows the pixel accuracies are consistently higher than mean accuracies at the same point in the x-axis. Pixel accuracy measures the proportion of correctly labelled pixels in the entire pixel population in the dataset. This metric does not allow us to see how accurate the segmentation is for each class but at the same time is affected by class population imbalance. High accuracy in the largest class will significantly mask poor performance in smaller classes. This is the case in our experiment as exemplified in Figure 7.5 The fact that the mean accuracy is lower than the pixel accuracy suggests that we have one dominant class that has significantly higher accuracy than the others.

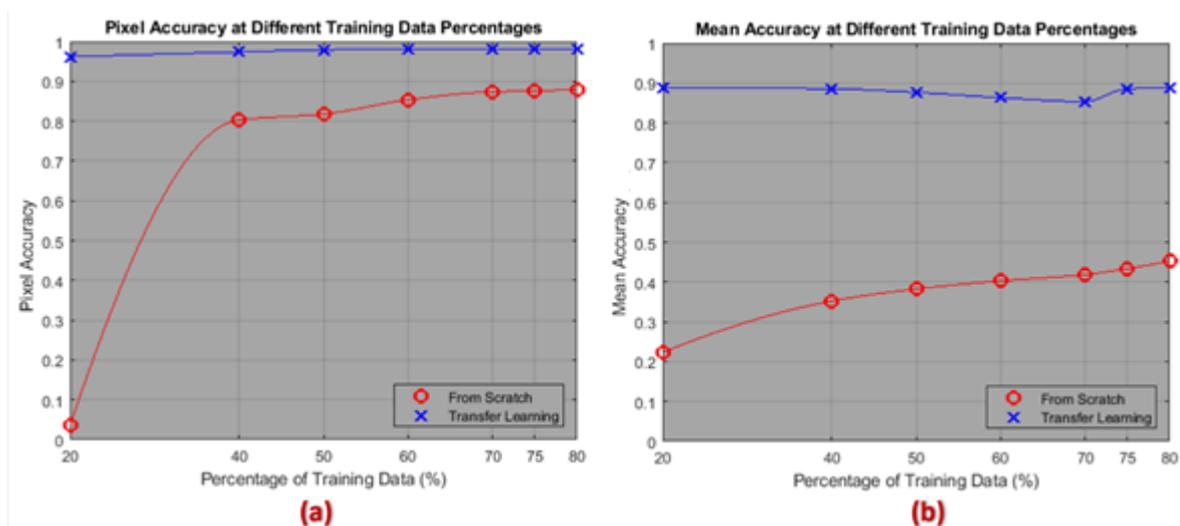


Figure 7.5: The plot of a) Pixel Accuracy and b) Mean Accuracy of the SegNet semantic segmentation results at different training data percentages.

The use of intersection-over-union metric gains popularity in the image segmentation research community because of one significant limitation in the accuracy metric. If there exists a class (or classes) of pixels that are not important to the calculation of the segmentation performance, one could design a classifier that increases the accuracy of other, more important, classes at the expense of the accuracy of the less important class, hence artificially boosts the overall segmentation accuracy. Segmentation algorithms that do this tend to produce a high number of false alarms or false positives. Moving away from this limitation, many image segmentation challenges such as Microsoft COCO challenge [146] introduced intersection-over-union (IoU) as a more representative metric. As with accuracy metric, there are also two versions of IoU metric that can be used. Frequency-weighted IoU is the equivalent of pixel accuracy whereas mean IoU is the equivalent of mean accuracy. The plot of mean IoU and Frequency-weighted IoU results are shown in Figure 7.6.

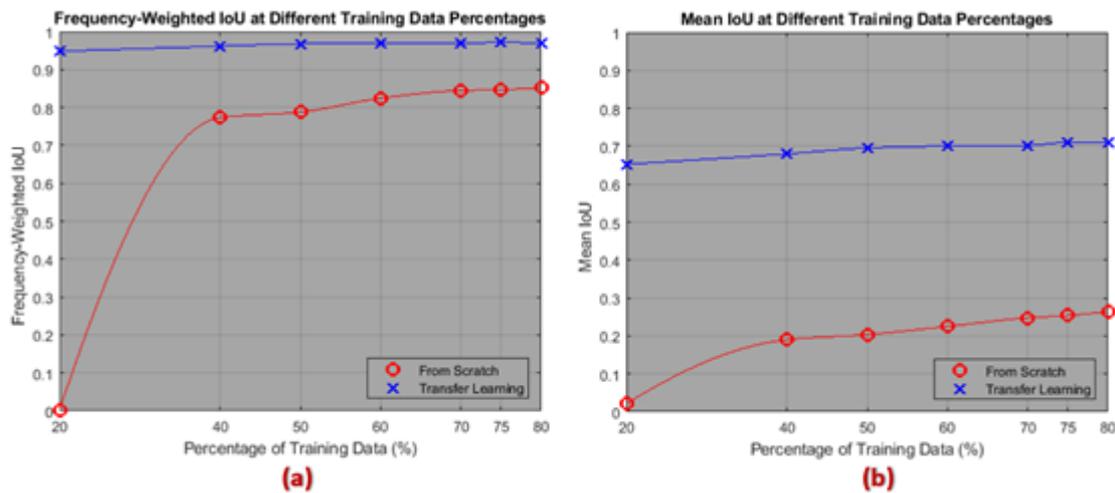


Figure 7.6: The plot of a) Frequency-Weighted IoU and b) Mean IoU of the SegNet semantic segmentation results at different training data percentages.

When we compare the pixel accuracy and frequency-weighted IoU results, we conclude that there is not much significant difference between them. This is expected since in our case, there is no class of pixels that are not used in the calculation of the segmentation performance. In

other words, the union of all pixels belonging to every class should comprise the total population of pixels under consideration. However, when we compare the mean accuracy and the mean IoU graphs, we can see that the segmentation performance measured using the latter metric is consistently lower than the former across all PTD values. To further investigate the underlying reasons for this phenomenon, we observe the individual class accuracies and class IoUs. These are given in Figure 7.7 and Figure 7.8.

