

CAPITAL UNIVERSITY OF SCIENCE AND  
TECHNOLOGY, ISLAMABAD



**Deep Learning-Based  
Segmentation, Measurements and  
Classification of Lumbar Spinal  
Canal Stenosis**

by

**Abdullah Farooq Abbasi**

A thesis submitted in partial fulfillment for the  
degree of Master of Science

in the

Faculty of Engineering

Department of Electrical Engineering

2023

Copyright © 2023 by Abdullah Farooq Abbasi

All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, by any information storage and retrieval system without the prior written permission of the author.

*Dedicated to all extremely hardworking clinicians and health-care professionals.*



## CERTIFICATE OF APPROVAL

### Deep Learning-Based Segmentation, Measurements and Classification of Lumbar Spinal Canal Stenosis

by

Abdullah Farooq Abbasi

(MEE213010)

### THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Syed Muhammad Tahir Zaidi	NUST, Islamabad
(b)	Internal Examiner	Dr. Nadeem Anjum	CUST, Islamabad
(c)	Supervisor	Dr. Imtiaz Ahmad Taj	CUST, Islamabad

  
Dr. Imtiaz Ahmad Taj

Thesis Supervisor

September, 2023

  
Dr. Noor Muhammad Khan

Head

Dept. of Electrical Engineering

September, 2023

  
Dr. Imtiaz Ahmad Taj

Dean

Faculty of Engineering

September, 2023

## *Author's Declaration*

I, **Abdullah Farooq Abbasi** hereby state that my MS thesis titled “**Deep Learning-Based Segmentation, Measurements and Classification of Lumbar Spinal Canal Stenosis**” is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.



**(Abdullah Farooq Abbasi)**

Registration No: MEE213010

## *Plagiarism Undertaking*

I solemnly declare that research work presented in this thesis titled “**Deep Learning-Based Segmentation, Measurements and Classification of Lumbar Spinal Canal Stenosis**” is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.



**(Abdullah Farooq Abbasi)**

Registration No: MEE213010

## *Acknowledgement*

First and foremost to the creator, the most gracious, the most beneficent, the Almighty **ALLAH S.W.T**, I owe it all to you, Thank you!

There have been many people who have walked alongside me, who have guided me through all these efforts. I would like to outstretch gratitude to each one of them. Topping the list is my supervisor **Dr. Imtiaz Ahmad Taj** to whom I owe my deepest gratitude for providing his valuable guidance to complete this research. Besides that, I am also very grateful to my teachers for their unconditional help as well as technical & motivational support thorough out the research journey.

Furthermore, I owe a great deal to my parents who shaped me into the person I am today. Their continuous support and encouragement made this work possible.

**(Abdullah Farooq Abbasi)**

# *Abstract*

Automated image understanding in medical imaging plays a crucial role in alleviating the time-consuming process of manual image handling for clinicians, while also providing a level of confidence in their diagnostic decision-making. Currently, medical imaging diagnosis heavily relies on manual image handling or semi-automated tools. However, these methods can have variations when performed by different clinicians. This thesis proposes a multi-task application that utilizes a comprehensive image dataset of axial views of lumbar spine disc magnetic resonance imaging (MRI). The dataset contains annotated labels and radiologists' remarks specifically focused on lumbar spinal stenosis. To create a comprehensive and fully automated image understanding application for lumbar spinal canal stenosis, the initial step involves segmenting the intervertebral bodies using deep learning models. Among these models, the Weighted Average Ensemble exhibits the highest performance in semantic segmentation, achieving a dice similarity score (DSC) of 0.98 and an intersection-over-union (IoU) of 0.976. Subsequently, the development of the application includes conducting spinal cord measurements, which encompass diameter and cross-sectional area measurements, to provide a thorough assessment of the condition. These measurements adhere to the clinical standards widely utilized by clinicians in practice. Moreover, an approach is introduced in this study to automatically evaluate lumbar central canal stenosis, offering valuable support to spinal surgeons in making objective decisions regarding necessary surgical interventions. The classification model based on deep learning demonstrates an accuracy of 98 % for lumbar central canal stenosis, whereas the machine learning-based classification achieves an accuracy of 96 %.



# Contents

<b>Author’s Declaration</b>	<b>iv</b>
<b>Plagiarism Undertaking</b>	<b>v</b>
<b>Acknowledgement</b>	<b>vi</b>
<b>Abstract</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xii</b>
<b>Abbreviations</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Outline . . . . .	1
1.2 Background . . . . .	1
1.3 Diagnosis of Lower-Back Pain . . . . .	3
1.4 MRI Scan or CT scan - Subjects’ Preference . . . . .	3
1.5 Body Planes . . . . .	4
1.6 Standard MR Image Sequences for Lumbar Spine . . . . .	5
1.7 MRI Exam Sequence . . . . .	7
1.8 Challenges in Diagnosing Lower Back Pain . . . . .	8
1.9 Significance of Lumbar Central Canal Measurements . . . . .	8
1.10 Challenges in Medical Image Analysis . . . . .	9
1.11 Summary . . . . .	10
<b>2 Literature Review, Problem Statement, and Research Contributions</b>	<b>11</b>
2.1 Overview . . . . .	11
2.2 Segmentation of Medical Image . . . . .	12
2.2.1 Conventional/Traditional Methods of Medical Image Segmentation . . . . .	12
2.2.2 Deep Learning-Based Methods for Medical Image Segmentation . . . . .	13
2.3 Supervised vs Unsupervised Learning . . . . .	14

---

2.4	Commonly Used Grading Systems for Lumbar Central Canal Stenosis	15
2.5	Lumbar Central Canal Measurements	17
2.5.1	Measurements Through Semi-Automated Methods	18
2.5.2	Measurements Through Fully Automatic Methods	18
2.6	Classification of Lumbar Spinal Stenosis	19
2.7	Problem Statement	20
2.8	Contributions	20
2.9	Summary	21
<b>3</b>	<b>Basic Lumbar Spine Anatomy with Pathophysiology and Clinicians' Overview</b>	<b>23</b>
3.1	Outline	23
3.2	Anatomy of Lumbar Spine	23
3.2.1	Main Vertebral Body (VB)	24
3.2.2	Vertebral Arch (VA)	24
3.2.3	Intervertebral Discs (IVDs)	25
3.3	Spinal Disorders (Pathophysiology)	25
3.3.1	Degenerative Disc Diseases and Related VB Changes	26
3.3.2	Common Clinical Features	27
3.4	Lumbar Spinal Stenosis	28
3.4.1	Lumbar Central Canal Stenosis	28
3.4.2	Lumbar Foraminal Stenosis	28
3.4.3	Lumbar Lateral Recess Stenosis	29
3.5	Clinical Evaluation / Analysis Perspectives	29
3.5.1	Radiologists' Viewpoint	30
3.5.2	Spinal Surgeons' Viewpoint	31
3.5.2.1	Lumbar Decompression	31
3.5.2.2	Lumbar Fusion	31
3.5.2.3	Non-Surgical Treatments for Lower Back Pain	33
3.6	In-depth Assessment of Intervertebral Disc Disorders	33
3.6.1	Assessment of Lumbar Stenosis	34
3.7	Summary	34
<b>4</b>	<b>Segmentation for Lumbar Spinal Stenosis</b>	<b>36</b>
4.1	Outline	36
4.2	Dataset	36
4.2.1	Raw Data	37
4.2.2	Extracted Data	37
4.2.3	Ground Truth	38
4.3	Segmentation Using Deep Learning Architectures	39
4.3.1	Deep Network Architectures Overview	40
4.3.2	Choice of Optimizer	44
4.3.3	Cross-Entropy Loss Function	46
4.3.4	Application of Fine-Tuning	46
4.3.5	Weighted Average Ensemble	47
4.4	Summary	49

---

<b>5</b>	<b>Automated Measurements of Lumbar Intervertebral Bodies</b>	<b>50</b>
5.1	Outline . . . . .	50
5.2	Proposed Methodology . . . . .	50
5.2.1	ROI by Thresholding . . . . .	51
5.2.2	Computation of Diameter and Transverse Distance . . . . .	52
5.2.3	Cross-sectional Area . . . . .	52
5.3	Summary . . . . .	53
<b>6</b>	<b>Classification of Lumbar Central Canal Stenosis</b>	<b>55</b>
6.1	Outline . . . . .	55
6.2	Proposed Automated Classification of Lumbar Central Canal Stenosis	55
6.2.1	SVM-Based Classification . . . . .	56
6.2.1.1	Feature Selection . . . . .	56
6.2.1.2	Data Augmentation . . . . .	57
6.2.1.3	Selection of Kernel Function . . . . .	57
6.2.1.4	Training the SVM . . . . .	58
6.2.1.5	Hyperparameter Tuning . . . . .	58
6.2.1.6	Evaluation and Prediction . . . . .	59
6.2.2	ResNet-Based Classification . . . . .	59
6.2.2.1	Image Copping . . . . .	59
6.2.2.2	Data Augmentation . . . . .	60
6.2.2.3	Data Split . . . . .	61
6.2.2.4	Optimizer . . . . .	61
6.2.2.5	Binary Cross Entropy Function . . . . .	62
6.2.2.6	Training Through Transfer Learning . . . . .	62
6.3	Summary . . . . .	63
<b>7</b>	<b>Results and Discussion</b>	<b>64</b>
7.1	Outline . . . . .	64
7.2	Performance Metrics . . . . .	64
7.3	Results . . . . .	67
7.3.1	Segmentation of Intervertebral Bodies . . . . .	67
7.3.2	Classification and Measurements of Lumbar Central Canal Stenosis . . . . .	70
7.4	Summary . . . . .	73
<b>8</b>	<b>Conclusion and Future Work</b>	<b>74</b>
8.1	Thesis Summary and Research Contributions . . . . .	74
8.2	Conclusion . . . . .	75
8.3	Future Work . . . . .	76
	<b>Bibliography</b>	<b>77</b>

# List of Figures

1.1	Human Body Reference Anatomical Planes [10]	5
1.2	(a) T1-Weighted (b) T2-Weighted	6
2.1	Lee Lumbar Central Canal Stenosis Grading System [42]	15
2.2	Sahizas Lumbar Central Canal Stenosis Grading System [43]	16
2.3	Miskin Spinal Stenosis Grading System [45]	17
2.4	Proposed Framework	21
3.1	Sections of Human Spine [56]	25
3.2	Herniated Disc [14]	26
3.3	Spinal Stenosis	27
3.4	Spondylolisthesis and Disc Bulge in Lumbar Spine	29
4.1	T2-Weighted Axial MRI Images of D3, D4 and D5 for Patient ID 001	37
4.2	Regions of Interest of Lumbar Spine in Axial Planes	38
4.3	Ground Truth Labels Images of D3, D4 and D5 for Patient ID 001	39
4.4	UNet Architecture [80]	41
4.5	ResNet Architecture	41
4.6	VGG16 Architecture	42
4.7	InceptionNet Architecture	43
4.8	Transfer Learning Through Fine Tuning Process	47
4.9	Weighted Average Ensemble	48
5.1	Measurements Acquisition (Block Diagram).	51
5.2	Results of Thresholding	52
5.3	Automated Measurements	53
6.1	SVM-Based Classification (Block Diagram).	57
6.2	ResNet-Based Classification (Block Diagram).	59
6.3	Cropped Images	60
6.4	Results of Data Augmentation	61
7.1	Qualitative Comparison of IVB Segmentation Results	68
7.2	Results of LSS Classification and Measurements	71
7.3	Comparison of Confusion Matrices for Both Classifiers	72

# List of Tables

1.1	Category Wise Number of MRI Scans Performed in Armed Forces Institute of Radiology (AFIRI), Rawalpindi, Pakistan from September 2022 to November 2022. . . . .	2
1.2	Comparison of Information Contents in T1- and T2-Weighted Images of Lumbar Spine . . . . .	7
7.1	Class-Wise Intersection Over Union (IoU) of Models . . . . .	67
7.2	Comparative Quantitative Analysis of Methods/Models used for Semantic Segmentation. Mean Pixel Accuracy (MPA), Mean Pixel Precision (MPP), Intersection-over-Union (IoU), Dice Similarity Coefficient (DSC) . . . . .	69
7.3	Comparison of Segmentation Results with Related Researchers' Work. Image Modality (IM), Mean Pixel Accuracy (MPA), Mean Pixel Precision (MPP), Intersection-over-Union (IoU), Dice Similarity Coefficient (DSC) . . . . .	69
7.4	Comparative Quantitative Analysis of Models Used for Classification.	71
7.5	Comparison of Classification Results with Related Researchers' Work.	71

# Abbreviations

<b>ALIF</b>	Anterior Lumbar Interbody Fusion
<b>AUC</b>	Area Under the Curve
<b>CNN</b>	Convolutional Neural Network
<b>CT</b>	Computed Tomography
<b>IVB</b>	Intervertebral Body
<b>IVD</b>	Intervertebral Body Disc
<b>LSCS</b>	Lumbar Spinal Canal stenosis
<b>LSS</b>	Lumbar Spinal Stenosis
<b>MIS</b>	Minimal Invasive Surgery
<b>MRI</b>	Magnetic Resonance Imaging
<b>PLIF</b>	Posterior Lumbar Interbody Fusion
<b>ResNet</b>	Residual Network
<b>ROC</b>	Receiver Operating Characteristic Curve
<b>TLIF</b>	Transforaminal Lumbar Interbody Fusion
<b>UNet</b>	U-Shaped Deep Learning Network
<b>VA</b>	Vertebral Arch
<b>VB</b>	Vertebral Body
<b>VGG</b>	Visual Geometry Group
<b>XLIF</b>	Extreme Lateral Interbody Fusion

# Chapter 1

## Introduction

### 1.1 Outline

This chapter gives an idea of the challenges involved in addressing the problem statement, along with background information. It begins by discussing the human spine and the specific area of the spine that is most susceptible to severe degenerative changes. The two diagnostic examinations recommended by clinicians to diagnose lower-back pain, computed tomography (CT) and magnetic resonance imaging (MRI), are also discussed. User preferences are identified in terms of safety, convenience, and suitability for each examination. The final part of the chapter elaborates on the challenges faced by clinicians while diagnosing patients with lower back pain, and the importance of medical image segmentation in developing an image understanding application.

### 1.2 Background

Lower back pain, which is a common ailment, is considered a chronic backache by radiologists and spinal surgeons [1, 2]. Lower back pain or lumbago is caused by an effect on the lumbar spine, which degrades the quality of life (QoL) of patients [3]. Low back pain is commonly caused by improper weightlifting and prolonged sitting in an incorrect posture. This type of pain is typically localized to the

back and occurs following an episode of acute backache. However, proper rest can provide relief from this condition. To relieve muscular spasms, clinicians may recommend physiotherapy augmented with radiation therapy or ultrasonography in some cases. Lower back pain can be caused by multiple reasons in chronic cases. Lower back pain can occur as a result of a fracture in either the vertebral body (VB) or vertebral arch (VA). Such fractures can lead to a reduction in the space occupied by intervertebral discs (IVDs) situated between the vertebral bodies [4]. The imbalance in load sharing between the intervertebral discs caused by bone fracture leads to the protrusion or herniation of the disc. Consequently, when discs protrude or slip, they apply pressure on the central canal and nerve roots, resulting in tingling pain in the legs. In severe cases of nerve damage, this pressure can lead to paralysis in the lower body [5]. The compression of nerves that causes pain radiating down one or both legs is often referred to as sciatica [6]. With age, the vertebral body (VB) can experience end-plate deformation known as osteophyte formation [7] [8].

TABLE 1.1: Category Wise Number of MRI Scans Performed in Armed Forces Institute of Radiology (AFIRI), Rawalpindi, Pakistan from September 2022 to November 2022.

MRI Scan Type	Months			Monthly	Percentage
	September	Octobe	November	Average	%
Brain	646	610	620	625	45.27
Lumbar Spine	305	321	294	307	22.2
Cervical Spine	285	225	164	225	16.26
Shoulder	36	27	29	31	2.22
Sacroiliac Joint	47	40	40	42	3.06
Knee	151	162	68	75	5.41
Elbow	7	5	3	57	4.15
Hip	6	5	42	18	1.28
Wrist	2	1	3	2	0.14
<b>Total</b>	<b>1485</b>	<b>1396</b>	<b>1263</b>	<b>1382</b>	

In this region of the vertebral body (VB), there is a possibility of impinging on the exiting nerve roots and resulting in compression. In such cases, the patient may be deemed suitable for a spinal surgical intervention, typically involving a



decompression procedure. This procedure aims to alleviate the pressure exerted on the nerves, ultimately improving the patient's quality of life (QoL).

### 1.3 Diagnosis of Lower-Back Pain

When a patient experiences back pain, especially in the lower back, healthcare professionals commonly suggest undergoing an MRI examination of the lumbar spine to assist with the diagnostic process. According to research conducted by Suri et al. [5], lumbar spine MRI exams are widely recognized as standard procedures for diagnosing lower back pain. In the study, 313 patients with lower back pain complaints were analyzed, with 77% of them undergoing MRI exams. Among those who underwent the examination, 73.8% exhibited abnormal findings that indicated the presence of lumbar spine disease. A total of 183 abnormal findings were detected out of 313 patients, which represents 58.5% of those diagnosed with lower back pain.

Additionally, data were collected on the number of MRI exams performed at a large radiology institute in Pakistan. Table 1.1 displays the data, which shows that the number of MRI exams performed on the spine region is relatively higher compared to other body regions, except the brain. Lumbar spine and cervical exams accounted for 38.47% of the total scans, while the majority (45.27%) were conducted to investigate brain-related diseases.

### 1.4 MRI Scan or CT scan - Subjects' Preference

The two main medical imaging techniques are CT scan and MRI, with patients generally preferring MRI as it does not expose them to ionizing radiation, making it the safest option. While CT scan has their own uses, medical researchers have utilized CT images for spine-related tasks. The main differences between MRI and CT scan are as follows:

- CT scan exposes the patient to more x-ray radiation than a single-plane x-ray scan. The reconstruction of images occurs post-scan and offers more information in multi-planar images. The image generation principle is similar to that of conventional x-ray images, where high-density tissues like bones appear brighter than low-density tissues such as lungs and kidneys.
- MRI examination, Unlike CT scans, MRI does not expose the patient to ionizing radiation. MRI scans take longer to perform but provide superior details compared to CT scans [9].
- Medical consultants recommend CT scans for bone-related disease investigations, while MRI is preferred for organ/muscle-related disease investigations.

Upon assessing the efficacy of MRI scans in examining lower back pain, the standard sequences for MRI of the lumbar spine are elucidated, highlighting the specific information conveyed within the images. Additionally, the MRI scan process relevant to the patient is detailed.

## 1.5 Body Planes

To understand the standard lumbar spine sequences, it is important to know about the standard reference planes and positions used in the human body. These planes are mainly used in medical imaging to locate and identify internal organs according to the area of interest [10]. The human body can be divided into three two-dimensional planes for this purpose as in Fig. 1.1

1. The sagittal plane refers to a vertical plane that divides the body into right and left sections when viewed from the side.
2. The axial plane, also known as the transverse plane, is a horizontal plane that divides the body into upper and lower sections, parallel to the ground. It provides an observer with a view of the body from either the top or bottom perspective.

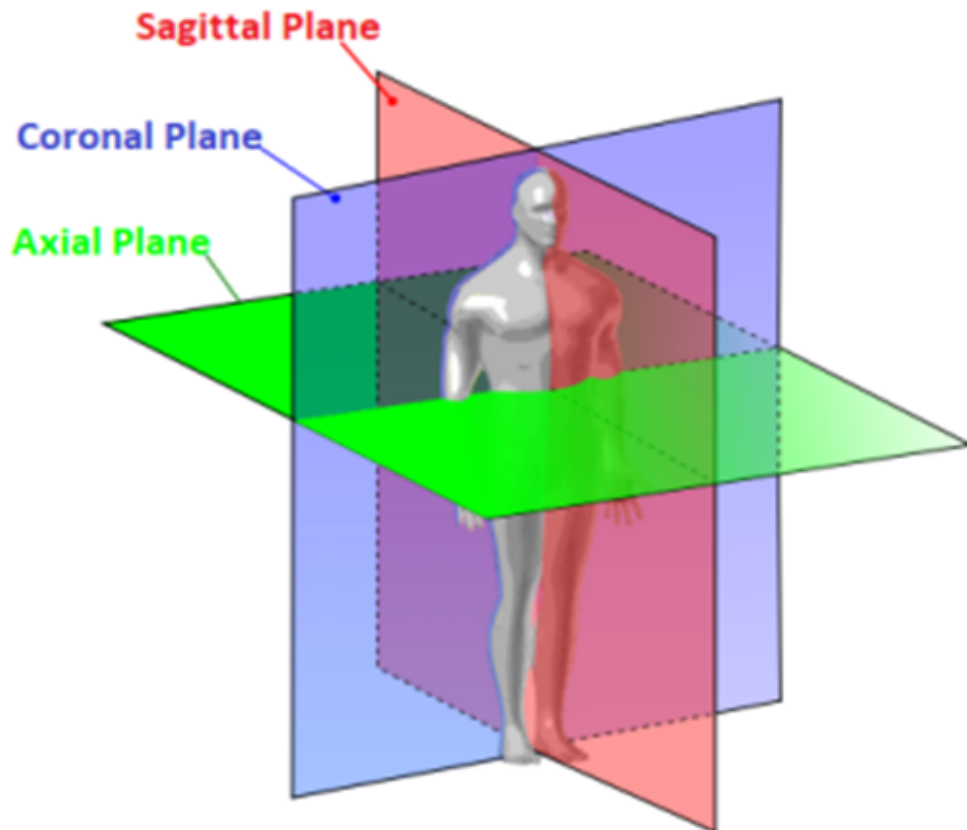


FIGURE 1.1: Human Body Reference Anatomical Planes [10]

3. The third is the coronal plane, which divides the body into front and back portions when the observer is facing it from the front.

