Image Analysis for the Diagnosis of MR Images of the Lumbar Spine

By
Sofia Michopoulou

A Thesis Submitted to University College London
for the Degree of Doctor of Philosophy
Department of Medical Physics and Bioengineering
University College London

I, Sofia Michopoulou confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature _____

Abstract

Intervertebral disc degeneration is related to chronic back pain and functional incapacity. Magnetic Resonance Imaging (MRI) is the modality of choice for diagnosing this condition, providing both morphological and biochemical information for the disc tissue. In clinical practice, grading schemes based on qualitative descriptions of disc image features such as the signal intensity and disc height are commonly used for disc degeneration severity evaluation. However, these grading schemes have a limited number of degeneration severity classes which impairs the detection of small changes. Additionally, this grading is susceptible to inter and intra observer variabilities.

To deal with these issues, this study introduces a system for the automated quantification and computer aided diagnosis of disc degeneration severity from spine MRI. The proposed system consists of a segmentation method, a quantification process, and a classification scheme. An atlas-based segmentation approach, combining prior anatomical knowledge provided by means of a probabilistic disc atlas with fuzzy clustering techniques, was designed for extracting the disc region from the images. In the quantification process, texture and shape descriptors are calculated from the segmented disc region aiming to capture structural and biochemical alterations of the tissue related to degeneration. Finally, the classification scheme exploits this information for differentiating between degeneration severity grades. The system is tested on a case sample of 255 discs from conventional T2-weighted MR images acquired by a 3 Tesla scanner.

Results indicate that the atlas-based method provides accurate disc segmentation, texture descriptors measuring intensity inhomogeneity can serve the quantification of degeneration severity, and the computer aided diagnosis scheme achieves high agreement to clinical diagnosis.

Concluding, the proposed system could be a valuable tool in hands of physicians to support clinical diagnosis of disc degeneration, track the evolution of disease progress and monitor the response to treatment in a simple, precise and repeatable manner.

Acknowledgements

My most sincere thanks are due to:

my supervisors Professor Andrew Todd-Pokropek, Professor Robert Speller, and Associate Professor Lena Costaridou;

the collaborators or this project Professor George Panayiotakis, Professor Elias Panagiotopoulos, Assistant Professor Marianna Vlychou;

the clinical experts Dr. Aleksandra Kazantzi, Dr. Nikos Karageorgos and Assistant Professor Aikaterini Vassiou;

Dr Vera Mann for her guidance in statistical analysis;

Dr Tryphon Lambrou for his advice in all image analysis aspects of this work;

Mr George Lambropoulos for his assistance in data acquisition;

my viva examiners Dr Gill Vivian and Dr Paul Taylor for their precious comments;

my colleagues and Orestis for their continuous support.

I am also grateful to the Greek State Scholarships Foundation for funding my studies.

Contents

List of Publications

Journal Publications

1. Michopoulou, S., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2009. Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. IEEE transactions on biomedical engineering 56, 2225-2231.

2. Michopoulou, S., Costaridou, L., Vlychou, M., Speller, R., Todd-Pokropek, A., 2010. Texture-Based Quantification of Lumber Intervertebral Disc degeneration from Conventional T2-weighted MRI. Accepted for publication in Acta Radiologica (August 2010).

Conference Publications

3. Michopoulou, S., Vlychou, M., Costaridou, L., Kazantzi, A., Vassiou, K., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2010. Lumbar Disc Herniation: Computer aided diagnosis from spine MRI. Computer Assisted Radiology and Surgery Conference (CARS). Geneva, Switzerland, published in the International Journal of Computer Assisted Radiology and Surgery; 5 (Suppl 1), 220-221.

4. Todd-Pokropek, A., Michopoulou, S., 2010. The current and future contribution of CAD in routine diagnostic imaging. UK Radiological Conference (UKRC), Birmingham, UK.

5. Michopoulou, S., Vlychou, M., Costaridou, L., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2010. Intervertebral Disc Herniation Quantification in Lumbar Spine MR Images Acquired by a 3T Scanner. UK Radiological Conference (UKRC), Birmingham, UK.

6. Michopoulou, S., Costaridou, L., Vlychou, M., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2009. Intensity Inhomogeneity Correction in MRI of Lumbar Spine at 3T. International Conference on Imaging Technologies in Biomedical Sciences (ITBS), Milos, Greece.

7. Michopoulou, S., Costaridou, L., Kazantzi, A., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2009. Computer Aided Diagnosis of Intervertebral Disc Degeneration in Spine MRI. Computer Assisted Radiology and Surgery (CARS). Berlin, Germany, published in the International Journal of Computer Assisted Radiology and Surgery; 4(Suppl 1), 187-188.

8. Michopoulou, S., Costaridou, L., Kazantzi, A., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A., 2009. Computer Aided Quantification of Lumbar Intervertebral Disc Degeneration in MRI. UK Radiological Conference (UKRC), Manchester, UK.
   Awarded IPEM President's Prize

9. Michopoulou, S., Costaridou, L., Speller, R., Panagiotopoulos, E., Todd-Pokropek, A., 2008. Segmenting lumbar intervertebral discs from spine MRI. British Machine Vision Association Symposium on Segmentation of Anatomical Soft Tissue Regions in Medical Data, London, UK.

10. Michopoulou, S., Costaridou, L., Panagiotopoulos, E., Speller, R., Todd-Pokropek, A., 2008. Segmenting degenerated lumbar intervertebral discs from MR images. IEEE Nuclear Science Symposium and Medical Imaging Conference Dresden, Germany.

List of Tables

List of Figures

Chapter 1. Introduction

The present study introduces an automated system for the quantification and computer aided diagnosis of intervertebral disc degeneration, from conventional T2-weighted magnetic resonance images of the lumbar spine. In addition, it works towards the segmentation of intervertebral discs as a step prior to degeneration quantification.

Section 1.1 gives an introduction to the clinical problem and briefly summarizes the current research in the fields of intervertebral disc segmentation, quantification and computer aided diagnosis. The aims and objectives of this study are analyzed in detail in section 1.2.

1.1 Introduction and brief summary of current research

Intervertebral disc degeneration is an age associated condition related to chronic back pain and functional incapacity [MODI2007]. It affects a significant proportion of the population and imposes a tremendous socioeconomic burden. In the UK, low back pain is the number one reason for absence from work, while the public healthcare cost of low back pain treatment is estimated at £1.6 billions per year [MANI2000]. The mechanism of degeneration is complex and involves structural disruption and cell mediated changes in disc composition [ADAM2006].

Various modalities are used for imaging the spine, with Magnetic Resonance Imaging (MRI) being a good modality for evaluating intervertebral disc degeneration. MRI is non invasive and does not use ionizing radiation, but more importantly it offers good soft tissue contrast which allows visualization of the disc's internal structure [PARI2007]. Clinical imaging relies on multislice 2-Dimensional protocols which offer fast acquisition time but with relatively low interslice resolution. 3D protocols provide high resolution volumetric data but their clinical use for spine imaging is limited, due to their long acquisition times and subsequent susceptibility to motion artefacts [WOLA2005]. However, the lack of 3D data is not a problem when clinically evaluating disc degeneration, as the disease severity can be assessed by the mid-sagittal slice of T2-weighted images [LUOM2000, PFIR2001]. T2-weighted images provide both biochemical and morphological information for the disc tissue. Here, the term biochemical tissue alterations refers to the disc's water and proteoglycan content which is reflected in MR signal intensity of T2-weighted images. On the other hand disc shape information allows the assessment of disc narrowing, fissuring or herniation [TERT1991, PARI2007].

In clinical practice, the evaluation of disc degeneration severity is based on grading schemes utilizing qualitative descriptions of disc features such as the signal intensity, height and distinction between the disc's nucleus and annulus. However, this qualitative evaluation is susceptible to inter and intra observer variabilities [RAIN1995, KETT2006]. In addition, the limited number of degeneration severity classes used by grading schemes (typically 3-5 classes)

impairs the detection of small changes in the intervertebral disc [NIEM2008]. More importantly, as disc pathology treatment is shifting from surgical procedures towards treating the underlying aetiologic process the clinical requirements for disc degeneration diagnosis are changing [MWAL2008]. The development of emerging treatment technologies, such as nucleus replacement, cell therapy, and growth factor therapy, creates a need for more objective diagnostic methods that would allow tracking the evolution of disease and monitoring the response to treatment [AUER2006].

MRI quantification has great potential as a tool for the diagnosis of disc pathology. Quantitative measurements can offer a more objective and reproducible way to evaluate degeneration. Moreover, the continuous nature of these measures renders them more sensitive to detecting small changes, a task particularly important when monitoring the effect of treatment [MWAL2008, NEIM2008].

Currently, the most common method for disc degeneration quantification relies on the measurement of disc's mean signal intensity from mid-sagittal T2-weighted MR Images. The decrease of disc's mean signal intensity may be the earliest degenerative change seen in MRI and is a sensitive and reliable measure of degeneration severity. The disc signal intensity value needs to be normalized by using an intrabody reference, most commonly a region within the cerebrospinal fluid. The great advantage of the adjusted mean signal intensity quantification method is that it utilizes conventional T2 images for quantification, and thus can be easily applied in clinical routine [LUOM2001, BENN2005, NIEM2008].

Alternative approaches to disc degeneration quantification include the measurement of T1, T1ρ and T2 relaxation times and of the apparent diffusion coefficient [MWAL2008]. These methods provide additional information to the conventional T2-weighted images, regarding both the biochemical composition and structural integrity of the intervertebral discs. However, this does not come without a cost. These quantification methods require specific image acquisition protocols which are not widely available and require relatively long acquisition times implicating their application in clinical routine [KERT2001b, AUER2006].

Before quantifying disc information, the intervertebral disc needs to be segmented from the image. Disc segmentation, apart from being a step prior to the quantification process, it could additionally serve computer assisted spine surgery [SEIF2006]. So far, studies dealing with the quantification of disc features for diagnostic or surgical purposes have been based on manually segmented data [TSAI2002, NIEM2008]. However, manual segmentation is a tedious and time-consuming process which lacks repeatability. A limited number of studies deal with the automatic segmentation of intervertebral discs, while most of them focus on the automated detection of the disc rather than the segmentation process itself [SEIF2006, CHEV2007, SHI2007]. Moreover, to the best of our knowledge the segmentation of degenerated intervertebral discs remains an open issue. Disc quantification and computer aided diagnosis require a segmentation method capable of accurately delineating intervertebral discs at various stages of disease (from normal to severely degenerated). With respect to the required segmentation accuracy, the clinically acceptable mean error for both diagnostic and surgical purposes has been indicated to be 1-2mm [personal communication with Elias Panagiotopoulos, Professor of Orthopaedic Surgery, University of Patras, Greece].

Moving on to the Computer Aided Diagnosis (CAD) task, CAD systems are designed for assisting physicians in the interpretation of medical images. CAD has been exploited for a variety of diagnostic problems, with common applications in the fields of breast cancer diagnosis from x-ray mammography and lung cancer diagnosis from computer tomography of the chest. Nowadays CAD systems are intended to support clinical diagnosis when being used as second readers. The clinical requirement is for a CAD system to increase diagnostic accuracy and improve agreement between observers [DOI2007, GIGE2008]. Spine imaging is a new field for CAD applications, with most studies appearing in the last decade [TSAI2002, CHAR2004, KOMP2006]. Moreover, there is a single very recent study from the Computer Aided Diagnosis and Interventions Research Group at the University of Buffalo, focusing specifically on computer aided diagnosis of intervertebral disc pathology from MR images [ALOM2010].

1.2. Aims and Objectives

The main aims of this study are presented in Table 1.1. The first aim is to develop a disc segmentation method which is intended to serve the quantification task by extracting the intervertebral disc regions from the MR images. A novel atlas based segmentation method has been developed for this purpose and its accuracy has been tested on both normal and degenerated discs. The proposed segmentation method is developed in 2D and tested on the midsagittal slices of conventional T2-weighted image data. This fulfils the clinical requirements for evaluating disc degeneration, but a 3D method could also be useful for assessing specific disc pathologies such as lumbar disc herniation or even surgical planning purposes. However, as 3D data were not available, a pseudo-3D segmentation approach applied on interpolated 2D data was developed to demonstrate the feasibility of expanding the atlas-based segmentation method to 3D.

The second aim is the quantification of intervertebral disc degeneration from clinical MR Images. An alternative quantification approach to the standard mean signal intensity quantification is presented here. Disc image texture and shape features are exploited for degeneration severity quantification, aiming to provide additional information with respect to the degenerative process and provide a more precise description of the stage of disease. MR Image texture is known to be particularly sensitive for the assessment of pathology [LERS1993], while shape information is related to the disc's structural integrity [MODI2007]. Our working hypothesis is that the quantification of such information can aid in diagnosing disc degeneration and evaluating the progression of this disease.

The final aim of this study is to develop a computer aided diagnosis system intended to serve as a second reader to support clinical diagnosis of disc degeneration severity. The method presented here exploits texture and shape features extracted from the segmented disc region in the quantification step of this study for training a hierarchical classification scheme which assigns the disc to one of four) degeneration severity classes. Since spine imaging is a new field for CAD applications only a small number of studies have been published so far and

to the best of our knowledge the present study is the first to report disc degeneration severity classification.

Table 1.1. Aims

1. To develop a segmentation method for delineating the intervertebral discs (both normal and degenerated) from magnetic resonance images of the lumbar spine.
2. To quantify intervertebral disc degeneration severity exploiting texture and shape properties of the segmented disc regions.
3. To design a computer aided diagnosis system putting together the segmentation and quantification steps along with a classification scheme for categorizing the discs to degeneration severity classes.

Table 1.2 presents a detailed analysis of research objectives which can be seen as smaller steps working towards the aims of this study. Finally, Table 1.3 presents the hypotheses that are to be tested. The theory behind theses hypotheses is described later in the thesis.

Table 1.2. Research Objectives

Sample Collection / Clinical Grading of Disc Degeneration
1. To collect clinical datasets consisting of conventional T2-weighted MR images of the lumbar spine presenting various degrees of degeneration severity
2. To select / design in collaboration with clinical experts a grading scheme for the description of qualitative image features related to disc degeneration. To have the images diagnosed by clinical experts according to this scheme. To test the reliability of clinical grading through the agreement between expert readers. Also obtain a severity grading ground truth based on clinical experts consensus to serve the disc degeneration quantification task

Intervertebral Disc Segmentation

3. To develop computerized methods for segmenting the intervertebral discs (both normal and degenerated).
4. To evaluate the segmentation accuracy of these methods against a gold standard produced by manual disc segmentation

Disc Degeneration Quantification

5. To calculate image features describing textural and shape properties of the disc region in order to quantify disc degeneration severity.
6. To investigate the ability of quantitative features in evaluating disc degeneration by calculating their association to degeneration severity as defined by radiologists consensus grading.
7. To compare textural and shape features' performance to the performance of the mean signal intensity quantification method in the task of disc degeneration quantification.
8. To test the repeatability of quantitative features measurements.

Computer Aided Diagnosis of Disc Degeneration

9. To design a computer aided diagnosis system putting together the segmentation and quantification steps with a classification scheme for categorizing the discs to degeneration severity grades.
10. To develop methods for training and testing the CAD system using radiologists consensus as ground truth, and evaluating its classification performance.
11. To apply the CAD system in a new dataset and test its influence on the radiologist's opinion when the system is used as a second reader.

Table 1.3. Research Hypotheses

Intervertebral Disc Segmentation
1. Atlas based methods would provide improved segmentation accuracy in comparison to methods relying solely on grey level intensity. This is because the incorporation of shape information in the segmentation process would help control boundary leakage towards surrounding structures.
2. The correction of inhomogeneity field in MRI would facilitate the segmentation process and improve accuracy.
Disc Degeneration Quantification
3. Textural and shape image features are suitable for describing intervertebral disc degeneration severity. This would be confirmed if these features values are statistically significantly correlated to the clinical diagnosis (grading) of disease.
4. Textural features are more suitable than shape features for evaluating the early stages of degeneration. This is because in the early stages of disease only the internal disc structure is affected but its outline remains almost intact. Moreover, shape features are more suitable that texture features for assessing the more advanced stages of disease, when internally the disc has become fibrotic.
Computer Aided Diagnosis of Disc Degeneration
5. The CAD system can help improve diagnostic accuracy when used as a second reader. This would be supported if the disagreement between different clinical experts decreases when using the CAD.

1.3. Thesis Layout

The layout of this thesis is as follows:

Chapter 2 analyzes the medical background of the current study. It starts with an overview of spinal anatomy and continues with a description of intervertebral disc degeneration. Section 2.4 discusses the modalities used for imaging the spine, and Section 2.5 gives a literature review on methods developed for the evaluation of disc degeneration from MR images.

Chapter 3 presents the clinical data collected, together with methods used for the evaluation of degeneration severity. Finally, the results of qualitative grading of degeneration are given in Section 3.3.

Chapter 4 presents the segmentation methods and results of the current study. It starts with an overview of medical image segmentation followed by a short literature review on intervertebral disc segmentation methods. Section 4.3 discusses the requirements and challenges for lumbar disc segmentation, while Section 4.4 analyses the methods developed for segmenting the intervertebral discs in the present study. Finally, Section 4.5 presents the segmentation results and discusses the merits and weaknesses of each segmentation method.

Chapter 5 presents the methods developed in this study for disc degeneration quantification. It starts with an overview of texture and shape quantification methods and describes how these methods were exploited in the present study for disc degeneration quantification. Sections 5.3 and 5.4 present the quantification results and discuss on the suitability of texture and shape analysis methods for evaluating degeneration.

Chapter 6 presents the Computer Aided Diagnosis system designed for disc degeneration severity evaluation. The chapter starts with an introduction to CAD systems their design and classification algorithms followed by a short literature review on spine CAD. Section 6.3 presents the CAD system developed in the present study followed by the CAD results and discussion in Section 6.4.

Chapter 7 gives an overall conclusion summarizing the contributions and limitations of this study and offering suggestions for future research.

Chapter 2. Medical Background

This chapter gives a short introduction to the medical knowledge related to the present study, to support reading for non-medical experts. Section 2.1 introduces the anatomical terms of location, while Section 2.2 presents a basic description of spine anatomy. Section 2.3 discusses on the current understanding of intervertebral disc degeneration. Section 2.4 presents methods used for imaging the spine, discussing on the suitability of each method for evaluating intervertebral disc degeneration. Finally, section 2.5 gives a literature review on methods developed for the evaluation of disc degeneration from MR images

Readers with a medical background shall move forward to Section 2.5.

2.1. Body Planes

Before describing the anatomy of the spine it is important to introduce the terms used for specifying relative anatomical locations. Figure 2.1 gives a diagram of body planes and anatomical terms of location. A position within the body is identified by three dimensions. The terms superior and inferior are used for describing the position along the main body axis (representing top and bottom). Horizontally, the terms anterior-posterior are used for the front and back of the body while right and left is used for description of lateral points. In addition, two-dimensional planes are used for describing cross-sections of the body. The sagittal plane is parallel to the body axis and divides the body into left and right sections. The coronal plane is also parallel to the body axis and it divides the body to anterior and posterior sections. Finally, the axial plane also known as transverse plane is perpendicular to the body axis and divides it to superior and inferior sections.

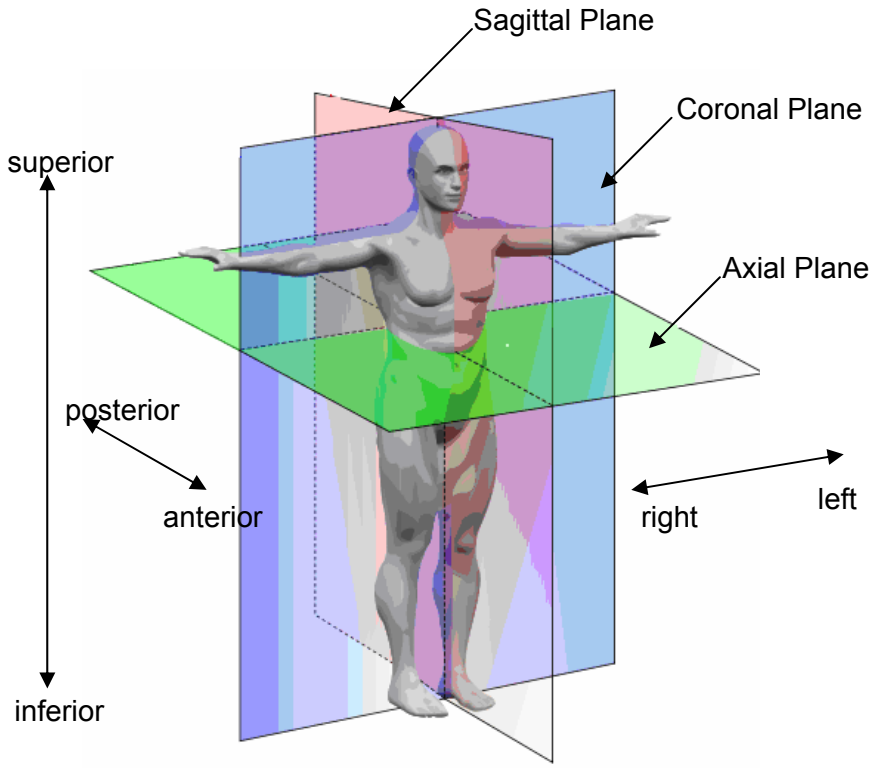


Figure 2.1. The body planes and anatomical terms of location – adapted from wikimedia commons [WIKI2010].