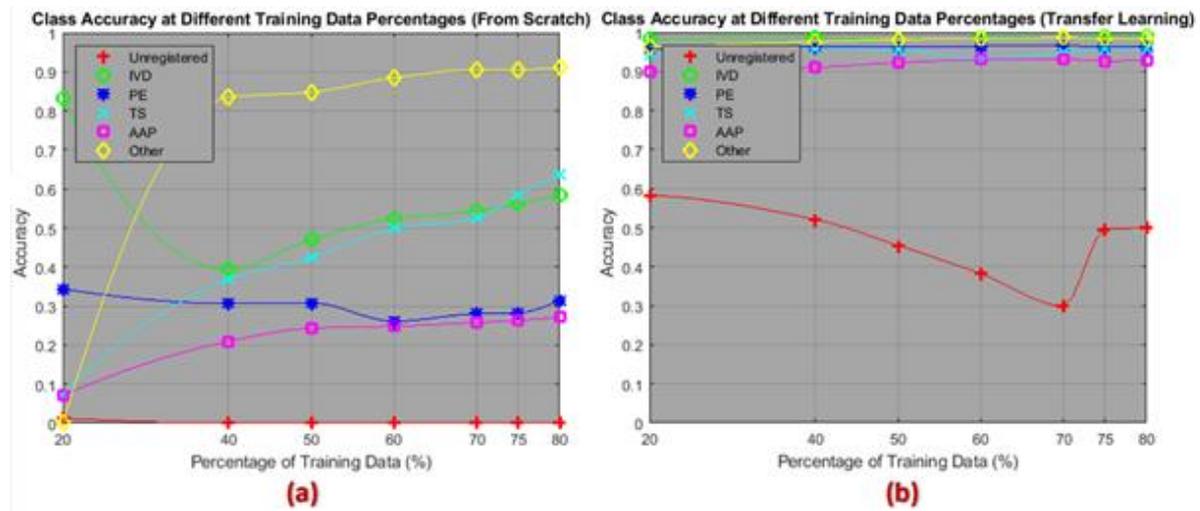


Figure 7.7: The plot of the class accuracy of the semantic segmentation results using a) SegNet-FS and b) SegNet-TL at different training data percentages.

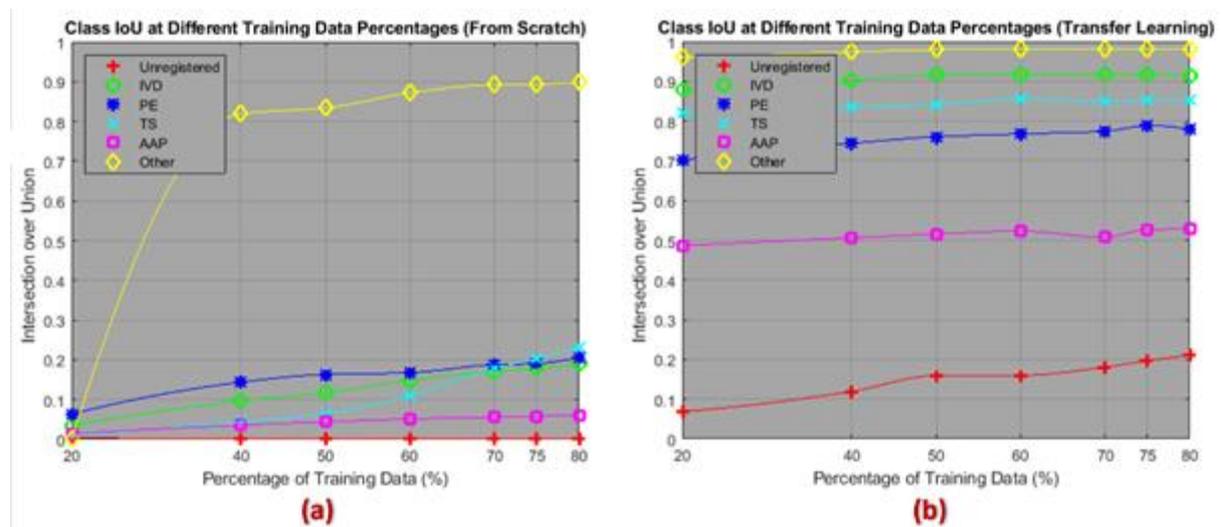


Figure 7.8: The plot of class IoU of the semantic segmentation results using a) SegNet-FS and b) SegNet-TL at different training data percentages.

One of the most straightforward findings from our observation of the individual class performance is that the ‘other’ class is consistently better than all other classes – across all PTD values and regardless of whether the network is trained from scratch or using transfer learning. Using this view, we can clearly identify which of the six classes is the worst performer. We found the ‘unregistered’ class to be consistently the worst region to classify. This is expected since the population of pixels that belongs to this class is the lowest and also at the same time not necessarily present in all images. In fact, we believe this is the reason behind the strangely shaped class-accuracy plot of the unregistered class in Figure 7.7 (b). We suspect that by chance, the randomised selection of the training data below the 75% mark had not picked up a sufficient number of unregistered pixels for training.

7.4 Chapter Summary

After observing all the results we have obtained, we concluded that SegNet-TL produces better segmentation than SegNet-FS. Additionally, we also concluded that the results get marginally better the higher the training percentage is used. Therefore, we decided to set the best semantic segmentation network to use to be SegNet-TL trained using 80% PTD. The segmentation performance of this network is summarised in table 7.3.

Table 7.3: The performance of the best semantic segmentation network (SegNet-TL trained using 80% PTD)

Regions (label/class)	a_c	iou_c
Unregistered	0.50	0.21
Intervertebral Disc	0.99	0.92
Posterior Element	0.96	0.78
Thecal Sac	0.96	0.85
AAP	0.93	0.53
Other	0.98	0.98

Regarding the different performance metrics, we found that the class IoU metric is the best image segmentation metric to use due to its ability to differentiate much more clearly each class's individual performances. However, it is important to note that our objective is to achieve delineation of important boundaries in the MRI images as illustrated in Figure.7.2 Therefore, we also need to analyse the delineation results to measure how the selected best SegNet performs in that regard.