## 1.6 Standard MR Image Sequences for Lumbar Spine

To understand the standard lumbar spine sequences, it's important to know the basics of MRI image construction. In MRI, the human body's tissues, which contain higher hydrogen nuclei that behave like tiny magnets, are exposed to an external magnetic field that causes them to align with the field, a phenomenon called precession. [11] The protons in the aligned hydrogen nuclei are then excited by a radio frequency of a specific frequency known as the Larmor frequency, and the protons de-excite, generating an MR signal that's subsequently used to construct the image using spatial encoding.

The relaxation of protons and emittance of MR signal depends on the tissue type and are categorized into two types: T1 relaxation and T2 relaxation.[12] Standard practice involves generating sagittal and axial views for T1- and T2-weighted images, but in certain cases, coronal views and T2-weighted fat-saturated sagittal slice sequences are also generated. [13] Fig. 1.2 shows a side-by-side view of both T1- and T2-weighted images of the lumbar spine, depicting end-plate deformation in the VB, disc bulge, and contrast difference of hydrated and dehydrated IVD. T1-weighted images provide a bright (white) representation of fat-containing struc-

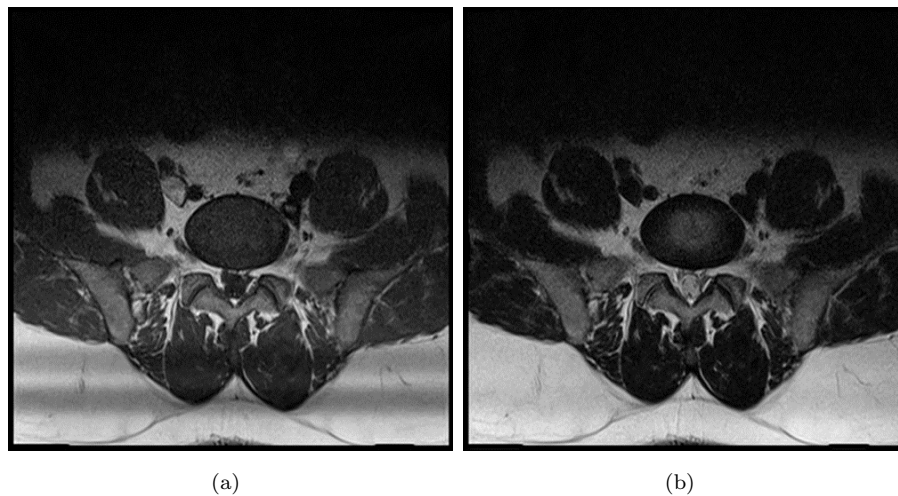


FIGURE 1.2: (a) T1-Weighted (b) T2-Weighted

tures, while T2-weighted images provide a bright (white) representation of both fat and fluid-based contents. T1 images are generally more suited for anatomical depiction, while T2 images are considered for pathological evaluation. For the purpose of this thesis, the image acquisition modes are limited to basic T1- and T2-weighted images only, although certain variations are used specifically for the anatomical region under study.

Table 1.2 summarizes the variation in the information presented by both T1 and T2-weighted images for the lumbar spine region. The easiest way to differentiate between a T1 and T2-weighted image is to identify cerebrospinal fluid (CSF), which gives bright intensity levels in T2-weighted images and dark intensity in T1-weighted images. Additionally, the hydrated IVDs appear bright in T2-weighted images, whereas in T1-weighted images, both hydrated and dehydrated IVDs appear dark [14–16].

## 1.7 MRI Exam Sequence

As previously mentioned, MRI scans provide more detailed information compared to CT scans, making them a better option for investigating the lumbar spine region. A patient is recommended for an MRI examination of the lumbar spine by a specialist consultant to determine the cause of lower-back pain or specific complaints. The scan, which can take between 20-45 minutes depending on the MRI machine's magnetic strength, typically produces sagittal and axial slices. However, for patients with scoliosis, coronal scans may also be included in the MRI exam [17].

TABLE 1.2: Comparison of Information Contents in T1- and T2-Weighted Images of Lumbar Spine

	<b>Dark</b>	<b>Bright</b>
<b>T1</b>	IVDs (All)	Epidural Fat
	CSF	Sub-cutaneous tissue fat
	Muscle Mass	
	Spinal Cord	
	Blood Vessels	
	VB End-plates	
	Nerves	
<b>T2</b>	IVDs (Dehydrated)	IVDs(Hydrated)
	Muscle Mass	*Epidural Fat
	Spinal Cord	*Sub-cutaneous tissue fat
	Blood Vessels	CSF
	VB End-plates	
	Nerves	

\*In case of non-fat saturated image

Once the scan is complete, the Digital Imaging and Communications in Medicine (DICOM) [18] images are processed through the Picture Archiving and Communications System (PACS) [19], where a radiologist reviews and evaluates the images to generate a report. The radiologist analyzes the MRI images without physically examining the patient, and the patient receives the report along with the MRI scans for review by the referring consultant. A spinal surgeon then evaluates the patient's clinical findings and establishes a correlation between their physical

symptoms and the MRI scan. Based on this, a decision is made on whether to proceed with conservative treatment or an appropriate surgical intervention to treat the condition.

## 1.8 Challenges in Diagnosing Lower Back Pain

Diagnosing lower back pain is frequently a subjective and time-consuming duty. Different clinicians may evaluate the same patient differently due to variations in their experiences and skill sets, which are further compounded by the variability in clinical findings. To address this issue, there is a need for quantitative methods to establish a concrete analysis of the lumbar spine region. Currently, spinal surgeons and radiologists rely on manual or software-assisted image understanding methods, which can be laborious, time-consuming, and subjective. These methods can be subject to intra- and inter-medical specialty variations, such as those within radiologists and spinal surgeons. Quantitative methods such as spinal alignment measurements, spinal deformation attributes, and spinal balance readings can be used to perform a more objective spinal assessment.

## 1.9 Significance of Lumbar Central Canal Measurements

Accurate measurement of the lumbar central canal plays a crucial role in diagnosing and treating various spinal conditions. The central canal, a narrow passage within the spinal cord, accommodates the spinal cord and nerves. Any alterations or abnormalities in its size can result in nerve compression, leading to symptoms such as pain, numbness, or weakness in the lower extremities. Precise assessment of the central canal is essential for diagnosing conditions such as spinal stenosis, herniated discs, and degenerative disc disease.

Additionally, central canal measurements can be used to assess the effectiveness of treatments such as surgery or physical therapy. Therefore, lumbar central canal

measurements are an essential component of the diagnosis and management of spinal disorders [19].

## 1.10 Challenges in Medical Image Analysis

The field of medical image analysis has undergone significant developments in recent times to address challenges in segmentation, identification, and labeling. The exponential growth of medical imaging data with varying characteristics such as high volume, speed, and diversity has increased the need for dynamic and robust techniques for extracting and evaluating image information.

However, undesired artifacts and noise, such as inhomogeneities in intensity and motion-related noise, present in MRI scans can make it challenging for researchers and clinicians to accurately identify diseases. To overcome these difficulties, pre-processing and post-processing techniques are used.

MRI is generally preferred for soft-tissue imaging due to the dark appearance of bones in MR images. However, as bony structures visible in MRI scans are a combination of bone, fat, and water, the overlapping regions of bones and soft tissues pose an additional challenge for segmentation using conventional methods based on edge-based or intensity-based segmentation. Additionally, the content of the image can vary significantly from one subject to another, adding to the difficulty.

As previously mentioned, radiologists focus on identifying the presence of spinal disease, while spinal surgeons evaluate patients based on their symptoms and attempt to correlate MR images with pain symptoms. To establish this correlation, spinal quantitative assessment is done either manually or with software-assisted methods. Therefore, there is a need for an automated quantitative assessment tool for the lumbar spine region to offset inter- and intra-clinician variations by obtaining reproducible automated readings [20, 21].

## **1.11 Summary**

Clinicians currently rely on subjective manual assessment when evaluating patients with lumbago, which can vary depending on the skill and experience of the clinician. There is a need to develop an automated tool to analyze MR images and establish a quantitative correlation between perceived symptoms and spinal disorders. Overcoming the challenges in MR imaging requires the implementation of rigorous image processing methods, including image segmentation and subsequent automated spinal measurements.



## Chapter 2

# Literature Review, Problem Statement, and Research Contributions

### 2.1 Overview

This chapter presents a concise literature review on the development of a system to identify the thecal sac compression by using a ground truth image dataset for lumbar spine disc images, methods and techniques of various researchers to segment the region of interest, which is the Thecal sac in the lumbar spine disc region for this thesis. Additionally, Section 2.5 discusses the methods and techniques used for spinal canal measurements that are clinically significant before surgical intervention. A brief analysis of the reviewed work is also presented to determine its suitability for research continuity. Finally, Section 2.7 analyzes existing research gaps and the efforts made in this thesis to address them.

## 2.2 Segmentation of Medical Image

In this section of the research thesis, the discussion revolves around prior research on medical image segmentation, including both traditional and machine learning-based methods. Sub-section 2.2.1 presents conventional approaches, while Sub-section 2.2.2 provides a review of deep learning techniques used for medical image segmentation tasks. In Sub-section 2.3, a brief comparison is provided between unsupervised and supervised deep learning methods for segmentation. Previous methods for segmentation can be categorized as semi-automatic and fully automatic. Semi-automatic methods involve some user intervention, such as placing landmarks, while fully automatic methods require no user involvement. A summary of previous research is provided, although it is not exhaustive.

### 2.2.1 Conventional/Traditional Methods of Medical Image Segmentation

Smyth et al. [22] utilized Active Shape Model (ASM) to quantify the VB shape on a dataset consisting of 84 bone density scans or dual-energy X-ray absorptiometry (DXA) images. They manually labelled the vertebrae on each image, including 10 vertebrae total, 6 thoracic from T7-T12 and 4 lumbar from L1-L4. Aslan et al. [23] used a universal shape model for CT image segmentation. Carballido et al. [24] employed normalized cuts for spinal MRI segmentation and sought user input to select the vertebrae to keep, finding that the segmentation results were reduced if the mid-sagittal slice was not precisely chosen. Generalized Hough transform (GHT) [25] template matching was used to locate lumbar vertebrae in X-ray images. In their research, Zhu et al. [26] employed Gabor filter banks to identify and estimate spinal curves and intervertebral disc (IVD) features. They further proceeded to segment the IVDs using a dataset that consisted of T2 images from 37 patients at a hospital in China. Their findings revealed a localization accuracy of 98.23% and a Dice Similarity Coefficient (DSC) of 0.9237 for the segmentation process.

Clustering-based fuzzy c-means algorithm segmentation was performed using GHT on a dataset comprising cervical radiograph scans, with claimed accuracy of 96.88% [27]. K-means clustering and GHT (template matching) were used in another study [28]. Spinal canal segmentation and extraction was performed using k-means clustering in T2-weighted images of sagittal slices in a study by Bampis et al [29]. Glocker et al [30]. utilized classification random forests to locate (finding centroids) based on a CT image dataset and reported an 81% identification rate with an overall median localization error of less than 6mm.

## 2.2.2 Deep Learning-Based Methods for Medical Image Segmentation

Lu et al. [31] employed the U-Net architecture to segment sagittal MR images with an input size of 512 x 512 pixels. Their evaluation focused on accurately detecting the number of vertebral bodies (VBs) and ensuring there was no overlap between the detected sacrum bone and the lumbar area, in comparison to ground truth (GT) images. Their results showed successful detection of 188 out of 200 VBs, achieving an accuracy of 94%. The dice similarity coefficient (DSC) was measured at 0.93 with a standard deviation of 0.02. Additionally, they calculated the mean error distance between the center of GT images and the detected center, resulting in 0.79 mm with a standard deviation of 0.44 mm.

Janssens and Zheng [32] utilized the Fully Connected Network (FCN) to segment lumbar spine vertebrae in CT images. Their approach involved using a localization net based on FCN to crop the region of interest (Lumbar Spine), followed by the Segmentation-Net (FCN) to segment the vertebrae. They achieved a dice similarity coefficient of  $0.9577 \pm 0.81$  and an average symmetric surface distance of  $0.37 \pm 0.06$  mm.

Lessmann et al. [33] employed CNN for vertebrae segmentation, attaining an average dice similarity coefficient of 94.9% across various datasets, including CT and MRI images.

Tang et al. [34] proposed the use of the Dual Densely Connected U-Net architecture for segmenting axial scans extracted from CT images. Their method achieved

high accuracy metrics, including pixel accuracy, mean pixel accuracy, mean Intersection over Union, and frequency-weighted IoU. Benjdira et al. [35] employed the U-Net architecture to perform semantic segmentation of Ultrasound Images of the Lumbar Spine. This segmentation approach is particularly beneficial for spinal surgeons during and after laminectomy procedures. Hassan et al. [36] conducted a comparative analysis of different deep-learning frameworks for retinal lesion segmentation. Their study revealed that RAGNet exhibited the most promising outcomes due to its ability to retain contextual information of the lesions during image decomposition, making it well-suited for retinal images. Siriwardhana et al. [37] introduced the usage of the U-Net architecture to segment T2 axial MRI images, achieving an mIOU (mean Intersection over Union) of 88.62% with this model. The dataset employed in their study consisted of 1545 patients with lower back pain. Narasimharao Kowlagi et al. [38] use combined FPN + ResNet34 and achieve mIOU  $97.5 \pm 0.1$ . The dataset used in this study was 1500 MRI with a resolution of 512x512. Merve Apaydin et al. [39] utilized the U-Net architecture for Intervertebral Disc Segmentation in T2-weighted axial MRI. A dataset of 515 patients MRI is used and analyzed the performance on different data splits achieves 0.99 pixel accuracy and 0.92 mIoU when 90% train and 10% test data. Abhinav shukla et al. [40] employed convolutional neural network based 2D U-Net architecture to segment intervertebral bodies in T1-weighted Axial MRI. The dataset used in this study was 1545 MRI with a resolution of 128x128. They achieved a mIoU of 0.714.

This thesis explores the utilization of pre-trained networks for segmentation tasks. The focus is on training different models using fine-tuning, a highly effective transfer learning technique. More information about the application of fine-tuning can be found in Chapter 4, specifically in Section 4.3.4.

## 2.3 Supervised vs Unsupervised Learning

Joyce et al. proposed the use of generative adversarial networks (GAN) [41] for the segmentation and reconstruction of cardiac images, including both CT scan and MR images, by utilizing labels from a different dataset with the same anatomy.

Their study suggests that supervised methods have shown promising results, as they were able to achieve a dice similarity score of 0.84 on MR images and 0.87 on CT scan images with the help of supervised methods using the U-Net architecture. In contrast, unsupervised methods achieved a dice similarity coefficient of 0.66 for MR images and 0.51 for CT scan images.

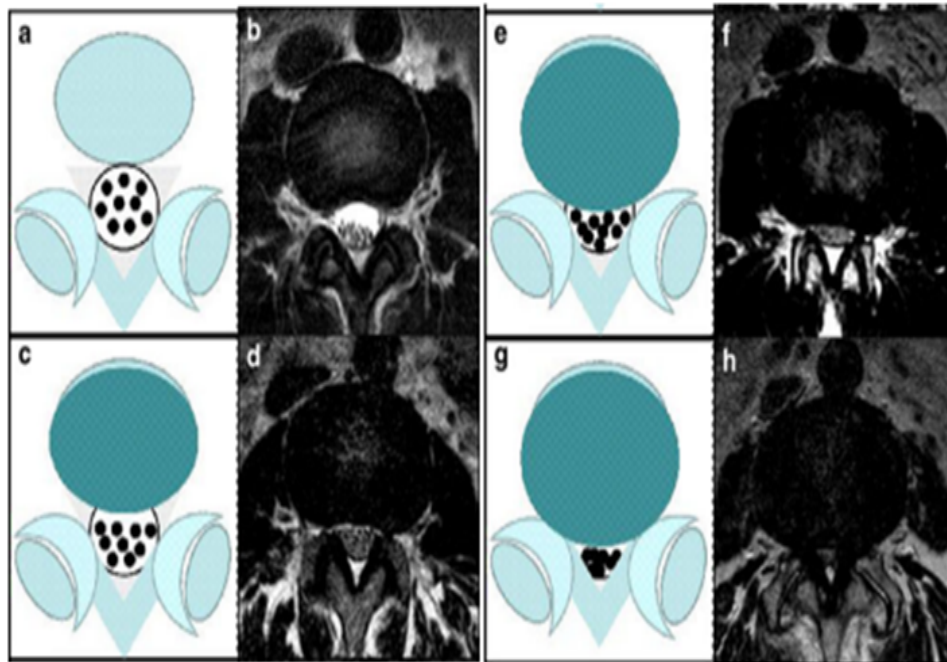


FIGURE 2.1: Lee Lumbar Central Canal Stenosis Grading System [42]

## 2.4 Commonly Used Grading Systems for Lumbar Central Canal Stenosis

The Lee grading system is an MRI-based classification system used to assess the severity of central canal stenosis in the lumbar spine. Lee [42] categorized central canal stenosis into four grades, considering the degree of compression on the dural sac, which houses the spinal cord and nerve roots. Additionally, the system takes into account factors such as disc herniation, facet joint osteoarthritis, and ligamentum flavum hypertrophy, all contributing to the level of central canal stenosis. The reliability between different readers was found to be almost perfect (ICC reliability=0.730–0.953), while the intra-reader reliability was also nearly perfect (kappa

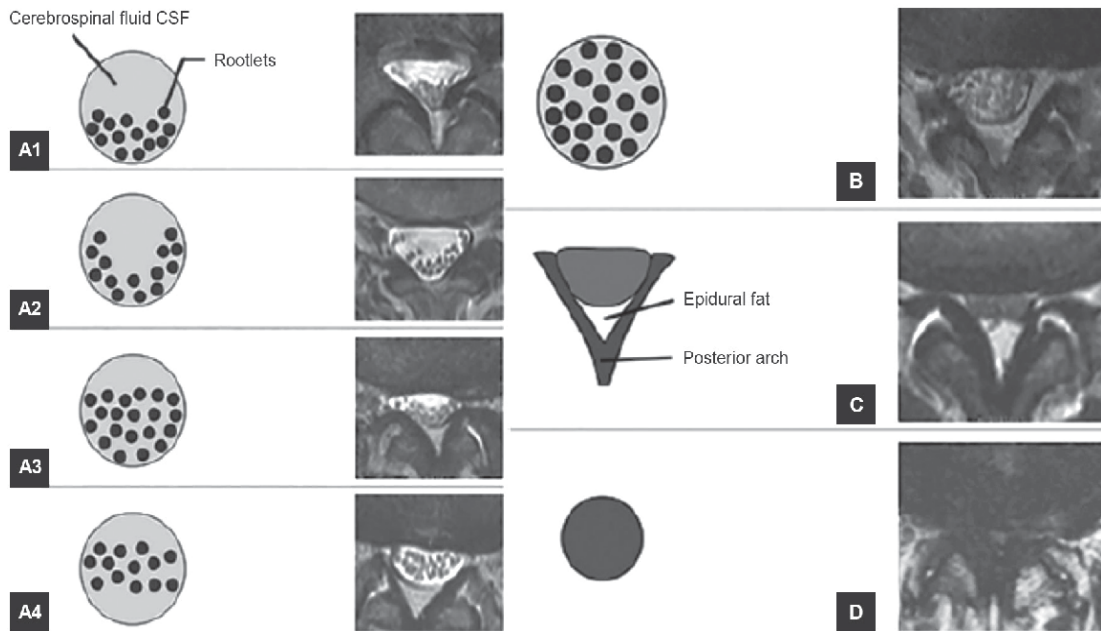


FIGURE 2.2: Sahizas Lumbar Central Canal Stenosis Grading System [43]

value=0.863–0.900). On the other hand, the Schizas grading system [43] uses a 7-grade classification based on the morphology of the dural sac observed in T2 axial magnetic resonance images. This system evaluates the rootlet/cerebrospinal fluid ratio and relates the morphological grading to digital measurements obtained by measuring the dural sac cross-sectional area (DSCA), AP diameter, and transverse distance using OSIRIX software. However, these measurements pose challenges due to their time-consuming nature and the variability in the number of nerve roots depending on the spinal level. The average intra- and inter-observer agreement showed substantial and moderate results, respectively ( $k = 0.65$  and  $0.44$ ).

This part of the thesis discusses common grading methods that clinicians and radiologists use to grade stenosis in their professional life. The Lee grading system is a classification system used to grade the severity of central canal stenosis in the lumbar spine based on magnetic resonance imaging (MRI) scans.

Average intra- and interobserver agreement were substantial and moderate, respectively ( $k = 0.65$  and  $0.44$ ). In the study by Yeon-jeong Ko et al. [44], five experts in the field evaluated both the Lee and Schizas grading systems. The evaluation involved two clinical fellows, one novice radiology resident, one neurosurgeon, and one orthopedic surgeon who were unaware of these grading systems. The analysis

included 70 patients, with a total of 280 disc levels assessed using T2-weighted axial MRI for grading lumbar central canal stenosis according to both the Lee and Schizas grading systems. The inter-observer agreements revealed similar results for both systems. ICC ranged of Schizas grading system from 0.827 to 0.983 and Lee grading system from 0.840 to 0.983. The Miskin grading system [45] addresses

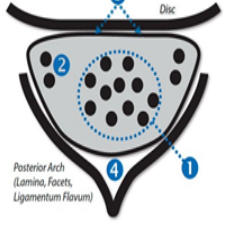



 <p><b>Normal</b></p>	 <p><b>Mild Spinal Stenosis</b></p>	 <p><b>Moderate Spinal Stenosis</b></p>	 <p><b>Severe Spinal Stenosis</b></p>
<p>1 • Free distribution of nerve roots, without crowding</p>	<p>• Slight crowding of nerve roots</p>	<p>• Crowding of nerve roots, resulting in a "speckled" appearance of CSF interspersed with nerve roots</p>	<p>• Complete effacement of CSF, resulting in nerve roots not being individually distinguishable</p>
<p>2 • Traversing dorsal and ventral nerve roots in the lateral recesses are distinct</p>	<p>• Traversing dorsal and ventral nerve roots in the lateral recesses remain distinct, but there may be abutment of the nerve roots in the lateral recesses</p>	<p>• Difficult to differentiate the traversing nerve roots in the lateral recesses</p>	<p>• Cannot discretely identify nerve roots in the lateral recesses</p>
<p>3 • Anterior margin of the thecal sac is flat or convex</p>	<p>• Anterior margin of the thecal sac is flat or concave</p>	<p>• Anterior margin of the thecal sac is concave</p>	<p>• Anterior margin of the thecal sac is concave</p>
<p>4 • Posterior epidural fat is preserved (dependent on level)</p>	<p>• Posterior epidural fat is preserved (dependent on level)</p>	<p>• Posterior epidural fat is preserved (dependent on level)</p>	<p>• Posterior epidural fat may be partially or completely effaced</p>

FIGURE 2.3: Miskin Spinal Stenosis Grading System [45]

the classification of three types of stenosis: Spinal stenosis, Foraminal Stenosis, and Lateral Recess Stenosis. In the case of Spinal stenosis, Miskin simplifies the classification proposed by Schizas, categorizing it as normal, mild, moderate, and severe. Intermediate grades such as mild-moderate and moderate-severe are also permitted, although not explicitly described. The evaluation of the grading system was conducted using axial T2-weighted images. For spinal stenosis, the readers demonstrated substantial agreement, with a kappa value of 0.702.