2.2. Spinal Anatomy

The spine is a bony structure in the middle of the back which extends from the skull to the pelvis and normally consists of 33 vertebrae. It is divided into five sections, the cervical, thoracic, lumbar, sacral and coccygeal spine, as shown in Figure 2.2. This study focuses on the lumbar spine which is in the low back area and consists of five 5 vertebrae (named L1 – L5). The vertebrae are joined axially by intervertebral discs which provide stability and flexibility to the spine, while the column is additionally stabilized by ligaments and muscles. The spinal cord and nerve roots pass through a longitudinal cavity called spinal canal, which is also the draining passage of the cerebrospinal fluid [BODG1997, DRAK2004]. Figure 2.3 illustrates a spinal segment showing the positions of intervertebral discs and the spinal cord relative to the vertebral bodies. The following sections describe in brief the main components of the spine, giving a more detailed description for the intervertebral discs.

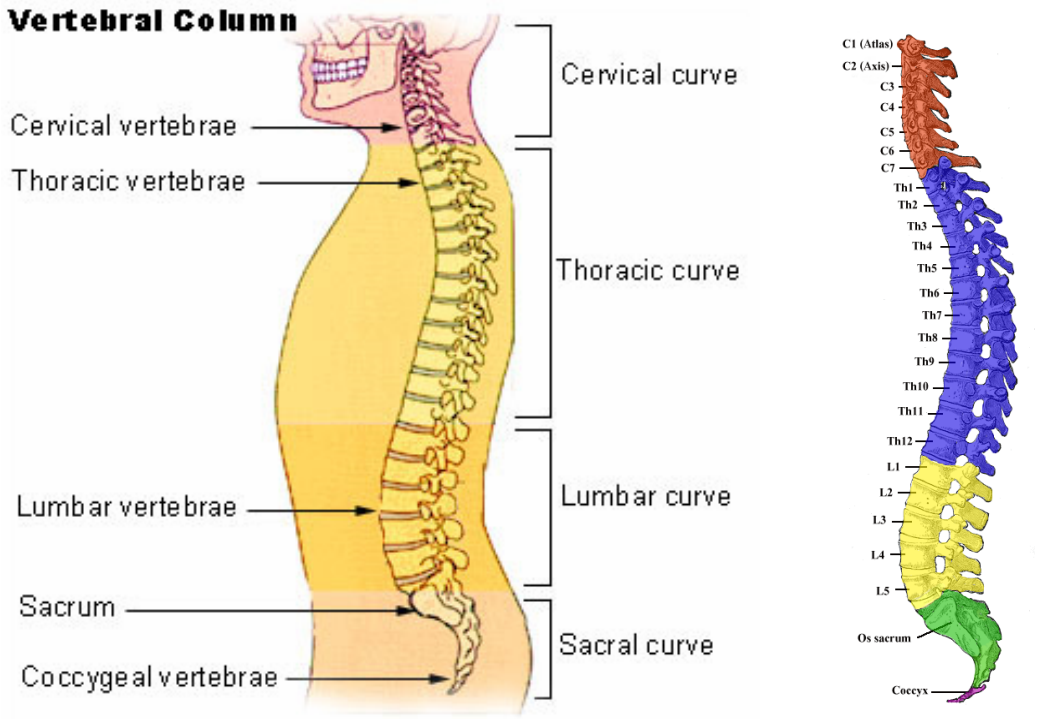


Figure 2.2. The human spine divided in five spine segments (left) and the corresponding abbreviations of the vertebrae (right) [WIKI2010].

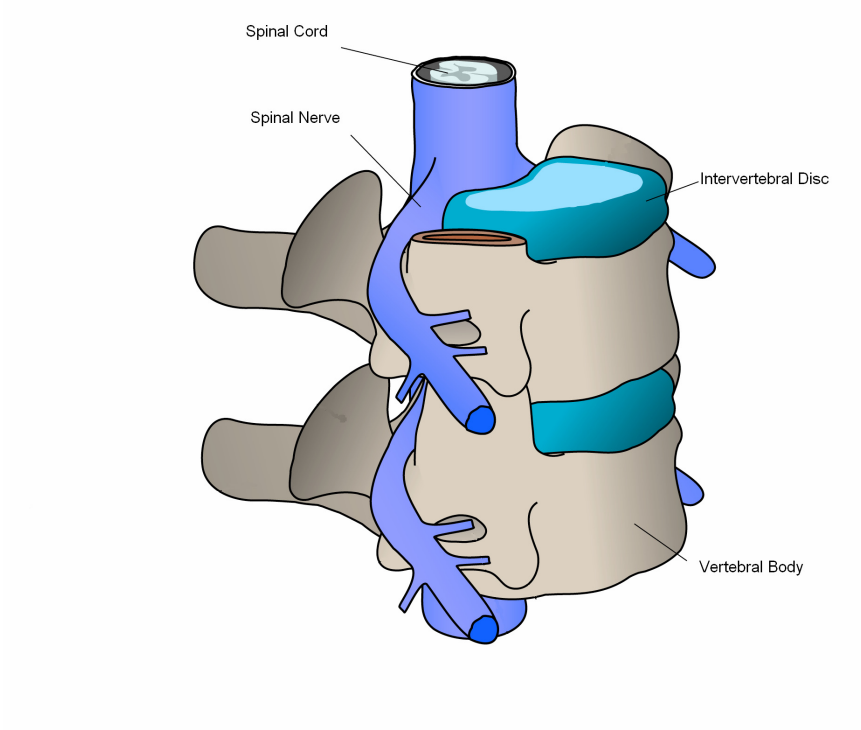


Figure 2.3. A spine segment, illustrating the vertebral bodies and intervertebral discs together with the spinal cord and nerves [WIKI2010].

2.2.1. The vertebrae: Anatomy and Function

The main components of typical vertebrae are their vertebral body and the vertebral arch. Figure 2.4 illustrates typical vertebrae in lateral and superior views. The vertebral body is the anterior portion of the vertebra, it has a nearly cylindrical shape with flat superior and interior surfaces. It consists of an external shell of cortical bone surrounding a core of trabecular bone. It is the main load bearing structure, supporting most of the axial compression of the body weight. The vertebral arch is the posterior portion of the vertebra with its processes articulating with those of adjacent vertebrae and creating joints that guide and limit spinal motion. The vertebral arch together with the posterior surface of the vertebral body forms the spinal canal, a bony ring that protects the spinal cord from injury [DRAK2008, BELK2007].

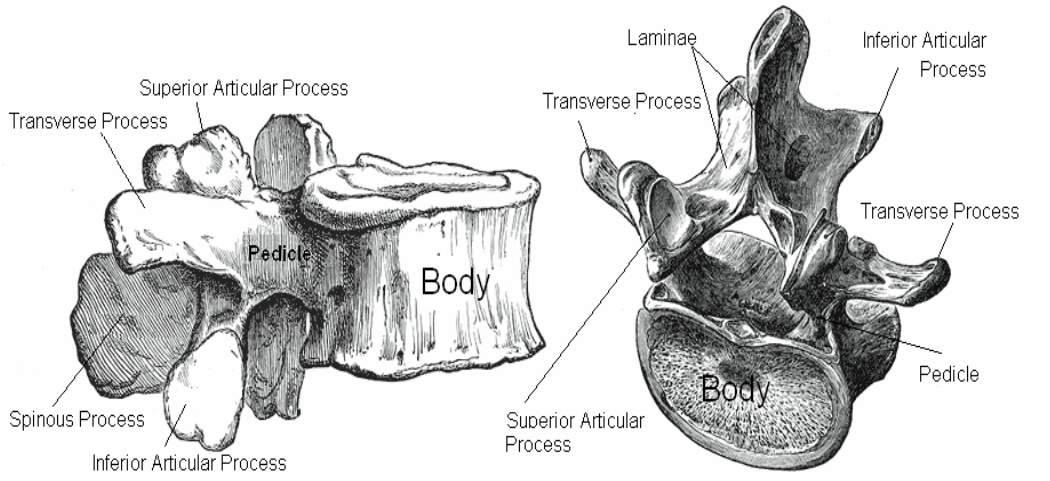


Figure 2.4. Lithographies of vertebrae from the 1918 version of Gray's anatomy [GRAY1918].

2.2.2. The intervertebral discs: Anatomy and Function

The intervertebral discs are cylindrical pads of fibrocartilage lying in between the vertebral bodies. Each disc is fused to the adjacent vertebral bodies through cartilaginous endplates and together they form a functional unit known as spinal segment. Figure 2.5 gives a schematic representation of a spinal segment in sagittal view. The discs connect adjacent vertebra providing stability and flexibility to the spine. They are responsible for load distribution and they also act as shock absorbers.

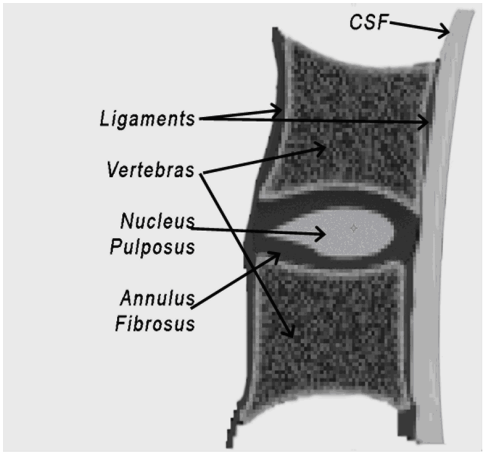


Figure 2.5. Schematic representation of a mid-sagittal cut of a spinal segment demonstrating its major anatomical components.

Each intervertebral disc consists of two major components: the nucleus pulposus, a hydrogel in the centre of the disc, and the annulus fibrosus, a collagenous ring surrounding the nucleus, as shown above in Figure 2.5. The nucleus pulposus consists primarily of water, proteoglycans and collagen molecules. Water accounts for over 80% of the nucleus weight in children and young adults. The proteoglycans are the macromolecules responsible for attracting and retaining water thus keeping the nucleus pulposus hydrated. [KERT2001a, URBA2003]. The annulus fibrosus is formed of multiple concentric collagenous layers called lamellae which surround the nucleus. These lamellae are composed by aligned collagen fibres in alternating orientations as shown in Figure 2.6. The primary component of the annulus is also water, but on a lower concentration than in the disc's nucleus (60-70%). On the other hand the annulus has a higher concentration of collagen (about 50% dry weight), which is responsible for the disc's strength. [KERT2001a].

The nucleus and annulus work together as a functional unit to evenly distribute the compressive load of the body weight, and resist shears and tensile loads resulting from flexion extension and rotation [ADAM2006, BELK2007]. With aging the intervertebral disc undergoes alterations in its composition and structure which greatly affect its functionality. A detailed description of these biochemical and structural disc alterations is given in section 2.3

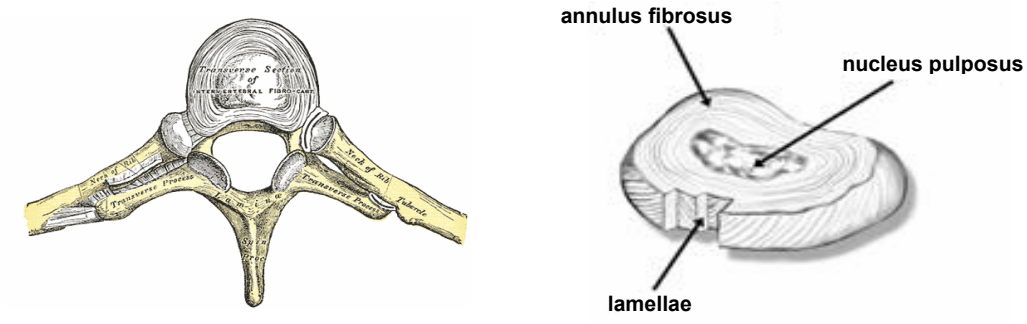


Figure 2.6. Lithography illustrating an axial section of the spine showing the position of the intervertebral disc (left), and a simple intervertebral disc model showing the alternating orientation of collagen fibres in the lamellas (right) [GRAY1918, NEUR2010].

2.2.3. The spinal cord and spinal nerves

The spinal cord is a nearly cylindrical bundle of nerves within the spinal canal. It extends from the brain down to the lumbar spine and it terminates on the first segment of the coccyx, as shown in Figure 2.7. From a functional point of view, the spinal cord forms part of the central nervous system and it is the main pathway that connects the brain to the peripheral nervous system.

The spinal nerves branch out from the spinal cord connecting the central nervous system to the peripheral nervous system. Each nerve is connected to the spinal cord by either anterior of posterior roots as shown in Figure 2.8. Nerves connected to posterior roots carry sensory information to the central nervous system. Sensory neurons transmit information from body segments called dermatomes. The anterior roots form the starting points of most motor neurons, which carry signals away from the central nervous system and innervate the muscles in order to control motion [DRAK2004].

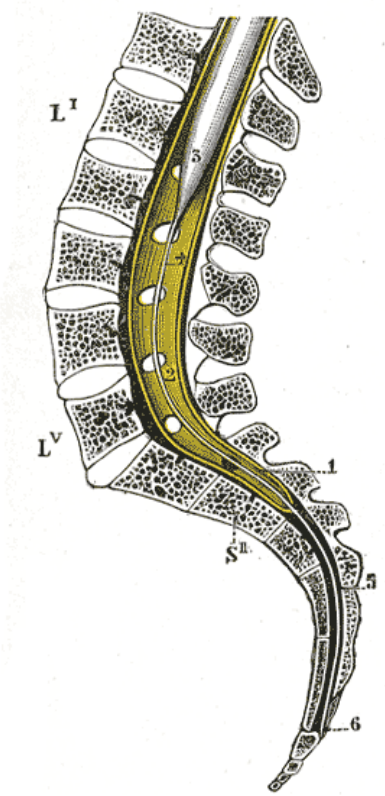


Figure 2.7. Lithographies depicting: a sagittal section of the vertebral canal, showing the lower end of the spinal cord [GRAY1918].

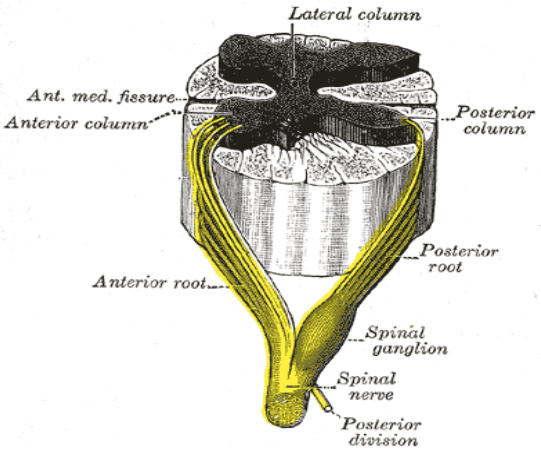


Figure 2.8. Lithography depicting an axial section of the spinal cord along with the anterior and posterior neural roots [GRAY1918].

2.2.4. The spinal ligaments: Anatomy and Function

The ligaments of the spine are composed of collagen and elastin fibres and contribute to spinal stability. They guide the joint motion and provide flexibility, while at the same time limiting excess motion that could harm the spinal cord. The major ligaments of the lumbar spine are depicted in Figure 2.9 [DRAK2004, BELK2007].

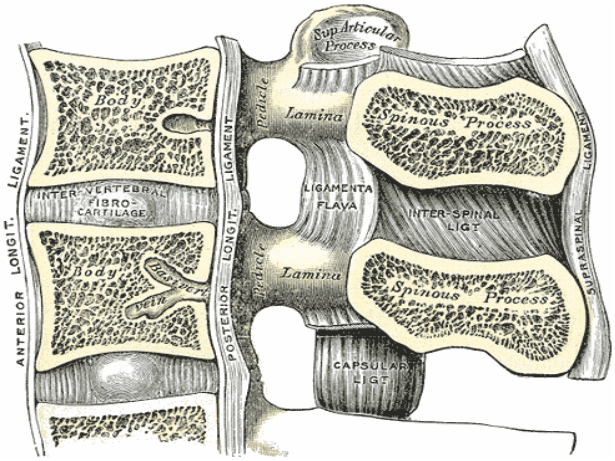


Figure 2.9. Lithography of a midsagittal section of the lumbar spine depicting two vertebrae together with their corresponding ligaments [GRAY1918].

2.2.5. The spinal muscles

The spinal muscles are a complex muscle group extending from the pelvis to the skull called deep muscles of the back. They work together with the ligaments to hold and support the spine and to control movement. They are categorized with respect to their function to forward and lateral flexors, extensors and rotators [DRAK2004].

2.3 Spine Pathology: Intervertebral Disc Degeneration

2.3.1 Introduction

Intervertebral disc degeneration was recently defined as an "aberrant cell mediated response to progressive structural failure" [ADAM2006]. The degenerative process involves dehydration of the disc's nucleus, and weakening of the annulus. This results in reduced load bearing and shock absorbing abilities, diminishing the disc's functionality.

2.3.2 Aetiology

The understanding of disc degeneration and its aetiology has changed tremendously within the last decade. In the past injuries related to physical loading were sought as the main causes of disc degeneration. However, newer studies on the effect of genetic influences to the progression of disc degeneration have demonstrated that heredity plays a dominant role [BATT1995, BATT2004]. In addition, it has become clear that degenerative changes are part of the aging process [AN2004]. Degenerative changes appear as early as the second decade of life, with 20% of teenagers having mild signs of degeneration. The progress of degeneration increases steeply with age and 60% of all discs are severely degenerated by the age of 70 [URBA2003]. Other influences affecting the progress of disc degeneration include mechanical, occupational, traumatic and nutritional factors [LUOM2000, URBA2003, MODI2007].

2.3.3 Biochemical Alterations

The degenerative process begins when catabolism exceeds synthesis of matrix proteins [AN2004]. Moreover, proteoglycan fragmentation results in smaller fragments which can leak out of the disc tissue. The reduction of disc's

proteoglycan content is the most significant change due to degeneration since it directly affects the disc's ability to retain water [ADAM2006].

2.3.4 Morphology and Function

According to the Nomenclature Project guidelines, the term Intervertebral Disc Degeneration includes any or all of the following: "real or apparent desiccation, fibrosis, narrowing of the disc space, diffuse bulging of the annulus beyond the disk space, extensive fissuring (i.e. numerous annular tears) and mucoid degeneration of the annulus, defects and sclerosis of the endplates, and osteophytes at the vertebral apophyses" [MILE1997, FARD2001].

These alterations of the disc tissue are related to the biochemical changes described in the previous section. The loss of proteoglycan content, results in disc dehydration which causes shrinkage of the nucleus. This leads to narrowing of the disc space and bulging of the annulus. At the same time the annulus dehydration makes the tissue stiffer and weaker. The annulus progressively encroaches the nucleus resulting in loss of the nucleus-annulus distinction. The disc becomes less elastic and thus more prone to fissuring [ADAM2006, PARI2007].

Figure 2.10 illustrates some of the most common morphological alterations related to disc degeneration. These macroscopic changes are manifestations of the architectural distortion and structural failure of the intervertebral disc. Disc narrowing is the direct result of dehydration, and volume reduction of the nucleus. Annular tears also know as fissures are cracks in the annulus fibrosus. Disc bulging is a broad displacement of disc material, covering 50-100% of its circumference, while disc herniation as defined by the nomenclature project is a localized displacement of disc material beyond the limits of the intervertebral disc space [FARD2001].

Finally, from the functional point of view, dehydration adversely affects the disc's ability to withstand compressive loads. The disc becomes stiffer and its shock absorbing ability diminishes. Disc space narrowing reduces the space between adjacent neural arches. A greater proportion of the compressive load

needs to be taken by the arches and this is related with osteoarthritis of the zygapophyseal joints. In addition, the alteration of stress distribution, results in the formation of bony spurs called osteophytes around the vertebral bodies' margins. Moreover, lumbar disc herniations can cause compression of the nerve roots resulting in pain radiating to the lower extremities [FUJI1999, PARI2007].

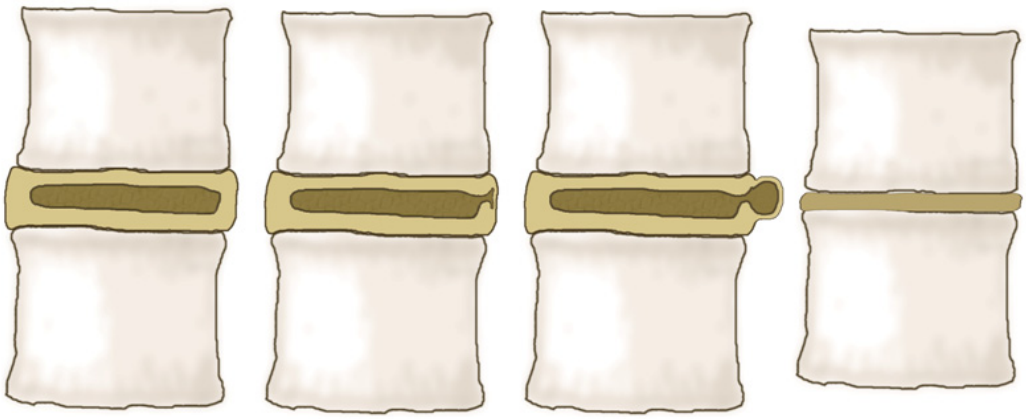


Figure 2.10. Schematic representation of disc's midsagittal cuts illustrating the most common morphological alterations affecting the intervertebral discs: (a) a normal case, (b) an annular tear, (c) a disc herniation and (d) disc space narrowing.