Chapter 8 Conclusion and Future Work

8.1 Conclusion

The progress in image processing, computer vision, artificial intelligence, computer graphics, and physics simulation has moved rapidly in the past decade. However, they have not been utilized in any significant way to improve Computer-Aided Diagnostic techniques, in particular, the way Chronic Lower Back Pain cases are diagnosed and managed. Most physicians still rely on a long and laborious task to visually identify abnormalities from the patient's Magnetic Resonance Images. Furthermore, currently, there is no solution to assist physicians to understand how patients' physiological characteristics, posture, and position affect the way pain at lower back spine is generated. This research proposed a novel methodology that utilises state-of-the-art computing technologies, which can be used as a computer-aided diagnosis tool, to help physicians in their efforts to diagnose and manage CLBP cases. The proposed method will be able to solve some of the existing problems with the current Chronic Lower Back Pain diagnosis and management procedures.

The presented method is to aid clinicians in performing lumbar spinal stenosis detection through the delineation of important boundaries in MRI images. The method starts by applying semantic segmentation to the MRI images to locate four regions of interest, namely the Intervertebral Disc (IVD), Posterior Element (PE), Thecal Sac (TS), and Area between Anterior and Posterior (AAP) elements. SegNet has been used as one of the best performing deep neural networks in the literature to date, as the pixel classifier. Due to the limitation in size and suitability of the currently existing open-access lumbar spine dataset, it has been decided to develop our own dataset. Our dataset contains clinical lumbar spine MRI studies of 515 patients with symptomatic back pains. Each study is annotated by expert radiologists with notes regarding the observed characteristics, condition of the lumbar spine, or presence of diseases,

these include bone marrow disease, endplate degeneration, IVD bulges, Thecal Sac compressing, central or foraminal stenosis, annular tears, scoliosis, endplate defects (Modic type), facet joint and Ligamentum Flavum hypertrophy, and spondylolisthesis. From this dataset, a ground-truth label image dataset is developed. It can be used to train and test an image segmentation model. Due to the lack of appropriate methodologies in the literature to assess the quality of ground-truth datasets, in this research two novel metrics have been developed to assess the accuracy and variability of a ground truth dataset, namely the confidence and consistency metrics, respectively. These metrics are derived from the widely used intersection-over-union metric to measure the accuracy of image segmentation algorithms.

SegNet models have been trained using a combination of different training set-ups. The first set-up trained the model from scratch, i.e., using random valued initial weights. The second set-up uses pre-trained initial weights provided by the VGG16 network. The experiment has been done using a variety of training-to-testing-percentage ratios on each of the above set-ups.

Analysis of the results of the semantic segmentation and the delineation results was carried out using a comprehensive set of contour-based and region-based performance metrics including accuracy, intersection-over-union, and BF score. Our experiment shows that the different pixel classifiers produce varying levels of performances, but in general, the model that uses the VGG16 pre-trained initial weights, as opposed to initial random weights, is the best. Moreover, using 80:20 ratio of training to testing percentages provides the best performance across the board. Therefore, it has been concluded that SegNet-TL80 is the best segmentation model to use.

The performance of this model is also analyzed in two ways. First, by measuring and comparing inter-rater agreement, rater-ground-truth agreement, and model-ground-truth agreement. The

research concluded that 1) the model's performance is within the range of manual labelling performance and 2) the ground-truth dataset has an excellent inter-rater agreement score. Qualitatively, this research presented two representative boundary delineation results. The results are selected from the entire dataset based on the worst and best contour-based metric score because they provide an indication of the range of quality of the boundary delineation results. Through visual inspection of these results, this research can confidently claim that our proposed method is sufficiently accurate, and the results are suitable for computer-aided-diagnosis purposes.

8.2 Research Contributions

The finding presented in this research is to develop a computer-assisted diagnosis of chronic lower back pain which was detailed in this thesis and in our previous publication [147]. The MRI study dataset, the ground-truth label dataset, and the MATLAB source code used in this research are made available freely for the benefit of the research community. The MRI images are stored in Siemens IMA (DICOM) format, the label ground truth and all extracted slices are stored as greyscale images in PNG format with lossless compression to preserve the quality.

The methodology proposed is to aid clinicians in performing disc herniation and lumbar spinal stenosis detection through machine learning delineate the magnetic resonance imaging (MRI) scans of the lumbar spine. Our dataset contains MRI studies of 515 patients with symptomatic back pains. Each study is annotated by expert radiologists with notes regarding the observed characteristics and condition of the lumbar spine. The research has developed a ground truth dataset, containing image labels of four important regions in the lumbar spine, to be used as the training and test images to develop classification models for segmentation. Two novel metrics have developed, namely confidence, and consistency, to assess the quality of the ground truth dataset through a derivation of the Jaccard Index. An experiment with semantic segmentation of our dataset using SegNet has been performed. The evaluation of the

segmentation and the delineation results show that our proposed methodology produces a very good performance as measured by several contour-based and region-based metrics. In addition, using the Cohen's kappa and frequency-weighted confidence metrics, it can show that 1) the model's performance is within the range of the worst and the best manual labelling results and 2) the ground-truth dataset has an excellent inter-rater agreement score. The research also presented two representative delineation results of the worst and best segmentation based on their BF-score to show visually how accurate and suitable the results are for computer-aided-diagnosis purposes. In addition to the delineation of the area of interest, the produced labelled images have been used to help the clinician diagnosing patients with spinal canal stenosis by allocating the specific points in the segmented area as advised by our external advisor. Then, these points have been used to calculate the distances between the centre of the disc and the bottom of the spinal canal and compare the calculated distances with the approved distance for the healthy lumbar spine. Meanwhile, the result of the comparison will be used as an indicator for patients with spinal canal stenosis. Table 8.1 shows the research contributions and the achievement of each one of them.