## 2.5 Lumbar Central Canal Measurements

This study did not review the current manual techniques used for spinal measurements, which involve physical tools such as inclinometers, flexible rulers, and spinal mice, and radiography images. Instead, we provide a thorough review of



semi-automated measurement techniques that require human intervention via software assistance and fully automatic methods that require no human intervention [46].

### 2.5.1 Measurements Through Semi-Automated Methods

In the past and currently, DICOM viewers with integrated measurement tools have been widely utilized by spinal surgeons and radiologists for spinal measurements. However, certain researchers have utilized computer software for evaluating spinal stenosis measurements and have relied on manual marking of landmarks to make measurements [47–49]. The manual landmarks in the process involve identifying the centroid of the intervertebral disc and thecal sac. These landmarks serve as reference points for estimating other spinal measurements like the anterior to posterior diameter and the cross-sectional area of the dural sac. In the study conducted by Bharadwaj et al. [50], the MD.ai software was utilized to delineate the boundaries of intervertebral bodies and obtain measurements such as the dural sac and intervertebral disc diameter, as well as the cross-sectional area of the dural sac and intervertebral disc. These measurements were then used to grade lumbar central canal stenosis by employing a binary classifier to differentiate between normal and stenotic cases. The proposed metrics, namely DDRDIA and DDRCA, achieved an accuracy of 96.2% and 94.6%, respectively.

### 2.5.2 Measurements Through Fully Automatic Methods

Lately, efforts have been made towards creating computer-based methods for comprehending images that rely on both fully and semi-automated diagnosis. This would aid clinicians in their manual diagnosis process.

Jiawei Huang et al. [51] introduced the utilization of the U-Net architecture for segmenting the lumbar spine using 486 T2-weighted sagittal images. The segmentation achieved a mean intersection-over-union of 94.7% for vertebra body and 92.6% for disc segmentation. The agreement between the vertebra and disc measurements obtained through the segmentation and ImageJ was excellent, with ICC



values ranging from 0.81 to 1.00.

Friska Natalia et al. [52] employed the Seg-Net architecture to segment T2-weighted axial MRI images of the lumbar spine. After segmentation, nine landmark points were identified on the segmented image, and the software connected these points to determine the anterior-posterior (AP) diameter and foraminal widths. The average error of the calculated AP diameter and foraminal widths, when compared to expert calculations, was 0.90mm and 0.28mm, respectively.

Siriwardhana et al. [37] utilized the U-Net architecture for the segmentation of T2-weighted axial MRI images. The model achieved a dice coefficient of 99.52% and an intersection-over-union (IOU) of 88.62%. Additionally, edge detection techniques were applied to measure the minimum distance between the intervertebral disc and the posterior element.

## 2.6 Classification of Lumbar Spinal Stenosis

Tackeun Kim et al. [53] proposed a transfer learning algorithm based on CNN for classifying lumbar spinal stenosis (LSS) and normal cases. The VGG19 architecture was employed as the backbone model, trained on a dataset of 12,442 images. The algorithm achieved an accuracy of 82.8% and an AUROC of 90.0%.

Jen-Tang Lu et al. [31] developed a multi-input, multi-task, and multi-class CNN for grading lumbar central canal and foraminal stenosis using both axial and sagittal images. The study utilized a dataset consisting of 22,796 disc levels extracted from 4,075 patients. The classification accuracy for multi-class grading was found to be 78.6%, and the AUROC for the binary classifier was 0.97.

Dongkyu Won [54] use the CNN model for the detection of ROI and then further train a VGG architecture-based model for four grades of lumbar spinal stenosis. Dataset of 542 T2 weighted axial L4-L5 images. Grading agreement between experts achieves an accuracy of 77.5% and F1 score of 75%.

Upasana Upadhyay Bharadwaj et al. [50] present the use of a binary decision tree for the classification of central canal stenosis and achieve AUROC of 0.95 Alessandro Siccoli et al. [55] utilized various machine learning methods, and among them,

the results of the XGBoost machine were promising, achieving an accuracy of 0.85, an AUC of 0.92, a sensitivity of 0.86, and an F1 score of 0.87.

## 2.7 Problem Statement

Currently, diagnosing medical conditions related to the lumbar spine disc is subjective and lacks quantitative assessment. As shown in the above Sections, clinicians depend on software-assisted methods for making measurements, which can be time-consuming and subject to variation based on the skill and expertise of the clinician. Therefore, there is a need to develop automated image understanding tools that can facilitate decision-making and improve the diagnostic ability of both radiologists and spinal surgeons in measuring spinal disorders and classifying diseases.

## 2.8 Contributions

Addressing to bridge the gap, the following efforts have been made in this research thesis:

1. In Section 4.3, a conventional fully automated image segmentation is suggested for segmenting the intervertebral disc. Additionally, existing deep learning methods that are currently used for segmentation related tasks are extensively tested on the same dataset as well, as described in Section 4.2. A quantitative comparison is made between the results obtained using the proposed conventional method and those obtained using deep learning methods.
2. In Section 5.2, a mathematical model is presented to extract important measurements of the lumbar spinal disc and thecal sac that hold clinical significance. These measurements encompass the anterior-posterior diameters and cross-sectional area of the disc, as well as the anterior-posterior diameters and cross-sectional area of the dural sac.

3. Additionally in Section 6.2, proposed a fully automated lumbar spinal stenosis classification on the basis of the quantitative measurements.

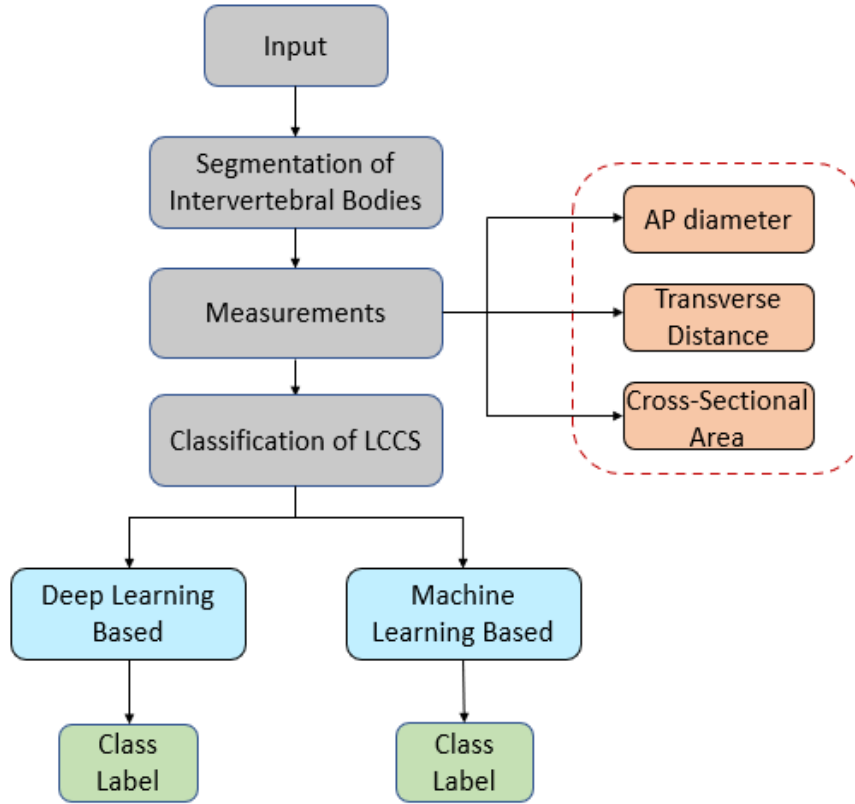


FIGURE 2.4: Proposed Framework

The proposed framework of the thesis is presented in Fig 2.4, showing the input image being the first step followed by IVB segmentation including the identification of the Intervertebral disc and thecal sac. The third step being the measurements related to lumbar spinal disc geometry while the last step in which lumbar spinal stenosis classification is performed through automated means.

## 2.9 Summary

Deep learning networks have shown promising outcomes in semantic segmentation, surpassing conventional machine learning-based methods and algorithms. Creating a tool for automated measurements not only saves time in acquiring relevant spinal measurements but also adds confidence to the readings obtained with the aid of computer-assisted software. To develop an automated tool for a specific anatomical site, it is critically essential to understand the clinical requirements

and preferences of both radiologists and spinal surgeons to make the tool clinically beneficial. Chapter 3 covers the clinical preferences of both radiologists and spinal surgeons while diagnosing lower-back related diseases and disorders.

# Chapter 3

## Basic Lumbar Spine Anatomy with Pathophysiology and Clinicians' Overview

### 3.1 Outline

This section presents a brief summary of the anatomical structure of the lumbar spine and related conditions. It aims to familiarize the reader with the fundamental factors that radiologists and spinal surgeons take into account when diagnosing individuals with lower-back pain. Towards the end of the chapter, there is a description of several spinal measurements that are essential for spinal surgeons to determine before performing spinal intervention procedures. These measurements are performed manually or with the assistance of software and are crucial in restoring the normal curvature of the lumbar spine in individuals with excessive inward curvature in that area.

### 3.2 Anatomy of Lumbar Spine

The human spine, commonly referred to as the backbone, is comprised of five main regions: cervical, thoracic, lumbar, sacrum, and coccyx. In total, there are usually

33 vertebrae, as shown in Fig. 3.1, which presents a left view in the sagittal plane and a subsequent section axial view on the right. This research thesis focuses only on the lumbar spine region, which begins after the last thoracic vertebra (typically T12) and ends at the fused sacrum bone. The lumbar spine is composed of five vertebrae, namely L1 to L5, each consisting of a main vertebral body (VB) in the anterior or front-side, and a vertebral arch (VA) in the posterior or back-side, when viewed laterally or in sagittal view.

### **3.2.1 Main Vertebral Body (VB)**

The primary load-bearing component of the lumbar spine is the main axial structure situated in the front region. The gaps between neighboring vertebral bodies are occupied by intervertebral discs (IVDs), which are commonly referred to by their respective levels, such as L1-L2, L2-L3, L3-L4, L4-L5, and L5-S1. These IVDs are made up of fibrocartilage and act as cushions to the vertebral column, providing both stability by holding adjacent vertebral bodies together and allowing mobility of the spine. The outer part of the IVD is called anulus fibrosus, while the inner part is referred to as nucleus pulposus. Any change in the shape, size, or structure of the IVDs can cause an imbalance in load sharing and functional issues.

### **3.2.2 Vertebral Arch (VA)**

Each vertebral arch (VA) consists of a pair of transverse and articular processes on the left and right sides, along with a single spinous process, formed by the joining of the laminae on both sides. The VA protects the spinal canal, which is the cavity through which the spinal cord, along with cerebrospinal fluid (CSF), passes. It also provides stability to the spine during excessive flexion and shear forces [56].

### 3.2.3 Intervertebral Discs (IVDs)

The intervertebral discs consist of two main components: the annulus fibrosus, which is an outer fibrous ring, and the nucleus pulposus, a gel-like substance located inside. The annulus fibrosus plays a crucial role in enclosing the nucleus pulposus and preserving the shape of the disc. Meanwhile, the nucleus pulposus functions as a shock absorber, evenly distributing the load throughout the vertebral column. The intervertebral discs are susceptible to damage, which can lead

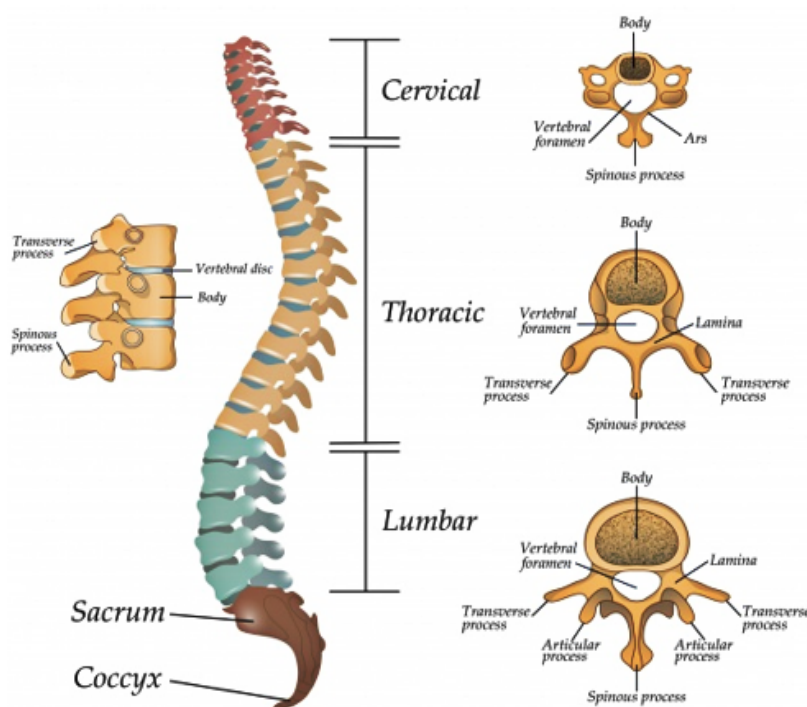


FIGURE 3.1: Sections of Human Spine [56]

to a variety of spinal conditions including herniated discs, degenerative disc disease, and spinal stenosis. For the purpose of this research thesis not consider the posterior element but focus on thecal sac and disc because measurements related to lumbar central canal stenosis involve these regions of IVD.

## 3.3 Spinal Disorders (Pathophysiology)

This section presents the issues and disorders related to the vertebral bodies (VB) and intervertebral discs (IVD), along with the clinical features that demonstrate the significance of these disorders.

### 3.3.1 Degenerative Disc Diseases and Related VB Changes

According to [57], degenerative disc diseases are the most common ailment that requires surgical intervention. Over 90% of spinal surgeries are performed to correct degenerative disc diseases and related issues [106]. Patients with lower back pain often have some form of degenerative disease, either due to abnormal wear, known as spondylosis, in VB/IVD, or reduction of cartilage covering the bones, known as osteoarthritis [58]. These changes can result from aging or trauma. Some

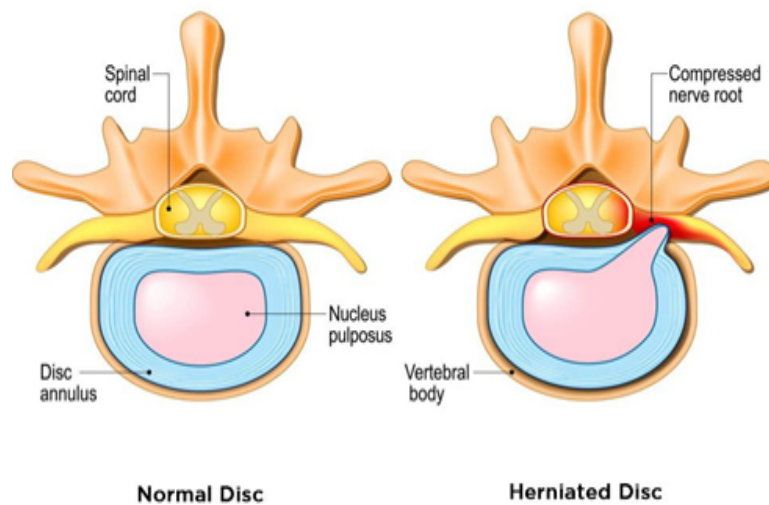


FIGURE 3.2: Herniated Disc [14]

degenerative changes that result from aging may not cause pain [14], but they can dehydrate the center of IVD, limiting the disc's ability to absorb shocks and making it more vulnerable to permanent damage. Fractures in VB can also contribute to degenerative changes in VB and IVDs. In some cases, end-plate deformation or VA deformation can result in the formation of bone spurs, known as osteophytes [56], which can significantly pressure existing nerves and cause stenosis. Some of the changes in the spine caused by degenerative diseases of IVDs are as follows:

1. Dehydration of the nucleus pulposus leads to disc narrowing [56], reducing shock absorption, altering spinal balance, and making the discs vulnerable.
2. Annular tear refers to cracks in the outer ring of fiber, known as annulus fibrosis [59].



3. Disc herniation refers to the movement of disc material outside the normal boundaries of the annulus fibrosus, while disc bulge refers to a more generalized displacement of disc material in a particular direction [60]. Fig. 3.2 illustrates examples of both hydrated and dehydrated intervertebral discs.

### 3.3.2 Common Clinical Features

The following are the most common clinical characteristics related to the lumbar spine and VB/IVD:

- Sciatica is a form of pain that commonly extends from the back to the lower leg. It is frequently the result of a herniated disc exerting pressure on the nerve root, resulting in compression.
- Spinal stenosis [61] refers to a condition in which the spinal canal is narrowed, as shown in Fig. 3.3. This is typically caused by pressure from a protruding disc in the front and the thickening of the yellow ligament in the back. As a result of this pressure, people may experience a tingling pain sensation that usually occurs after walking a certain distance and is relieved with rest.

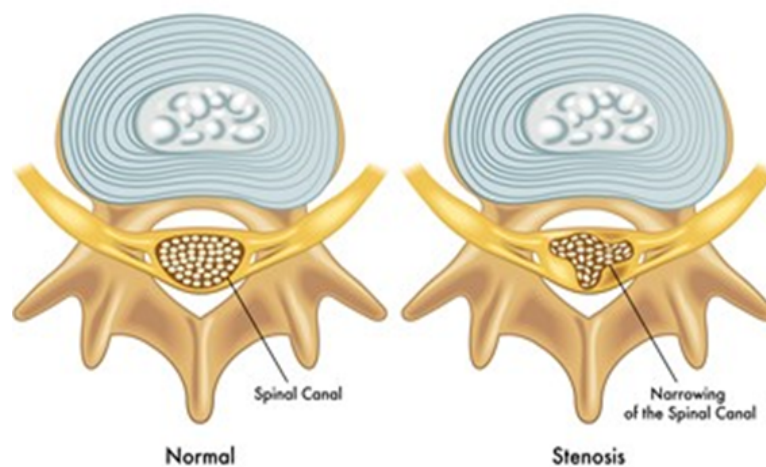


FIGURE 3.3: Spinal Stenosis

- Spondylolisthesis is a condition in which a VB becomes dislocated either towards the anterior or posterior side in relation to its counterpart with which the VB articulates. This malalignment is caused by a fracture of the

posterior element (VA) that allows the affected vertebra to slip. This is shown in Fig. 3.4.

## **3.4 Lumbar Spinal Stenosis**

Lumbar spinal stenosis is a condition characterized by the narrowing of the spinal canal in the lumbar region, which can cause compression of the nerves and lead to symptoms such as pain, weakness, and numbness in the lower back, buttocks, and legs. There are three main types of lumbar spinal stenosis:

### **3.4.1 Lumbar Central Canal Stenosis**

Central canal stenosis is a form of lumbar spinal stenosis characterized by the narrowing of the spinal canal in the lower back, resulting in compression of the spinal cord or cauda equina nerves. Several factors can cause this condition, including thickened ligaments, protruding or herniated discs, or bone spurs. Patients suffering from central stenosis may experience symptoms such as leg or buttock pain, numbness or weakness, and difficulties with bladder or bowel control [61].

### **3.4.2 Lumbar Foraminal Stenosis**

Foraminal stenosis is a form of lumbar spinal stenosis that specifically affects the foramen, which are openings in the vertebrae that allow the spinal nerves to exit the spinal cord and extend to other parts of the body. As these openings become narrower, they can exert pressure on the nerves within the spinal canal, causing a range of symptoms such as leg or foot pain, numbness, tingling, or muscle weakness. The causes of foraminal stenosis usually involve degenerative changes in the spine, such as the development of bone spurs, herniated discs, or thickening of ligaments. In addition, it can also be caused by spinal injuries, infections, or tumors [62].