2.3.5 Disc Degeneration and Low-Back Pain

The sequelae of disc degeneration are considered to be major causes of low back pain [MODI2007]. Despite increasing knowledge about the mechanisms of degeneration the actual relationship between degeneration and pain is still unclear [AN2004]. Multiple studies show relations between degenerative findings, such as disc space narrowing, disc herniation and modic changes in adjacent vertebrae, and the occurrence of low back pain [PAAJ1997, BRAI1998, LUOM2000]. From a clinical point of view, the degenerative process is considered to be responsible for over 90% of surgical spine procedures [AN2004]. However, this does not imply that disc degeneration is an indication of pain, or that it requires treatment. On the contrary, many degenerative findings are common among asymptomatic individuals [JENS1994].

Looking into the socioeconomic aspect of pain, low back pain has been described as "a major public health problem in western industrialized societies" [URBA2003]. It affects a significant proportion of the population and imposes a tremendous economic burden due to patient treatment costs and reduced productivity. Its annual incidence in the adult population is between 10% and 15%, while it affects most individuals at some point in their lifetime [ANDE1999, KATZ2006]. It is reported that the cost for low-back pain health care is about £1.6 billion pounds per year in the UK. And this is only a small proportion of the actual cost, since the cost related to production losses is multiple times higher (estimated between £5 and £10 billions in the UK). Low back pain is the largest cause for absence from work in the UK accounting for 12.5% of all sick-days, while similar figures are presented in studies from Sweden [ANDE1999, MANI2000].

2.3.6 Treatment

Currently, the management of disc degeneration aims at reducing the pain and reversing disability. Disc degeneration treatment is conservative and palliative. Possible treatment options include: use of analgesic and anti-inflammatory drugs, manipulation therapies and exercise for strengthening the back muscles. Moreover, current guidelines recommend activity and discourage bed rest [KOES2001, VANT2007].

In cases where conservative treatment is not effective, and severe pain accompanied by neurological symptoms persists lumbar surgery might be undertaken. Interventional options include: spinal decompression, spinal fusion and total disc replacement. Surgery has been shown to improve patient outcomes, but is recommended only for a small number of patients since the majority of cases are improving with conservative treatment. A disadvantage of surgical procedures is that they alter spine biomechanics and could result in adjacent levels degeneration [KOES2001, HILI2004].

New treatment methods such as growth factor and gene therapy aim into repairing the intervertebral disc, instead of only dealing with its' symptoms. Growth factor therapy works by modulating the activity of intervertebral disc cells. Growth factors are proteins which when injected in the intervertebral disc they can stimulate an increase in proteoglycan synthesis. However, this therapy is only short term since growth factor levels in the disc decrease over time. On the other hand gene therapy can provide a more sustainable solution. Gene therapy involves the transfer of genetic material responsible for producing the growth factor of interest into target cells. Consequently these cells are implanted into the disc, producing growth factors and increasing the synthesis of proteoglycan. Currently, these approaches are experimental and have a lot of barriers to overcome before being used in clinical practice. A specific treatment might currently be out of reach, yet these approaches hold good potential into becoming the future treatments for disc degeneration [URBA2003, VADA2007].

2.4 Clinical Imaging of the Lumbar Spine

2.4.1 Introduction

The diagnosis of disc degeneration involves clinical investigation and imaging of the spine. A variety of imaging modalities, such as X-rays, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are used for this purpose. Additional imaging approaches include myelography and discography, combined with either X-rays or CT [PARI2007]. This section gives a brief summary on the use of these modalities for imaging disc degeneration and discusses their merits and weaknesses.

2.4.2 Plain radiographs

In plain radiographs, the intervertebral disc has no apparent optical density due to its soft tissue nature and thus cannot be directly visualized. However, the adjacent vertebral bodies provide indirect information regarding the disc tissue. Specifically, plain radiographs are used for evaluating a variety of conditions related with disc degeneration such as disc space narrowing, endplate sclerosis and the development of osteophytes on adjacent vertebrae [MODI1988, PARI2007]. Disc space narrowing was sought as a good indicator of progressing disc degeneration. However its validity is questioned from studies showing that the disc space is narrower in young than middle-aged men [LUOMA2001, ROBE1997]. Figure 2.11 presents a lateral radiograph of a lumbar spine with extensive presence of osteophytes.

Plain radiographs are very useful when assessing the vertebrae and offer a non invasive and cost effective way to image the spine. However, they have a limited role when assessing intervertebral disc changes. Radiographs can show gross morphological changes of the disc affecting surrounding structures, but are rather insensitive to early degenerative changes. Moreover, there is no firm relationship between radiographic findings and low back pain [MODI1989, PARI2007].

Figure 2.11. Lateral radiograph of a lumbar spine [WIKI2010].

2.4.3 Computed Tomography

Computed tomography provides excellent visualization of the bony structures of the spine, and it additionally depicts the intervertebral discs. In spine CT the disc's outer annulus appears slightly denser than the nucleus and inner annulus probably due to its higher collagen content. The degenerative findings depicted in CT are: disc space narrowing, bulging, herniation, calcification, the vacuum phenomenon, as well as osteophytes on adjacent vertebrae. CT provides good visualization of disc herniation as shown in Figure 2.12. Moreover, being a volumetric imaging modality, it allows visualization of the spine in any plane. Spine CT is a non invasive and fast method for imaging the spine in 3D. Although it is used less frequently nowadays due to the advances in MR imaging, it still has an important role in evaluating disc disease. CT is particularly useful for imaging patients who cannot have an MR scan such as patients with claustrophobia, implants or pacemakers.

On the down side, the major limitation of spine CT with respect to evaluating disc degeneration is the limited soft tissue contrast of this modality. CT cannot accurately depict the internal disc structure, and thus is insensitive to early degenerative changes that do not affect the disc's configuration. In addition, it utilizes ionizing radiation with an exposure dose of 6mSv per scan (equivalent to 60 chest X-rays or 2 years background radiation). Finally, with respect to low back pain, spine CT (as well as X-rays and MRI) cannot distinguish between symptomatic and incidental findings [MODI1988, TALL1998, KERT2001a, PARI2007, RADI2010].

Figure 2.12. Spine CT with an intervertebral disc herniation (arrow) [NEUR2010].

2.4.4 Myelography

Myelography is performed by injecting contrast material in the subarachnoid space using fluoroscopy for guidance. In the past myelography was obtained by fluoroscopic examination, or plain radiographs, but nowadays it is combined with CT for 3D visualization of the spine structures. Myelography provides good visualization of the spinal cord and nerve roots. It is particularly useful for the diagnosis of disc herniations causing nerve root compromise. However, it is an invasive method using ionizing radiation. In recent years the use of Myelography declines, although it is still indicated for selected patient cases and especially when MRI findings do not justify the clinical symptoms [BART1998, SHAF1999].

2.4.5 Discography

In discography, contrast medium is initially injected percutaneously to the centre of the intervertebral disc followed by a CT scan and assessment of pain provocation. Discography depicts the disc's internal structure and provides good visualization of fissures and herniations thus assisting the evaluation of disc integrity. In addition, the injected medium can reproduce pain helping to define the disc levels responsible for the patient's symptoms. Thus pain provocation is additionally assessed and discography is the only method directly relating the disc's integrity to patient's symptoms [MODI1988, PARI2007].

Looking into the disadvantages of discography, it is an invasive method which carries a risk of infection and utilizes ionizing radiation with a moderate dose exposure. In addition, it may induce long-term symptoms, especially in patients with chronic back pain. Its reliability is questioned since abnormal discographies have been reported in normal asymptomatic volunteers. More importantly, a recent matched cohort study with 10 years follow-up provided evidence that disc injections accelerate the degenerative progress. Due to its controversial nature and the advances of other imaging modalities discography has fallen out of favour [PARI2007, CARR2009a].

2.4.6. Magnetic Resonance Imaging

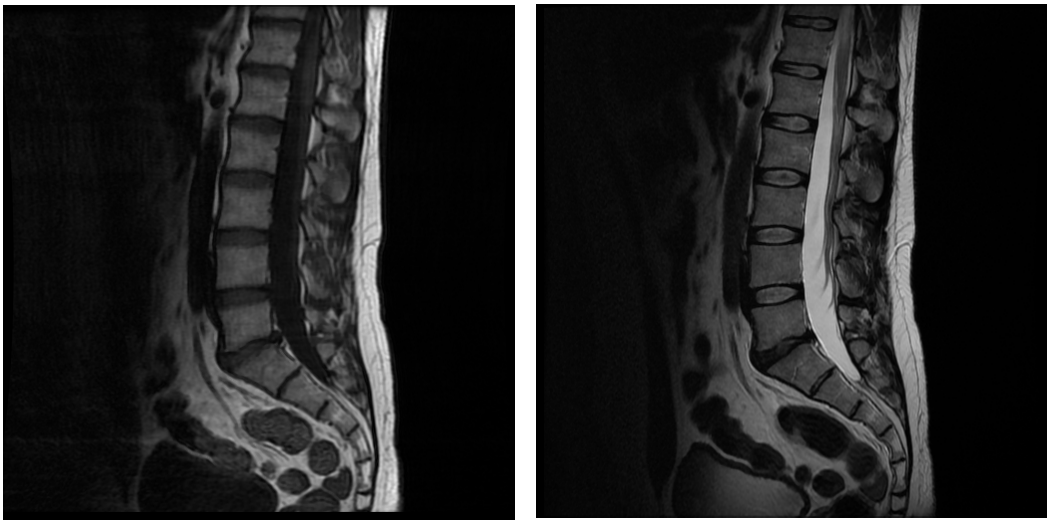
MRI is a good modality for imaging the spine and the most sensitive method for evaluating intervertebral disc degeneration. It provides high soft tissue contrast, allowing good visualization of the intervertebral disc as well as the surrounding nerves, ligaments and muscles. In addition, it is non-invasive, does not use ionizing radiation, and is capable of multiplanar imaging [MODI1988, HAUG2006, MODI2007, PARI2007].

MRI provides both morphological and biochemical information of the internal disc structure facilitating the evaluation of disc degeneration. Specifically, morphological alterations such as the loss of nucleus annulus distinction, disc space narrowing, disc fissuring, bulging and herniation are clearly depicted in

MRI. Moreover, biochemical information is encoded in MR signal intensity offering an indirect evaluation of the disc's water and proteoglycan content [KERT2001a, MODI2007].

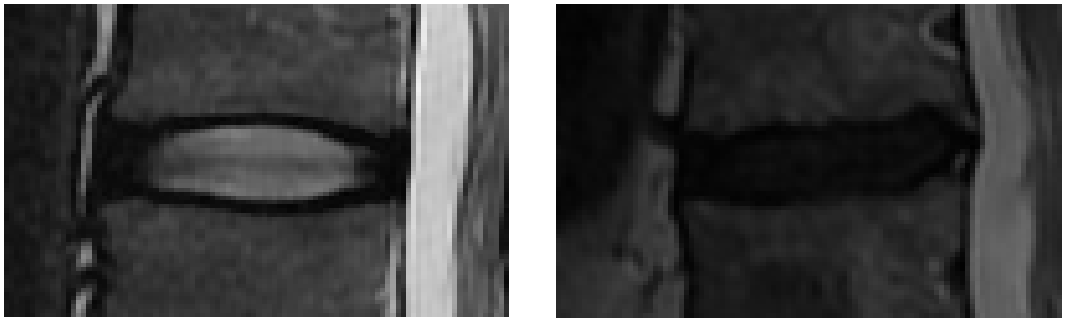
In clinical practice the standard MRI protocols for imaging the lumbar spine consist of sagittal T1 and T2 weighted images, supplemented by axial T1 or T2 weighted images at selected levels [WOLA2005]. Figure 2.13 illustrates two mid-sagittal images of the lumbar spine. T1-weighted images show low signal intensity from the cerebrospinal fluid (CSF) and cortical bone and high signal intensity from the bone marrow. They clearly depict spine anatomy and are used for inspecting the vertebral bodies (Figure 2.13a). On the other hand, T2-weighted images show high signal intensity from structures with high water content such as the CSF and the disc's nucleus. T2-weighted images are used for inspecting the intervertebral discs and assessing their structural integrity (Figure 2.13b). Moreover, they are commonly used for classifying the disc according to the severity of degeneration.

In sagittal T2-weighted images, a normal intervertebral disc appears as a bright ellipse the nucleus surrounded by a dark ring the annulus while the end plates and ligaments also appear dark as shown in Figure 2.14a. Looking in more detail into the natural course of disc degeneration, a young healthy disc is well hydrated and appears with a bright homogeneous nucleus in MRI. With aging the



(a) (b)

Figure 2.13. Examples of (a) T1-weighted and (b) T2-weighted midsagittal MR images of the lumbar spine.



(a) (b)

Figure 2.14. Examples of (a) a normal and (b) a degenerated intervertebral disc as depicted in T2 weighted MR images.

fibrous transformation of nucleus matrix creates an intranuclear cleft which appears as a dark band in T2-weighted images. Progressive degeneration involves further dehydration of the disc and encroachment of the annulus into the nucleus. A degenerated disc appears darker since dehydration results in decreased signal intensity as shown in Figure 2.14b. End plate damage is also clearly depicted in MRI, while the excellent visualization of disc herniation and bulging allows accurate diagnosis of these conditions. Moreover, MRI can show vertebral endplate changes (also known as Modic changes) which are related to fissuring and disruption of end plates, as well as degeneration of the bone marrow. This is actually the only modality depicting Modic changes which are a useful indication of low-back pain. [BRAI1998, KERT2001a, PFIR2001, KJAE2006].

Although MRI provides accurate depiction of disc morphology, its value in low back pain diagnosis is limited by its low specificity. This is due to the high prevalence of degenerative findings such as dehydration, fissuring and bulging in asymptomatic individuals. Consequently, MRI results should be evaluated in conjunction with patients' clinical symptoms, to avoid over treatment. The same is true for CT and X-ray imaging of the spine. Nevertheless there are specific MR imaging finding, such as disc extrusions and modic changes which are rare among asymptomatic individuals and highly associated to low back pain. An additional disadvantage of MRI is its relatively high cost [BODE1996, JENS1994, MODI2007].

Recent advances in MRI include magnetic resonance spectroscopy, dynamic imaging under load bearing conditions, as well as quantification of T1 and T2 relaxation times and the apparent diffusion coefficient [HAUG2006]. Disc

spectroscopy and dynamic imaging are briefly described in the remaining of this section, while disc quantification techniques are analyzed in detail in section 2.5.2.

Magnetic resonance spectroscopy can offer a biochemical evaluation of the disc tissue. It measures the amount of specific metabolites, such as lactic acid, which are related to the degenerative process. However, the application of this method in vivo is limited due to its low signal to noise ratio, and further technological developments are required before it is applied in clinical practice [HAUG2006, MAJU2006].

Dynamic imaging schemes utilize MR units which scan the patient standing upright. This allows imaging the spine under weight-bearing conditions, as well as during flexion or extension. Such schemes can detect occult herniations, which do not normally appear on recumbent position. Having the patient in a clinically relevant position could potentially help better correlate image findings with patient symptoms [JINK2006].

2.5. Evaluating Disc Degeneration form Spine MRI

2.5.1. Qualitative Grading

In clinical practice, the evaluation of disc degeneration severity relies on qualitative descriptions of image features such as the signal intensity, and shape of the intervertebral disc. A variety of qualitative schemes have been proposed aiming to assist the clinicians it the task of grading disc degeneration severity in a standardized manner [PFIR2001, BENN2005, KJAE2005]. These grading schemes differ with respect to the imaging modality used, the disc features evaluated, as well as the number of degeneration severity classes.

The task of reporting degenerative findings related to lumbar disc pathology suffers from great inconsistencies in the terminology used by the scientific community. In an effort to establish consensus, the American Society of Neuroradiology together with the North American Spine Society and the American Society of Spine Radiology published a recommendation on the nomenclature and classification of lumbar disc pathology [FARD2001]. However, as pointed out by J. Ross in his recent editorial in Radiology, the inconsistencies in terminology and the variety of classification schemes used keep complicating the communication between clinicians [ROSS2010].

A major issue in grading disc degeneration is the inter- and intra-observer reliability. This is because of the subjective nature of imaging features' description which sometimes results in great variance between the radiologists' gradings. In a review of disc degeneration grading schemes, the authors compare the different schemes with respect to their reliability, given by Cohen's Kappa or intraclass correlation coefficients, as well as with respect to their clinical feasibility and relevance. They recommend using grading systems with at least substantial interobserver agreement (Kappa >0.6) and with 3 to 5 classes of degeneration severity. Specifically for the case of magnetic resonance evaluation of lumbar disc degeneration the authors recommend the system developed by Pfirrmann [KETT2006, PFIR2001]. This system is based on the qualitative description of disc's signal intensity, height as well as the distinction between the

nucleus and annulus as described in detail in section 4.3, and has been shown by multiple studies to provide good reliability [PFIR2006, CARR2009b, ARAN2010].

Apart from suffering from interobserver variability, qualitative grading schemes have an additional drawback. They only use a limited number of degeneration severity classes (typically 3 to 5) and thus are not suitable for describing small changes in intervertebral discs [NEIM2008]. Although in current clinical practice a rough description of the degenerative status is sufficient, the need for new objecting diagnostic methods that would be sensitive to small changes is rapidly increasing. This is due to the development of emerging technologies for disc degeneration treatment, which require precise monitoring [MWAL2008].

2.5.2. Short Review on Disc Degeneration Quantification

The quantification of disc properties can provide an objective and reproducible evaluation of degeneration severity and potentially assist the physicians in disease diagnosis. Moreover, the continuous nature of quantitative measures renders them more sensitive to small changes. Thus disc quantification could be employed for tracking the evolution of disease. Moreover, it could be used for monitoring the effect of the new treatment methods for disc pathology, such as nucleus replacement, cell therapy, and growth factor therapy [VIDE2003, MWAL2008, NEIM2008]. In addition, through computerized image analysis it is possible to design more comprehensive rating systems that are not limited by the time that the physician can spend on each patient [DAM2007]. This section provides a short review of disc degeneration quantification methods.

The simplest and most widely used method for degeneration quantification relies on the measurement of disc's mean signal intensity from mid-sagittal T2-weighted MR Images [LUOM2001, BENN2005, NEIM2008]. Signal intensity reflects tissue biochemical composition and is correlated to disc's proteoglycan content [TERT1991, BENN2005]. The decrease of disc's mean signal intensity may

be the earliest degenerative change seen in MRI and is a sensitive and reliable measure of degeneration severity. In addition, it is the most highly specific MR measure associated with age [NEIM2008]. A great advantage of adjusted mean signal intensity is that it relies on conventional T2 images for quantification, and thus can be easily applied in clinical routine. On the down side, this method requires an intra-body reference for disc signal intensity adjustment. A region within the cerebrospinal fluid (CSF) adjacent to the disc level is normally used as a reference, but such a region is sometimes unavailable. This is usually due to flow artefacts or narrowing of the dural sack as demonstrated in Figure 2.15, which hinder the use of the adjusted mean signal intensity method [LUOM2001, NEIM2008].

Alternative approaches to disc degeneration quantification make use of specific imaging protocols for measuring physical properties of the disc tissue such as the relaxation times and diffusion coefficients [MWAL2008].

The T2 relaxation time, also known as transverse relaxation time can be measured using a multiecho sequence. This sequence acquires multiple images at different time points of the same signal decay (echoes). The relaxation time can be calculated by fitting the decay of the signal intensity from the different echoes to an exponential model. In the case of intervertebral discs, multiple studies have been conducted measuring the disc's T2 relaxation time both on low and high field MRI. Cadaveric studies looking at the correlation between the T2 relaxation

(a) (b)

Figure 2.15. Examples of cases where a CSF reference sample in unavailable due to (a) flow artefacts resulting in locally decreased signal in CSF and (b) narrowing of the dural sac.

time and the disc's biochemical composition show a decrease of the T2 time with decreasing water and proteoglycan content [MARI2009]. This is in agreement with in vivo studies showing a decrease in T2 with progressing degeneration [PERR2006, BLUM2010]. In addition, the T2 time has been shown to be influenced by tissue anisotropy such as the orientation of collagen fibres and is suggested that it could consequently be useful for evaluating the structural integrity of the disc tissue [PERR2006].

The T1ρ relaxation time reflects the interactions between macromolecules and bulk water, and has been exploited for the quantification of disc degeneration. A cadaveric study indicated that the T1ρ time is significantly correlated to disc's proteoglycan content [JOHA2006]. Moreover, an in vivo study on 10 volunteers demonstrated a significant correlation between the T1ρ values and disc clinical grading of degeneration severity [AUER2006]. This association was also verified by a recent study on 16 patients, where the authors concluded that T1ρ time would be a useful biomarker for assessing early degenerative changes [BLUM2010]. The T1ρ relaxation times are longer than the corresponding T2 and thus the T1ρ quantification approach has an increased dynamic range. However, its main drawback is the long acquisition time (about 30 minutes) which is near the patients' tolerance level [AUER2006].

The Apparent Diffusion Coefficient (ADC) of intervertebral discs has also been studied as a biomarker for evaluating disc degeneration. The intervertebral disc lacks vascularity and thus diffusion is its main source of nutrient, with diffusion patterns playing a crucial role in disc's integrity [ADAM2006]. In vivo measurements on healthy and degenerated discs have shown a decrease in ADCs values over the x and y axes indicating a different diffusion pattern between the two disc groups on the axial plane. Moreover, the ADC values have been shown to correlate with disc water and proteoglycan content in this direction dependant manner. It has been suggested that decreased ADC values reflect the lost integrity of the disc and thus ADC could also be used for early diagnosis of degeneration [KERT2001b, MWAL2008].