Table 8.1: The research contributions and how it has been achieved.

Research Contribution	Achievement
<p>1. Presenting a method to aid clinicians in performing CLBP detection.</p>	<p>This is done through two steps:</p> <ol style="list-style-type: none"> 1. Developing a method to aid clinician detecting the disc herniation using the disc segmentation and centroid distance function. (Ch5) 2. Developing a method to aid clinician detecting lumbar spinal stenosis detection via segmentation of Area between Anterior and Posterior (AAP). (Ch6)
<p>2. Develop our own dataset.</p>	<p>Developing a dataset containing clinical lumbar spine MRI studies of 515 patients with symptomatic back pains. From this dataset, a ground-truth label image dataset is developed. It can be used to train and test an image segmentation model. (Ch4)</p>
<p>3. The development of two novel metrics to assess accuracy and variability.</p>	<p>Developing confidence and consistency metrics. These metrics are derived from the widely used intersection-over-union metric to measure the accuracy of image segmentation algorithms.</p>

8.3 Summary and Future Research

Chronic Lower Back Pain (CLBP) is one of the major types of pain that is affecting many people around the world. It is estimated that 28.1% of US adults suffer from this illness and 2.5 million of the UK population experience this type of pain every day. Most CLBP cases do not happen overnight and it is usually developed from a less serious but acute variant of lower back pain. An acute type of lower back pain can develop into a chronic one if the underlying cause is serious and left untreated. The longer a person is disabled by back pain, the less chance he or she will return to work and the more health care cost he or she will require. It is therefore important to identify the cause of back pains as early as possible in order to improve the chance of patient rehabilitation. The speed of early diagnosis can depend on many factors including referral time from a general practitioner to the hospital, waiting time for a specialist appointment, time for an MRI scan and time for the analysis result to come out. Currently diagnosing the lower back pain is done by visual observation and analysis of the lumbar spine MRI images by radiologists and clinicians and this process could take up much of their time and effort. This, therefore, rationalizes the need for a new method to increase the efficiency and effectiveness of the imaging diagnostic process.

As a future plan will work to develop a new methodology to automatically aid clinicians in performing diagnosis of CLBP, the method will be based on the current accepted medical practice of manual inspection the MRI scans of the patient's lumbar spine. There are different possible causes of CLBP and identifying one requires a different procedure from another. To identify lumbar spinal stenosis, for example, the procedure involves identifying boundaries between specific regions in MRI images, measuring distances between specific points in these boundaries and comparing the distances with the range of normal values. Another cause of CLBP, namely intervertebral disc degeneration, can be identified through observing the pixel intensity variations within the disc. Our method will use deep learning to perform semantic

segmentation on these regions and perform boundary delineation process, meanwhile, developing a machine learning model of the segmented intervertebral disc regions to detect the level of degeneration. The preliminary work has been carried out in the form of data gathering and ground truth data development. This data have been used to test the feasibility of the methodology by using it to detect lumbar spinal stenosis, one of the possible causes of CLBP. In this research, a further extending of this method to detect and classify other possible causes of CLBP will include: developing a new algorithm for Intervertebral Disc Herniation Detection, developing a new algorithm for Intervertebral Disc Degeneration Classification, developing a new algorithm for Lumbar Spine Fracture Detection and to develop a software prototype demonstrating the capability of the algorithms in assisting the diagnosis process of chronic lower back pain. Another direction for the proposed research is to use the current methodology to help the clinicians in diagnosing brain tumor using brain MRI images. Automatic tumor segmentation will help in speed up the diagnosing process and highlight the patients at risk which will contribute in saving people's lives and reduce the overall cost by early diagnosing the patient's with this kind of diseases. Moreover, the proposed research can also contribute to developing a CAD that support image classification and boundary detecting for the lung using any kind of Dicom images such as X-Ray, MRI, or CT. This will help in diagnosing chest diseases such as pulmonary diseases, lung infections, and cancer which cause a huge number of deaths every year. The need of having these kinds of systems is motivating me to look forward in developing an automatic methodology that helps clinicians in diagnosing a different kind of diseases to mitigate the severity of the disease for the patients and help healthcare service authority in reducing the cost and saving the time.