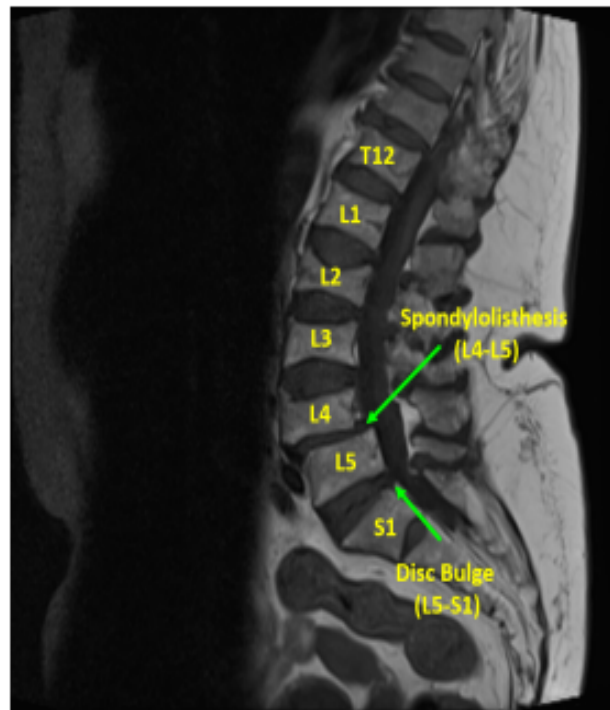


FIGURE 3.4: Spondylolisthesis and Disc Bulge in Lumbar Spine

### 3.4.3 Lumbar Lateral Recess Stenosis

Lateral recess stenosis is a condition that specifically impacts the lower back region of the spine and results from the narrowing of the openings located on the sides of the vertebrae, also called lateral recesses. As these openings get narrower, they press on the nerve roots that pass through them, leading to pain, numbness, or weakness in the legs or buttocks, and issues with balance or coordination. The causes of lateral recess stenosis can include herniated discs, bone spurs, or degenerative changes in the spine [63].

## 3.5 Clinical Evaluation / Analysis Perspectives

In the previous section, we discussed the basic anatomy of the lumbar spine and common pathological conditions. The present section focuses on the evaluation of lower back pain by radiologists and spinal surgeons to make a diagnosis. The preferences of both professionals are discussed, providing a comprehensive understanding of the responsibilities of spinal surgeons who not only conduct clinical assessments of patients but also analyze and correlate their findings with MRI

images. While this thesis elaborates on spinal disorders related to spinal geometry in detail, it provides a brief overview of other pathological conditions related to the lumbar spine.

### **3.5.1 Radiologists' Viewpoint**

According to radiology experts, the lumbar spine is more vulnerable to degenerative changes due to the impact of body weight in this area [14, 64]. They believe that the most affected region in the lumbar spine is the last three-disc levels L3-L4, L4-L5, and L5-S1 due to maximum load stress [56].

The examination of the lumbar spine by radiologists aims to pinpoint the specific site of spinal conditions, assisting spinal surgeons in determining the most suitable surgical intervention, if necessary [65]. While the diagnosis and analysis reports may vary based on individual patient scans, radiologists typically observe the presence of degenerative disc diseases, vertebral body deformations such as spondylosis, and osteoarthritis in the vertebral body or intervertebral discs. They also observe vertebral endplate deformations, which may result in the formation of bone spurs or osteophytes, a significant cause of nerve compression.

Additionally, they may identify dehydrated intervertebral discs, disc cracks, and herniation [56, 60, 66]. In addition, the diagnosis may involve identifying clinical symptoms related to sciatica, which is commonly characterized by pain that spreads from the back to the lower leg, grading the narrowing of the spinal canal known as spinal stenosis, and other related disorders. The diagnosis is made by creating a three-dimensional projection of the lumbar spine using sagittal and axial slices, and for individuals with scoliosis, a coronal scan is also included.

The comprehensive examination of the lumbar spine from multiple perspectives enables the radiologist to enhance their knowledge of its anatomy and detect any evident irregularities or distortions. If required, specific measurements are taken to further assess the condition. These measurements are crucial for establishing a correlation between the radiological findings and any surgical interventions required [57, 67].

## **3.5.2 Spinal Surgeons' Viewpoint**

The spinal surgeon conducts a thorough examination of the patient, carefully correlating the clinical observations with the information provided by the MRI scans and radiologist's analysis report. Additional physical assessments are conducted to address the specific symptoms reported by the patient. The objective of the surgeon is to identify the surgical site relevant to the disease and propose a corrective treatment that restores the patient's quality of life [68]. The choice of treatment, whether surgical or non-surgical, is determined by factors such as the severity of the spinal condition, the patient's pain tolerance, and their preference for conservative treatments. The selection of the most suitable surgical intervention procedure is based on a careful evaluation of the individual case and its unique circumstances. The surgical interventions for Lumbar Spine are typically categorized into two major types.

### **3.5.2.1 Lumbar Decompression**

The primary goal of lumbar decompression surgery is to alleviate the pressure created by the bulging intervertebral disc (IVD) and yellow ligament, leading to spinal stenosis [69]. During the surgical procedure, the surgeon will remove the specific portion of the disc, bone, or ligament that is responsible for the compression or constriction of the spinal canal. Microdiscectomy refers to the removal of the portion of the slipped IVD that is causing compression on the spinal canal, while laminectomy refers to the removal of the section of bone and yellow ligament that results in spinal stenosis, a condition characterized by the narrowing of the spinal canal.

### **3.5.2.2 Lumbar Fusion**

To alleviate pressure on spinal nerves caused by stenosis in the lumbar region, a surgeon may perform a fusion procedure by joining two adjacent vertebrae together to restrict movement between them. This is achieved by inserting a bone graft, which can be taken from the patient's own hip bone or a donor, between the

vertebrae to encourage fusion and restrict motion at the affected segment. The approach for the surgery may vary depending on the surgeon's preference and the specifics of the case. It could be done using a minimally invasive technique, where a small incision is made and specialized instruments are used to access the affected area, or through open surgery, where a larger incision is made and the surrounding tissue and muscle are moved aside to access the spine. The specific surgical technique used may also differ depending on the location and severity of the stenosis [70–72]. The following are descriptions of different types of spinal fusion procedures:

- **Posterolateral Gutter Fusion** involves taking a bone graft, usually from the pelvis, or using a substitute bone graft, and placing it in the posterolateral area of the spine, which is a vascular area that allows for proper blood flow to the grafted region [73]. The transverse process serves as a muscle attachment site, providing tension and muscle support over the grafted bone. The bone then grows and fuses the vertebrae, stopping segmental motion. Pedicle screws may be used for added support.
- **Posterior Lumbar Interbody Fusion (PLIF) Surgery** involves making an incision in the midline of the back and removing muscles to expose the vertebral column. Laminectomy is performed to visualize the nerve roots, and the facet joints are trimmed to create space for the nerve roots [72–74]. The intervertebral disc (IVD) material is removed, and a bone graft is inserted in the IVD space to fuse the superior and inferior vertebrae, restricting joint movement. Pedicle screws and rods may be used for additional support.
- **Anterior Lumbar Interbody Fusion (ALIF) Surgery** involves approaching the vertebral column from the front of the patient, and may be combined with posterolateral gutter fusion for added stability. The abdominal muscles and blood vessels are retracted, exposing the spine [73]. The IVD material is removed and a cage implant is inserted, followed by a bone graft to fuse adjacent vertebrae and restrict segmental motion. Pedicle screws and rods may also be used for additional support.

- **Transforaminal Lumbar Interbody Fusion (TLIF) Surgery** is similar to PLIF, but both anterior and posterior sides are fused through a single approach. The posterior side of the spine is fused as in PLIF, while the anterior side is fused by placing pedicle screws, rods, and bone grafts. This procedure avoids forceful retraction of spinal nerves to prevent nerve root damage [72].
- **Extreme Lateral Interbody Fusion (XLIF) Surgery** involves approaching the spine from the side and performing a fusion operation. A dilator and retractor are placed directly above the IVD to spread the muscles and tissues. A spacer (cage) is placed in the cleared space followed by a bone graft. This procedure is minimally invasive, and pedicle screws and rods may be used for additional reinforcement [72].

### **3.5.2.3 Non-Surgical Treatments for Lower Back Pain**

When surgery is not necessary, non-surgical treatments may be prescribed to alleviate pain. These treatments include medication such as pain relievers and muscle relaxants [75], along with specific exercises like the McKenzie Method. Other options include caudal epidural injection for pain management, acupuncture, spinal manipulation by chiropractors, and heat therapy through thermography [76].

## **3.6 In-depth Assessment of Intervertebral Disc Disorders**

To determine the appropriate surgical intervention procedure, spinal surgeons conduct a thorough assessment of central canal and nerve root compression. This assessment entails performing manual measurements on X-ray radiographs and MRI images to determine various parameters such as the diameter from front to back, the distance from side to side, and the cross-sectional area of the protective covering of the spinal cord known as the dural sac. Unlike radiologists, spinal surgeons heavily rely on detailed and comprehensive measurements, as discussed in

subsections 3.6.1. These measurements play a vital role in assessing the presence and severity of stenosis, which in turn helps determine the appropriate surgical approach for treatment.

### **3.6.1 Assessment of Lumbar Stenosis**

Radiologists employ various measurements derived from imaging studies like MRI or CT scans to assess the severity of lumbar spinal stenosis. These objective radiologic criteria involve analyzing the size of the central canal, which accommodates the spinal cord and nerves, and considering the presence of stenosis if the diameter falls below a specific threshold. Another measurement examines the front-to-back (Anterior to Posterior) diameter of the spinal canal, identifying stenosis if it is below a certain threshold [77]. Radiologists also calculate the cross-sectional area of the spinal canal to provide a more comprehensive assessment of stenosis. Moreover, they assess the impingement of nerve roots by measuring the area of the foramina, through which the nerve roots exit the spinal canal. Ligamentum flavum thickness is also measured as it can contribute to stenosis [78]. It's worth noting that specific cutoff values for these measurements may vary, and quantitative criteria alone do not determine the diagnosis. Clinical symptoms, patient history, and physical examination findings are also crucial in making an accurate diagnosis. Radiologists employ a combination of qualitative and quantitative assessments to comprehensively evaluate lumbar spinal stenosis.

## **3.7 Summary**

- Clinicians depend on software-assisted or manual measurements to assess spinal canal disorders.
- Assessment of stenosis is dependent on the structure of the intervertebral disc and thecal sac.
- Measurements including space between the intervertebral disc and posterior element or dimensions of the thecal sac to assess the central canal structure.



- Before choosing the appropriate spinal intervention procedures, a spinal surgeon relies on these quantitative measurements to establish a connection and assess the effectiveness of the surgical procedure, typically spinal fusion, in restoring alignment that causes compression of the central canal.

# Chapter 4

## Segmentation for Lumbar Spinal Stenosis

### 4.1 Outline

This chapter provides an overview of the dataset involve in the segmentation for lumbar spinal stenosis and a discussion about the generated ground truth images for lumbar disc MRI images covered in Section 4.2. When conducting experiments with deep learning techniques, different networks such as encoder-decoder architectures, scene parsing networks, and fully convolutional networks are utilized to achieve the same objective. Furthermore, the proposed weighted average ensemble method for segmentation is discussed in Section 4.3.5. The Python programming language is utilized to execute these networks. The details of deep learning implementation are covered in Section 4.3 of this Chapter.

### 4.2 Dataset

The Sudirman Lumbar Spine Dataset [79] is utilized in this research thesis. The dataset consists of a scientific study involving 515 patients who experienced back pain. Additional examinations may be included in each patient's data, involving images acquired from sagittal or axial views. Specifically, axial view images

primarily capture the last three intervertebral discs (IVDs) and the separation between the last vertebra and the sacrum. It is crucial to emphasize that this dataset is specifically focused on examining the lumbar region of the human spine.

### 4.2.1 Raw Data

The dataset comprises a total of 48,545 MRI slices, averaging approximately 60-95 slices per patient. These slices encompass Axial MRI slices of intervertebral discs (IVDs) D3, D4, and D5, as well as sagittal MRI slices of the lumbar spine. Both T1-weighted and T2-weighted images are included in the MRI scans, providing different tissue characteristics based on the arrangement of radiofrequency pulses. For instance, T1-weighted images represent fat as white, whereas both fat

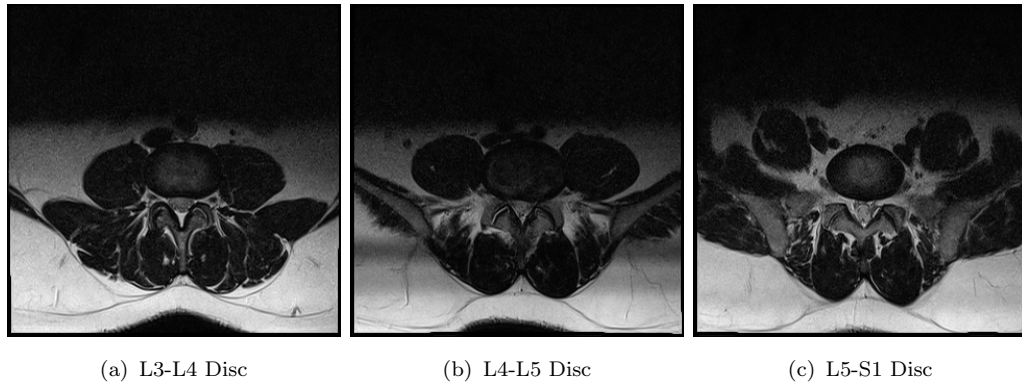


FIGURE 4.1: T2-Weighted Axial MRI Images of D3, D4 and D5 for Patient ID 001

and water appear white on T2-weighted images. To visualize the MRI slices for individual patients, a DICOM viewer can be employed.

### 4.2.2 Extracted Data

A set of 1,545 images comprising T1-weighted and T2-weighted Axial MRI slices of D3, D4, and D5 were extracted from the raw data [79]. Each patient has three images available. In Fig. 4.1, the T2-weighted Axial MRI images of D3, D4, and D5 for Patient ID 001 are displayed. These Axial view MRI images are utilized in the research for segmentation purposes, as they enable the creation of

a ground truth where specific regions can be identified. This ground truth forms the foundation for detecting lumbar spinal stenosis.

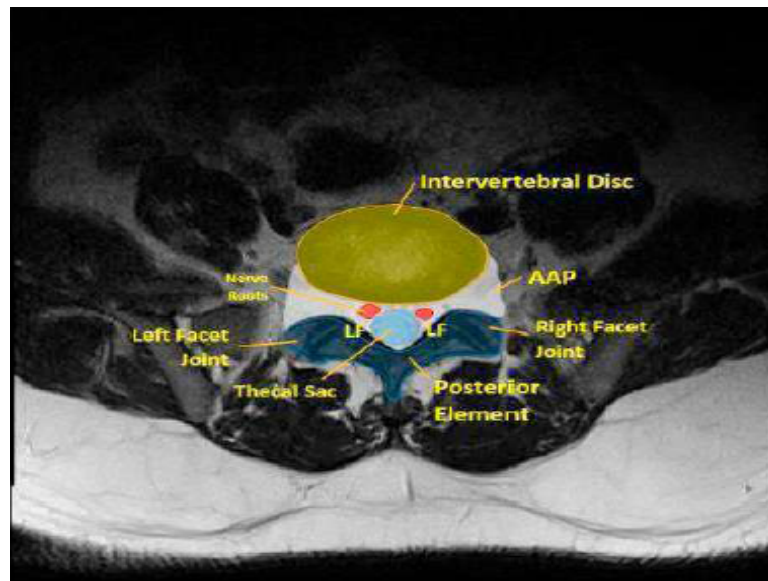


FIGURE 4.2: Regions of Interest of Lumbar Spine in Axial Planes

### 4.2.3 Ground Truth

The ground truth for the given dataset was generated using T1-weighted MRI images. It comprises labeled images that indicate distinct regions. A total of four regions have been identified.

- IVD (Intervertebral Disc)
- PE (Posterior Element)
- TS (Thecal Sac)
- AAP (Area between Anterior and Posterior Elements)

Fig. 4.2 provides a clear depiction of these mentioned regions of interest. To detect or identify the presence of spinal stenosis, the initial step involves segmenting the Axial MRI image within the designated regions depicted in Fig. 4.2. Fig. 4.3 displays the labeled image of the T2-weighted Axial MRI image of D3, D4, and D5

for patient ID 001. The primary goal of the research is to segment the provided T2-weighted MRI image into specific regions, which are classified as different classes.

CLASS 1: Inter Vertebral Disc (IVD)

CLASS 2: Posterior Element (PE)

CLASS 3: Thecal Sac (TS)

CLASS 4: Area between Anterior and Posterior Elements (AAP)

CLASS 5: Background

The objective of the study is to divide the given T2-weighted axial view MRI image



FIGURE 4.3: Ground Truth Labels Images of D3, D4 and D5 for Patient ID 001

into four separate regions, as depicted in the labeled images. The segmentation model's effectiveness is evaluated using IoU metrics. Subsequently, the next phase of the research expands on this work and concentrates on creating a system to detect and identify whether a patient is suffering from spinal stenosis or not.

### 4.3 Segmentation Using Deep Learning Architectures

In recent years, conventional image segmentation methods have demonstrated favorable outcomes in solving computer vision challenges. However, with the advent of deep learning and the application of neural networks in segmentation tasks, traditional approaches have been surpassed in terms of both quality and quantity.

Extensive experimentation was performed using traditional image processing techniques to address medical image segmentation. Before delving into the specifics of implementation, it is crucial to assess the appropriateness of chosen deep learning architectures for a particular task. The subsequent sections offer a concise summary of deep learning architectures and their desirability from a user's standpoint.

### 4.3.1 Deep Network Architectures Overview

As mentioned in the literature review section, several researchers have highlighted the efficacy of utilizing deep learning techniques for conducting segmentation tasks. In this research thesis, the focus is on segmenting the intervertebral disc using well-established deep learning methods such as encoder-decoder and scene parsing. Furthermore, a comparative analysis is conducted among various popular models, including UNet [80], ResNet [81], VGG16 [82], and InceptionNet [83]. Below is a brief description of these models:

- **UNet**

The U-Net architecture [80] was developed specifically for performing scientific image segmentation. This model is composed of two main components: the first is one encoder and the second is a decoder. The encoder is responsible for extracting features from the image and comprises a conventional convolutional layer followed by a max pooling layer. On the other hand, the decoder employs transposed convolutions, also known as the opposite convolutions, and is designed to be symmetrical to the encoder. The U-Net architecture is a fully convolutional network that solely consists of convolutional layers without any dense layers, allowing it to handle images of any size [80]. The structure of the classic U-Net network is depicted in Fig. 4.4, illustrating the deep learning architecture of U-Net.

- **ResNet**

Deeper networks have the capability to extract more valuable information from images, but this introduces certain challenges. As the network depth increases, the number of parameters becomes excessively large, leading to

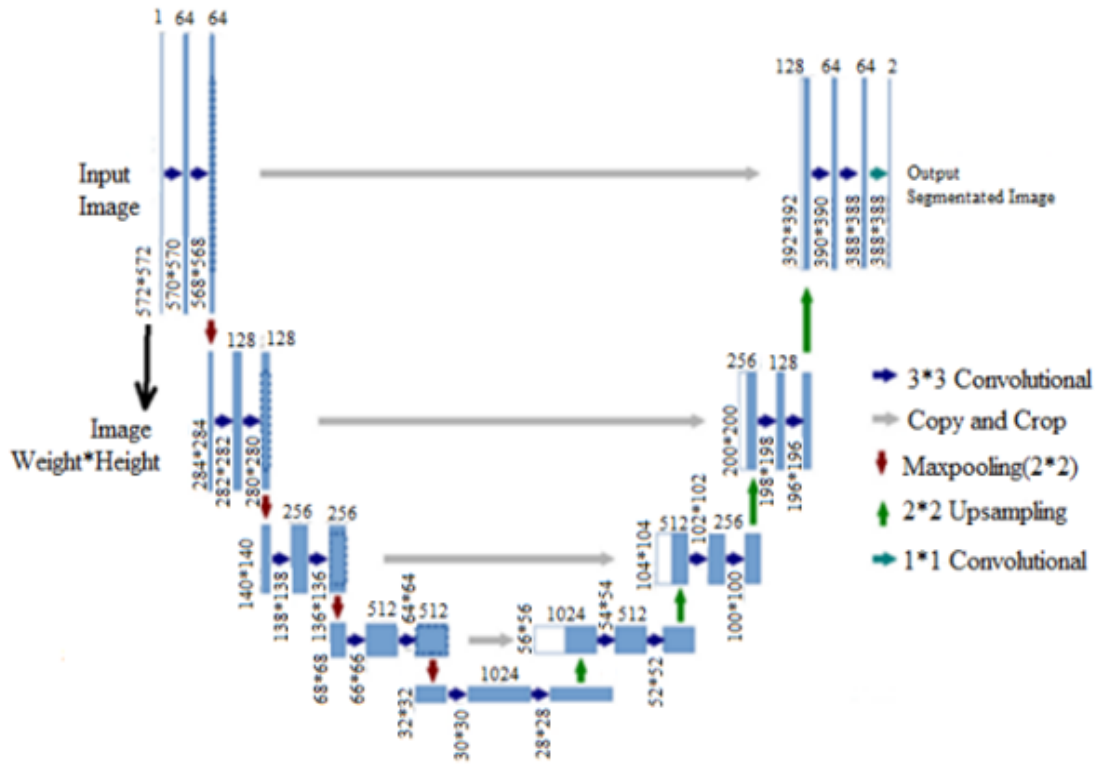


FIGURE 4.4: UNet Architecture [80]

computational complexities. Additionally, the weights tend to approach zero over the course of training, causing valuable information to become indiscernible from the image. ResNet addresses these challenges through a

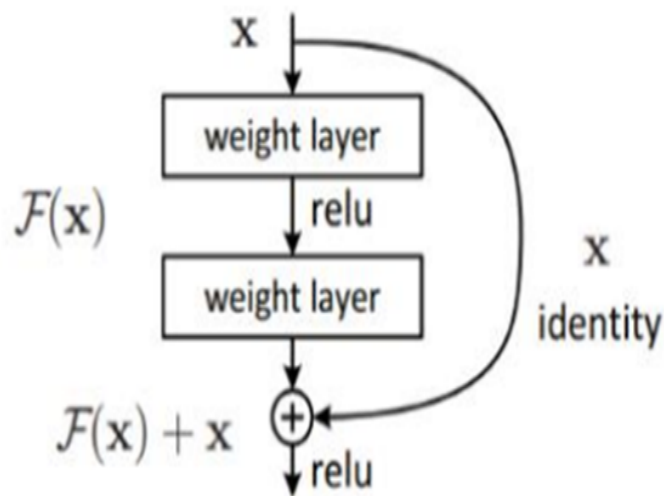


FIGURE 4.5: ResNet Architecture

straightforward approach known as skipping connections. In Fig. 4.5, the input directly connects to the output without passing through the weight layers, which are the convolution layers [81]. This approach employed by

ResNet effectively resolves the issues associated with deep networks.

$$H(x) = F(x) + x \quad (4.1)$$

In Eq. 4.1, when the value of  $F(x)$  is zero, the resulting output will also be equal to  $x$ , which is known as the identification of a residual block. Conversely, if  $F(x)$  is not zero, it is categorized as a convolution block. It is worth noting that in order for the convolution operation to be executed, the dimensions must match. As shown in Eq. 4.2, the input  $x$  is multiplied by a weight factor denoted as  $W_s$  [81].

$$H(x) = F(x) + W_s * x \quad (4.2)$$

- **VGG16**

VGG16, on the other hand, is a relatively simple network model that distinguishes itself by utilizing 2 or 3 convolutional layers. In the fully connected layer, a feature vector with  $7 \times 7 \times 512 = 4096$  neurons is obtained. The softmax performance for 1000 classes is computed based on the outputs of the two fully connected layers, utilizing approximately 138 million parameters. Similar to other models, the height and width dimensions of the matrices decrease from the image input to the end, while the number of channels increases [82]. Integrating VGG16 into the UNet model is straightforward.