Overall, these quantitative approaches are believed to provide additional information to the conventional T2-weighted images, regarding both the

biochemical composition and structural integrity of the intervertebral discs. However, this does not come without a cost. The T2, T1ρ and ADC quantification approaches require specific image acquisition protocols which are not widely available since they do not form part of the standardized clinical protocols. An additional disadvantage of these methods is that they require additional and in some times long acquisition times, which makes their application in clinical routine more time consuming [KERT2001b, AUER2006].

2.6. Summary

This chapter described the medical background related to this study. The key points presented in this chapter are summarized in the following list:

- The intervertebral discs are pads of fibrocartilage lying in-between the vertebral bodies and acting as shock absorbers while providing flexibility to the spine
- Disc degeneration is related to dehydration of the nucleus and weakening of the annulus which affect the disc functionality
- The relation between disc degeneration and low back pain is unclear, however degenerative alterations are considered responsible for over 90% of spine surgery.
- MRI is a good modality for evaluating disc degeneration because it is non invasive and provides good soft tissue contrast without using ionizing radiation.
- MRI clearly depicts the disc morphology and internal structure. Moreover, MR signal intensity is associated to the disc's water and proteoglycan content, thus providing biochemical tissue information. MRI is highly sensitive in detecting early degenerative changes.
- In clinical practice the evaluation of disc degeneration severity relies on qualitative grading of disc signal intensity and shape from midsagittal T2-weighted MR images, which is subjective and susceptible to interobserver variability.
- Disc quantification can provide a more objective and reproducible evaluation of degeneration severity, which is particularly useful for tracking the evolution of disease and monitoring the effect of treatment.

- The standard method for disc quantification is based on measuring the disc's mean signal intensity from conventional T2-weighted images. This intensity value is normalized using a reference region from the CSF.
- Alternative approaches to disc degeneration quantification include the measurement of T1, T1ρ and T2 relaxation times, as well as the measurement of the apparent diffusion coefficient. However, these approaches require specific imaging protocols and additional image acquisition time.

Chapter 3. Clinical Data Collection and Analysis

This chapter presents the data collected in the present study and discusses on the methods used for clinical diagnosis of degeneration. Section 3.1. provides details on the data acquisition protocols, while Section 3.2. describes the grading schemes utilized by the clinical experts for qualitatively evaluating disc degeneration severity. Finally, Section 3.3. presents the results of this qualitative analysis, discusses on interobserver agreement and establishes a severity grading ground truth to serve as basis for subsequent analysis, and testing the quantification results.

3.1. Summary of data sets collected

The present study aims in the quantification of intervertebral disc degeneration from conventional T2-weighted MR images. Sections 2.4.6 and 2.5.1 of the medical background chapter provide detailed descriptions about these images and the information they provide with respect to disc degeneration. Based on this information, the motivation for choosing T2-weighted images to serve the current quantification task is summarized by the following:

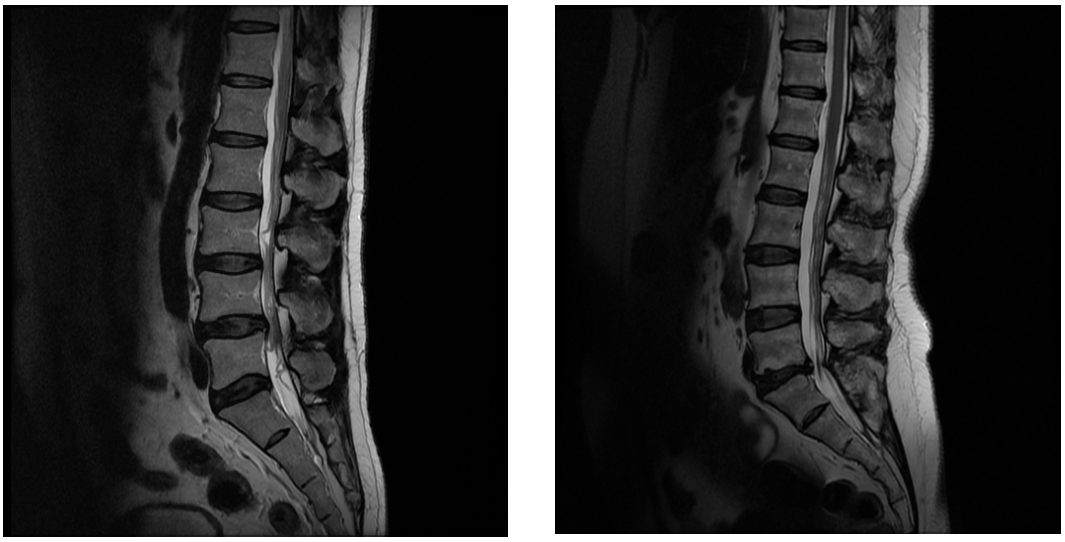
- T2-weighted images form part of the standard spine imaging protocols. Their wide availability means that the quantification method designed could be directly applicable without requiring specific imaging protocols or additional acquisition time.
- T2-weighted images form the basis of clinical diagnosis of disc degeneration and thus a qualitative severity grading ground truth can be established to be used for testing the quantification results.
- These images provide both morphological and biochemical (water and proteoglycan content) information of the disc and depict even the early stages of degeneration.

In this study lumbar spine MR scans were acquired from patients referred to the diagnostic centre with low-back pain symptoms. Specifically, two data sets were collected and analyzed independently, due to equipment and clinical time availability. The first dataset comprised **170** intervertebral discs from 34 midsagittal T2-weighted MR images acquired by a **1.5 Tesla** scanner (Signa, General Electric Medical Systems, Milwaukee, WI) using a Fast Relaxation Fast Spin Echo imaging sequence with the following acquisition parameters: TR =3100msec, TE=115msec, flip angle = 90°, field of view = 32cm, slice thickness = 4mm, slice spacing=0.8mm, matrix = 512x512, number of averages = 1.

The second dataset comprised **255** intervertebral discs from 51 midsagittal T2-weighted MR images acquired by a **3.0 Tesla scanner** (Signa, General Electric Medical Systems, Milwaukee, WI) using again a Fast Relaxation Fast Spin Echo imaging sequence with the following acquisition parameters: TR =

3000-3400ms, TE = 102-109ms, flip angle = 90°, field of view = 28 cm, slice thickness = 3 mm, slice spacing= 1 mm, matrix = 512x512, number of averages = 1. Figure 3.1 illustrates example images of the two data sets. Both datasets are acquired by multislice 2D protocols, with non-isotropic voxels. The two image acquisition protocols are very similar, but images acquired by the 3.0 Tesla scanner have slightly increased resolution in comparison to those acquired by the 1.5 Tesla one (3T voxel size =0.547x0.547x3mm^3, while 1.5T voxel size=0.625x0.625x4mm^3). The segmentation method presented in this thesis was initially developed for the 1.5T dataset and later translated to the 3T dataset. To test the validity of this translation, the accuracy of the proposed segmentation method was quantitatively assessed against golden standard segmentations for all discs of the 3T dataset (in addition to the 1.5T dataset segmentation accuracy verification). The corresponding results are summarized in Tables 4.5 and 4.6.

Although 3D data would be desirable not so much for the quantification of degeneration as for measuring with better accuracy disc morphology related pathologies such as narrowing and herniation, their use in clinical practice is limited, mostly due to the long acquisition times, which are near the patients' tolerance making the images more susceptible to motion artefacts [WOLA2005].



(a) (b)

Figure 3.1. Midsagittal T2-weighted images acquired (a) by the 1.5 Tesla scanner and (b) by the 3.0 Tesla scanner.

3.2. Clinical Grading of Disc Degeneration

The 1.5 Tesla dataset was obtained first and the images were initially reviewed by an experienced orthopaedist, who characterized the intervertebral discs as normal or degenerated according to the recommendations of the American Association of Neuroradiology [MILE1997, FARD2001]. More specifically, a disc was characterized as degenerated if it demonstrated at least one of the following imaging findings: desiccation, fibrosis, narrowing of the disc space, diffuse bulging of the annulus beyond the disc space, extensive fissuring (i.e. numerous annular tears), defects and sclerosis of the end-plates, or osteophytes at the vertebral apophyses. From a total of 170 lumbar intervertebral discs, 78 were characterized as degenerated, while the remaining 92 discs were characterized as normal. Next, the grading system shown in Table 3.1 was used for evaluating disc degeneration severity. This is a simplified version of the system introduced by Kjaer describing disc signal intensity in a qualitative manner [KJAER2005]. Two clinical experts (a radiologist and a neurosurgeon) reviewed the images and graded the discs as normal (0), mildly (1) or severely degenerated (2). This clinical grading was intended to serve as ground truth for training and testing the computer aided diagnosis system but was found to be unreliable. Detailed results of disc degeneration grading are presented in section 3.3. Cohen's Kappa with quadratic weighting was utilized for evaluating inter-observer agreement [KUND2003]. Results indicated only poor agreement between the two experts, which could be attributed to the subjectiveness of the grading scheme. This raised questions with respect to the reliability of the ground truth, and was not further exploited.

Table 3.1. The grading system used for evaluating disc degeneration severity of the 1.5 Tesla dataset

Disc Degeneration Grading	Qualitative Description
Grade 0 → normal disc	Hyperintense disc signal (higher than vb)
Grade 1 → intermediate degeneration	Intermediate signal intensity (similar to vb)
Grade 2 → severe degeneration	Hypointense disc signal (lower than vb)

vb: vertebral bodies

For the 3.0 Tesla dataset, three radiologists independently reviewed the images and qualitatively evaluated disc degeneration severity. Reviewer 1 (Aikaterini Vassiou, Assistant Professor of Anatomy, University of Thessaly) is a radiologist with 15 years of experience in reading MR images. Reviewer 2 (Marianna Vlychou, Assistant Professor or Radiology, University of Thessaly) is a specialized musculoskeletal radiologist who also has 15 years experience. Finally, Reviewer 3 (Alexandra Kazantzi, Trained Radiologist, University of Patras) has 6 years of experience in reading MRI. A different grading scheme was used this time, aiming to improve interobserver agreement. For this purpose the five grade scheme described in Table 3.2 was utilized. This scheme was introduced by Pfirrmann and gives a more explicit description of the intervertebral disc accounting for signal intensity, nucleus annulus distinction and disc height [PFIR2001]. It has been shown to provide repeatable results by two independent studies, and is also recommended by a review paper for disc degeneration grading from MRI [CARR2009b, ARAN2010, KETT2006]. Again Cohen's Kappa was used for evaluating interobserver agreement. The interpretation of inter-observer agreement results was based on the method introduced by Landis and Koch as given in Table 3.3 [LAND1977].

Table 3.2. Pfirrmann's scheme used for grading disc degeneration severity of the 3.0 Tesla data set

Disc Degeneration Grading	Qualitative Description
Grade 1 → normal disc	Hyperintense signal with homogenous bright white nucleus, clear nucleus annulus distinction, normal disc height
Grade 2 → normal disc	Hyperintense signal with visible intranuclear cleft, clear nucleus annulus distinction, normal disc height
Grade 3 → mild degeneration	Intermediate signal intensity, unclear nucleus annulus distinction, normal to slightly decreased disc height
Grade 4 → moderate degeneration	Intermediate to hypointense signal, lost nucleus annulus distinction, normal to moderately decreased disc height
Grade 5 → severe degeneration	Hypointense signal, lost nucleus annulus distinction, collapsed disc space

Table 3.3. The Landis and Koch's kappa values interpretation

κ	Interpretation
< 0	No agreement
0.0 — 0.20	Slight agreement
0.21 — 0.40	Fair agreement
0.41 — 0.60	Moderate agreement
0.61 — 0.80	Substantial agreement
0.81 — 1.00	Almost perfect agreement

The 3T clinical grading was found to be reliable and served as ground truth for evaluating disc degeneration severity quantification and also training and testing the computer aided diagnosis system. More specifically, readers' consensus (given by the majority of three gradings) was calculated, putting together the clinical knowledge and experience of all three radiologists. Radiologists' consensus was used for establishing a severity grading for each individual disc, which formed the ground truth of clinical evaluation.

3.3. Clinical Data Grading Results and Observers' Agreement

The 170 discs of the 1.5 Tesla dataset were reviewed by two clinical experts and discs were described as normal (0), mildly (1) or severely (2) degenerated. The resulting evaluations are presented as a contingency matrix in Table 3.4. The experts agreed in 50.5% of all cases, while they disagreed by 1 grade in 42.4 % and by 2 grades in 7.1%. The Kappa coefficient for interobserver agreement was found to be 0.380 indicating poor agreement between the two experts. This can be attributed to the subjective nature of the clinical grading scheme, which is based on the qualitative description of disc signal intensity as hyperintense, intermediate or hypointense. Moreover, the different clinical backgrounds of the two experts (a radiologist and a neurosurgeon) could be an additional reason for the low interobserver agreement. However, as the reliability of this grading is questioned due to the poor agreement between observers, a different grading scheme was utilized for the new 3.0 Tesla dataset as described in section 3.2. and the corresponding results are analyzed below.

Table 3.4. Contingency matrix for the 1.5 Tesla dataset clinical grading results of the 2 reviewers

Reviewer 1 / Reviewer 2	**Grade 0**	**Grade 1**	**Grade 2**
Grade 0	24	21	7
Grade 1	9	37	14
Grade 2	5	28	25

The 255 discs of the 3.0 Tesla dataset were reviewed by three radiologists and degeneration severity was evaluated utilizing Pfirrmann's grading scheme. Figure 3.2 illustrates representative examples of discs assigned to the different grades of degeneration severity. The resulting interobserver evaluations are presented in the form of contingency matrices in Tables 3.5., 3.6. and 3.7. The greatest proportion of disagreement between the three experts was observed between Grade 2 and Grade 3. The Kappa coefficients of inter-observer agreement are given in Table 3.8. The average Kappa value is 0.698, indicating substantial agreement between observers, and thus this grading scheme is considered reliable to serve as ground truth for testing the quantification results.

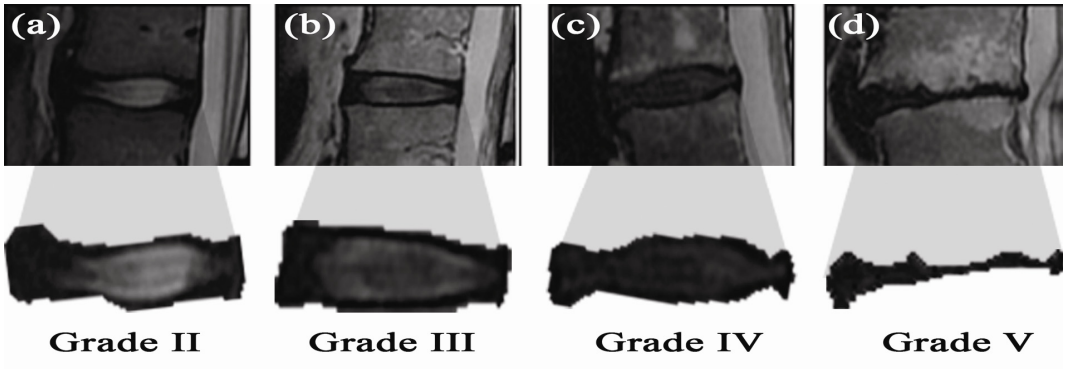


Figure 3.2. Examples of discs with different degrees of degeneration severity. From left to right, the discs become darker and the distinction between nucleus and annulus is reduced.

Table 3.5. Contingency matrix for the 3.0 Tesla dataset clinical grading results of reviewers 1 and 2

Reviewer 1 \ Reviewer 2	Grade I	Grade II	Grade III	Grade IV	Grade V
Grade I	0	0	0	0	0
Grade II	6	25	3	1	0
Grade III	0	23	66	5	0
Grade IV	0	0	12	82	2
Grade V	0	0	1	6	23

Table 3.6. Contingency matrix for the 3.0 Tesla dataset clinical grading results of reviewers 1 and 3

Reviewer 3 \ Reviewer 1	Grade I	Grade II	Grade III	Grade IV	Grade V
Grade I	0	0	0	0	0
Grade II	0	21	14	0	0
Grade III	0	46	45	3	0
Grade IV	0	9	53	29	5
Grade V	0	1	4	5	20

Table 3.7. Contingency matrix for the 3.0 Tesla dataset clinical grading results of reviewers 2 and 3

Reviewer 2 \ Reviewer 3	Grade I	Grade II	Grade III	Grade IV	Grade V
Grade I	0	4	2	0	0
Grade II	0	32	16	0	0
Grade III	0	31	48	3	0
Grade IV	0	10	47	32	5
Grade V	0	0	3	2	20

Also, this interobserver agreement value is comparable to that of the study by Carrino *et al*. (Kappa=0.66), where 4 expert radiologists graded 111 patients' discs utilizing Pfirrmann's grading scheme [CARR2009b].

The severity grading ground truth was established on the basis of radiologists' consensus, assigning to each disc the grade given by the majority of experts (Table 3.9.). This consensus grading served as basis for validating the disc degeneration quantification results, as well as training and testing the CAD system. Although, it would be desirable to have more clinical experts reviewing the data, clinical time availability was limited, yet interobserver agreement was found to be substantial providing confidence in the validity of clinical consensus grading.

As described in section 3.2 the three radiologists who reviewed the data shared different levels of experience in MR images interpretation. It is worth noting that the interobserver agreement results presented in Table 3.8. show the highest agreement between the two most experienced radiologists (Reviewers 1 and 2) and lower agreement for the less experienced (Reviewer 3).

Table 3.8. Kappa values for Interobserver Agreement

Observers Test	Kappa	95% Confidense Interval
Reviewer 1 – Reviewer 2	0.8587	[0.7156 , 1.0000]
Reviewer 1 – Reviewer 3	0.5895	[0.3596 , 0.8194]
Reviewer 2 – Reviewer 3	0.6470	[0.4339 , 0.8601]
Average Value	0.6980	

Table 3.9. Consensus grading of disc degeneration severity

Grade I	**Grade II**	**Grade III**	**Grade IV**	**Grade V**
0	52	90	88	25

3.4 Summary and Analysis of objectives

This chapter presented the datasets collected in this study together with the qualitative analysis methods used from clinically evaluating degeneration severity. Table 3.10. presents an analysis of objectives related to data collection and clinical grading summarizing this chapter.

Table 3.10. Research Objectives Summary and Discussion

Sample Collection / Clinical Grading of Disc Degeneration
1. To collect a clinical sample consisting of conventional T2-weighted MR images of the lumbar spine representative of various degrees of disc degeneration severity *Two datasets were collected, the first comprising 170 intervertebral discs from 34 images acquired by a 1.5 Tesla scanner, and the second comprising 255 discs from 51 images acquired by a 3.0 Tesla scanner. Both datasets represented various degrees of degeneration, from normal to severely degenerated discs.*
2. To select / design in collaboration with clinical experts a grading scheme for the qualitative description of image features related to disc degeneration. To have the images diagnosed by clinical experts according to the selected grading scheme. To test the clinical grading reliability through the agreement between expert readers. Also obtain a ground truth based on clinical experts consensus to serve the quantification and computer aided diagnosis tasks. *Two different schemes were selected. The first scheme (Table 4.1) was based on qualitative descriptions of disc signal intensity. It resulted in poor agreement between the clinical experts' gradings on the 1.5 T dataset (κ=0.380), and thus was not considered reliable to serve as severity grading ground truth. The second scheme (Table 4.2) was based on descriptions of disc intensity, nucleus annulus distinction and height and comprised five classes of degeneration severity. This scheme was utilized by three radiologists for grading the 255 discs of the 3.0 Tesla dataset. Interobserver*

repeatability among the three radiologists was found to be substantial (κ=0.698) in this case and radiologists' consensus was calculated to serve as ground truth for the quantification task.

Chapter 4. Intervertebral Disc Segmentation

This chapter presents the intervertebral disc segmentation methods and results. The chapter starts with an overview of segmentation methods used in medical imaging and continues with a literature review on methods focusing on intervertebral disc segmentation. Next the requirements and challenges of disc segmentation from conventional T2-weighted MR images are analyzed. In addition, the methods developed in the present study for disc segmentation are described in detail starting with the simple fuzzy clustering approach and moving on to atlas based segmentation. Finally, the disc segmentation results are presented and discussed.

4.1. Segmentation Methods Overview

Segmentation is the partitioning of an image into multiple regions that are homogeneous with respect to some characteristic, such as grey level or texture [HARA1985]. In medical images, these regions correspond to anatomical structures or structures demonstrating certain pathologies (e.g. tumours).

The task of segmentation is to provide a meaningful and easier to analyze image representation [SHAP2001]. In medical imaging applications, segmentation algorithms are utilized for accurate quantification and disease diagnosis, pathology localization, study of anatomical structures, treatment planning and computer aided surgery [PHAM2000].

Segmentation methods can be divided in two major categories with respect to the image-derived information exploited for segmentation. The first category includes methods solely based on grey level image information or methods utilizing textural information for image segmentation. The second includes model based methods, which incorporate spatial information using shape constrains in the segmentation process [REED1993, PHAM2000].

The selection of an appropriate segmentation method for a specific image type is considered to be a challenging task [PAL1993]. It is widely recognized that no single method is suitable for all images, while the performance of each method depends on the specific application [PAL1993]. In addition, different segmentation methods are often combined in order to achieve improved performance [PHAM2000].

This section provides a short overview to commonly used grey-level based segmentation methods and methods incorporating spatial information, while fuzzy clustering and atlas based segmentation approaches, which are utilized in the this study are presented in more detail.