References

- [1] J. Koh, R. S. Alomari, V. Chaudhary, and G. Dhillon, "Lumbar Spinal Stenosis CAD from ClinicalMRMand MRI Based on Inter-and Intra-Context Features with a Two-Level Classifie," *SPIE Med. Imaging*, vol. 7963, pp. 796304-796304-8, 2011.
- [2] G. Waddell and A. K. Burton, "Occupational health guidelines for the management of low back pain at work: Evidence review," *Occup. Med. (Chic. Ill).*, vol. 51, no. 2, pp. 124-135, 2001.
- [3] R. Gordon and S. Bloxham, "A Systematic Review of the Effects of Exercise and Physical Activity on Non-Specific Chronic Low Back Pain.," *Healthc. (Basel, Switzerland)*, vol. 4, no. 2, p. 22, 2016.
- [4] S. Dixon, "Diagnostic Imaging Dataset Annual Statistical Release 2013/14," pp. 1-27, 2014.
- [5] Royal College of Radiologists, "2015 Clinical Radiology UK Workforce Census," 2016.
- [6] K. P. Botwin and R. D. Gruber, "Lumbar spinal stenosis: anatomy and pathogenesis," *Phys. Med. Rehabil. Clin.*, vol. 14, no. 1, pp. 1-15, Jan. 2003.
- [7] H. S. Nguyen *et al.*, "Upright magnetic resonance imaging of the lumbar spine: Back pain and radiculopathy.," *J. Craniovertebral Junction Spine*, vol. 7, no. 1, pp. 31-37, 2016.
- [8] B. B. Hansen *et al.*, "Conventional Supine MRI With a Lumbar Pillow—An Alternative to Weight-bearing MRI for Diagnosing Spinal Stenosis?: A Cross-sectional Study," *Spine (Phila. Pa. 1976).*, vol. 42, no. 9, pp. 662-669, 2017.
- [9] M. Merkle *et al.*, "The value of dynamic radiographic myelography in addition to magnetic resonance imaging in detection lumbar spinal canal stenosis : A prospective study," *Clin. Neurol. Neurosurg.*, vol. 143, pp. 4-8, 2016.
- [10] Z. Gao *et al.*, "Automated Framework for Detecting Lumen and Media--Adventitia Borders in Intravascular Ultrasound Images," *Ultrasound Med. Biol.*, vol. 41, no. 7, pp. 2001-2021, 2015.
- [11] S. Folland, A. Goodman, and M. Stano, *The Economics of Health and Health Care*. 2013.
- [12] H. Jiang *et al.*, "Quantitative evaluation of lumbar disc herniation based on MRI image," in *Abdominal Imaging. Computational and Clinical Applications*, Springer, 2012, pp. 91-98.
- [13] K. McCamey and P. Evans, "Low back pain.," *Prim. Care*, vol. 34, no. 1, pp. 71-82, 2007.
- [14] "Pain Community Centre." [Online]. Available: <http://www.paincommunitycentre.org/article/low-back-pain-problem#ref6>.
- [15] G. Waddell, *The Back Pain Revolution*, 2nd ed. Churchill Livingstone (Elsevier), 2004.
- [16] M. Deane and D. Crick, "Outcome of low back pain in general practice. Evidence based practice can improve outcome.," *BMJ (Clinical research ed.)*, vol. 317, no. 7165. BMJ Publishing Group Ltd, p. 1083, 1998.
- [17] K. Burton and N. Kendall, "Musculoskeletal Disorders," *Bmj*, vol. 348, no. feb21 2, p. bmj.g1076-bmj.g1076, 2014.
- [18] G. J. Macfarlane, E. Thomas, P. R. Croft, A. C. Papageorgiou, M. I. V Jayson, and A. J. Silman, "Predictors of early improvement in low back pain amongst consulters to general practice: the influence of pre-morbid and episode-related factors," *Pain*, vol. 80, no. 1, pp. 113-119,

- 1999.
- [19] N. Maniadas and A. Gray, "The economic burden of back pain in the UK," *Pain*, vol. 84, no. 1, pp. 95–103, Jan. 2000.
- [20] NHS, "Quarterly Diagnostic Waiting Times," *NHS England*, 2017. .
- [21] National Office for Statistics, "NHS Imaging and Radiodiagnostic activity in England 2014," no. August, 2013.
- [22] R. M. Ellis, "Back pain.," *BMJ*, vol. 310, no. 6989, p. 1220, 1995.
- [23] P. Raj, H. Nolte, and M. Stanton-Hicks, "Anatomy of the Spine," *Illus. Man. Reg. ...*, pp. 1–5, 1988.
- [24] P. P. Proximal *et al.*, "Patellofemoral Pain: Proximal, Distal, and Local Factors, 2nd International Research Retreat," *J. Orthop. Sports Phys. Ther.*, vol. 42, no. 6, pp. 1–55, 2012.
- [25] M. C. Hay, "Anatomy of the lumbar spine," *Med. J. Aust.*, vol. 1, no. 23, pp. 874–876, 1976.
- [26] "The Healthy Spine | Spinal Simplicity." [Online]. Available: <http://www.spinalsimplicity.com/the-healthy-spine/>. [Accessed: 29-Mar-2016].
- [27] A. Raja'S, J. J. Corso, V. Chaudhary, and G. Dhillon, "Automatic diagnosis of lumbar disc herniation with shape and appearance features from MRI," in *SPIE Medical Imaging*, 2010, pp. 76241A--76241A.
- [28] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Lumbar spine disc herniation diagnosis with a joint shape model," in *Computational Methods and Clinical Applications for Spine Imaging*, Springer, 2014, pp. 87–98.
- [29] M. Allegri *et al.*, "Mechanisms of low back pain : a guide for diagnosis and therapy [version 2 ; referees : 3 approved] Referee Status :," vol. 5, pp. 1–11, 2016.
- [30] J. Hartvigsen, M. Hancock, A. Kongsted, and M. L. Ferreira, "Series Low back pain 1 What low back pain is and why we need to pay attention," vol. 6736, no. March, 2018.
- [31] T. F. De Campos, C. G. Maher, D. Steffens, J. T. Fuller, and M. J. Hancock, "a systematic review and meta-analysis," *J. Physiother.*, vol. 64, no. 3, pp. 159–165, 2018.
- [32] D. Sowah, R. Boyko, D. Antle, L. Miller, and M. Zakhary, "Occupational interventions for the prevention of back pain : Overview of systematic reviews," *J. Safety Res.*, vol. 66, pp. 39–59, 2018.
- [33] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Automatic Diagnosis of Lumbar Disc Herniation with Shape and Appearance Features from MRI," *Prog. Biomed. Opt. imaging*, vol. 11, pp. 76241A-76241A, 2010.
- [34] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Lumbar Spine Disc Herniation Diagnosis with a Joint Shape Model," *Clin. Appl. Spine Imaging*, vol. 17, pp. 87–98, 2014.
- [35] St John of God Subiaco Clinic, "SPINE WEST." [Online]. Available: <http://www.spinewest.com.au/spine-disorders/lumbar-disc-herniation/>.
- [36] R. Methods and B. Practices, "Appropriateness of Care : Use of MRI in the Investigation of Patient Low Back Pain Executive Summary," pp. 1–29.
- [37] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Automatic diagnosis of lumbar disc herniation with shape and appearance features from MRI," in *SPIE Medical Imaging*, 2010,