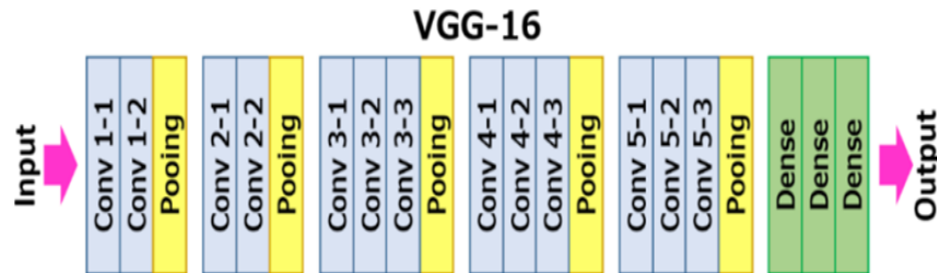


FIGURE 4.6: VGG16 Architecture

The high parameter count of VGG16 enables it to delve deeper into the image and extract detailed features. In fact, the ResNet and Inception models go even deeper. However, it's important to note that not every model is



suitable for every dataset, and the choice of model depends on the specific requirements. The structure of the VGG16 network is presented in Fig. 4.6.

### • InceptionNet

GoogLeNet, developed by Google, achieves increased depth while reducing the number of network parameters. This allows for an overlap with the network structure of LeNet, leading to the Inception network being referred to as the LeNet network [83]. The initial module of GoogLeNet typically consists of three different sizes of convolutional layers and a maximum pooling layer. After the convolution process, the output of the previous layer's channels are summed and subjected to non-linear fusion. This approach enhances the network's expressiveness, and adaptability to different scales, and helps prevent overfitting. The Inception network structure is depicted in Fig. 4.7 [83].

The models were trained and tested using Intel(R) Xeon(R) Silver 4110 CPU with 64 GB RAM and NVIDIA Quadro P4000 GPU, utilizing the Tensorflow Keras API and Python 3.9.

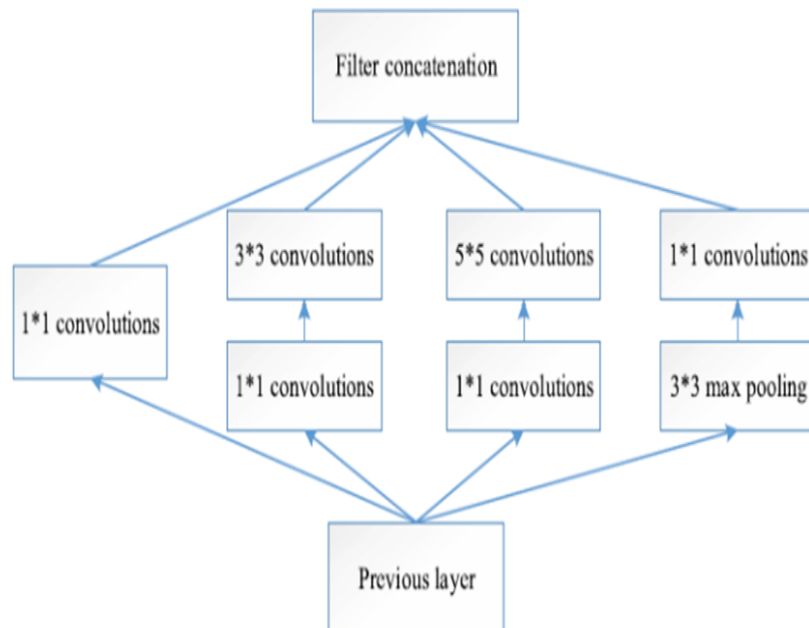


FIGURE 4.7: InceptionNet Architecture

### 4.3.2 Choice of Optimizer

The significance of the optimizer should be acknowledged as it plays a crucial role in appropriately adjusting the weights to facilitate accurate predictions for unknown samples in the future. Together with the loss function, the model parameters are updated to refine the network's performance. In this research, the Adam optimizer is employed to enhance convergence by dynamically adjusting the learning rate using a weighted exponential method. It is important to examine the weight update rule to understand why the Adam optimizer is chosen. The Adaptive Moment Estimation (Adam) is an optimization algorithm based on gradient descent, known for its effectiveness in dealing with large-scale problems that involve extensive amounts of data or parameters. It offers advantages such as reduced memory requirements and improved efficiency. Conceptually, Adam combines the benefits of both the 'gradient descent with the momentum algorithm and the 'RMSP' algorithm.

The Adam optimizer combines two gradient descent methodologies:

- **Momentum**

The objective of using this algorithm is to accelerate the gradient descent process by incorporating the concept of "exponentially weighted averages" of the gradients. This inclusion allows the algorithm to converge more quickly towards the minima, enabling faster optimization.

$$w_{t+1} = w_t - \alpha m_t \quad (4.3)$$

where

$$m_t = \beta m_{t-1} + (1 - \beta) * \left[ \frac{\partial L}{\partial w_t} \right] \quad (4.4)$$

Where  $m_t$  aggregate of gradients at time  $t$  (current) (initially,  $m_t = 0$ ),  $m_{t-1}$  aggregate of gradients at time  $t-1$  (previous),  $w_t$  weights at time  $t$ ,  $w_{t+1}$  weights at time  $t+1$ ,  $\alpha$  learning rate at time  $t$ ,  $\frac{\partial L}{\partial w_t}$  derivative of Loss Function with respect of  $w_t$  at time  $t$  and  $\beta$  is the Moving average parameter (const, 0.9)

- **Root Mean Square Propagation (RMSP)**

RMSprop, also known as root mean square propagation, is an adaptive learning algorithm that aims to enhance AdaGrad. Which accumulates the sum of squared gradients, RMSprop utilizes the concept of "exponential moving average" for this purpose.

$$w_{t+1} = w_t - \frac{\alpha_t}{(V_t + \epsilon)^{1/2}} * \left[ \frac{\partial L}{\partial W_t} \right] \quad (4.5)$$

$$V_t = \beta V_{t-1} + (1 - \beta) * \left[ \frac{\partial L}{\partial W_t} \right]^2 \quad (4.6)$$

Where  $V_t$  sum of square of past gradients. i.e., sum  $\left( \frac{\partial L}{\partial W_{t-1}} \right)$  initially,  $V_t = 0$ ,  $\epsilon$  is a small positive constant ( $10^{-8}$ ) The formula use for first moment and second moment is provided as follows.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) * \left[ \frac{\partial L}{\partial W_t} \right] \quad (4.7)$$

$$v_t = \beta_2 V_{t-1} + (1 - \beta_2) * \left[ \frac{\partial L}{\partial W_t} \right]^2 \quad (4.8)$$

As both  $\beta_1$  and  $\beta_2$  approach 1, it is observed that  $m_t$  and  $v_t$  tend to be biased towards 0, which can lead to issues. To address this, the Adam optimizer calculates bias-corrected versions of  $m_t$  and  $V_t$ . This correction helps in controlling the weights and prevents excessive oscillations when approaching the global minimum. The updated formulas for the bias-corrected weight parameters are:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (4.9)$$

Instead of using the original weight parameters  $m_t$  and  $v_t$ , the bias-corrected parameters  $\hat{m}_t$  and  $\hat{v}_t$  are used. By incorporating these bias-corrected parameters into the general equation, we ensure that the optimization process remains controlled and unbiased throughout, hence the name Adam. Hyperparameters used for the Adam optimizer are outlined below.

1.  $\epsilon$  a small positive constant to avoid division by 0 when ( $v_t = 0$ ). ( $\epsilon = 10^{-8}$ )

2.  $\beta_1$  and  $\beta_2$  decay rates of average gradients in the above two methods.  
( $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ )
3.  $\alpha$  is Step size parameter/learning rate (0.001)

### 4.3.3 Cross-Entropy Loss Function

The cross-entropy loss function is used to measure the performance of the classification model. Simply, the loss function will increase when the predicted label diverges from the actual class label. The expression cross-entropy loss function is given below.

$$l(\theta) = - \sum_{i=1}^n y_i \log \psi_i + (1 - y_i) \log(1 - \psi_i) \quad (4.10)$$

where,  $\theta$  represents the parameters of model,  $y_i$  is the actual class label,  $\psi_i$  is the predicted class label.

### 4.3.4 Application of Fine-Tuning

The widely used approach for training the model is known as Fine-Tuning, and it is also adopted in this research. Fine-Tuning involves experimenting with various encoder-decoder configurations by combining networks such as VGG-UNet [84], ResNet-UNet [85], and Inception-UNet [86]. The process of Fine-Tuning can be divided into the following these steps, as illustrated in Fig. 4.8:

- Selecting a pre-trained model from a source dataset such as ImageNet, which we will refer to as the source model.
- Creating a model that replicates the layers and parameters of the source model, except for the last layer, which is modified or removed.
- Introducing a new output layer with the desired number of classes specific to the custom dataset and initializing its parameters randomly.
- Appending the newly defined layer to the model.

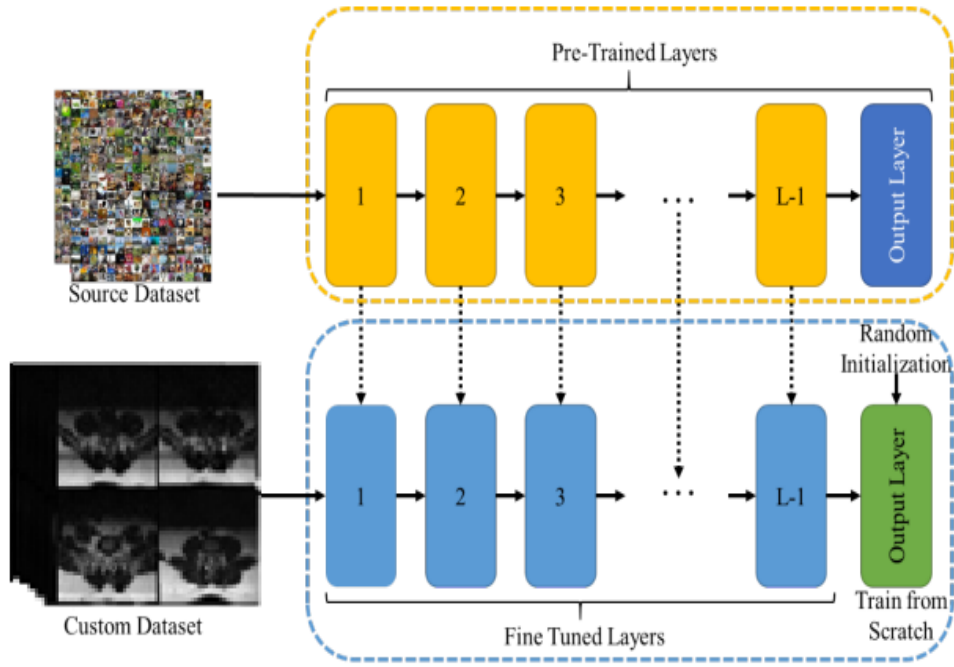


FIGURE 4.8: Transfer Learning Through Fine Tuning Process

- Training the modified model from scratch using the custom dataset. This involves updating the parameters of the last layers while fine-tuning the parameters of the remaining layers.

The total number of epochs for training; using transfer learning, is 50 with a batch size of 8 images.

### 4.3.5 Weighted Average Ensemble

An approach called weighted average ensemble of classifiers is utilized in machine learning to combine the predictions made by multiple individual classifiers into a single prediction, and it is also adopted in this research. Assign a weight to each classifier, and the final prediction is achieved by calculating the weighted average of the individual predictions. The process of a weighted average ensemble of classifiers can be divided into the following these steps, as illustrated in Fig. 4.9:

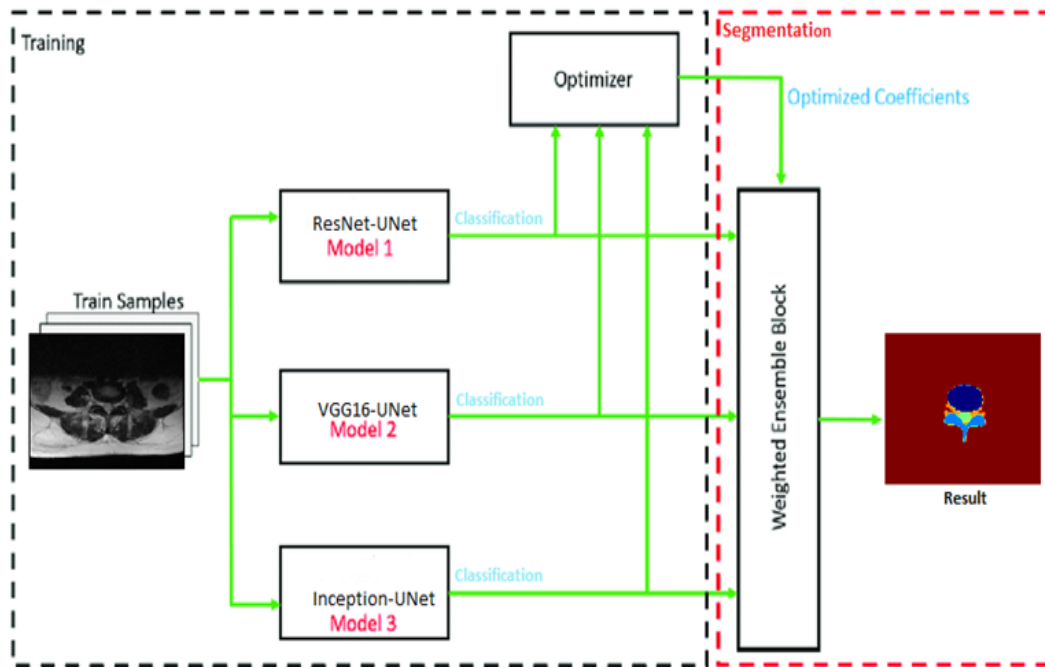


FIGURE 4.9: Weighted Average Ensemble

1. Train individual classifiers: Train several classifiers using different algorithms or variations of the same algorithm on the same dataset. Each classifier should be capable of making independent predictions.
2. Generate predictions: Once the classifiers have been trained, use each of them to make predictions on new, unseen data. Gather the individual predictions for each instance in the dataset.
3. Assign weights: Allocate a weight to each classifier in the ensemble. These weights reflect the significance or reliability of the predictions made by each classifier. The weights can be assigned based on their performance on a validation set or using other criteria like classifier accuracy or confidence level.
4. Compute weighted average: Calculate the weighted average prediction for each instance by summing up the individual predictions multiplied by their corresponding weights. This can be achieved by taking the dot product of the prediction vector and the weight vector.

5. Final prediction: Apply a decision rule to convert the weighted average predictions into a final prediction. For instance, in binary classification problems, a threshold can be used to determine whether the prediction is positive or negative.
6. Evaluate the ensemble: Assess the ensemble's performance by comparing its predictions to the ground truth labels. Use appropriate evaluation metrics such as accuracy, precision, recall, IoU, or F1 score. By employing the weighted average ensemble, different classifiers can contribute differently to the final prediction, assigning more significance to classifiers expected to perform better. The weights can be adjusted to optimize the ensemble's performance on a specific metric or to achieve the desired balance between accuracy and other factors. In this research evaluating the weighted average ensemble by using IoU metric and getting the satisfactory weights on which higher performance.

## 4.4 Summary

- Amongst the tested methods, the proposed deep learning network with ResNet as a base model and UNet as a classification network gave the best results for segmentation.
- Transfer learning is an effective learning mode well suited and less time-consuming.
- The weighted average ensemble also gives promising results by combining ResNet-UNet, VGG16-UNet, and Inception-UNet.

# Chapter 5

## Automated Measurements of Lumbar Intervertebral Bodies

### 5.1 Outline

After completing the segmentation task, automated measurements of the Intervertebral Bodies (IVDs) are performed. These measurements range from transverse distance to cross-sectional area measurements. These clinically relevant measurements help correlate the findings of clinicians with the symptoms observed in the subject patient, as already established in Section 3.6. In this chapter, Initially discuss the distance-related measurements.

### 5.2 Proposed Methodology

In this research thesis, completely automated measurements (without human intervention) related to the lumbar intervertebral disc and thecal sac regions are performed using Python 3.9. The steps for the process are presented in the flowchart shown in Fig. 5.1.

In this research thesis, the input image consists of a T2-weighted MRI of the L4-L5 intervertebral disc with a resolution of 320x320 pixels. Segmentation plays a crucial role in measurements since in measurement regions, pixels are essential.



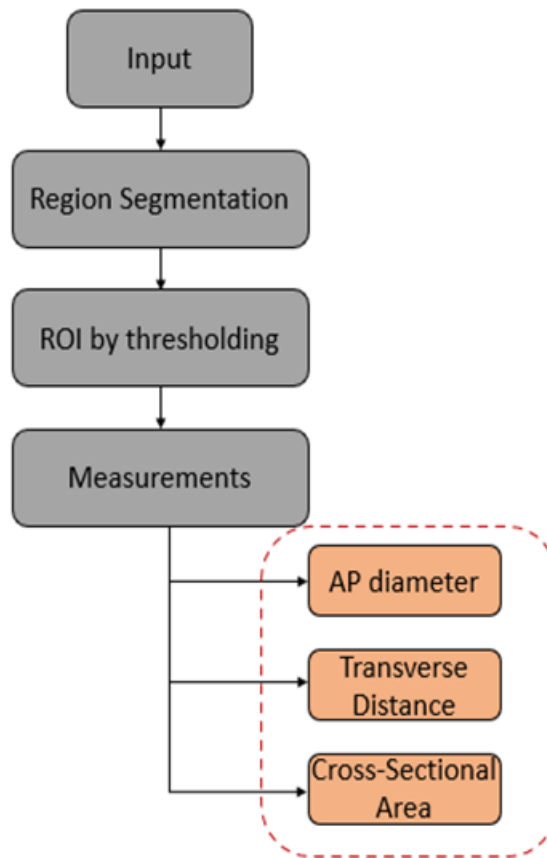


FIGURE 5.1: Measurements Acquisition (Block Diagram).

If pixels are missing, the measurements are not accurate. Therefore, automated segmentation is achieved by utilizing a weighted average ensemble approach to extract spinal measurements and establish correlations with the reports provided by radiologists

### 5.2.1 ROI by Thresholding

Initially, the process involves segmenting the region to isolate each individual component. Identification of IVDs and thecal sac from grey scale image is performed by making use of pixel intensity value because each region has its own intensity. After thresholding, a binary image is obtained. The subsample that emerged is depicted in Fig. 5.2.

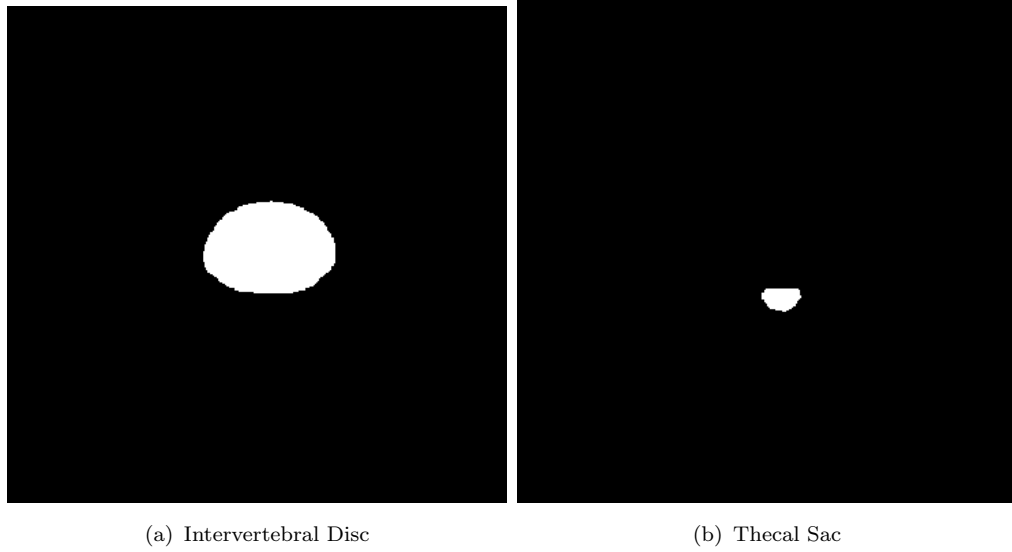


FIGURE 5.2: Results of Thresholding

### 5.2.2 Computation of Diameter and Transverse Distance

After identifying the intervertebral disc and thecal sac, the next step is to find the region properties, which include the centroid, height, and width of the whole region. The height of the thecal sac corresponds to the AP diameter, while the width represents the transverse distance. In Fig. 5.3 (a), the AP diameter is denoted by the blue line in millimeters, and the transverse distance is represented by the green line in millimeters. After calculating the diameter of the thecal sac and intervertebral disc, the next step involves further computation of the diameter ratio (DDRAP).

$$DDRAP = \frac{\text{Diameter of duralsac}}{\text{Diameter of anterior posterior Disc}} \quad (5.1)$$

### 5.2.3 Cross-sectional Area

The cross-sectional area of the thecal sac is important for determining lumbar spinal stenosis and measuring the area of the intervertebral disc. The thecal sac area utilizes the height and width of the region, but the cross-sectional area also depends on the shape of the object or region. Therefore, in this research thesis, the total number of pixels in the region is counted, and through this approach,

the area is calculated in millimeters square for both regions.

$$DDRCA = \sqrt{\frac{\text{Area of duralsac}}{\text{Area of anterior posterior Disc}}} \quad (5.2)$$

After getting the area and diameter of the thecal sac and intervertebral disc, calculating the ratio of area (DDRCA) and diameter (DDRAP) through this observing a significant change of diameter and area of the thecal sac with respect to the intervertebral disc as shown in Fig. 5.3 (b).