4.1.1. Grey Level Based Segmentation Methods

Thresholding and clustering are the two major categories of grey level based segmentation methods. In thresholding, an image is divided into groups of pixels having grey level values less than a threshold, and groups of pixels with values greater or equal to it. On the other hand, clustering methods partition an image into groups of pixels (clusters) having similar grey level values [ROGO2000].

4.1.1.1. Thresholding

Thresholding methods include simple approaches based on the image histogram and more complicated ones that adaptively calculate local thresholds [PRAT2001]. The simplest thresholding approach is global thresholding. In this approach a single image threshold is calculated and used for thresholding the image according to the following transformation:

$$S(x,y) = \begin{cases} 1 & if \ f(x,y) \geq T \\ 0 & if \ f(x,y) < T \end{cases} \quad (4.1)$$

where *f(x,y)* is a pixel of the initial image, *T* is the threshold value and *S(x,y)* is the thresholded image.

Image threshold values can be calculated through various methods, with the minimization of intraclass variance proposed by Otsu being a highly popular one [OTSU1979]. Global thresholding is a computationally inexpensive method that works well when the image objects have relatively uniform grey level values and good contrast to the background. However, global thresholding fails in cases of low contrast and noisy images [ROGO2000]. To deal with such cases adaptive thresholding methods have been developed. These methods either calculate local threshold values by splitting an image to sub regions or calculate a different threshold for each pixel taking into account the grey level values in the pixel's neighbourhood. Locally adaptive thresholding can deal with background and

contrast variations but is more computationally intensive than global thresholding [ROGO2000].

4.1.1.2. Clustering

Clustering methods partition image pixels into groups through an iterative procedure. The simplest clustering method is the k-means algorithm that clusters data by iteratively computing a mean intensity for each class and segmenting the image by classifying each pixel in the class with the closest mean grey-level value [COLE1979, PHAM2000].

A different approach to segmentation by data clustering is fuzzy clustering. This approach is based on the fuzzy sets theory, introduced as a way to represent vagueness. In contrast to conventional sets, whose members satisfy precise properties, fuzzy sets correspond to functions measuring the degree to which the sets' members satisfy imprecisely defined properties [BEZD1993].

Fuzzy clustering techniques are a common approach to soft image segmentation. Soft segmentation allows uncertainty in the location of object boundaries and permits regions to overlap. In contrast to hard segmentation where the characteristic function of an image takes binary values showing whether a pixel is either inside (1) or outside (0) a given object/region, in soft segmentation the characteristic function takes continuous values in the range [0-1]. Thus, soft segmentation techniques are particularly useful in addressing partial volume effects [PHAM2000], where there are pixels containing more than one tissue types, and thus partly belonging to more than one tissue classes.

Fuzzy connectedness goes a step further from fuzzy clustering as it is not a purely grey level based technique. The underlying assumption here is that regions which are close to each other and have similar grey-levels are more likely to belong to the same anatomical structure (object). This method calculates the strength of connectedness for each pair of pixels within the image. This strength is a measure putting together the pixels' proximity in space (fuzzy adjacency) and their grey level values' similarity (fuzzy affinity). Consequently, a fuzzy connectedness map is created encoding the strength of connectedness between a given image pixel and every other pixel within the image to be segmented.

Finally, object segmentation can be performed by thresholding the fuzzy connectedness map [UDUP1996].

4.1.2. Model based segmentation methods

In many segmentation applications grey level information is not sufficient to distinguish between various structures. This is often the case with low contrast images, or when there are overlapping grey-level values between neighbouring structures. Different anatomical structures often have similar grey level values and only differ from one another with respect to their location. In these cases spatial information needs to be incorporated in the segmentation process [ROHF2005]. Model based methods such as, active contours, statistical shape models and atlas based approaches put together grey level, edge and shape information for segmentation purposes

4.1.2.1. Active contours

"Active contours" or "Snakes" are curves on the image plane which deform under the influence of internal forces depending on the curve parameters and external forces computed from the image data [MCIN1996, XU1998]. They were introduced in 1988 [KASS1988] and have since been successfully used in many medical imaging applications [MCIN2000a]. In addition, various methods have been proposed for improving active contours by introducing topological flexibility [MCIN2000b].

4.1.2.2. Statistical shape models

Statistical shape models (also known as active shape models) are a highly popular segmentation method for both medical imaging and computer vision. A statistical shape model is constructed by analyzing the positions of a set of landmark points annotated by a human expert in a set of training images. More specifically, the shape variation over the training set is analyzed and a model that can mimic this variation is built. To segment a new image, the model parameters

are adjusted in order to match a model instance as closely to the image as possible [COOT1992, COOT2000]. Statistical shape models have been used in a variety of medical applications, and have proven to be particularly useful in segmenting regions with a well defined shape, such as bones and organs [COOT2000].

4.1.2.3. Atlas-based segmentation

Atlas-based methods utilize prior anatomical knowledge in the segmentation process. An atlas is a description of shapes, locations and/or neighbourhood relationships of anatomical structures within the image. Atlas-based segmentation (or segmentation by registration) is achieved by mapping the atlas to the image to be segmented in an anatomically correct way [ROHF2005].

Atlas-based segmentation methods utilize various atlas selection strategies and registration algorithms. The simplest selection strategy is based on a simple anatomical atlas constructed by manually segmenting a single image and labelling the different structures. Image segmentation and labelling is achieved by registering this atlas to the image to be segmented and transferring its pixel labels to the corresponding image pixels.

A more advanced method involves the use of multiple individual atlases originating from different images corresponding to multiple subjects [ROHF2005]. Each atlas is registered to the image to be segmented, and similarity criteria are used to find the atlas that best fits the image and thus, is likely to produce the best segmentation. An important advantage of this method is that by using atlases from different subjects it takes into account the anatomical variation within the population. However this method is computationally expensive since it requires multiple registration procedures.

A different approach is exploiting a probabilistic atlas. Such an atlas is usually constructed on the basis of multiple manually segmented image templates. First, one template is selected as reference and the remaining templates are registered to the reference. The average template is constructed by superimposing all registered templates [SLUI2005]. This average template is called the "probabilistic atlas" because its pixel values indicate the probability of a given pixel belonging to a given object within an image [PARK2003]. When used for

segmentation, the probabilistic atlas has to be registered to the image to be segmented and finally, the segmented image is obtained by thresholding the outcome. The main disadvantage of this method is the requirement of multiple manually segmented image templates for constructing the atlas. However, thanks to these templates this method manages to capture normal anatomical variation, which can help improve segmentation accuracy [SLUI2005]. In addition, it is not computationally expensive since it only requires a single registration per segmentation.

In conclusion, atlas-based methods are particularly useful in cases where grey level information is not sufficient for segmenting an image, since they incorporate prior anatomical knowledge in terms of shape in the segmentation process by means of the atlas model. In the present study, a probabilistic disc atlas is utilized to introduce spatial information in the segmentation process aiming to deal with overlapping grey-level values between neighbouring tissue classes.

4.2. Literature Review on Lumbar Disc Segmentation Methods

Lumbar discs segmentation is the first step towards the quantification of disc features for diagnostic or surgical purposes. Up to date, studies dealing with this task rely on manual segmentation which is a tedious and time-consuming process [TSAI2002, NIEM2008]. Moreover, manual segmentation depends on user's experience and lacks repeatability. On the other hand, automated segmentation methods provide repeatable results, while alleviating the need and reducing the time of manual interaction. Segmentation methods are becoming crucial elements of medical image analysis particularly in the fields of computer aided diagnosis, image guided surgery and radiotherapy planning [PHAM2000].

This section presents a short literature review of methods designed for lumbar disc segmentation from MR images, in chronological order.

In a 1997 study, Roberts *et al.* present a fully automatic lumbar disc segmentation algorithm based on watershed techniques [ROBE1997]. This algorithm combined information from sagittal Proton Density (PD) weighted and T2 weighted images. Proton density images can serve the delineation of intervertebral discs due to the signal voids on the vertebral bodies' endplates. Moreover, the T2 weighted images give high signal intensity for the discs, which can be used for guiding the segmentation algorithm. In the prepossessing step of this segmentation algorithm the PD and T2 images were filtered by means of the top hat algorithm and combined by subtraction. Next the watershed algorithm was used for delineating the local maxima within the image and a set of heuristic rules was employed for selecting the ones that correspond to lumbar discs. The method was tested on 40 patients' datasets. The authors report that the proposed segmentation method works well on non-degenerated discs, but not so good on degenerated ones as it requires high signal intensity for identifying the discs. However, this study does not include any evaluation of segmentation accuracy.

Shi *et al.* present a study on the automatic detection and segmentation of intervertebral discs from sagittal whole spine MR images [SHI2007]. The main focus of this study is the localization of discs, while the segmentation process is only briefly described. The proposed segmentation algorithm comprises a

preprocessing step, a spinal cord extraction process, and a scheme for intervertebral discs' localization. Image preprocessing is preformed using a median filter for noise reduction. Next the edge image is calculated using the Sobel operator and the line of the spinal cord is detected by means of the Hough transform. Finally, the intervertebral discs are localized using a self adaptive rectangle window moving along the spinal cord curve. Intervertebral discs segmentation is performed using edge detection within the localization window. The algorithm is tested on 50 patient whole body MR images acquired by a two station sequence with a combined field of view of 70x35cm2. The method achieved a disc detection rate of 96%, while no evaluation of disc segmentation accuracy is reported.

Chevrefils *et al*. report on the segmentation of intervertebral discs for scoliotic patients [CHEV2007, CHEV2009]. This is a challenging task because of the complex spinal curves of patients with scoliosis. The group has published two studies on this task, which will be summarized in the remaining of this paragraph. The first study presents a segmentation method utilizing watershed techniques. In this approach, first morphological operations together with thresholding techniques are utilized for detecting local maxima, which are assumed to correspond to the intervertebral discs. Next, these regions are segmented using the watershed transform. The method is tested on magnetic resonance images acquired using the Multi Echo Data Image Combination sequence, which obtains volumetric data (voxel size = $1mm^3$). The authors report that this segmentation method provides fast and accurate disc segmentation, but it suffers from over-segmentation problems, which are augmented in images of scoliotic spines [CHEV2007]. To deal with the problem of over-segmentation, the authors designed a classification scheme to find which of the segmented regions correspond to intervertebral discs. This classification scheme is presented in detail in their second study [CHEV2009]. This study extracts a set of textural features from each closed region in order to characterize the underlying tissue properties and uses a k-Nearest Neighbour classifier for distinguishing between intervertebral discs and background regions. The classification scheme was tested on 505 images of three patients (multiple slices per patient) and achieved disc detection accuracies in the

range of 80%-88%. The classification scheme is reported to reduce the over-segmentation error. Again this study does not provide any evaluation of segmentation accuracy for the proposed method, while the images provided indicate that only disc nucleus is actually segmented and not the entire disc structure.

It should be noticed, that although these methods work towards the segmentation of intervertebral discs, they focus more on the localization/ identification of discs than the actual segmentation process. On the other hand, a study conducted by Seifert and Wachter reports more extensively on the segmentation process [SEIF2006, SEIF2009]. This study presents a method for segmenting the intervertebral discs, spinal cord and trachea from 3D MRI data of the cervical spine, and providing a framework for reconstruction of the spine. The study is published in two papers, the first reporting on the segmentation algorithm [SEIF2006], while the second additionally presenting a validation of segmentation results [SEIF2009]. The algorithm starts with a disc localization step and next active contours are employed for identifying the boundaries between the trachea, spinal column and spinal cord. Finally, a combination of active shape models (where the disc shape is modelled as an ellipse) and fuzzy connectedness methods is used for segmenting the structures of interest. The method was tested on a dataset of 9 patients comprising both T1 and T2 weighted images of the cervical spine acquired using 3D imaging protocols with a slice thickness of 1.16mm. The authors report that the algorithm demonstrated robust performance. The segmentation accuracy is evaluated on a single midsagittal image with the ground truth provided through manual segmentation by an expert.

This section presented a literature review on lumbar disc segmentation methods from MRI. Out of the four studies working on this task, only the last one reports on the evaluation of segmentation accuracy, but this is only for a single spine image. Moreover, none of these studies reports on the segmentation of degenerated discs, which to the best of our knowledge remains an open issue.

4.3 Lumbar disc segmentation from clinical T2-weighted MR Images: Requirements and challenges

The present study focuses on segmenting intervertebral discs from MR images. This segmentation is intended to serve our quantification task, and the specific requirements are described in the following list.

- Midsagittal T2-weighted images need to be segmented, and thus a 2D segmentation method is required.
- The method should be able to segment both normal and degenerated discs
- Segmentation accuracy is highly important, and the target is a mean segmentation error of 1-2mm

The requirement for segmenting the discs from T2-weighted MR images is in accordance with clinical standards since disc degeneration diagnosis relies on the evaluation of these images [PFIR2001, LUOM2001, NIEM2008]. According to these studies the midsagittal slice is utilized for grading degeneration severity and thus a 2D segmentation method is required.

Moreover, since this project aims at disc quantification, the segmentation method should work well with discs at various degenerative stages. Thus, a single method that can effectively segment both normal and degenerated discs while no prior knowledge of the disc's state is required.

Finally, in order for the segmentation method to serve well the diagnostic task under question it should provide adequate accuracy. Specifically, as discussed in the introduction of this thesis, a collaborating orthopaedic surgeon indicated that a mean error of 1-2 mm would be acceptable for both diagnostic and surgical purposes [personal communication with Elias Panagiotopoulos, Professor of Orthopaedic Surgery, University of Patras, Greece].

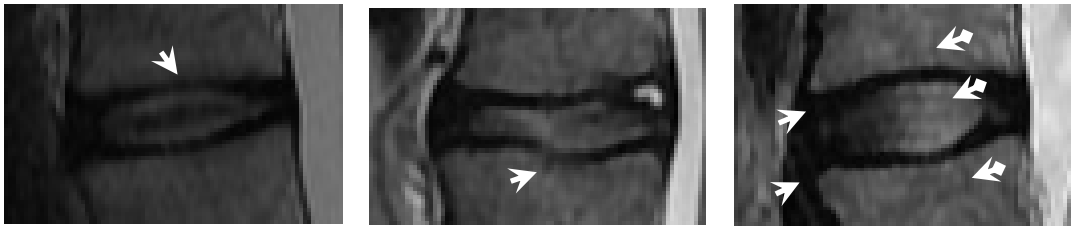
The disc segmentation task is challenging because clinical images suffer from:

- Partial volume effects
- Grey- level overlapping between neighbouring tissues

Clinical MR images are acquired using multislice 2D protocols with relatively large slice thickness (3-4mm). These protocols obtain data with adequate signal to noise ratio and short acquisition times. However, this results in images of relatively low resolution and with non-isotropic voxels (a typical voxel size is $0.625 \times 0.625 \times 4$ mm^3). The voxel size and shape plays an important role in the segmentation task due to the partial volume effect. The term partial volume effect is used for describing the presence of more than one kinds of tissue within the same voxel. In images with large non-isotropic voxels partial volume effects are augmented resulting in blurring of intensity across object boundaries (Figure 4.1a, b) [PHAM2000].

In addition, in spine MRI there is no unique correspondence of grey level ranges to different tissue classes. In other words, these images suffer from overlapping grey level values between neighbouring tissues [PARV2006]. More specifically, the annulus fibrosus has grey level values similar to those of the anterior and posterior ligaments, while the grey level values of the disc's nucleus lay within the range of the vertebral bodies' grey level values (Figure 4.1c). Consequently, simple grey-level based techniques are not adequate for segmenting the discs, since grey-level overlapping would result in boundary leakage towards neighbouring tissues.

The next section describes the methods developed in the present study for segmenting the intervertebral discs and dealing with the abovementioned challenges.



(a) (b) (c)

Figure 4.1. Examples of intervertebral discs demonstrating (a and b) partial volume effects causing blurring of intensity across boundaries, shown by white solid arrows (c) overlapping grey levels between neighbouring tissue classes: solid white arrows in point to the annulus fibrosus and anterior longitudinal ligament which have similar grey level values, while diamond tailed white arrows point to the nucleus pulposus and vertebral bodies.

4.4. Methods: What Do we Propose for disc segmentation?

4.4.1. Proposed Methods Overview

In this study, atlas based segmentation together with fuzzy clustering techniques are exploited for segmenting intervertebral discs from midsagittal T2-weighted MRI. The proposed segmentation methods are developed in 2D in accordance to clinical requirements for degeneration evaluation, and tested on midsagittal T2-weighted MR images. However, a pseudo-3D approach (applied on interpolated multislice 2D-data) is additionally presented to demonstrate the feasibility of expanding the current methods to 3D.

As described in the previous section, a single method able to segment both normal and degenerated discs is required. Disc appearance in MRI is characterized by increased variability both in terms of signal intensity and boundary shape, as shown in Figure 4.2. Normal discs have a bright nucleus, whereas in degenerated discs the nucleus is darker and merges with the annulus. However, in all cases the annulus always appears as a low intensity ring surrounding the disc in T2-weighted MRI and this can be exploited for segmenting the disc structure. More specifically, by detecting the external boundary of the annulus fibrosus, which is also the boundary of the intervertebral disc, accurate disc segmentation can be achieved independently of the degenerative stage of the disc.

A variety of segmentation methods were tested in the present study, starting with clustering techniques and moving on to fuzzy connectedness and atlas based segmentation. A constructive approach is followed here demonstrating how each method is built upon the previous one aiming to improve segmentation accuracy.

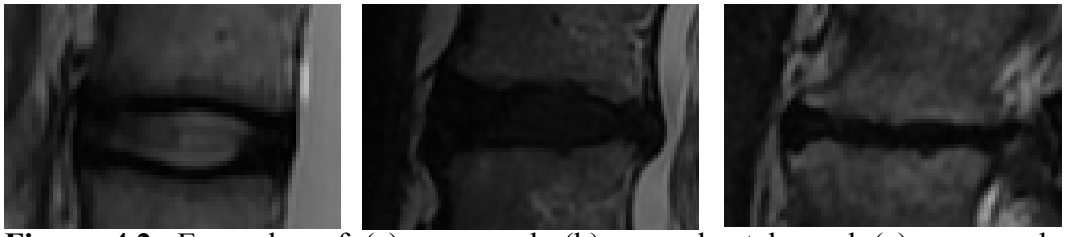


Figure 4.2. Examples of (a) a normal, (b) a moderately and (c) a severely degenerated disc.

Initially, k-means clustering was employed to distinguish between bone, disc and cerebrospinal fluid, while Fuzzy c-means (FCM) was utilized aiming to compensate for blurred edges due to partial volume effects. However, clustering methods suffered from border leakage towards surrounding structures especially across the anterior disc boundary.

In an effort to reduce border leakage, fuzzy connectedness was tested to additionally account for spatial relationships between pixels. However, this proved problematic because in many cases the disc annulus and surrounding structures near the anterior disc boundary (the anterior longitudinal ligament and neighbouring vessels) are in close proximity and have very similar grey-level values (see Figure 4.3). Consequently, the strength of connectedness between the disc and surrounding tissues in that area was no different than that among the disc pixels. As a result fuzzy connectedness was no more successful in disc delineation than fuzzy c-means clustering, while unnecessarily complicating the segmentation task.

An alternative approach towards reducing border leakage was undertaken next. Atlas-based methods were tested aiming to introduce prior anatomical knowledge in the segmentation process. Specifically, a probabilistic disc atlas acting as shape prior was utilized combined with fuzzy clustering in a method we called Atlas-FCM. Next, smoothness constrains were introduced by the robust-fuzzy clustering approach combined together with the atlas to help improve robustness to noise (Atlas-RFCM). Finally, inhomogeneity correction was applied as a pre-processing step in MR images to correct for the bias field, which can adversely affect the performance of automated segmentation [VOVK2007].

The present study focuses on disc segmentation accuracy since the segmentation method needs to serve well the task of disc degeneration quantification. For this reason, segmentation accuracy here is quantitatively evaluated against manually obtained gold standard.

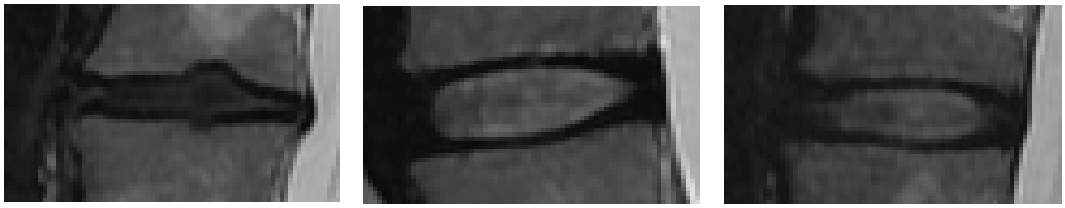


Figure 4.3. Disc examples the anterior disc border is unclear.

This is in contrast to previous disc segmentation studies which focused more on discs' automatic localization rather than segmentation performance [ROBE1997, SHI2007, CHEV2007]. However, as disc localization is not a requirement in the present study, the methods developed here work for a user defined disc location.

The FCM, Atlas-FCM and Atlas-RFCM segmentation methods are presented in detail in the remaining of this chapter. These methods were initially tested on the 1.5 Tesla dataset while the corresponding results were presented in the IEEE Medical Imaging Conference 2008, and later published in the IEEE Transactions on Biomedical Engineering Journal. Moreover, the optimal segmentation method was applied on the 3.0 Tesla dataset, where the inhomogeneity correction approaches were also tested. Finally, the pseudo-3D approach was tested on a subset of the same sample.