- pp. 76241A--76241A.
- [38] J. Koh, R. S. Alomari, V. Chaudhary, and G. Dhillon, "Lumbar Spinal Stenosis CAD from ClinicalMRMand MRI Based on Inter-and Intra-Context Features with a Two-Level Classifie," vol. 7963, pp. 796304-796304–8, 2011.
 - [39] S. W. Atlas, *Magnetic resonance imaging of the brain and spine*, 5th ed., vol. 1. Lippincott Williams & Wilkins, 2016.
 - [40] D. A. Lisin, M. A. Mattar, M. B. Blaschko, M. C. Benfield, and E. G. Learned-miller, "Combining Local and Global Image Features for Object Class Recognition," *CVPR Work.*, 2005.
 - [41] J. H. Freidman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Softw.*, vol. 3, no. 3, pp. 209–226, 1977.
 - [42] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *Proc. 5th ACM-SIAM Symp. Discret. Algorithms*, vol. 1, no. 212, pp. 573–582, 1994.
 - [43] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1000–1006, 1997.
 - [44] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2227–2240, 2014.
 - [45] S. Abbasi, F. Mokhtarian, and J. Kittler, "Curvature scale space image in shape similarity retrieval," *Multimed. Syst.*, vol. 7, no. 6, pp. 467–476, 1999.
 - [46] W. Y. Ma and B. S. Manjunath, "NeTra: A toolbox for navigating large image databases," *Multimed. Syst.*, vol. 7, no. 3, pp. 184–198, 1999.
 - [47] M. D. Saleh and C. Eswaran, "An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 186–196, 2012.
 - [48] K. Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential," *Comput. Med. Imaging Graph.*, vol. 31, no. 4–5, pp. 198–211, Mar. 2007.
 - [49] M. Lootus, T. Kadir, and A. Zisserman, "Vertebrae Detection and Labelling in Lumbar MR Images," *Lect. Notes Comput. Vis. Biomech.*, vol. 17, 2014.
 - [50] S. Su, Z. Hu, Q. Lin, W. K. Hau, Z. Gao, and H. Zhang, "An artificial neural network method for lumen and media-adventitia border detection in IVUS," *Comput. Med. Imaging Graph.*, vol. 57, pp. 29–39, 2017.
 - [51] S. Su, Z. Gao, H. Zhang, Q. Lin, W. K. Hau, and S. Li, "Detection of lumen and media-adventitia borders in IVUS images using sparse auto-encoder neural network," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, 2017, pp. 1120–1124.
 - [52] X. Du *et al.*, "Deep regression segmentation for cardiac bi-ventricle MR images," *IEEE Access*, vol. 6, pp. 3828–3838, 2018.
 - [53] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, 2010.
 - [54] S. Wang *et al.*, "Multiple Sclerosis Identification by 14-Layer Convolutional Neural Network With Batch Normalization , Dropout , and Stochastic Pooling," *Front. Neurosci.*, vol. 12, no.

- November, pp. 1–11, 2018.
- [55] C. P. Loizou, V. Murray, M. S. Pattichis, I. Seimenis, M. Pantziaris, and C. S. Pattichis, "Multiscale Amplitude-Modulation Frequency-Modulation (AM–FM) Texture Analysis of Multiple Sclerosis in Brain MRI Images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 119–129, Jan. 2011.
- [56] Y. Zhang, C. Pan, J. Sun, and C. Tang, "Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU," *J. Comput. Sci.*, vol. 28, pp. 1–10, 2018.
- [57] H. Jiang *et al.*, "Quantitative evaluation of lumbar disc herniation based on MRI image," in *Abdominal Imaging. Computational and Clinical Applications*, Springer, 2012, pp. 91–98.
- [58] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [59] D. P. H. Pedro F Felzenszwalb, "Pictorial Structures for Object Recognition," 2004.
- [60] X. He, H. Zhang, M. Landis, M. Sharma, J. Warrington, and S. Li, "Unsupervised boundary delineation of spinal neural foramina using a multi-feature and adaptive spectral segmentation," *Med. Image Anal.*, vol. 36, pp. 22–40, 2017.
- [61] K. O. Babalola *et al.*, "An evaluation of four automatic methods of segmenting the subcortical structures in the brain," *Neuroimage*, vol. 47, no. 4, pp. 1435–1447, Oct. 2009.
- [62] G. Kevin, *An Introduction to Neural Networks*. London: CRC Press, 1997.
- [63] M. Kantardzie and A. N. SRIVASTAVA, "Data Mining: Concepts, Models, Methods, and Algorithms," *Journal of Computing and Information Science in Engineering*, vol. 5, no. 4. p. 394, 2005.
- [64] A. Mahdi, "The Application of Neural Network in Financial Time Series Analysis and Prediction Using Immune System." 2010.
- [65] Y. Tsukamoto and A. Namatame, "Evolving neural network models," pp. 689–693, 2002.
- [66] D. Reid, A. J. Hussain, and H. Tawfik, "Financial time series prediction using spiking neural networks," *PLoS One*, vol. 9, no. 8, pp. 1–13, 2014.
- [67] R. A. Aliev and Rashad Rafik Aliev, "Soft Computing and Its Applications Classical and New Methods Nonlinear Mathematical Models Symmetry and Invariance Principles." World Scientific Publishing Company, 2001.
- [68] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network," in *ICET2017*.
- [69] MathWorks, "Introducing Deep Learning with MATLAB," 2018.
- [70] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [71] M. Alam, D. Thapa, J. I. Lim, D. Cao, and X. Yao, "Automatic classification of sickle cell retinopathy using quantitative features in optical coherence tomography angiography," *Biomed. Opt. Express*, vol. 8, no. 9, p. 4206, 2017.
- [72] W. K. Härdle, D. Prastyo, and C. Hafner, "Support Vector Machines with Evolutionary Feature Selection for Default Prediction," *Ssrn*, 2017.
- [73] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Supervised Machine Learning: A Review of