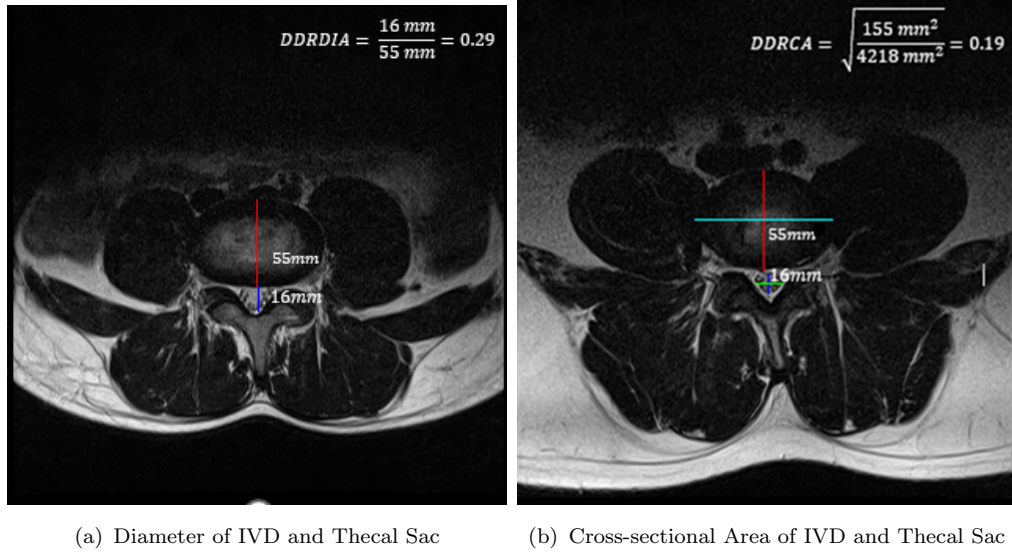


FIGURE 5.3: Automated Measurements

### 5.3 Summary

- In the proposed lumbar spinal stenosis measurements acquisition methodology, height and width are initially automatically computed in the binary images containing the IVD and thecal sac, followed by distance-based measurements.
- These measurements are performed through an automated way to give support to the clinicians to acquire these measurements certainly conserving time as well as giving quantitative bias to their manually performed measurements.

- All these measurements are performed to support the subsequent lumbar central canal stenosis classification methodologies as proposed in this thesis.

# Chapter 6

## Classification of Lumbar Central Canal Stenosis

### 6.1 Outline

After extracting the spinal cord profiles, including diameter measurements section and cross-section area estimation, classification methodologies for lumbar central canal stenosis i.e., the deep learning base classification and machine learning base classification. A binary classifier that identifies the normal and abnormal stenosis using a novel proposed method. In Section 6.2 classification methodology for lumbar central canal stenosis is presented.

### 6.2 Proposed Automated Classification of Lumbar Central Canal Stenosis

Numerous studies in the literature review portion have highlighted the effectiveness of utilizing deep-learning and machine techniques for accomplishing classification tasks. In this research thesis, the focus is on the classification of lumbar central

canal stenosis by using well-established deep learning and machine learning methods. Furthermore, a comparative analysis is conducted among various popular models, including ResNet and SVM (support vector machine)

The training and testing of all models were conducted on an Intel(R) Xeon(R) Silver 4110 CPU with 64 GB RAM and NVIDIA Quadro P4000 GPU, using the Tensorflow Keras API and Python 3.9.

### 6.2.1 SVM-Based Classification

Support Vector Machine (SVM) is a widely recognized and efficient algorithm used for classification purposes. The objective of SVM-based classification is to identify an optimal hyperplane that can successfully distinguish data points from different classes in a high-dimensional space. The aim is to maximize the distance or margin between the hyperplane and the closest data points belonging to each class that's why in this research thesis, SVM based classifier is used to predict lumbar spinal stenosis. The following is a general outline of SVM-based classification, and Fig. 6.1 depicts a block diagram illustrating the classifier.

#### 6.2.1.1 Feature Selection

The initial step involves gathering and preprocessing the data. It is crucial to ensure that the data is in a suitable format, preferably numeric, as SVMs perform best with such data. Additionally, normalizing or standardizing the data is important to bring the features to a comparable scale. Class labels are retrieved from the same data set that discusses in Section 4.2. The Excel sheet with that dataset is open source in which patient id and radiologist remarks from these remarks get the class labels (0 or 1). Furthermore, relevant features that provide the most information for classification need to be selected. Appropriate feature selection can enhance the performance and efficiency of the SVM model. Model train by using the 7 features and performance is better on these input features.

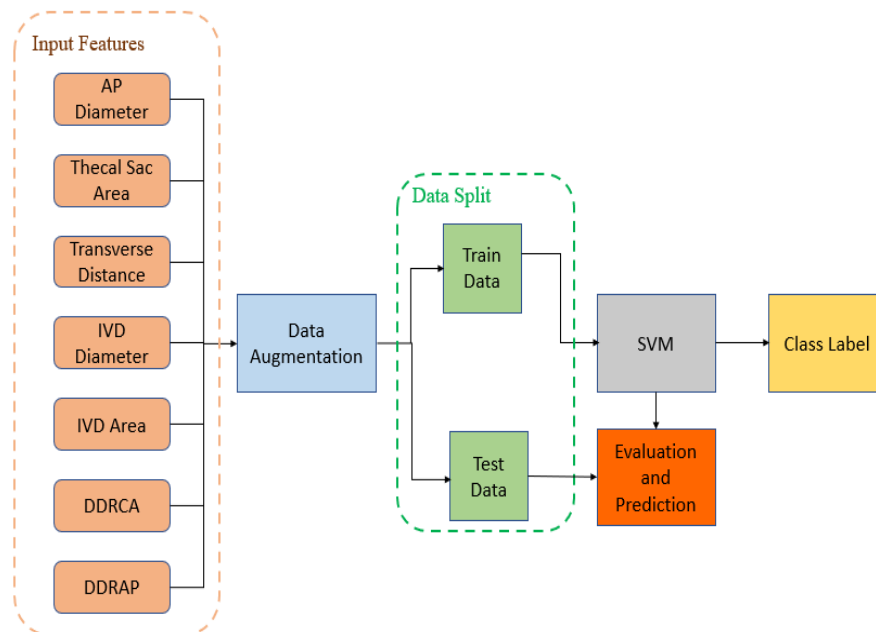


FIGURE 6.1: SVM-Based Classification (Block Diagram).

### 6.2.1.2 Data Augmentation

Data augmentation is a widely used technique to solve the data imbalance problem. If the dataset is imbalanced then model predictions will be generalized and biased to a class in which the number of the samples is comparatively greater. The data set used in this research thesis, samples of a normal class is less than abnormal class because MRI performed for those patients that have the problem of low backbone pain. The upsampling technique is used to address the class imbalance in a dataset by artificially increasing the number of samples in the minority class. This technique is particularly useful when the dataset contains significantly more samples from one class compared to another, leading to biased training and potentially poor performance on the minority class.

### 6.2.1.3 Selection of Kernel Function

SVMs employ kernel functions to transform the input data into a higher-dimensional space, allowing for linear separation of the data points. The choice of the kernel function depends on the characteristics of the data and the specific problem being addressed. Common kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid. In this scenario, the RBF kernel is selected for training

the model. The RBF kernel is defined by the equation given in Eq. 6.1.

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (6.1)$$

Where  $x$  and  $x'$  represent input feature vectors,  $\|\cdot\|$  denotes the Euclidean distance, and  $\gamma$  is the gamma parameter.

The gamma parameter is a crucial factor in shaping the decision boundary and determining the influence of each training example. A smaller gamma value widens the decision boundary, resulting in a smoother model, whereas a larger gamma value narrows the decision boundary, enabling more complex models that could potentially overfit. Support Vector Machines (SVMs) utilizing the RBF kernel are capable of learning intricate decision boundaries and effectively dealing with non-linearly separable data.

#### 6.2.1.4 Training the SVM

In this stage, the SVM model learns the optimal hyperplane by identifying support vectors. Support vectors refer to the data points that are in proximity to the decision boundary or hyperplane. The training process involves solving an optimization problem to determine the hyperplane that maximizes the margin while minimizing classification errors.

#### 6.2.1.5 Hyperparameter Tuning

SVMs have various hyperparameters that require tuning to achieve the best possible performance. The hyperparameters involved in the SVM model include the regularization parameter (C), which controls the trade-off between maximizing the margin and minimizing classification errors, along with kernel specific parameters. Techniques like cross-validation, such as grid search, can aid in finding the best combination of hyperparameters. It is crucial to select suitable values for the C parameter and gamma. The selected C value is 100 and the gamma is 0.5 to optimize the hyperplane



### 6.2.1.6 Evaluation and Prediction

After training the SVM model, its effectiveness can be evaluated using a separate validation or test dataset. Accuracy, precision, recall, and F1-score are widely employed evaluation metrics for assessing the performance of classification tasks. Once the model's performance has been assessed, it can be utilized to predict the class labels of new, unseen data points. SVM-based classification finds applications in diverse domains, including image classification, text categorization, and bioinformatics. Its ability to handle high-dimensional data and robustness against noise make it a popular choice for diverse classification problems.

## 6.2.2 ResNet-Based Classification

In this research thesis, a binary classifier is proposed by using the ResNet50 architecture. The ResNet architecture was previously discussed in Section 4.3.1. This model gives promising results due to its residual feature. The block diagram as shown in Fig. 6.2.

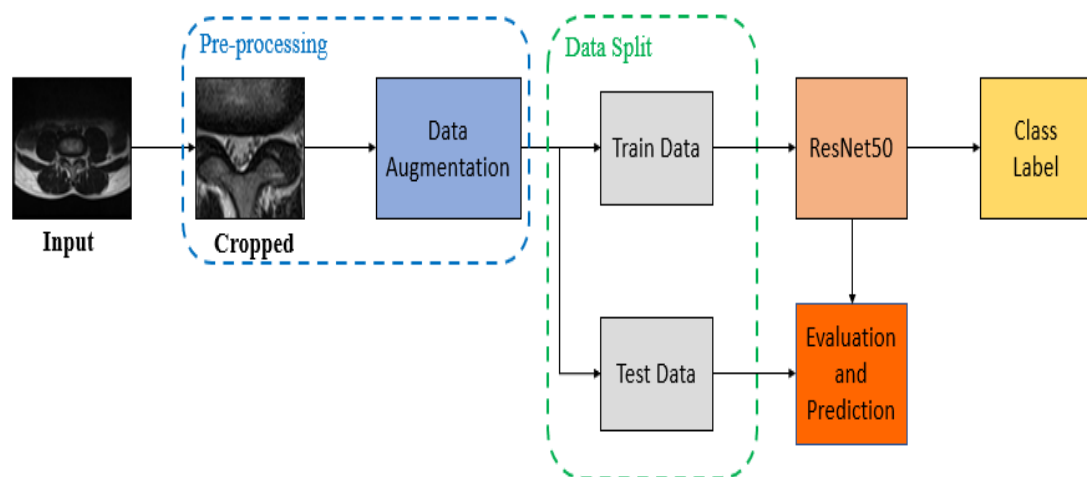


FIGURE 6.2: ResNet-Based Classification (Block Diagram).

### 6.2.2.1 Image Copping

The input of this model is a T2-weighted MRI of the L4-L5 IVD dataset previously described in section 4.2. Then apply pre-processing in which first crop the original

image by fixed dimensions because the region of interest (Thecal sac) does not have too much variation this approach decreases the input size (70\*70) Result subsample comprising showcased in Fig. 6.2. and cropping operation does not cut the area of the thecal sac.

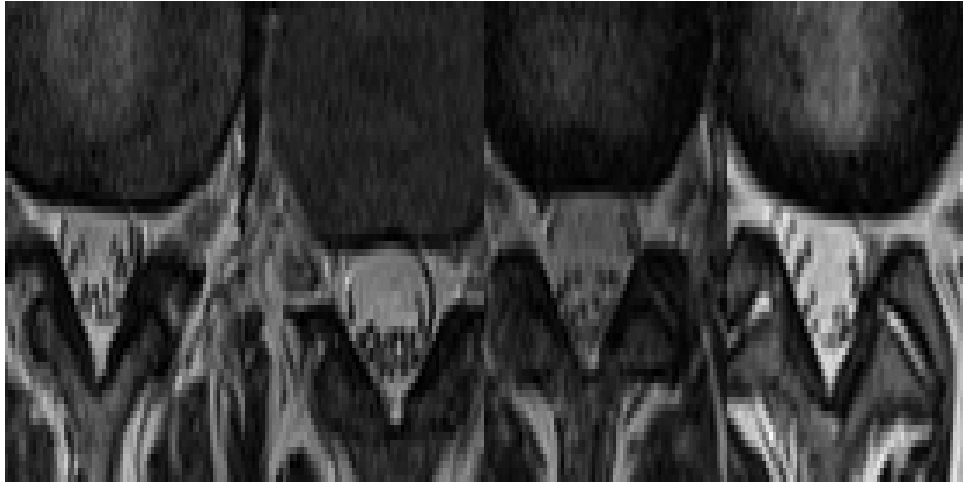


FIGURE 6.3: Cropped Images

#### 6.2.2.2 Data Augmentation

Data augmentation is a method employed in deep-learning to expand the training dataset by applying diverse transformations or modifications to existing data samples. This process generates new versions of the data, aiming to enhance the generalization and performance of deep learning models by introducing realistic variations that resemble real-world situations. Popular techniques include altering images, flipping or mirroring them, rotating or shearing them, scaling or zooming them, introducing noise, adjusting colors, and applying elastic deformations. During training, data augmentation is implemented in real-time, generating augmented versions of the original data to enlarge the dataset and upgrade the model's capacity to generalize. In this study, the dataset is imbalanced therefore data augmentation for class normal is required following transformations or modifications applied to specific class (normal) samples. Results of data augmentation as shown in Fig. 6.4.

- Rotate the image between -10 to 10 degrees.

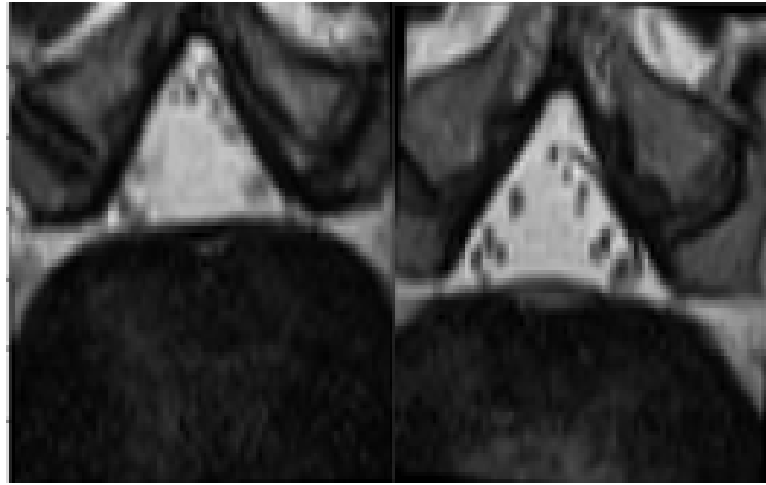


FIGURE 6.4: Results of Data Augmentation

- Flip images vertically with a 50% probability.
- Apply Gaussian blur with sigma between 0 to 1

### 6.2.2.3 Data Split

In the area of deep learning, data splitting refers to the act of dividing a dataset into distinct subsets: training, validation, and testing. The training set is employed to train the model, the validation set aids in tuning hyperparameters, and the test set is utilized to assess the model's performance. Ensuring that the splits accurately represent the data and preserve class distribution is crucial. This procedure facilitates model training, hyperparameter optimization, and evaluation of the model's ability to generalize to unseen data. Typically, around 60-80% of the data is allotted for training, and around 10-20% of the data is allotted for testing.

### 6.2.2.4 Optimizer

The importance of recognizing the optimizer cannot be overstated, as it plays a vital role in adjusting the weights to ensure accurate predictions for future unknown samples. Alongside the loss function, the model parameters are modified to improve the performance of the network. In this research thesis, the Adam optimizer is utilized to enhance convergence by dynamically adjusting the learning rate through a weighted exponential approach. It is crucial to understand the

rationale behind choosing the Adam optimizer by examining the weight update rule that was previously discussed in Section 4.3.2. Adaptive Moment Estimation (Adam) is an optimization algorithm used for gradient descent, known for its effectiveness in handling large-scale problems with extensive data or parameters. It provides several advantages, including improved efficiency and reduced memory requirements. Essentially, Adam combines the benefits of both the momentum algorithm and the 'RMSP' algorithm, resulting in enhanced performance.

#### 6.2.2.5 Binary Cross Entropy Function

A commonly employed loss function is binary cross entropy for binary classification tasks. It assesses the disparity between predicted probabilities and actual binary labels. The loss is computed by comparing the logarithmic difference between the predicted probability and the true label for both positive and negative categories. The objective is to minimize this loss during training in order to enhance the precision of the binary classifier.

$$L(y, \hat{y}) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (6.2)$$

Where,  $y$  is the true binary label (either 0 or 1),  $\hat{y}$  is the predicted probability of the positive class (ranging from 0 to 1).

#### 6.2.2.6 Training Through Transfer Learning

Transfer learning is a widely used approach for training the model, and it is also adopted in this study. In Section 4.3.4 a detailed description of transfer learning is presented. ResNet 50 is used as a backbone model for training freeze the all intermediate layers and the output layer is trained from scratch. The selected input size is 70x70 due to the limited dataset, and the training hyperparameters include a batch size of 6 and 70 epochs. The ResNet-based model outperforms the machine learning-based model in terms of performance.

Furthermore, the deep learning-based model exhibits high performance on this dataset. However, it will be tested on another dataset in the future, and the variance of that dataset may be different, compromising its performance. On the other hand, the machine learning-based model delivers better practical performance, and it is also more accessible to professionals.

### **6.3 Summary**

- Traditionally as practiced by clinicians, the classification of Lumbar central canal stenosis is performed visually without making any dedicated measurements which is subjective and prone to variability.
- Deep learning base classifier performance is better than machine learning-based but clinicians prefer machine learning methods because qualitative data is used.
- Machine learning-based classifiers utilize all seven qualitative measurements, with particular emphasis on the accurate and robust ratio-based metrics such as DDRDIA and DDRCA.

# Chapter 7

## Results and Discussion

### 7.1 Outline

This chapter begins by introducing the metrics utilized to validate the performance of automated disc and thecal sac extraction, including segmentation, automated measurements, and lumbar central canal stenosis classification. These metrics are widely employed in the research community. Subsequently, a thorough examination of experimental results is provided, accompanied by qualitative analysis of ground truth annotations and intervertebral bodies (IVBs) segmentation. Additionally, this chapter explores the quantitative analysis of the implemented segmentation techniques, automated spinal measurements, and the methodologies utilized for classifying lumbar central canal stenosis.

### 7.2 Performance Metrics

To evaluate the performance of semantic segmentation, the effectiveness is measured using mean pixel accuracy (MPA) and mean pixel precision (MPP) metrics, as denoted by Eq. 7.1 and Eq. 7.2, respectively. The accuracy metric measures the percentage of correctly classified pixels in the overall context of the intervertebral bodies (IVBs) and background, while precision evaluates the accuracy of

classification on a global scale.

$$MPA = \frac{TP + TN}{TP + TN + FP + FN} \quad (7.1)$$

In this context,  $TP$  represents the true positives at the pixel level, indicating accurate extraction of intervertebral discs (IVDs).  $TN$  refers to the true negatives, which indicates the accurate identification of pixels as background.  $FP$  represents the false positives, indicating the incorrect labeling of background pixels as IVBs (intervertebral bodies), while  $FN$  represents the false negatives, signifying the incorrect labeling of IVB pixels.

$$MPP = \frac{TP}{TP + FP} \quad (7.2)$$

The mean intersection-over-union (IoU) score, which is also referred to as the Jaccard similarity coefficient, is employed to assess the level of agreement between the segmented mask and the ground truth mask. The IoU score is calculated using Eq. 7.3 to determine the level of similarity.

$$IoU = \frac{TP}{TP + FP + FN} \quad (7.3)$$

Moreover, the dice similarity coefficient (DSC) score is calculated, which measures the similarity between the segmented mask and the ground truth mask by taking into account the doubling of true positives ( $TP$ ). The DSC score is determined using Eq. 7.4, as given below.