4.4.2. The Fuzzy Clustering Method (FCM)

The first segmentation method presented here is based on the Fuzzy c-Means algorithm (FCM). The FCM is commonly used for medical image segmentation [BEZD1993, PHAM2000]. It was selected to serve the current segmentation task, due to its ability to account for partial volume effects. Section 4.4.2.1 provides a short description of the FCM algorithm, while section 4.4.2.2 describes how this method was applied on intervertebral disc segmentation.

4.4.2.1 Fuzzy c-Means

The FCM method assigns tissue class membership values in each pixel within the image [BEZD1984]. The membership values indicate the amount of each tissue type within the voxel or equivalently they indicate the probability that the given voxel could belong to the specified tissue class [PARV2006]. This algorithm is designed to minimize an overall objective function F with respect to the class centres and membership functions. Equation 4.1 provides the overall objective function, while Equation 4.2 provides the function which upgrades

cluster centre values in every iteration and Equation 4.3. gives the function for calculating membership values.

$$F = \sum_{i=1}^{w}\sum_{k=1}^{s}(\mu_{ik})^m \left(\left\|x_i - v_k\right\|\right)^2 \tag{4.1}$$

$$v_k = \sum_{i=1}^{w}(\mu_{ik})^m x_i \Big/ \sum_{i=1}^{w}(\mu_{ik})^m \tag{4.2}$$

$$\mu_{ij} = \left[\sum_{k=1}^{s}\frac{\left\|x_i - v_j\right\|^{\frac{2}{m-1}}}{\left\|x_i - v_k\right\|}\right]^{-1} \tag{4.3}$$

where, x_i is the intensity of pixel i, s is the number of clusters or classes, w is number of pixels within the image and the parameter m determines the amount of fuzziness of the resulting classification. In addition, μ_{ik} is the membership values of a pixel i to each class k, while v_k is the value of the cluster centres of class k.

4.4.2.2 Using FCM for Intervertebral Disc Segmentation

Figure 4.4. outlines the application of FCM for disc segmentation. Three classes were assumed for this purpose, corresponding to the three tissue types within the image: bone, disc and cerebrospinal fluid. In the beginning, the user is asked to manually define the approximate disc centre within the image for disc localization. Next, a fixed dimensions ROI, centred on the user defined disc centre is cropped from the image and given as input to the FCM algorithm. The dimensions of the disc-ROI have been chosen to be larger than the largest disc within the sample (150×100pixels^2). The FCM algorithm calculates three class membership values (μ_{bone}, μ_{disc}, μ_{CSF}) for each pixel within the disc-ROI and constructs the corresponding membership matrices (M_{bone}, M_{disc}, M_{CSF}) shown in Figure 4.4. The segmented disc template is obtained by thresholding the M_{disc}

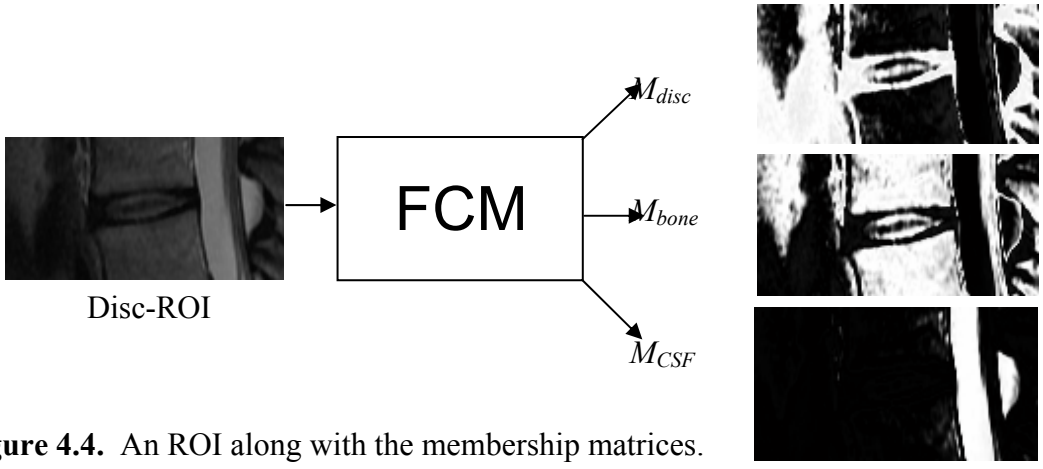


Figure 4.4. An ROI along with the membership matrices.

matrix, with a threshold value automatically calculated through minimization of the intraclass variance according to the method introduced by Otsu [OTSU1979]. Finally, a flood-fill operation is employed for filling holes within the disc template and morphological opening (using a circular object with 2 pixels radius) is utilized to smooth the resulting boundary.

4.4.3. Atlas Based Segmentation

Two atlas based methods were developed aiming to improve segmentation accuracy by introducing prior anatomical knowledge in terms of disc shape in the segmentation process. A probabilistic disc atlas was exploited to compensate for boundary leakage resulting from FCM segmentation due to grey-level overlapping between neighbouring tissues. The first atlas based method was named Atlas-FCM as it combined the probabilistic atlas with the output of the FCM algorithm. The second method was called Atlas-RFCM (where RFCM stands for Robust-FCM) and it exploited a modified fuzzy clustering approach to improve the segmentation robustness to noise. Section 4.4.3.1. describes the atlas design while sections 4.4.3.2. and 4.4.3.3. provide the details of the Atlas-FCM and Atlas RFCM methods' design.

4.4.3.1 Building a Probabilistic Disc Atlas

A probabilistic atlas was designed to serve as a shape prior aiming to assist the segmentation process. The choice of a probabilistic atlas, instead of a single reference model was made in an effort to take into account the normal variation in tissue shape [SLUI2005]. The atlas was built on the basis of 50 normal intervertebral discs superimposed through a landmark-based rigid registration process.

First, one disc image was selected to serve as reference. To avoid bias towards a specific disc type the orthopaedist was asked to select as reference the image that best represented the population [PARK2003]. Next, the remaining 49 discs were registered to the reference through the registration process described in section 4.4.3.2. Finally, the registered intervertebral discs were manually

delineated and the resulting binary templates were superimposed to construct the probabilistic disc atlas, which is shown as part of the segmentation method in Figure 4.5.

A limitation of the atlas construction is that only normal discs were selected to serve as templates. Thus shape variation related to degenerative alterations was not modelled in the atlas, since such an effort would require a very large database to model all types of degenerative alterations.

4.4.3.2 Rigid Landmark Based Registration

A landmark based registration process was employed for the construction of the probabilistic atlas, and for wrapping the atlas to the MR image during segmentation. In this process, the leftmost and rightmost disc points along the disc's middle axis are interactively selected by the user in the image to be registered, to serve as landmarks for disc registration. Next, an automatic landmark refinement process is utilized in order to increase segmentation tolerance to erroneous positioning of landmarks within a small scale. In this process the normalized cross-correlation is calculated for a rectangular region of interest surrounding the initial landmark and the corresponding one in the reference image. The refined landmark position in the image to be registered corresponded to the position that maximized cross-correlation [KOVE2000]. Finally, a rigid transformation (Equations 4.4, 4.5) involving translation, rotation and scaling is used for registering the disc image to the reference image [BROW1992].

$$\overline{p} = \bar{t} + s \cdot R \cdot \overline{p}_0 \tag{4.4}$$

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{4.5}$$

where $\overline{p}_0$ is the initial coordinate vector, $\overline{p}$ is the transformed coordinate vector, $\bar{t}$ is the translation vector, s is the scaling factor and R is the rotation matrix, with θ being the rotation angle.

The rigid registration method was chosen here on the merit of simplicity and fast processing time. An elastic registration method was also tested in cascade

to rigid registration as part of the segmentation process. Elastic deformation was performed by assuming a locally affine and globally smooth transformation according to the method developed by Periaswamy and using their publicly available algorithm [PERI2003]. Incorporating this elastic registration approach in the current study, offered a very small but not statistically significant increase in segmentation accuracy (please see IEEE TBME paper). However, as the increase in processing time when using elastic registration was disproportional to the benefit, this method was considered impractical for clinical application.

4.4.3.3. The Atlas-FCM segmentation approach

The Atlas-FCM approach is an atlas based segmentation method which combines the probabilistic disc atlas with fuzzy clustering. A flowchart of this method is outlined in Figure 4.5. First the user is asked to define two landmarks for wrapping the atlas to the MR image according to the process described in section 4.4.3.2. Next, a ROI surrounding the disc is automatically defined on the basis of these landmarks and the FCM algorithm is used to calculate the membership matrices for this disc-ROI. The registered probabilistic disc atlas is multiplied pixel by pixel with the M_{disc} membership matrix of the disc-ROI in order to calculate the combined probability matrix. The combined (joint)

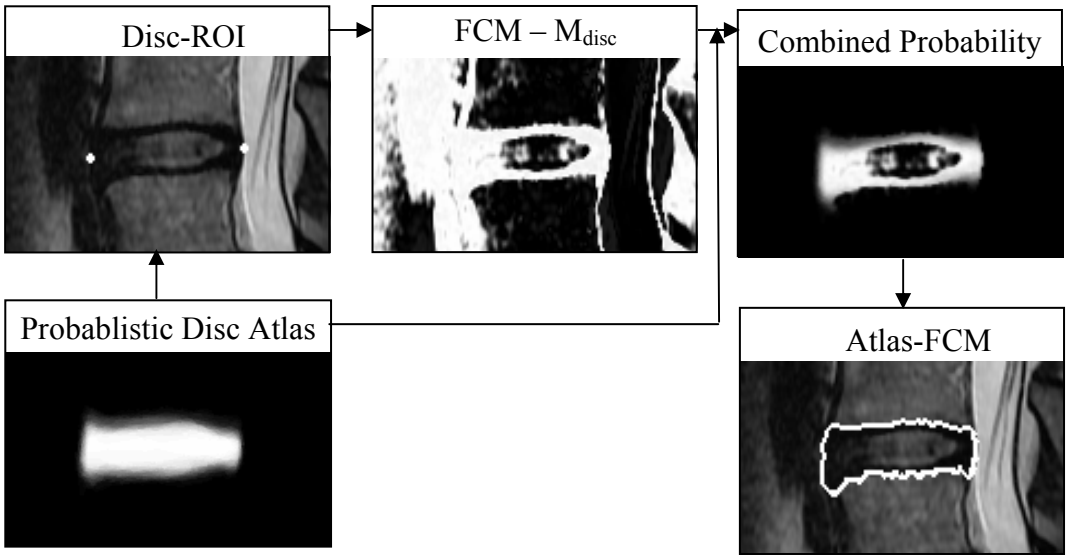


Figure 4.5. An outline of the Atlas-FCM segmentation method. Bright pixels in the M_{disc} and combined probability matrices indicate high probability, while dark pixels correspond to low probability.

probability matrix is given by the multiplication of the disc atlas (representing spatial probability) and the M_{disc} matrix corresponding to tissue class membership probability. Finally, the segmented disc is obtained by thresholding the combined probability matrix, using again the method proposed by Otsu and the same morphological operations as described in the FCM method [OTSU1979].

4.4.3.4. Atlas combined with Robust Fuzzy Clustering (Atlas-RFCM)

This method is similar to the Atlas-FCM approach, the only difference being that Bezdek's FCM algorithm is replaced by Pham's RFCM [PHAM1999, PHAM2001]. RFCM stands for Robust Fuzzy C-Means. It is an extension of the FCM method, which incorporates a penalty term in the objective function forcing the membership values to be similar to neighbouring values. In this way, the RFCM produces spatially smooth membership functions and achieves robustness to image noise. The RFCM objective function is given by equation 4.6. The strength of the penalty term is determined using a cross-validation criterion to calculate parameter β [PHAM2001]. A flowchart of the Atlas-RFCM method is provided in Figure 4.6.

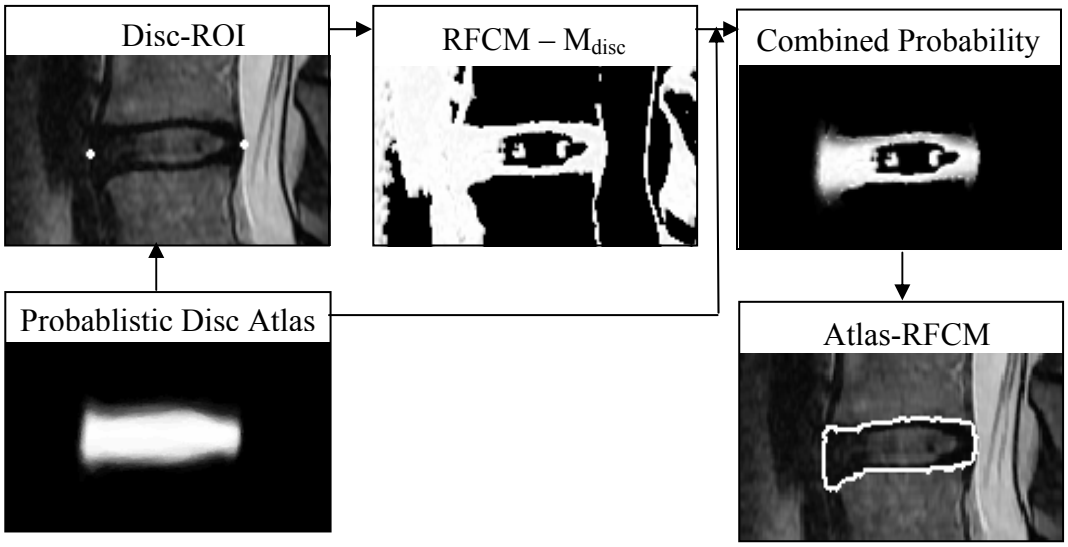


Figure 4.6. An outline of the Atlas-RFCM segmentation method. Again bright pixels in the M_{disc} and combined probability matrices indicate high probability of the pixel belonging to the disc tissue, while dark pixels correspond to low probability.

$$F_{RFCM} = \sum_{i=1}^{w}\sum_{k=1}^{s}(\mu_{ik})^m(\|x_i - v_k\|)^2 + \frac{\beta}{2}\sum_{i=1}^{w}\sum_{k=1}^{s}(\mu_{ik})^m \sum_{n\in N_i}\sum_{p\in P_k}\mu_{np}^m \quad (4.6)$$

where, again x_i is the intensity of pixel i, s is the number of clusters or classes, w is number of pixels within the image and the parameter m determines the amount of fuzziness of the resulting classification. In addition, μ_{ik} is the membership values of a pixel i to each class k, while v_k is the value of the cluster centres of class k. Finally, β is the parameter controlling the smoothness of membership functions, N_i is the set of neighbours of pixel i and P_k=[{1…s} excluding k].

4.4.4. Pseudo-3D Disc Segmentation

Both data samples used in this study are acquired by multislice 2D protocols. More specifically, for the 3.0 Tesla sample, which is of higher resolution each dataset consists of 8-13 slices with 3mm slice thickness each and 1mm spacing. The pixel size for each image is 0.5469x0.5469 mm, which means that the in-plane resolution is approximately 7 times higher than the between-slices resolution (i.e. equivalent voxel size 1x1x7). This anisotropy is the main limitation for applying 3D segmentation methods in this data sample.

However, in order to test the applicability of atlas based segmentation to segment 3D data, a pseudo-3D approach was developed by expanding the 2D segmentation method and testing it on interpolated data. More specifically, linear data interpolation was employed for producing isotropic 3D disc volumes from the 2D slices. Interpolated data cannot really make up for the lack of true 3D data, but can serve as a proof of concept by accommodating our pseudo-3D segmentation approach.

The pseudo-3D segmentation approach combines fuzzy clustering applied on the interpolated disc volume, with spatial information propagation between neighbouring slices. In this approach, the midsagittal slice is first segmented by combining fuzzy membership information with the 2-D probabilistic Atlas. Next, the probabilistic atlas is modified by addition of the segmented midsagittal disc template. The result is a “biased-atlas” which gives higher probability weighting to pixels belonging to the midsagittal disc template. The “biased-atlas” is used for

segmenting the neighbouring slices. This process is repeated until the entire disc volume is segmented as shown in Figure 4.7.

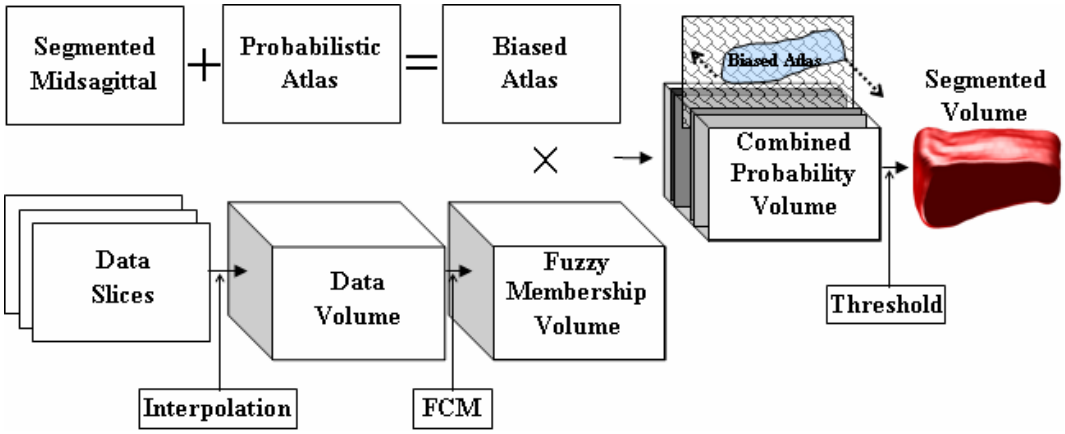


Figure 4.7. Layout of the Pseudo-3D approach.

4.4.5. Inhomogeneity Correction

Intensity inhomogeneity is a smooth intensity variation in MR images caused by poor RF coil uniformity, eddy currents and patient anatomy [SLED1998]. In clinical MR scanners, this intensity variation is often subtle enough that it is difficult to be detected by visual inspection, and has little impact on visual image interpretation and diagnosis. However, it can adversely affect the performance of automated image analysis techniques and particularly segmentation techniques that assume homogeneity of intensity within each tissue class [VOVK2007].

A variety of algorithms have been designed for intensity inhomogeneity correction, while different approaches have been shown to perform better on different types of anatomical images and acquisition protocols [VOVK2007]. It has been suggested that our study might benefit from inhomogeneity correction, which could help improving the algorithms' segmentation accuracy. Thus, three well-established methods were tested and quantitatively compared in order to identify the most suitable one for correcting inhomogeneity in lumbar spine MRI in order to facilitate the lumbar disc segmentation process. The three methods tested are:

- Entropy Minimization, which assumes that intensity inhomogeneity introduces additional information to the image that can be removed by constrained entropy minimization [LIKA2000].
- N3 (Nonparametric Nonuniform intensity Normalization), which assumes a smooth multiplicative model for the intensity inhomogeneity field and corrects the image by maximizing its high frequency content [SLED1998].
- Homeomorphic Filtering, which removes low frequency information corresponding to the smoothly varying inhomogeneity field [BRIN1998].

The three methods were quantitatively tested on a randomly selected set of five T2-weighted mid-sagittal spine images from the 3.0 Tesla dataset (25 intervertebral discs). Direct evaluation was performed by measuring intensity variation, while indirect evaluation was performed by investigating the effect of inhomogeneity correction on intervertebral discs segmentation accuracy. The most suitable method was identified and used as a pre-processing step for intervertebral disc segmentation.

4.4.5.1. Intensity Variation Based Evaluation of inhomogeneity Correction

Intensity inhomogeneity correction is expected to reduce the intensity distribution variance for a tissue of interest within the image. In addition, intensity scaling, due to inhomogeneity correction can be expressed by the change in mean tissue intensity. The Coefficient of Variation (CV) combines these measures and can be used to quantitatively evaluate the effect of an inhomogeneity correction strategy on a specific image. In addition, the Coefficient of Joint Variation is used for estimating the grey-level overlap between different tissue classes (i.e. t_1=disc and t_2=bone) [VOVK2007].

Coefficient of Variation: $$CV(t_1) = \frac{\sigma(t_1)}{\mu(t_1)} \tag{4.7}$$

Coefficient of Joint Variation: $$CJV(t_1, t_2) = \frac{\sigma(t_1) + \sigma(t_2)}{|\mu(t_1) - \mu(t_2)|} \tag{4.8}$$

where σ and μ are the standard deviation and mean value of the selected tissue class t.

4.4.5.2. Segmentation Based Evaluation of Inhomogeneity Correction

Intensity inhomogeneity correction can be evaluated indirectly by its effect on image segmentation [VOVK2007]. In this case the performance of the segmentation algorithm is evaluated on initial (uncorrected) and inhomogeneity corrected images by means of the quantitative measures described in the next section. The comparison of these measures values offers an estimate of the effect of inhomogeneity correction on image segmentation.

4.4.6. Segmentation Accuracy Evaluation

Segmentation accuracy of all methods was quantitatively evaluated against manually segmented data, considered as the ground truth. Manual segmentations of 120 discs from the 1.5 Tesla data sample as well as the 255 discs from the 3.0 Tesla data sample were utilized, while all discs were segmented twice (by two independent users) in order to additionally evaluate inter-observer accuracy.