- Classification Techniques,” *Artif. Intell. Rev.*, vol. 26, pp. 159–190, 2006.
- [74] X. Gao, L. Fan, and H. Xu, “Multiple rank multi-linear kernel support vector machine for matrix data classification,” *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 2, pp. 251–261, 2018.
- [75] G. Postma, B. Lionel, van S. Frederik-Jan, and B. Lutgarde, “16 Nonlinear Supervised Techniques.” p. 285, 2018.
- [76] P. Tan, M. Steinbach, and V. Kumar, “Classification : Basic Concepts , Decision Trees , and,” *Introduction to Data Mining*, vol. 67, no. 17. pp. 145–205, 2006.
- [77] S. O. N. G. Yan-yan and L. U. Ying, “Decision tree methods: applications for classification and prediction,” *Shanghai Arch. Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.
- [78] O. Maimon and L. Rokach, “The Data Mining and Knowledge Discovery Handbook. A complete Guide for Researchers and Practitioner.” p. 1436, 2005.
- [79] J. Aspnes, E. Blais, M. Demirbas, R. O’Donnell, A. Rudra, and S. Uurtamo, “K+ decision trees,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6451 LNCS, pp. 74–88, 2010.
- [80] G. Ke *et al.*, “LightGBM : A Highly Efficient Gradient Boosting Decision Tree,” *Adv. Neural Inf. Process. Syst. 30 (NIPS 2017)*, no. Nips, pp. 3149–3157, 2017.
- [81] A. B. A. Christopher and S. Appavu Alias Balamurugan, “Prediction of warning level in aircraft accidents using data mining techniques,” *Aeronaut. J.*, vol. 118, no. 1206, pp. 935–952, 2014.
- [82] C. Shahabi, M. R. Kolahdouzan, and M. Sharifzadeh, “A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases Categories and Subject Descriptors,” *Network*, pp. 255–273, 2003.
- [83] R. T. Ionescu and M. Popescu, *Knowledge transfer between computer vision and text mining : similarity-based learning approaches. .*
- [84] “SpineWeb: Collaborative Platform for Research on Spine Imaging and Image Analysis,” 2016. [Online]. Available: <http://spineweb.digitalimaginggroup.ca/>. [Accessed: 27-Apr-2018].
- [85] J. Yao, J. E. Burns, H. Munoz, and R. M. Summers, “Detection of Vertebral Body Fractures Based on Cortical Shell Unwrapping,” *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 15, no. 0 3. pp. 509–516, 2012.
- [86] M. S. Aslan, A. Shalaby, and A. A. Farag, “Clinically desired segmentation method for vertebral bodies,” *Proc. - Int. Symp. Biomed. Imaging*, pp. 840–843, 2013.
- [87] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, “Shape Representation for Efficient Landmark-Based Segmentation in 3-D,” *IEEE Trans. Med. Imaging*, vol. 33, no. 4, pp. 861–874, 2014.
- [88] R. Korez, B. Ibragimov, B. Likar, F. Pernus, and T. Vrtovec, “A Framework for Automated Spine and Vertebrae Segmentation,” *IEEE Trans. Med. Imaging*, vol. 34, no. 8, pp. 1649–1662, 2015.
- [89] Y. Cai, S. Osman, M. Sharma, M. Landis, and S. Li, “Multi-Modality Vertebra Recognition in Arbitrary Views Using 3D Deformable Hierarchical Model,” *IEEE Trans. Med. Imaging*, vol. 34, no. 8, pp. 1676–1693, 2015.
- [90] J. Yao *et al.*, “A multi-center milestone study of clinical vertebral CT segmentation,” *Comput. Med. Imaging Graph.*, vol. 49, no. February, pp. 16–28, 2016.

- [91] S. Sudirman, A. Al Kafri, F. Natalia, H. Meidia, N. Afriliana, and M. Al-Rashdan, Wasfi Bashtawi, Mohammad Al-Jumaily, "Lumbar Spine MRI Dataset," *Mendeley Data*, 2019. [Online]. Available: <https://data.mendeley.com/datasets/k57fr854j2/2>.
- [92] S. Sudirman, A. Al Kafri, F. Natalia, H. Meidia, N. Afriliana, and M. Al-Rashdan, Wasfi Bashtawi, Mohammad Al-Jumaily, "Radiologists Notes for Lumbar Spine MRI Dataset," *Mendeley Data*, 2019. [Online]. Available: <http://dx.doi.org/10.17632/s6bgczr8s2.2>.
- [93] S. Sudirman, A. Al Kafri, F. Natalia, H. Meidia, N. Afriliana, and M. Al-Rashdan, Wasfi Bashtawi, Mohammad Al-Jumaily, "Label Image Ground Truth Data for Lumbar Spine MRI Dataset," *Mendeley Data*, 2019. .
- [94] E. Alpaydin, *Introduction to Machine Learning, third edition*. 2014.
- [95] P.-N. Tan, M. Steinbach, and V. Kumar, "Classification : Basic Concepts , Decision Trees , and," *Introd. to Data Min.*, vol. 67, no. 17, pp. 145–205, 2006.
- [96] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- [97] N. Bhatia and C. Author, "Survey of Nearest Neighbor Techniques," *IJCSIS) Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, 2010.
- [98] C. C. Aggarwal, *Data classification : algorithms and applications*. 2014.
- [99] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and a. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," *2008 19th Int. Conf. Pattern Recognit.*, no. March 2016, pp. 8–11, 2008.
- [100] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans*, vol. 40, no. 1, pp. 185–197, 2010.
- [101] D. Zhang and G. Lu, "A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval," *J. Vis. Commun. Image Represent.*, vol. 14, no. 1, pp. 39–57, 2003.
- [102] D. Zhang and G. Lu, "Generic Fourier descriptor for shape-based image retrieval," *Proc. - 2002 IEEE Int. Conf. Multimed. Expo, ICME 2002*, vol. 1, no. July, pp. 425–428, 2002.
- [103] S. Bhonsle, A. Klinzmann, F. Park, and E. Esser, "Centroid Distance Function and the Fourier Descriptor with Applications to Cancer Cell Clustering," *UC Irvine*, 2009.
- [104] E. R. Davies, *Machine vision: theory, algorithms, practicalities*. Elsevier, 2004.
- [105] D. Zhang and G. Lu, "A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures," *Int. Conf. Intell. Multimed. Distance Educ.*, vol. 1, pp. 1–9, 2001.
- [106] P. Coupé, J. V Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *Neuroimage*, vol. 54, no. 2, pp. 940–954, 2011.
- [107] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843–2851.
- [108] M. Freiman *et al.*, "Learning a sparse database for patch-based medical image segmentation," in *International Workshop on Patch-based Techniques in Medical Imaging*, 2017, pp. 47–54.
- [109] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for