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (7.4)$$

Precision, recall, and F1-score are widely used evaluation metrics in binary classification tasks to assess the performance of a classification model. These metrics are based on the fundamental principles of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Precision evaluates the proportion of accurately predicted positive instances among all instances predicted as positive. It specifically emphasizes the accuracy of positive predictions. The corresponding expression is provided in Eq. 7.5. High precision indicates a low false positive

rate, meaning the model is good at correctly identifying positive instances and minimizing false positives. Recall (also known as sensitivity or true positive rate)

$$Precision = \frac{TP}{(TP + FP)} \quad (7.5)$$

Recall assesses the ratio of accurately predicted positive instances to all actual positive instances. It emphasizes the coverage of positive instances, which is measured through Eq. 7.6.

$$Recall = \frac{TP}{(TP + FN)} \quad (7.6)$$

A high recall value suggests a low rate of false negatives, indicating that the model effectively captures most positive instances and minimizes the number of missed positive cases. The F1-score is a metric that combines precision and recall by calculating their harmonic mean. It offers a balanced assessment of both precision and recall, providing a unified measure of the model's performance in capturing true positives and minimizing false negatives. The expression is given as Eq. 7.7. The F1-score combines precision and recall, assigning equal importance to both metrics. It is particularly useful in scenarios where there is an imbalance between the positive and negative classes.

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (7.7)$$

The Area Under the Receiver Operating Characteristic Curve (AUC ROC) is a commonly used metric to assess the effectiveness of binary classification models. By utilizing the ROC curve, which illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds, the Receiver Operating Characteristic Area Under the Curve (ROC AUC) metric evaluates the performance of a binary classification model. The TPR signifies the proportion of accurately predicted positive instances, while the FPR represents the proportion of erroneously predicted positive instances. The Area Under the Curve (AUC) offers a comprehensive evaluation of the classifier's performance, with larger values indicating better performance.

The ROC AUC metric is valuable in assessing the discriminatory ability of a binary



classifier, particularly in situations involving imbalanced data or when the costs associated with false positives and false negatives differ. It allows for comparisons between different models to determine which one performs better in distinguishing between positive and negative classes.

## 7.3 Results

For better presentation, the respective results are categorized in the same order according to each experiment. Details are as under:

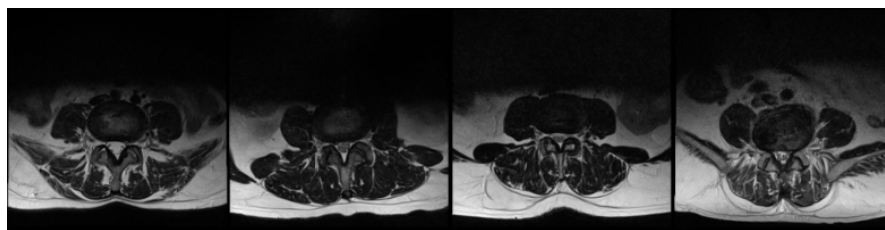
### 7.3.1 Segmentation of Intervertebral Bodies

Fig. 7.1 presents the qualitative results, accompanied by detailed quantitative assessments in Table 7.2. Additionally, Table 7.1 provides a class-wise IoU metric

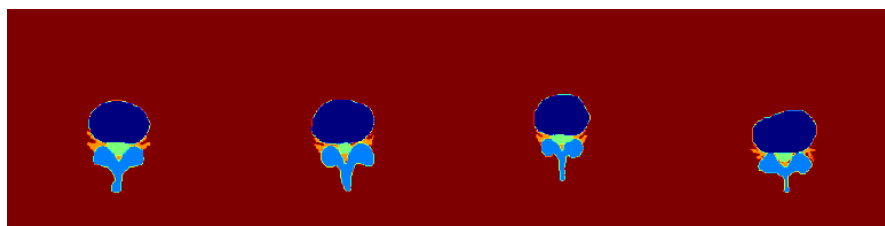
TABLE 7.1: Class-Wise Intersection Over Union (IoU) of Models

Class	ResNet-UNet	Inception-UNet	VGG-UNet	WAE
	IoU	IoU	IoU	IoU
Intervertebral Disc	0.99	0.98	0.98	<b>0.99</b>
Posterior Element	0.98	0.94	0.93	<b>0.98</b>
Thecal Sac	0.98	0.92	0.93	<b>0.98</b>
AAP	0.94	0.83	0.83	<b>0.95</b>
Background	0.99	0.99	0.99	<b>0.99</b>

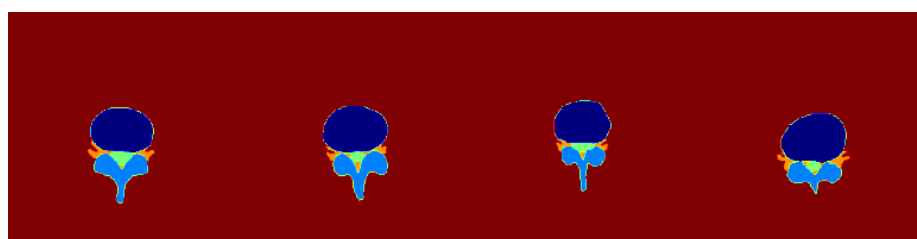
assessment. Each row in the figure corresponds to a specific image type, with sub-captions provided below. The comparison is depicted between the results obtained by a weighted average ensemble and segmentation outcomes from various deep learning networks. Each column presents the outcomes achieved using distinct models or approaches for a specific sample. Notably, ResNet-UNet demonstrates superior performance compared to other deep learning networks, as observed from the presented results. Upon reviewing the quantitative data provided in Table 7.1, it is evident that the Area between Anterior Posterior Element (AAP) yields the lowest IoU metrics. Comparing ResNet-UNet and WAE, both models exhibit similar performance across



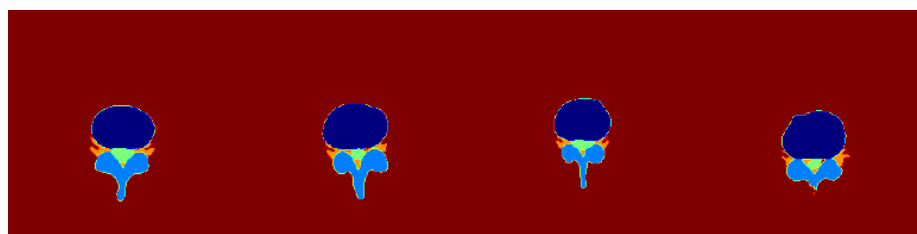
(a) Input Images.



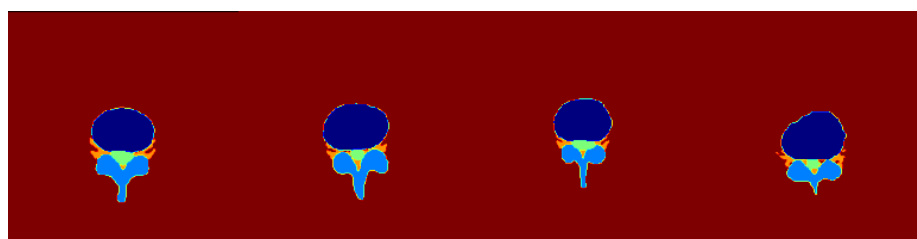
(b) Ground Truth Images.



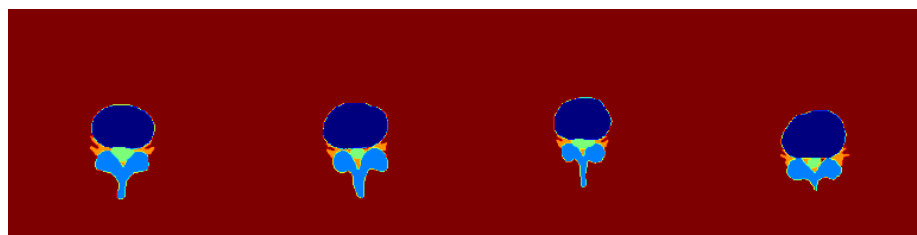
(c) ResNet-UNet Architecture.



(d) Inception-UNet Architecture.



(e) VGG-UNet Architecture.



(f) Weighted Average Ensemble

FIGURE 7.1: Qualitative Comparison of IVB Segmentation Results

TABLE 7.2: Comparative Quantitative Analysis of Methods/Models used for Semantic Segmentation. Mean Pixel Accuracy (MPA), Mean Pixel Precision (MPP), Intersection-over-Union (IoU), Dice Similarity Coefficient (DSC)

	<b>MPA</b>	<b>MPP</b>	<b>mIoU</b>	<b>DSC</b>
ResNet-UNet	99.9	98.4	0.96	0.98
Inception-UNet	99	94.1	0.89	0.94
VGG-UNet $\downarrow$	98.1	93.5	0.88	0.93
WAE $\uparrow$	<b>99.9</b>	<b>98.4</b>	<b>0.976</b>	<b>0.98</b>

most classes, except for AAP where WAE outperforms ResNet-UNet. Similarly, Table 7.2 shows that VGG-UNet $\downarrow$  achieves the lowest quantitative metrics, but with the advantage of the shortest training time. The results of ResNet-UNet and WAE $\uparrow$  are comparable to VGG-UNet $\downarrow$  in terms of IoU and DSC scores. In Table 7.2, MPA represents mean pixel accuracy, MPP represents mean pixel precision, IoU stands for intersection over union, and DSC indicates the dice similarity coefficient score.

TABLE 7.3: Comparison of Segmentation Results with Related Researchers' Work. Image Modality (IM), Mean Pixel Accuracy (MPA), Mean Pixel Precision (MPP), Intersection-over-Union (IoU), Dice Similarity Coefficient (DSC)

	Method	IM	<b>MPA</b>	<b>MPP</b>	<b>IoU</b>	<b>DSC</b>
Proffered Method	WAE	MR	<b>99.9</b>	98.4	<b>0.976</b>	<b>0.989</b>
Lu et al.[31]	UNet	MR	94	-	0.91	0.93
Lessmann et al.[33]	FCN	MR/CT	-	-	-	0.94/0.96
Huang et al.[51]	UNet	MR	-	-	0.94	-
Tang et al.[34]	DDUNet	CT	90.1	-	0.83	-
Benjdira et al.[35]	FC-DenseNet103	USS	<b>98.9</b>	<b>98.9</b>	0.94	0.98
Siriwardhana et al.[37]	CNN-UNet	MR	-	-	0.88	-
Narasimharao Kowlagi et al.[38]	FPN + ResNet34	MR	-	-	0.97	-
Merve Apaydin et al.[39]	CNN-UNet	MR	99	-	0.92	-
Abhinav shukla et al.[40]	2D UNet	MR	-	-	0.71	-
Upasana Upadhyay Bharadwaj et al.[50]	V-Net CNN	MR	-	-	-	0.93/0.94

The most favorable results are highlighted in bold, while the worst results in the deep learning category are denoted by  $\downarrow$ , specifically observed with VGG-UNet. It is important to note that MPA and MPP values range from 0 to 100, with higher values indicating better performance. Similarly, IoU and DSC values range between 0 and 1, where values closer to 1 indicate better model/method performance. In Table 7.3, a comparison of results obtained with the weighted average ensemble (WAE) architecture used in this thesis with the results of other researchers for segmentation task is presented. The image modality (IM) and the method or technique adopted by respective researcher is also mentioned. Bilel Benjdira et al.[35]

segment IVD achieving a superior MPP of 98.9 making use of FC-DenseNet103 as represented in Table 7.3. Most researchers achieved good results with FCN and UNet architectures [31, 33, 51]. Narasimharao Kowlagi et al. [38] obtained excellent results in terms of IoU score for vertebra segmentation using the FPN + ResNet34 architecture on a dataset consisting of 250 annotated MRI images, with a training and testing split of 80/20. Upasana Upadhyay Bharadwaj et al.[50] use V-Net CNN architecture for segmentation of dural sac and intervertebral disc and achieve Dice scores of 0.93 and 0.94. In this study 200 MRI of patients randomly selected from 30619 patients in which 75/10/15 split was used for training, validation and testing purpose. Tang et al.[34] used dual densely connected UNet (DDUNet) to achieve IoU score of 0.83.

### 7.3.2 Classification and Measurements of Lumbar Central Canal Stenosis

Fig. 7.2 displays the qualitative results, and Table 7.4 provides a detailed quantitative assessment. According to the radiologist's report, out of a total of 515 patients, 357 were diagnosed with thecal sac compression. The proposed SVM-based automated disease classification system achieved an accuracy of 96%, while the ResNet-based classification system achieved an accuracy of 98%. Sub-images in Fig. 7.2 showcase selected results where the classification model correctly identifies and labels central canal disorders, along with displaying measurements such as 'DDRDIA, DDRCA, and TSDIA'. Additionally, evaluating the quantitative metrics of the SVM-based classifier, it exhibits the lowest training time. The results of the ResNet-based classifier are comparable in terms of achieved AUROC (Area under the ROC Curve) and Recall scores. The best results are highlighted in bold. It is worth noting that accuracy and AUROC values range from 0 to 100, with higher values indicating better performance. Similarly, Recall, Precision, and F1-score values range between 0 and 1, with values closer to 1 representing better model performance. While deep learning approaches showcase excellent accuracy, machine learning methods also perform well, particularly in terms of Precision. Healthcare professionals prefer machine learning methods due to their incorporation of

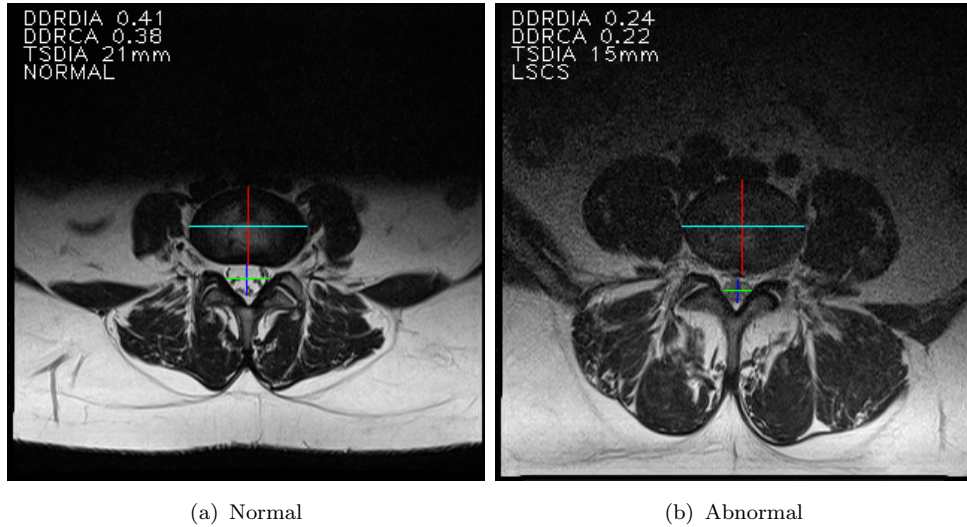


FIGURE 7.2: Results of LSS Classification and Measurements

TABLE 7.4: Comparative Quantitative Analysis of Models Used for Classification.

	Accuracy	Precision	Recall	F1-score	AUROC
SVM	96	1	0.915	0.95	95.7
ResNet	<b>98</b>	0.975	<b>0.987</b>	<b>0.98</b>	<b>97.5</b>

measurements. Consequently, this study employs a machine learning-based and deep learning-based classifier to effectively classify cases of lumbar central canal stenosis. The SVM model was tested on 149 previously unseen samples, including 71 actual positive cases and 78 actual negative cases. Similarly, the ResNet model was evaluated on 131 unseen samples, consisting of 79 actual positive cases and 52 actual negative cases, as depicted in Fig 7.3. The SVM model demonstrates accurate performance on negative samples, while the ResNet model excels in accurately classifying positive cases. In Table 7.5, a comparison is presented between the re-

TABLE 7.5: Comparison of Classification Results with Related Researchers' Work.

	Method	Accuracy	Precision	Recall	F1-score	AUROC
Proposed Method	ResNet50	<b>98</b>	0.975	<b>0.987</b>	<b>0.98</b>	<b>0.975</b>
	SVM	96	1	0.915	0.95	0.957
Jen-Tang Lu et al.[31]	CNN	-	-	-	-	<b>0.97</b>
Alessandro Siccoli et al.[55]	XGBoost	85	-	0.86	0.87	0.92
Tackeun Kim et al.[53]	VGG	82.2	-	-	-	0.90
Dongkyu Won [54]	VGG	77.5	-	-	75	-
Upasana Upadhyay Bharadwaj et al.[50]	DT	-	-	-	-	0.95

sults obtained using the SVM and ResNet architectures employed in this thesis

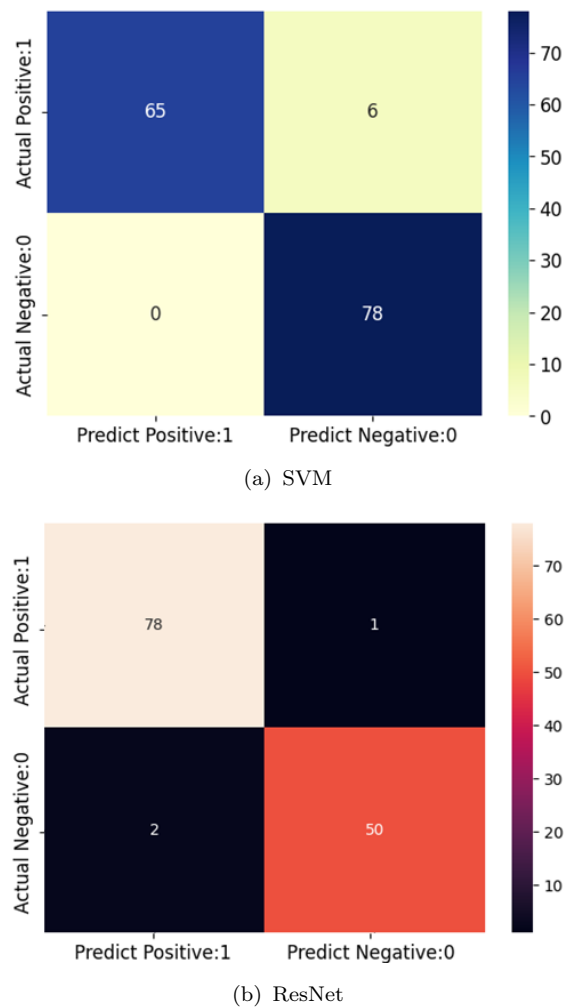


FIGURE 7.3: Comparison of Confusion Matrices for Both Classifiers

and the outcomes achieved by other researchers for the classification task. Nevertheless, machine learning-based classifiers are preferred by clinicians due to their visibility and ease of identifying abnormalities. The method or technique adopted by the respective researcher is also mentioned. Upasana Upadhyay Bharadwaj et al.[50] achieved a superior AUROC score for the classification of central canal stenosis with a binary decision tree (DT) architecture and train on a dataset of 750 patients T2 axial MRI labeled by a board of neuroradiologist with 25 years of experience. Tackeun Kim et al.[53] use VGG architecture for the classification of central canal stenosis and achieve high accuracy with an AUROC of 90 % in this study 12,442 MRI are used to train the model. The dataset is labeled by 4 expert radiologists. Jen-Tang Lu et al.[31] achieved a superior AUROC by utilizing 2D CNN and 3D CNN for the classification of lumbar spinal canal stenosis by using both sagittal and axial MRI. For this study, a dataset of 15,957 disc images was

utilized, employing a split of 70% for training, 15% for validation, and 15% for testing purposes.

## **7.4 Summary**

- Currently, clinicians employ subjective grading methods to assess the severity of stenosis. In contrast, this research introduces an innovative radio-based metric to accurately identify the presence of this disorder.
- Clinicians diagnose lumbar central canal stenosis based on transverse and anterior to posterior distances, which are influenced by the posterior element's structure. In contrast, this thesis proposes a method that utilizes the ratios of dural sac to disc cross-sectional area and dural sac-to-disc anterior-posterior diameters, offering improved detection of spinal cord compression.

# Chapter 8

## Conclusion and Future Work

### 8.1 Thesis Summary and Research Contributions

The central theme of this thesis revolved around development of automated lumbar central canal assessment toolkit. Usefulness of such an application is gauged by establishing its clinical relevance as discussed in Section 3.5. Quite evidently, the effectiveness and reliability of automated assessment framework is directly linked with the precision of intervertebral bodies segmentation which later is used for extraction of central canal measurements. For achieving this goal, selection of suitable dataset is most important for further analyzing the problem statement. The available dataset containing 2D intervertebral body discs labelled masks in axial plane is used (Section 4.2).

Extensive non-exhaustive review of related literature is carried out and the identified research gaps are further explored in order to plug the gaps. As a first measure, a intervertebral body segmentation is proposed (Section 4.3) based on deep learning methods. Deep learning architectures are then extensively used with the suggested topology as reflected in the literature reviewed. The segmentation results obtained using the deep learning networks clearly surpassed. In an effort to improve the segmentation task, weighted average ensemble technique is used. The most refined and less time-consuming process of transfer learning fine-tuning



has also proved to be extremely useful once avoiding complexity of training the deep learning models from scratch (Section 4.3.4).

After completing the segmentation task, clinically significant and highly relevant central canal stenosis measurements are extracted from the L5-S1 disc, as outlined in Section 5.2. Finally, effectively using the extracted measurements, classification methodologies deep learning-based and machine learning-based (Section 6.2) for Lumbar central canal stenosis are proposed.

## 8.2 Conclusion

In a nutshell, an effort was made in this research thesis, to present an automated image understanding of lumbar spine disc images with the perspective of spinal surgeon who is clinically evaluating the candidate patient through physical examination. The decision for most appropriate surgical intervention procedure is based on certain disc and thecal sac measurements which are presently done through laborious manual measurements.

Adoption of the automated measurement as proposed in this research thesis will certainly save the valuable time of the spinal surgeon as well as provide confidence to the decision for suitability of shortlisted/selected surgical intervention procedure. Moreover, by dynamically extracting spinal disc measurements, this approach assists clinicians in objectively evaluating the spinal pathology of patients. Additionally, this thesis presents a fully automated classification system for lumbar central canal stenosis.

It is important to note that the purpose of this autonomous lumbar spine toolkit is not to replace clinicians but to enhance their manual diagnosis by providing increased confidence and reliability. With the implementation of this toolkit, clinicians can utilize dependable quantitative metrics, thereby enhancing precision in the selection and planning of surgical interventions.