An area and a boundary metric were utilized for evaluating segmentation accuracy. The area metric is called Dice Similarity Index (DSI) based on the kappa coefficient theory [ZIDJ2004]. DSI measures the pixel-by-pixel classification agreement between manually (M) and automatically (A) segmented data, correcting for agreement by chance (equation 4.9). The boundary metric is D_{mean} and it is a measure assessing the difference in boundary shape (equation 4.10) [KORF2007]. Moreover, the same metrics are utilized for comparing the manually segmented data of the two users to each other, to get an estimate of interobserver variability [CHAL1997]. Finally, Student's t-test is employed to investigate whether the differences in the performance of the segmentation methods are statistically significant. In addition, the discs were separated in classes of degeneration severity following the clinical grading and segmentation accuracy was independently calculated for early and severe degeneration classes.

$$DSI = \frac{2 \cdot |M \cap A|}{|M| + |A|} \times 100\% \tag{4.9}$$

$$D_{mean} = \frac{\sum_{q=1}^{p} D_q}{p} \tag{4.10}$$

where M is the area of the manually segmented disc, A is the area of the computer segmented disc, D_q is the distance of pixel q on A boundary to the closest one on M boundary and p is the total number of boundary pixels on A.

4.5. Segmentation Results and Discussion

This section presents the disc segmentation results, and discusses on the merits and weaknesses of each one of the three segmentation methods developed in this study (FCM, Atlas-FCM, Atlas-RFCM). Section 4.5.1 discusses on the performance of each method illustrating representative segmentation examples, while section 4.5.2 presents the quantitative evaluation results. Finally, section 4.5.3 presents results of the pseudo-3D approach

4.5.1. Disc Segmentation Results- Image Examples

The FCM method was rather successful in discriminating between the disc, bone and CSF tissues. However, FCM was unable to separate the disc from the longitudinal ligaments and other surrounding soft tissues such as vessels. As a result the FCM suffered from severe border leakage, mostly towards the anterior disc side as shown in Figure 4.8. This border leakage resulted from overlapping grey level values of the intervertebral disc's annulus with those of neighbouring ligaments and muscles.

The second method (Atlas-FCM method) combined the fuzzy clustering approach with a probabilistic disc atlas in order to deal with overlapping grey level values and control border leakage. A combined probability matrix was employed putting together anatomical information by means of the disc atlas with tissue class membership information from the FCM, thus providing higher weighting to the disc pixels. The Atlas-FCM method managed to control border leakage in the anterior disc side improving the segmentation outcome, as shown in Figure 4.8.

The third segmentation method is called Atlas-RFCM and it is a simple modification of the Atlas-FCM approach, where the FCM algorithm was replaced by the Robust-FCM one. This Robust-FCM algorithm introduces smoothness constraints in the fuzzy membership function rendering the segmentation method more robust to border leakage related to image noise.

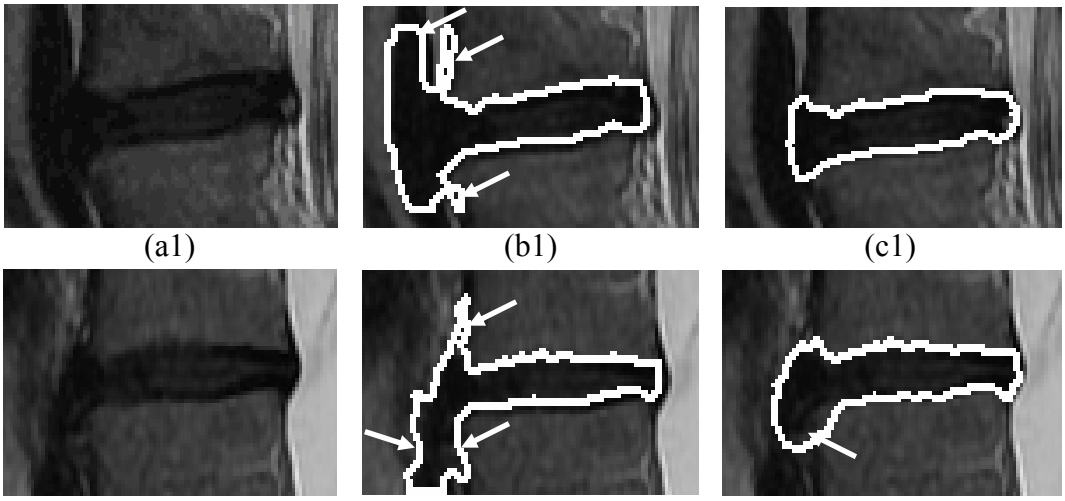


Figure 4.8. Comparative segmentation examples: the initial image (a) followed by the segmentation results of the FCM (b) and the Atlas-FCM (c) methods. The white arrows point to regions with boundary leakage.

Figures 4.9. and 4.10. illustrate the segmentation results of the Atlas-FCM and Atlas-RFCM methods together with manual segmentation for a normal and a degenerated disc. Both atlas based methods perform well with the resulting segmentation being very close to the manual ground truth, while they also accurately delineate the disc herniation shown by the black arrow in Figure 4.10.

However, small segmentation errors are still present. As shown in Figure 4.9., the Atlas-FCM method results in a slight leakage of the bottom disc boundary. This boundary is accurately segmented by the Atlas-RFCM method possibly due to its robustness to noise. Overall, the Atlas-RFCM provided further improvement in disc segmentation performance.

As seen in Figures 4.8.-4.10. the final segmentation boundaries are not very smooth. Although more intense boundary smoothing could be applied (i.e. by using a larger element for morphological opening) this could result in loss of information as demonstrated in Figure 4.11. A small amount of smoothing is beneficial (see Figures 4.11.c1-c2), but when more smoothing is applied the endplates shape changes and disc herniations almost disappear (see Figures 4.11.d1-d2). Thus, a compromise between boundary smoothness and retaining diagnostic information needs to be made.

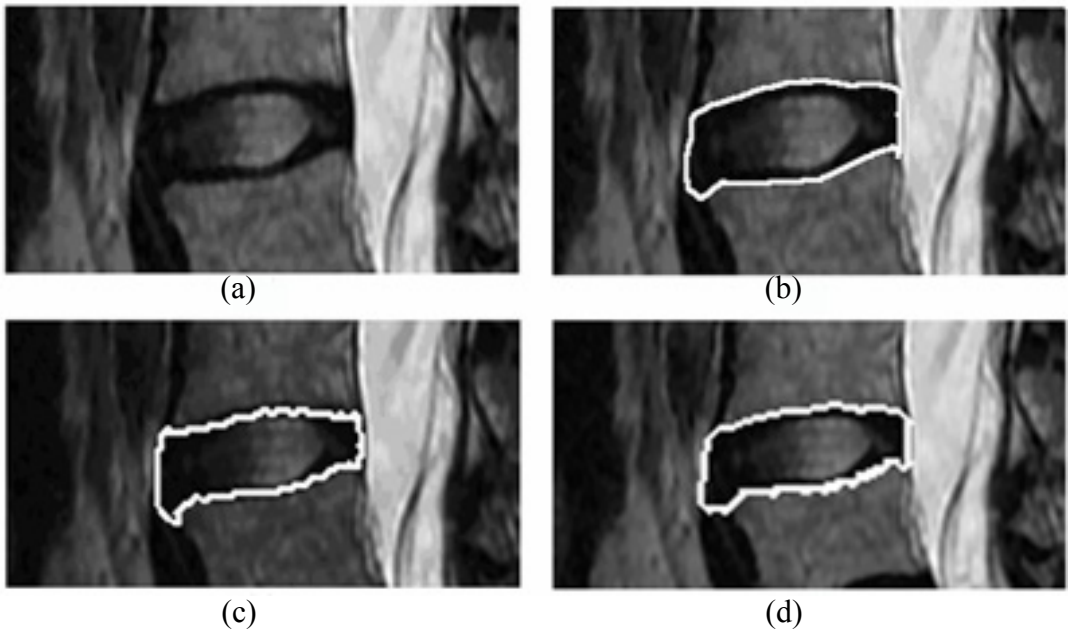


Figure 4.9. Segmentation example for a normal intervertebral disc. The initial image is given in the top left corner (a), followed by ground truth provided by manual segmentation (b), as well as the Atlas-FCM (c) and Atlas-RFCM (d) segmentation results.

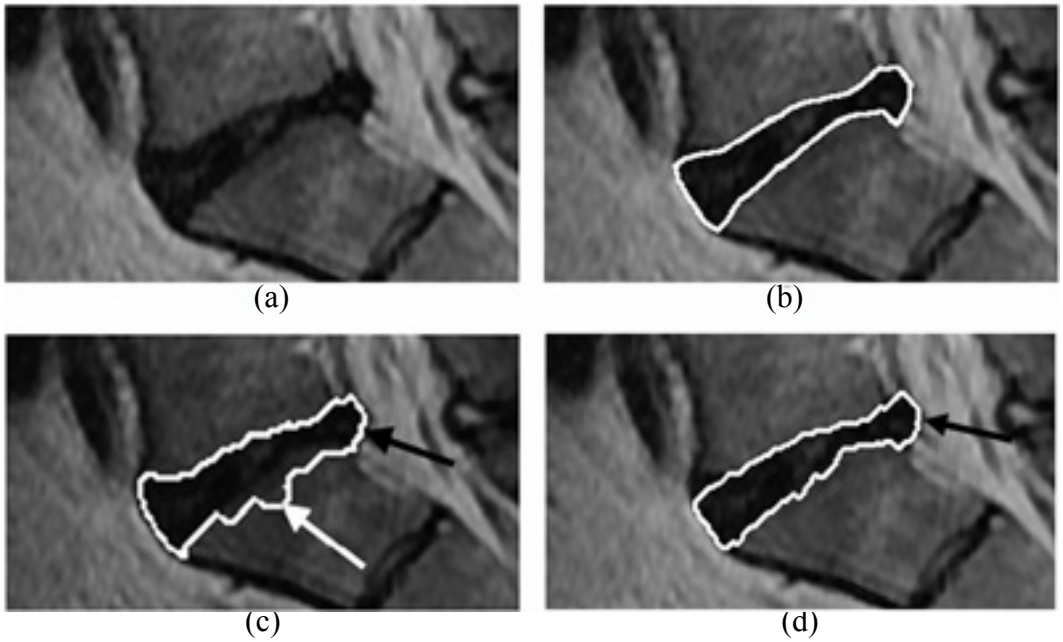


Figure 4.10. Segmentation example for a degenerated intervertebral disc with a posterior herniation. Again the initial image is given in the top left corner (a), followed by ground truth provided by manual segmentation (b), as well as the Atlas-FCM (c) and Atlas-RFCM (d) segmentation results. The black arrows point to the disc herniation, while the white arrow points a region with boundary errors.

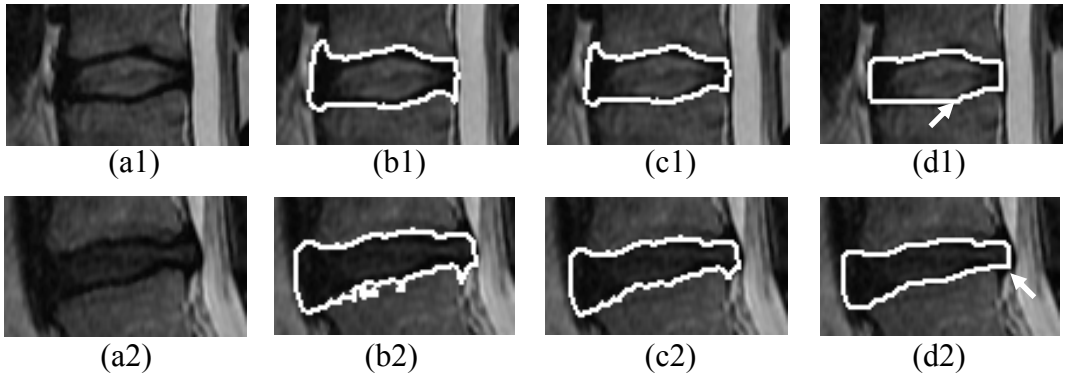


Figure 4.11. Examples of disc segmentation with increasing boundary smoothing: (a) initial disc image, (b) no boundary smoothing, (c) a small amount of boundary smoothing (morphological opening with a circular element with 2 pixels radius) and (d) a lot of smoothing (morphological opening with a circular element with 6 pixels radius). White arrows point to areas where too much smoothing results in loss of information, such as the shape of the endplates and the disc herniation.

Results presented so far are for the 1.5 Tesla dataset. Figures 4.12.–4.14. illustrate results of the Atlas-RFCM method applied on the 3.0 Tesla dataset, together with the manual delineation. Specifically, Figure 4.12. illustrates the segmentation on of a severely degenerated disc (Grade V), which is completely flattened. Moreover, Figure 4.13. illustrates a disc with a large posterior herniation shown by the black arrow. The Atlas-RFCM approach provides good delineation of the disc hernia. Finally, Figure 4.14. shows a disc case where the upper disc boundary, on the interface with the vertebral body is not very clear. As shown by the white arrows the segmentation method fails to detect this border and leaks towards the upper vertebra.

(a) (b) (c)

Figure 4.12. Segmentation example for a severely degenerated disc illustrating the initial image (a), the manual segmentation results (b), as well as the Atlas-Atlas-RFCM segmentation result (c).

(a) (b) (c)

Figure 4.13. Segmentation example for a herniated intervertebral disc illustrating the initial image (a), the manual segmentation results (b), as well as the Atlas-Atlas-RFCM segmentation result (c). The black arrow points to the disc herniation, while the white arrow points to a region with boundary errors.

(a) (b) (c)

Figure 4.14. Example of mis-segmentation of the disc-vertebra boundary illustrating the initial image (a), the manual segmentation results (b), as well as the Atlas-RFCM segmentation result (c). The white arrows points to mis-segmented areas in the disc-vertebra interface.

4.5.2. Quantitative Results and Analysis

This section presents results from the quantitative evaluation of segmentation accuracy, versus a ground truth provided by manually segmented data and discusses on methods' performance. Section 4.5.2.1. compares the three segmentation methods performances to each other, as well as to the interobserver

accuracy, for the 1.5 Tesla dataset. Moreover, Section 4.5.2.2. outlines the segmentation results of the optimal method for the 3.0 Tesla dataset. Section 4.5.2.3. outlines the effect of inhomogeneity correction on Atlas-RFCM segmentation accuracy for the Homeomorphic Filtering, Entropy Minimization and N3 methods (described in Section 4.4.5.).

4.5.2.1. Segmentation methods comparison through quantitative accuracy evaluation

This section presents results from the quantitative evaluation of intervertebral disc segmentation accuracy for the 1.5 Tesla dataset. Out of the 170 discs of this dataset, 50 were used in the design of the probabilistic disc atlas, and the remaining 120 were employed for validating the segmentation performance.

Table 4.1. lists the average segmentation results for the three methods along with the interobserver results for the same 120 discs. Results show that the FCM method has the worst overall performance and a rather large mean boundary error, which can be attributed to boundary leakage towards surrounding structures. On the other hand the Atlas-FCM and Atlas-RFCM methods provide improved segmentation performance, with the Atlas-RFCM yielding the highest overall accuracy. The Atlas-RFCM method performs statistically significantly better than FCM and Atlas-FCM methods, as indicated by Student's paired t-test ($p<<0.01$).

Comparing the Atlas-RFCM segmentation results to interobserver accuracy it should be noted that the mean DSI value is 6% lower for the proposed method, while the mean boundary error is approximately 0.6 pixels higher. The clinically acceptable error in boundary detection is 1-2mm as indicated by an expert orthopaedist. The Atlas-RFCM method results in mean boundary error

Table 4.1. Average segmentation performance of the three methods and interobserver accuracy for the 120 discs of the 1.5Tesla dataset

All Discs	**DSI %**		**Dmean (pixels)**	
	mean	**std**	**mean**	**Std**
FCM	66.7	15.2	7.79	3.90
Atlas-FCM	86.1	5.1	1.58	0.59
Atlas-RFCM	88.4	3.5	1.28	0.35
Interobserver	94.4	2.8	0.69	0.25

which is well within this range (D_{mean}=1.28 pixels corresponding to 0.8mm), showing that its accuracy is within clinically acceptable limits.

Tables 4.2. and 4.3. list the average results for normal and degenerated discs separately (following the grading of the nomenclature project). It can be observed that segmentation accuracy is lower in the case of degenerated discs compared to normal discs. This can be attributed to the increased shape variability of degenerated discs, which often differ a lot from the normal discs used for the atlas construction. In addition, the inter-observer DSI was found to be slightly lower for degenerated discs (93.9%) in comparison to normal discs (95.3%). Although this difference is not large, it is statistically significant ($p<0.001$) and in agreement with both physicians' comments that manual segmentation of degenerated discs was more difficult than normal discs' segmentation.

Table 4.2. Average segmentation performance of the three methods and interobserver accuracy for 42 normal discs

Normal Discs	**DSI %**		**Dmean (pixels)**	
	Mean	**std**	**mean**	**std**
FCM	67.2	17.4	7.72	4.54
Atlas-FCM	89.2	4.5	1.23	0.53
Atlas-RFCM	91.6	3.2	0.98	0.37
Interobserver	95.3	1.2	0.61	0.24

Table 4.3. Average segmentation performance of the three methods and interobserver accuracy for 78 degenerated discs

Degenerated Discs	**DSI %**		**Dmean (pixels)**	
	Mean	**std**	**mean**	**std**
FCM	66.5	13.8	7.83	3.51
Atlas-FCM	84.4	5.4	1.77	0.62
Atlas-RFCM	87.2	3.6	1.44	0.34
Interobserver	93.9	3.0	0.74	0.25

The DSI results are graphically represented in the scatter chart in Figure 4.15. This chart summarizes the segmentation performance results, showing the progressive improvement of segmentation accuracy when moving from the simple FCM to the Atlas-FCM and the Atlas-RFCM method. Finally, Table 4.4. gives the manual segmentation time for a trained radiologist and the corresponding computer processing times. The automated segmentation methods provide important time savings, by greatly reducing the manual interaction time and relieving the user from the tedious manual delineation task.

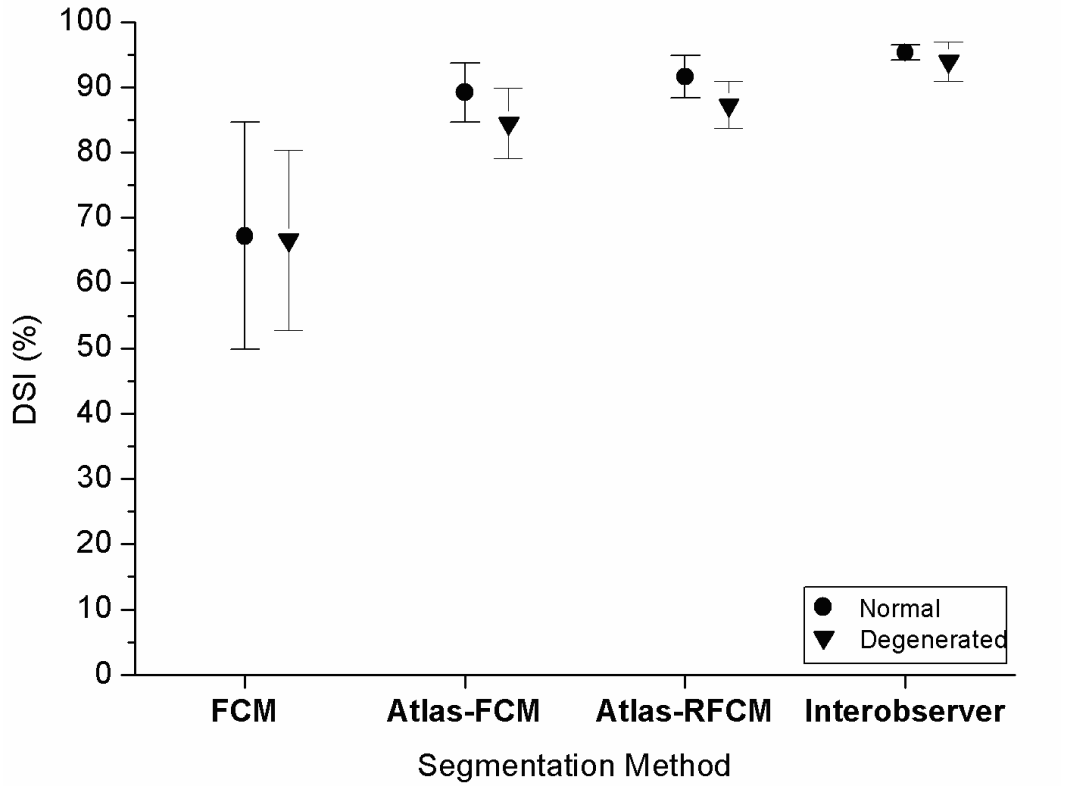


Figure 4.15. Scatter chart summarizing the DSI mean values and standard deviations of both normal and degenerated discs for the developed segmentation methods and the interobserver accuracy.

Table 4.4. The manual interaction and processing times for disc segmentation Processing times estimated for a PC (1.8GHz Dual Core CPU and 3GB RAM)

	Manual Interaction Time (s)	**Processing Time (s)**
Manual Segmentation	124	-
FCM	2	1
Atlas-FCM	6	1
Atlas-RFCM	6	2

4.5.2.2. Segmentation accuracy of the Atlas-RFCM on the 3.0 Tesla dataset

The accuracy of the Atlas-RFCM, which is the optimal segmentation method was additionally tested on a set of 255 discs comprising the 3.0 Tesla data sample. The average segmentation results for this dataset are listed in Table 4.5. together with the interobserver accuracies, while Table 4.6. presents results for discs separated in two groups of degeneration severity.