- Semantic Image Segmentation,” *CoRR*, vol. abs/1706.0, 2017.
- [110] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *CoRR*, vol. abs/1411.4, 2014.
- [111] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [112] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *CoRR*, vol. abs/1502.0, 2015.
- [113] M. Khalaf *et al.*, “Recurrent Neural Network Architectures for Analysing Biomedical Data Sets,” *2017 10th Int. Conf. Dev. eSystems Eng.*, pp. 232–237, 2017.
- [114] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.
- [115] A. de Brébisson and P. Vincent, “An Exploration of Softmax Alternatives Belonging to the Spherical Loss Family,” *CoRR*, vol. abs/1511.0, 2015.
- [116] S. Ruder, “An overview of gradient descent optimization algorithms,” *CoRR*, vol. abs/1609.0, 2016.
- [117] P.-N. Tan, M. Steinbach, and V. Kumar, “Classification : Basic Concepts , Decision Trees , and,” *Introd. to Data Min.*, vol. 67, no. 17, pp. 145–205, 2006.
- [118] P. Fergus, S. Iram, D. Al-Jumeily, M. Randles, and A. Attwood, “Home-Based Health Monitoring and Measurement for Personalised Healthcare,” *J. Med. Imaging Heal. Informatics*, vol. 2, pp. 35–43, 2012.
- [119] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.
- [120] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, “RUSBoost: A hybrid approach to alleviating class imbalance,” *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans*, vol. 40, no. 1, pp. 185–197, 2010.
- [121] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and a. Napolitano, “RUSBoost: Improving classification performance when training data is skewed,” *2008 19th Int. Conf. Pattern Recognit.*, no. March 2016, pp. 8–11, 2008.
- [122] D. Al-Jumeily, S. Iram, F.-B. Vialatte, P. Fergus, and A. Hussain, “A Novel Method of Early Diagnosis of Alzheimer’s Disease Based on EEG Signals,” *Sci. World J.*, vol. 2015, pp. 1–11, 2015.
- [123] M. Khalaf, A. J. Hussain, D. Al-jumeily, R. Keight, A. S. Al Kafri, and I. O. Idowu, “A Performance Evaluation of Systematic Analysis for Combining Multi-class Models for Sickle Cell Disorder Data Sets,” vol. 10362, pp. 115–121, 2017.
- [124] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- [125] C. C. Aggarwal, *Data classification : algorithms and applications*. 2014.
- [126] A. S. Al-Kafri *et al.*, *Lumbar spine discs labeling using axial view MRI based on the pixels coordinate and gray level features*, vol. 10363 LNAI. 2017.
- [127] N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm,” *Procedia Comput. Sci.*, vol. 54,

- pp. 764–771, Jan. 2015.
- [128] H. A. Vrooman *et al.*, “Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification,” *Neuroimage*, vol. 37, no. 1, pp. 71–81, 2007.
- [129] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, “Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system,” *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287–300, May 2013.
- [130] J. Shotton, M. Johnson, and R. Cipolla, “Semantic texton forests for image categorization and segmentation,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [131] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
- [132] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.
- [133] G. Lin, A. Milan, C. Shen, and I. Reid, “Refinenet: Multi-path refinement networks for high-resolution semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, vol. 1, no. 2, p. 3.
- [134] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in *International Conference on Learning Representations*, 2015.
- [135] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [136] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248–255.
- [137] D. G. Anasuya, A. Jayashree, N. L. N. Moorthy, and S. Madan, “Anatomical Study of Lumbar Spinal Canal Diameter on Mri To Assess Spinal Canal Stenosis,” *Int. J. Anat. Res.*, vol. 3, no. 3, pp. 1441–1444, 2015.
- [138] G. Wang *et al.*, “Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning,” *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
- [139] A. S. Al-Kafri *et al.*, “Segmentation of Lumbar Spine MRI Images for Stenosis Detection using Patch-based Pixel Classification Neural Network,” *IEEE Congr. Evol. Comput.*, p. to appear, 2018.
- [140] V. Vapnik, E. Levin, and Y. Le Cun, “Measuring the VC-dimension of a learning machine,” *Neural Comput.*, vol. 6, no. 5, pp. 851–876, 1994.
- [141] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [142] S. Krig, “Ground Truth Data, Content, Metrics, and Analysis,” in *Computer Vision Metrics: Survey, Taxonomy, and Analysis*, Berkeley, CA: Apress, 2014, pp. 283–311.
- [143] F. Natalia *et al.*, “Development of Ground Truth Data for Automatic Lumbar Spine MRI Image Segmentation,” *3rd Int. Work. Internet Things Big Data Healthc.*, p. to appear, 2018.

- [144] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI.," *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 153–165, 2000.
- [145] R. V Gilchrist, C. W. Slipman, and S. M. Bhagia, "Anatomy of the intervertebral foramen.," *Pain Physician*, vol. 5, no. 4, pp. 372–378, 2002.
- [146] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014, pp. 740–755.
- [147] A. S. Al Kafri *et al.*, "A Framework on a Computer Assisted and Systematic Methodology for Detection of Chronic Lower Back Pain Using Artificial Intelligence and Computer Graphics Technologies," in *Intelligent Computing Theories and Application*, vol. 9771, no. July, Springer International Publishing, 2016, pp. 843–854.