### **8.3 Future Work**

In the future, there is potential to expand the proposed framework for objective measurement of Lumbar Foraminal stenosis and Lumbar Lateral recess stenosis at each disc level (L3-L4, L4-L5, and L5-S1), providing a deeper understanding of central canal disorders. Similar to other frameworks, the medical image understanding framework requires substantial amounts of data. The available data may appear insufficient due to the novel challenges researchers face when analyzing medical images. Supplementing the dataset with 3D volumetric scans, particularly whole spine MRI images, would be highly advantageous.

Another avenue to explore is assessing the effectiveness of the automated spinal toolkit in assisting spinal surgeons in choosing less invasive methods over extensive invasive approaches.

# Bibliography

- [1] S. H. Roy, C. J. De Luca, and D. A. Casavant, “Lumbar muscle fatigue and chronic lower back pain,” *Spine*, vol. 14, no. 9, pp. 992–1001, 1989.
- [2] F. Raciborski, R. Gasik, and A. Kłak, “Disorders of the spine: A major health and social problem,” *Reumatologia*, vol. 54, no. 4, pp. 196–200, 2016.
- [3] K. Hasegawa, M. Okamoto, S. Hatsushikano, H. Shimoda, M. Ono, and K. Watanabe, “Normative values of spino-pelvic sagittal alignment, balance, age, and health-related quality of life in a cohort of healthy adult subjects,” *European Spine Journal*, vol. 25, no. 11, Nov. 2016.
- [4] H. Shemshaki, M. Etemadifar, M. Fereidan-Esfahani, M. Mokhtari, and S.-M. Nourian, “What is the source of low back pain?” *Journal of Craniovertebral Junction and Spine*, vol. 4, no. 1, pp. 21–24, Jun. 2013.
- [5] P. Suri, E. J. Boyko, J. Goldberg, C. W. Forsberg, and J. G. Jarvik, “Longitudinal associations between incident lumbar spine MRI findings and chronic low back pain or radicular symptoms: Retrospective analysis of data from the longitudinal assessment of imaging and disability of the back (laidback),” *BMC Musculoskeletal Disorders*, vol. 15, no. 1, pp. 1–10, Dec. 2014.
- [6] J. W. Frymoyer, “Back pain and sciatica,” *New England Journal of Medicine*, vol. 318, no. 5, pp. 291–300, 1988.
- [7] K. Gelse, S. Söder, W. Eger, T. Diemtar, and T. Aigner, “Osteophyte development—molecular characterization of differentiation stages,” *Osteoarthritis and Cartilage*, vol. 11, no. 2, pp. 141–148, 2003.

- [8] R. A. Deyo and Y.-J. Tsui-Wu, “Descriptive epidemiology of low-back pain and its related medical care in the united states,” *Spine*, vol. 12, no. 3, Apr. 1987.
- [9] S. S. Eun, H. Y. Lee, S.-H. Lee, K. H. Kim, and W. C. Liu, “MRI versus CT for the Diagnosis of Lumbar Spinal Stenosis,” *Journal of Neuroradiology*, vol. 39, no. 2, pp. 104–109, 2012.
- [10] S. A. Mirjalili, S. L. McFadden, T. Buckenham, B. Wilson, and M. D. Stringer, “Anatomical planes: Are we teaching accurate surface anatomy?” *Clinical Anatomy*, vol. 25, no. 7, pp. 819–826, 2012.
- [11] P. Armstrong and S. F. Keevil, “Magnetic resonance imaging–1: Basic principles of image production.” *BMJ: British Medical Journal*, vol. 303, no. 6793, pp. 35–40, 1991.
- [12] R. B. Buxton, R. R. Edelman, B. R. Rosen, G. L. Wismer, and T. J. Brady, “Contrast in rapid mr imaging: T1- and t2-weighted imaging,” *Journal of Computer Assisted Tomography*, vol. 11, no. 1, pp. 7–16, 1987.
- [13] L. J. Wolansky, D. D. Parikh, K. J. Shah, R. Yalamanchili, and J. Farkas, “Magnetic resonance imaging protocols for cervical disc disease: What is your neighbor up to?” *Journal of Neuroimaging*, vol. 15, no. 2, pp. 183–187, Apr. 2005.
- [14] M. C. Jensen, M. N. Brant-Zawadzki, N. Obuchowski, M. T. Modic, D. Malkasian, and J. S. Ross, “Magnetic resonance imaging of the lumbar spine in people without back pain,” *New England Journal of Medicine*, vol. 331, no. 2, pp. 69–73, Jul. 1994.
- [15] D. A. Seehusen, M. Reeves, and D. Fomin, “Cerebrospinal fluid analysis,” *American Family Physician*, vol. 68, no. 6, pp. 1103–1108, 2003.
- [16] S. Y. Hahn, Y. H. Lee, and J.-S. Suh, “Detection of vertebral metastases: A comparison between the modified dixon turbo spin echo t2 weighted MRI and conventional t1 weighted mri: A preliminary study in a tertiary centre,” *British Journal of Radiology*, vol. 91, no. 1085, pp. 1–9, Mar. 2018.

- 
- [17] P. Suri, D. J. Hunter, J. N. Katz, L. Li, and J. Rainville, "Bias in the physical examination of patients with lumbar radiculopathy," *BMC musculoskeletal disorders*, vol. 11, no. 3, pp. 1–8, 2010.
- [18] P. Mildemberger, M. Eichelberg, and E. Martin, "Introduction to the DICOM standard," *European Radiology*, vol. 12, no. 4, pp. 920–927, 2002.
- [19] R. E. Cooke Jr, M. G. Gaeta, D. M. Kaufman, and J. G. Henrici, "Picture archiving and communication system," U.S. Patent 6,574,629, Jun. 3, 2003.
- [20] M. Aiello, C. Cavaliere, A. D'Albore, and M. Salvatore, "The challenges of diagnostic imaging in the era of big data," *Journal of Clinical Medicine*, vol. 8, no. 3, pp. 316–327, Mar. 2019.
- [21] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019.
- [22] P. Smyth, C. Taylor, and J. Adams, "Automatic measurement of vertebral shape using active shape models," *Image and Vision Computing*, vol. 15, no. 8, pp. 575–581, Aug. 1997.
- [23] M. S. Aslan, A. Shalaby, and A. A. Farag, "Vertebral body segmentation using a probabilistic and universal shape model," *IET Computer Vision*, vol. 9, no. 2, pp. 234–250, Apr. 2015.
- [24] J. Carballido-Gamio, S. J. Belongie, and S. Majumdar, "Normalized cuts for spinal MRI segmentation," in *CARS 2002 Computer Assisted Radiology and Surgery*. Berlin, Heidelberg: Springer, 2002, pp. 1054–1054.
- [25] A. Bundy and L. Wallen, "Generalised hough transform," in *Catalog of Artificial Intelligence Tools*, 1st ed. Berlin, Heidelberg: Springer, 1984, ch. 4, pp. 42–43.
- [26] X. Zhu, X. He, P. Wang, Q. He, D. Gao, J. Cheng, and B. Wu, "A method of localization and segmentation of intervertebral discs in spine MRI based on gabor filter bank," *BioMedical Engineering OnLine*, vol. 15, no. 1, pp. 1–5, Dec. 2016.

- 
- [27] A. Mehmood, M. U. Akram, and M. Akhtar, “Vertebra localization using shape based analysis and unsupervised clustering from X-Ray images,” *PJ-CIS*, vol. 1, no. 1, pp. 13–24, 2016.
- [28] M. A. Larhmam, S. Mahmoudi, and M. Benjelloun, “Semi-automatic detection of cervical vertebrae in X-ray images using generalized hough transform,” in *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, Oct. 2012, pp. 396–401.
- [29] C. G. Bampis, A. C. Bovik, M. K. Markey, and K. M. Webb, “Segmentation and extraction of the spinal canal in sagittal MR images,” in *2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE, Mar. 2016, pp. 5–8.
- [30] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, “Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans,” in *2012 Medical Image Computing and Computer-Assisted Intervention (MICCAI) 15th International Conference*. Springer, Oct. 2012, pp. 590–598.
- [31] J.-T. Lu, S. Pedemonte, B. Bizzo, S. Doyle, K. P. Andriole, M. H. Michalski, R. G. Gonzalez, and S. R. Pomerantz, “Deep spine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning,” in *Machine Learning for Healthcare Conference*. PMLR, Nov. 2018, pp. 403–419.
- [32] R. Janssens and G. Zheng, “Deep learning based segmentation of lumbar vertebrae from ct images,” *CAOS*, vol. 2, no. 5, pp. 94–97, Jul. 2018.
- [33] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Išgum, “Iterative fully convolutional neural networks for automatic vertebra segmentation and identification,” *Medical Image Analysis*, vol. 53, no. 5, pp. 142–155, Apr. 2019.
- [34] H. Tang, X. Pei, S. Huang, X. Li, and C. Liu, “Automatic lumbar spinal CT image segmentation with a dual densely connected U-Net,” *IEEE Access*, vol. 8, pp. 89 228–89 238, Oct. 2020.

- 
- [35] B. Benjdira, K. Ouni, M. M. Al Rahhal, A. Albakr, A. Al-Habib, and E. Mahrous, “Spinal cord segmentation in ultrasound medical imagery,” *Applied Sciences (Switzerland)*, vol. 10, no. 4, p. 1370, Feb. 2020.
- [36] T. Hassan and N. Werghi, “Exploiting the transferability of deep learning systems across multi-modal retinal scans for extracting retinopathy lesions,” in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, Oct. 2020, pp. 577–581.
- [37] Y. Siriwardhana, D. Karunaratna, and I. U. Ekanayake, “Segmentation and significance of herniation measurement using lumbar intervertebral discs from the axial view,” in *2022 Moratuwa Engineering Research Conference (MER-Con)*. IEEE, 2022, pp. 1–6.
- [38] N. Kowlagi, H. H. Nguyen, T. McSweeney, S. Saarakkala, J. Määttä, J. Karpinen, and A. Tiulpin, “A stronger baseline for automatic pfirrmann grading of lumbar spine mri using deep learning,” in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, Apr. 2023, pp. 1–5.
- [39] M. Apaydin, M. Yumus, A. Degirmenci, S. Kesikburun, and O. Karal, “Deep convolutional neural networks using U-Net for automatic intervertebral disc segmentation in axial mri,” in *2022 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2022, pp. 1–6.
- [40] A. Shukla, S. Bhardwaj, and M. Singh, “Segmentation for lumbar spinal stenosis using convolutional neural networks,” *Procedia Computer Science*, vol. 218, no. 6, pp. 2210–2223, 2023.
- [41] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, no. 8, pp. 2672–2680, 2014.
- [42] Y. L. Guen, W. L. Joon, S. C. Hee, O. Kyoung-Jin, and S. K. Heung, “A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method,” *Skeletal radiology*, vol. 40, no. 2, pp. 1033–1039, 2011.

- [43] C. Schizas, N. Theumann, A. Burn, R. Tansey, D. Wardlaw, F. W. Smith, and G. Kulik, “Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images,” *Spine*, vol. 35, no. 21, pp. 1919–1924, 2010.
- [44] Y.-j. Ko, E. Lee, J. W. Lee, C. Y. Park, J. Cho, Y. Kang, and J. M. Ahn, “Clinical validity of two different grading systems for lumbar central canal stenosis: Schizas and lee classification systems,” *PLoS One*, vol. 15, no. 5, pp. 15–25, 2020.
- [45] N. Miskin, Z. Isaac, Y. Lu, M. C. Makhni, D. L. Sarno, T. R. Smith, J. M. Zampini, and J. C. Mandell, “Simplified universal grading of lumbar spine MRI degenerative findings: Inter-reader agreement of non-radiologist spine experts,” *Pain Medicine*, vol. 22, no. 7, pp. 1485–1495, 2021.
- [46] B. S. Russell, K. A. Muhlenkamp, K. T. Hoiriis, and C. M. DeSimone, “Measurement of lumbar lordosis in static standing posture with and without high-heeled shoes,” *Journal of Chiropractic Medicine*, vol. 11, no. 3, pp. 145–153, Sep. 2012.
- [47] P. Rajnics, V. Pomeroy, A. Templier, F. Lavaste, and T. Illes, “Computer-assisted assessment of spinal sagittal plane radiographs,” *Clinical Spine Surgery*, vol. 14, no. 2, pp. 135–142, 2001.
- [48] R. Vialle, N. Levassor, L. Rillardon, A. Templier, W. Skalli, and P. Guigui, “Radiographic analysis of the sagittal alignment and balance of the spine in asymptomatic subjects,” *Journal of Bone and Joint Surgery*, vol. 87, no. 2, pp. 260–267, 2005.
- [49] P. Roussouly, S. Gollogly, E. Berthonnaud, and J. Dimnet, “Classification of the normal variation in the sagittal alignment of the human lumbar spine and pelvis in the standing position,” *Spine*, vol. 30, no. 3, pp. 346–353, 2005.
- [50] U. U. Bharadwaj, M. Christine, S. Li, D. Chou, V. Padoia, T. M. Link, C. T. Chin, and S. Majumdar, “Deep learning for automated, interpretable classification of lumbar spinal stenosis and facet arthropathy from axial MRI,” *European Radiology*, vol. 33, no. 5, pp. 3435–3443, 2023.



- [51] J. Huang, H. Shen, J. Wu, X. Hu, Z. Zhu, X. Lv, Y. Liu, and Y. Wang, “Spine explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine mr images,” *The Spine Journal*, vol. 20, no. 4, pp. 590–599, 2020.
- [52] F. Natalia, H. Meidia, N. Afriliana, J. C. Young, R. E. Yunus, M. Al-Jumaily, A. Al-Kafri, and S. Sudirman, “Automated measurement of anteroposterior diameter and foraminal widths in MRI images for lumbar spinal stenosis diagnosis,” *PLoS One*, vol. 15, no. 11, pp. 1–27, 2020.
- [53] T. Kim, Y.-G. Kim, S. Park, J.-K. Lee, C.-H. Lee, S.-J. Hyun, C. H. Kim, K.-J. Kim, and C. K. Chung, “Diagnostic triage in patients with central lumbar spinal stenosis using a deep learning system of radiographs,” *Journal of Neurosurgery: Spine*, vol. 37, no. 1, pp. 104–111, 2022.
- [54] D. Won, H.-J. Lee, S.-J. Lee, and S. H. Park, “Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks,” *Spine*, vol. 45, no. 12, pp. 804–812, 2020.
- [55] A. Siccoli, M. P. de Wispelaere, M. L. Schröder, and V. E. Staartjes, “Machine learning-based preoperative predictive analytics for lumbar spinal stenosis,” *Neurosurgical Focus*, vol. 46, no. 5, pp. 1–9, 2019.
- [56] J. Kim, S.-J. Yang, H. Kim, Y. Kim, J.-B. Park, C. DuBose, and T.-H. Lim, “Effect of shear force on intervertebral disc (ivd) degeneration: an in vivo rat study,” *Annals of Biomedical Engineering*, vol. 40, no. 9, pp. 1996–2004, Sep. 2012.
- [57] C. Colosimo, M. Pileggi, A. Pedicelli, G. Perotti, and A. M. Costantini, “Diagnostic imaging of degenerative spine diseases: The technical approach,” in *Minimally Invasive Surgery of the Lumbar Spine*, 1st ed. London: Springer, 2014, ch. 2, pp. 21–47.
- [58] A. P. Goode, T. S. Carey, and J. M. Jordan, “Low back pain and lumbar spine osteoarthritis: how are they related?” *Current rheumatology reports*, vol. 15, pp. 1–8, 2013.

- [59] O. Osti, B. Vernon-Roberts, R. Moore, and R. Fraser, “Annular tears and disc degeneration in the lumbar spine: A post-mortem study of 135 discs,” *The Journal of bone and joint surgery. British volume*, vol. 74, no. 5, pp. 678–682, Sep. 1992.
- [60] K. Luoma, H. Riihimäki, R. Luukkonen, R. Raininko, E. Viikari-Juntura, and A. Lamminen, “Low back pain in relation to lumbar disc degeneration,” *Spine*, vol. 25, no. 4, pp. 487–492, 2000.
- [61] J. L. Melancia, A. F. Francisco, and J. L. Antunes, “Spinal stenosis,” in *Handbook of Clinical Neurology*. Elsevier, 2014, vol. 119, ch. 35, pp. 541–549.
- [62] S. Orita, K. Inage, Y. Eguchi, G. Kubota, Y. Aoki, J. Nakamura, Y. Matsuura, T. Furuya, M. Koda, and S. Ohtori, “Lumbar foraminal stenosis, the hidden stenosis including at L5/S1,” *European Journal of Orthopaedic Surgery & Traumatology*, vol. 26, no. 5, pp. 685–693, 2016.
- [63] C. K. Lee, W. Rauschnig, and W. Glenn, “Lateral lumbar spinal canal stenosis: classification, pathologic anatomy and surgical decompression,” *Spine*, vol. 13, no. 3, pp. 313–320, 1988.
- [64] S. D. Glassman, K. Bridwell, J. R. Dimar, W. Horton, S. Berven, and F. Schwab, “The impact of positive sagittal balance in adult spinal deformity,” *Spine*, vol. 30, no. 18, pp. 2024–2029, 2005.
- [65] J. A. Carrino, J. D. Lurie, A. N. Tosteson, T. D. Tosteson, E. J. Carragee, J. Kaiser, M. R. Grove, E. Blood, L. H. Pearson, J. N. Weinstein *et al.*, “Lumbar spine: reliability of MR imaging findings,” *Radiology*, vol. 250, no. 1, pp. 161–170, 2009.
- [66] B. Roudsari and J. G. Jarvik, “Lumbar spine MRI for low back pain: Indications and yield,” *American Journal of Roentgenology*, vol. 195, no. 3, pp. 550–559, Sep. 2010.

- [67] T. J. Errico, B. S. Lonner, and A. W. Moulton, "Radiologic imaging of spinal deformities," in *Surgical Management of Spinal Deformities*, 2nd ed. Elsevier Health Sciences, 2008, ch. 4, pp. 45–56.
- [68] M. Haefeli and N. Boos, "Outcome assessment in spinal surgery," in *Spinal Disorders: Fundamentals of Diagnosis and Treatment*, 2nd ed. London: Springer Berlin Heidelberg, 2008, ch. 40, pp. 1123–1142.
- [69] P. P. Raj, "Intervertebral disc: anatomy-physiology-pathophysiology-treatment," *Pain Practice*, vol. 8, no. 1, pp. 18–44, 2008.
- [70] I. A. Harris, A. Traeger, R. Stanford, C. G. Maher, and R. Buchbinder, "Lumbar spine fusion: what is the evidence?" *Internal Medicine Journal*, vol. 48, no. 12, pp. 1430–1434, Dec. 2018.
- [71] G. Logroscino and W. Lattanzi, "Bone substitution in spine fusion: the past, the present, and the future," in *Minimally Invasive Surgery of the Lumbar Spine*, 1st ed. London: Springer, 2013, ch. 15, pp. 311–331.
- [72] R. J. Rothrock, I. T. McNeill, K. Yaeger, E. K. Oermann, S. K. Cho, and J. M. Caridi, "Lumbar lordosis correction with interbody fusion: Systematic literature review and analysis," *Spine Deformity*, vol. 6, no. 5, pp. 594–602, Oct. 2018.
- [73] R. J. Mobbs, K. Phan, G. Malham, K. Seex, and P. J. Rao, "Lumbar interbody fusion: techniques, indications and comparison of interbody fusion options including plif, tlif, mi-tlif, olif/atp, llif and alif," *Journal of spine surgery*, vol. 1, no. 1, pp. 2–18, 2015.
- [74] T. Harrington and S. Javedan, "Posterior lumbar interbody fusion," *Neurosurgery Quarterly*, vol. 10, no. 3, pp. 228–238, 2000.
- [75] T. P. Yamato, C. G. Maher, B. T. Saragiotto, M. J. Hancock, R. W. Ostelo, C. M. Cabral, L. C. Menezes Costa, and L. O. Costa, "Pilates for low back pain," *Sao Paulo Medical Journal*, 2016.

- [76] C. E. Kosasih, T. Solehati, and A. Cakrahayat, “Non-invasive treatment to reduce low back pain among treatment acupuncture, massage, spinal manipulation, yoga, and tai chi a systematic review,” *Journal of Maternity Care and Reproductive Health*, vol. 1, no. 1, pp. 42–54, 2018.
- [77] S. Budithi, R. Dhawan, A. Cattell, B. Balain, and D. Jaffray, “Only walking matters—assessment following lumbar stenosis decompression,” *European Spine Journal*, vol. 26, no. 7, pp. 481–487, 2017.
- [78] S. Y. Lee, T.-H. Kim, J. K. Oh, S. J. Lee, and M. S. Park, “Lumbar stenosis: a recent update by review of literature,” *Asian spine journal*, vol. 9, no. 5, pp. 818–828, 2015.
- [79] S. Sudirman, A. Al Kafri, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi, and M. Al-Jumaily, “Lumbar spine MRI dataset,” *Mendeley Data*, vol. 2, 2019.
- [80] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [81] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. El-Mahallawy, “An efficient brain tumor image segmentation based on deep residual networks (resnets),” *Journal of King Saud University - Engineering Sciences*, vol. 33, no. 6, pp. 404–412, 2021.
- [82] A. A. Pravitasari, N. Iriawan, M. Almuhyar, T. Azmi, I. Irhamah, K. Fithriasari, S. W. Purnami, and W. Ferriastuti, “UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 3, pp. 1310–1318, 2020.
- [83] N. Dong, L. Zhao, C. Wu, and J. Chang, “Inception v3 based cervical cell classification combined with artificially extracted features,” *Applied Soft Computing*, vol. 93, no. 106311, pp. 1–7, 2020.

- 
- [84] C. Balakrishna, S. Dadashzadeh, and S. Soltaninejad, “Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 encoder,” *CoRR*, vol. abs/1806.07554, 2018.
- [85] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Trans. Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [86] İ. Delibaşoğlu and M. Çetin, “Building segmentation with inception-unet and classical methods,” in *28th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2020, pp. 1–4.