The probabilistic atlas used for segmentation was built on templates taken from the 1.5 Tesla dataset. However, this does not seem to adversely affect the segmentation outcome, since the Atlas-RFCM accuracy for the 3.0 Tesla dataset is actually slightly higher (1%) than the corresponding accuracy for the 1.5 Tesla dataset. A direct comparison is inappropriate, due to the heterogeneity of the two datasets with respect to clinical sample characteristics, image resolution and acquisition parameters.

Table 4.5. Average segmentation performance of the Atlas-RFCM method and interobserver accuracy for the 255 discs of the 3.0 Tesla dataset

All Discs	**DSI %**		**D_{mean}**	
	Mean	std	mean	std
Atlas-RFCM	89.4	6.9	1.30	0.85
Interobserver	93.5	2.6	0.82	0.27

Table 4.6. Average disc segmentation performance in classes of degeneration severity

	Normal and Mild Degeneration (Grades II and III)				Moderate and Severe Degeneration (Grades IV ad V)			
	DSI %		**D_{mean} (px)**		**DSI %**		**D_{mean} (px)**	
	mean	std	mean	std	mean	std	mean	std
Atlas-RFCM	90.3	6.5	1.19	0.49	88.3	7.3	1.45	1.14
Interobserver	94.4	1.8	0.75	0.23	92.3	3.1	0.91	0.30

4.5.2.3. Segmentation accuracy with and without Inhomogeneity correction

As described in Section 4.4.5. three different inhomogeneity correction methods were tested in order to remove the field bias prior to segmentation. These are the Entropy Minimization, Homeomorphic Filtering and N3 methods. The performance of these methods in correcting image intensity inhomogeneity was evaluated on a randomly selected disc subset (comprising 25 discs), derived from

5 patient scans, using the coefficient of variation and coefficient of joint variation, while the corresponding results are given in Table 4.7. All three methods resulted in slightly improved image homogeneity, as indicated by the reduced values of the coefficients of variation. Moreover, the separability between disc and bone tissues increased yielding reduced values for the coefficients of joint variation.

In addition, the effect of inhomogeneity correction on disc segmentation accuracy was tested using the D_{mean} and DSI measures and the corresponding results are summarized in Table 4.8. The N3 method provided the best overall performance with respect to facilitating the segmentation algorithm when used as a preprocessing step and thus providing improved segmentation accuracy. The effect of inhomogeneity correction on disc segmentation performance is very small (0.9%). This can be attributed to the small size of the disc structure, which is only slightly affected by the slowly varying inhomogeneity field. However, this marginal accuracy improvement is statistically significant ($p<0.001$). Thus to exploit this effect, the N3 method was selected to be used as a preprocessing step for correcting image inhomogeneity prior to disc segmentation.

Table 4.7. Indicative intensity variation values results for the uncorrected image and the corresponding inhomogeneity corrected images

Correction Method	**Intensity Variation**		
	CV_{disc}	**CV_{bone}**	**CJV**
Initial Images	0.297	0.556	0.774
Entropy Minimization	0.294	0.534	0.773
N3	0.279	0.552	0.760
Homeomorphic Filtering	0.274	0.537	0.760

Table 4.8. Mean segmentation evaluation results for initial and inhomogeneity corrected images

Correction Method	**Segmentation Accuracy**		
	DSI %	**D_{mean} (px)**	**Paired t-test**
Initial Images	92.3	1.08	-
Entropy Minimization	92.3	1.08	p=0.11
N3	93.2	0.95	p<0.001
Homeomorphic Filtering	92.6	1.03	p<0.001

4.5.3. Segmentation accuracy of the Pseudo-3D segmentation method

The pseudo-3D segmentation method is applied on 3D volumes constructed by interpolating our 2D data as described in section 4.4.4. This method was validated against a data set of 10 randomly selected intervertebral discs and the corresponding accuracy evaluation results are given in Table 4.9. Moreover, representative results of the pseudo 3D segmentation method for multiple slices of interpolated volumes are illustrated in Figures 4.16-4.17.

Specifically, Figure 4.16 illustrates a herniated disc demonstrating the adaptability of the algorithm to the disc shape which results in a rather good delineation of the hernia. White arrows in the leftmost slice point to areas where the disc boundary is unclear, as this slice corresponds to the end of the disc and thus contains a superimposition of disc and surrounding tissues.

Table 4.9. Pseudo-3D segmentation performance

	DSI %		**D_{mean} (px)**	
	mean	std	mean	std
Pseudo-3D	89.4	7.7	1.42	1.06

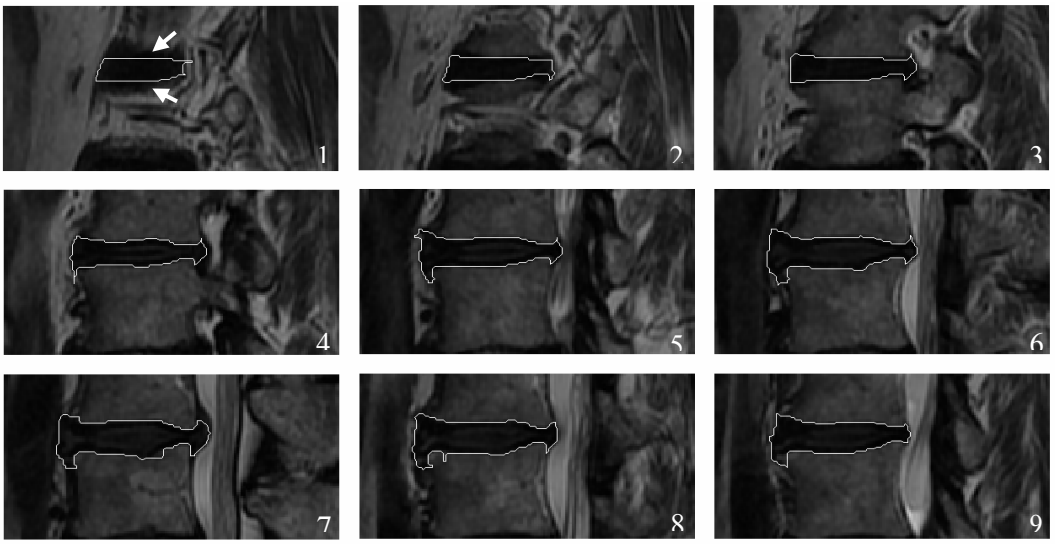


Figure 4.16. The pseudo 3D segmentation results for a normal intervertebral disc. The images correspond to equidistant slices of the interpolated data volume, with 1 being the leftmost slice and 9 being the rightmost one. White arrows in the leftmost slice show an area where the disc boundary appears unclear.

Figure 4.17 also illustrates segmentation results in multiple slices of an interpolated volume, with white arrows pointing to mis-segmented areas. Finally, Figure 4.18 illustrates the surface rendering of the segmented disc volume displayed in Figure 4.17.

The pseudo-3D approach was designed in order to test if the proposed atlas-based method can be expanded to serve disc segmentation in 3D. As shown in Table 4.9, the mean DSI segmentation accuracy is 89.4% which is similar to the accuracy of the 2D atlas based segmentation. Moreover the mean boundary error is 1.48pixels (which is equivalent to 0.85mm) and this accuracy meets clinical requirements. Thus this method appears capable to serve 3D segmentation, however further investigation and testing on volumetric data is required for the complete evaluation of this approach.

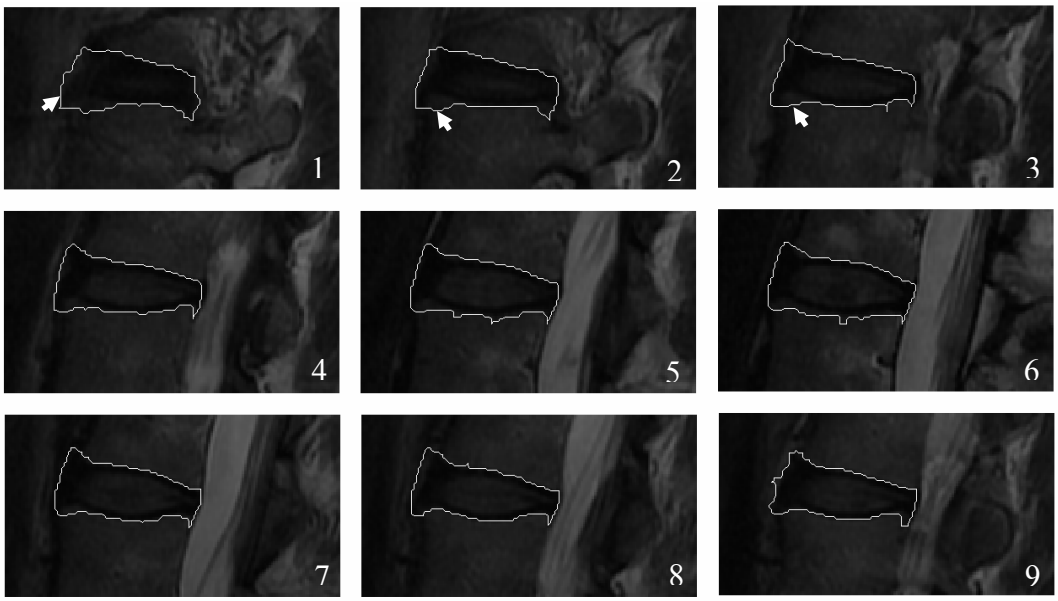


Figure 4.17. Pseudo-3D segmentation results for a normal intervertebral disc. Again, the images correspond to equidistant slices of the interpolated data volume, with 1 being the leftmost slice and 9 being the rightmost one, while white arrows point to mis-segmented disc areas.

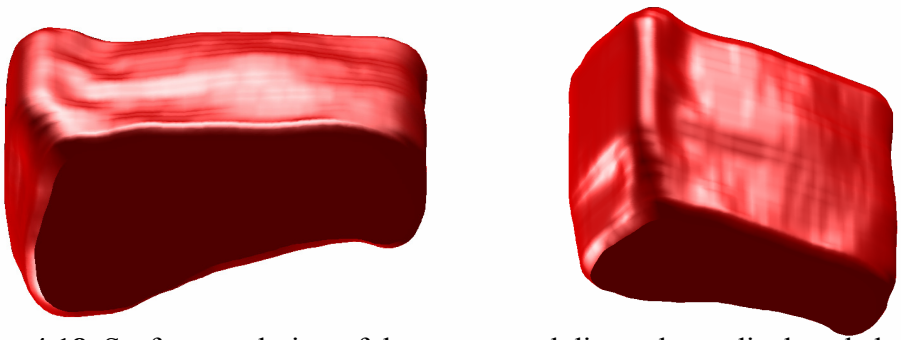


Figure 4.18. Surface rendering of the segmented disc volume displayed above.

4.6. Segmentation Summary and Conclusion

This chapter presented the methods developed for segmenting the intervertebral discs from spine MR images. In the present study, disc segmentation is aimed to serve the quantification of degeneration severity, and thus 2D segmentation methods were developed for extracting the discs from the midsagittal slice of conventional T2-weighted MR images.

Three segmentation methods were presented here, the FCM, Atlas-FCM and Atlas-RFCM, each one building upon the previous method to correct errors and improve accuracy. Moreover, three inhomogeneity correction approaches where tested in order to remove the bias MR field and facilitate the segmentation process. Finally, a pseudo-3D approach was presented expanding the atlas-based method and demonstrating its' ability to serve 3D disc segmentation.

As shown by the quantitative analysis of segmentation accuracy, the Atlas-RFCM method provided the best overall performance for both the 1.5 Tesla and 3.0 Tesla datasets. The following list presents some key points of the Atlas-RFCM performance.

- This method resulted in accurate segmentation for both normal and degenerated discs, and good delineation of specific disc pathologies such as disc narrowing and herniation. However, there are cases where the Atlas-RFCM method misses the disc-vertebra interface and this can be related to low image contrast and shape bias introduced by the atlas.
- Quantitative results evaluation yielded a mean boundary error < 1mm, which meets clinical requirements
- This automated segmentation offers important interaction time savings in comparison to the manual delineation process.

With respect to intensity inhomogeneity, among the three methods tested the N3 approach was shown to provide the optimal bias field correction. When this method is used prior to disc segmentation it facilitates this process resulting in a small but statistically significant improvement in segmentation accuracy.

A direct comparison of the proposed method's accuracy with disc segmentation methods presented in the literature is not feasible. The majority of these methods focus on disc localization and they do not carry out any evaluation of segmentation accuracy [ROBE1997, CHEV2007, SHI2007]. However, a recent study by Seifert and Wachter reports quantitative results on the segmentation of cervical intervertebral discs [SEIF2009]. Evaluation was performed on a midsagittal slice of the T1w cervical spine image acquired with a 3D protocol. Direct comparison would be inappropriate, due to the data heterogeneity, yet the quantitative values are given just as an indication. The study by Seifert reports a mean DSI value equal to 91% with mean boundary error of 1.51 pixels (2.4 mm) averaged over nine normal intervertebral discs of a single dataset.

In conclusion, the proposed Atlas-RFCM approach incorporates prior anatomical information in the segmentation process, using a probabilistic disc atlas, combined with fuzzy clustering techniques and smoothness constraints. This method provides accurate segmentation for both normal and degenerated intervertebral discs, which meets clinical requirements, and offers important interaction time savings.

4.7. Analysis of Objectives and Hypotheses for Disc Segmentation

This section analyses the aims and objectives of the present study with respect to the lumbar disc segmentation task. Table 4.10 presents a list of objectives and discusses on the extent to which these objectives were met. In addition, Table 4.11 discusses on the hypotheses of this study.

Table 4.10. Research Objectives Summary : Intervertebral Disc Segmentation

3. To develop computerized methods for segmenting the intervertebral discs *Three segmentation methods were developed, with each new method building upon the previous one aiming to provide accurate segmentation of both normal and degenerated discs. The optimal method is Atlas-RFCM which combines spatial information by means of a probabilistic atlas with fuzzy clustering techniques. In addition, the N3 inhomogeneity correction method was used for correcting the MR bias field and facilitating the segmentation process. Disc segmentation was developed in 2D and tested on midsagittal MR images, since this method is intended to serve disc degeneration quantification. Moreover, a pseudo-3D approach has been tested demonstrating how the Atlas-RFCM can be extended to 3D.*
4. To utilize quantitative methods for evaluating segmentation accuracy of the computerized methods against a gold standard produced by manual disc segmentation *The gold standard was obtained by manual delineation of all data by two independent users. Disc segmentation accuracy was evaluated against manually segmented data by means of the DSI and D_{mean} measures. Results indicated high segmentation accuracy for the Atlas-RFCM approach meeting clinical requirements (mean boundary error <1mm). Although the automated segmentation accuracy is not as high as manual interobserver accuracy (the DSI is on average 4.1% lower for the 3T dataset), it meets clinical needs while offering important interaction time savings.*

Table 4.11. Research Hypotheses Intervertebral Disc Segmentation

1. Atlas based methods would provide improved segmentation accuracy in comparison to methods relying solely on grey level intensity. This is because the incorporation of shape information in the segmentation process would help control boundary leakage towards surrounding structures.

   Confirmed: *the Atlas-FCM and Atlas-RFCM methods managed to control boundary leakage and provided statistically significantly improvement in segmentation accuracy in comparison to the FCM method as shown in Tables 4.1. and 4.4.*

2. The correction of the inhomogeneity field in MRI would facilitate the segmentation process and improve accuracy.

   Confirmed: *Inhomogeneity correction prior to segmentation using the N3 method resulted in a very small but statistically significant improvement of segmentation accuracy as shown in Table 4.7. ($p<0.001$).*

Chapter 5. Disc Quantification

This chapter presents the disc quantification methods and results. It starts with an overview of texture and shape quantification methods, followed by a discussion on the requirements and challenges of disc quantification. The methods developed for disc quantification in the present study are analyzed in Section 5.2, while the corresponding results are presented and discussed in Sections 5.3 and 5.4. Finally, Section 5.5 analyzes the research objectives and hypotheses.

5.1. Quantification Methods Background

5.1.1. Introduction to Quantification

Texture and shape are among the most informative visual cues for the diagnostic interpretation of medical images [BANK2000]. The different anatomical regions are characterized by different textures and shapes, which are qualitatively evaluated by the physicians in order to discriminate between normal and abnormal state [CAST2004]. Image analysis systems are designed to quantify texture and shape information and assist physicians to make a decision on the underlying pathology [COST2005].

5.1.2. Texture Quantification

Image texture is a rich source of visual information and a key component in image analysis and understanding in humans [DURG1996]. In digital images, the notion of texture is associated with the spatial distribution of pixel intensity variations, and is related to image properties such as smoothness, coarseness and regularity. Texture analysis is ultimately concerned with automated methods that can quantify image information [TOUR1999]. Image texture has been utilized for distinguishing between healthy and pathological tissue in a variety of clinical problems and a wide range of imaging modalities starting from X-ray and moving to CT, MRI and Ultrasound [CAST2004, TOUR1999]. Common applications include the classification of normal and benign tissue from X-ray and MRI mammograms [KARA2007, CHEN2007], the detection and quantification of lung nodules and interstitial pneumonia patterns in lung CT [KIM2003, KORF2009], as well as and the characterization of intracranial tumours in MRI [SCHA1993, KJAE1995]. In musculoskeletal imaging, texture analysis has been successfully used for the quantification of osteoarthritis from hip X-rays and knee MRI [BONI2006, QAZI2006].

In the present study, texture analysis algorithms are utilized for extracting quantitative image information from lumbar spine MRI, aiming to quantify

intervertebral disc degeneration severity. The remaining part of this section describes commonly used features for texture characterization, while the equations defining these features are given in Appendix 1.

5.1.2.1. First order features

First order textural features are extracted from the image histogram and thus encode information regarding the frequency of appearance of each grey-level within the image. More specifically, these features quantify the mean and standard deviation of grey-level values within the image, as well as the skewness and kurtosis of the histogram [THEO2003].

5.1.2.2. Co-occurrence features

Co-occurrence features are perhaps the most widely used method of texture analysis, while their popularity is attributed to their good performance [LERS1993, NIXO2002]. Co-occurrence matrices (also known as grey-level spatial dependence matrices) describe the frequency in which two grey-levels of an image appear together at a certain distance (D) and in a particular direction [HARA1979]. The features extracted from these co-occurrence matrices describe different characteristics of texture such as image homogeneity, contrast or the level of randomness within the image. However, it is not always easy to identify which specific textural characteristic is represented by each of the co-occurrence features, as not all of them convey a clear physical or perceptual sense of texture [HARA1973, RANG2005].

5.1.2.3. Run-length matrix features

These features were introduced by Galloway and are calculated from run-length matrices extracted from the image [GALL1975]. A run is a set of consecutive pixels that have the same grey-level value, aligned in a given orientation. The "grey-level run length matrix" $r_\theta(i,j)$ shows the frequency of each run with length j of each grey-level i in the image, along the direction θ. Features extracted from these run-length matrices describe the distribution of grey-level runs within the image. Thus these features quantify properties such the presence of linear structures and grey-level uniformity in the image.

5.1.2.4. Laws texture energy features

These features are extracted from a set of Texture Energy Images, which are computed by applying small convolution kernels to the ROI [LAWS1980]. Texture Energy Images are considered to be associated to the inherent microstructure of the ROI. Features extracted from Texture Energy Images can be used for quantifying image microstructure. First order statistics are most commonly used for this purpose, although run length features can also be extracted aiming to capture additional image information [BONI2007b].

5.1.3. Shape Features

In clinical practice, qualitative evaluation of shape information and semi-quantitative measurements (such as the intervertebral disc height) are generally utilized for discriminating between normal and abnormal status. The automatic quantification of shape information could provide accurate and reproducible measures to aid clinical diagnosis and planning of surgical treatment [BANK2000]. Simple techniques for 2D quantification of shape information, such as region measures, statistical moments and Fourier descriptors are summarized in the following sections. The equations defining shape features are given in Appendix 1.

5.1.3.1. Fourier Descriptors

Fourier descriptors are calculated from the boundary of an object by means of the Discrete Fourier Transform (DFT). The Fourier description of an object gives a set of spatial frequencies that fit this boundary. Low order descriptors represent information such as the object centroid and size, while higher order descriptors are associated with higher spatial frequencies and thus describe increasingly higher detail [NIXO2002].

5.1.3.2. Statistical Moments

Statistical moments can be used for representing spatial image distributions. Their values depend on the position, orientation and size of the object within the image. However, many applications require an object representation that is invariant to translation, scale and rotation [BANK2000]. Position invariance is achieved by defining the moments with respect to the image centroid, and the resulting measures are called central moments. Moreover, scale invariance is achieved by reducing the object to unit axis, while the resulting measures can be combined to produce rotation invariance moments according to the method introduced by Hu [HU1962].

5.1.3.3. Region Properties

The simplest shape features are region properties calculated from a binary template of the region of interest. Some basic region properties are listed in table 5.1. [NIXO2002, MATL2010].

Table 5.1. The basic region properties together with a short description for each

Property	Description
Area	The number of pixels within the ROI
Perimeter	The number of pixels surrounding the ROI
Eccentricity	The eccentricity of the ellipse that has the same second moments as the ROI
Major Axis Length	The major axis length of the ellipse that has the same normalized second central moments as the ROI
Minor Axis Length	The minor axis length of the ellipse that has the same normalized second central moments as the ROI
Convex Area	The number of pixels of the smallest convex polygon that can contain the ROI
Solidity	The proportion of ROI pixels in the convex area calculated as Area/Convex Area

5.2. Methods: What Do we Propose for Disc Quantification?

Currently, the standard method for disc quantification relies on the measurement of the disc mean signal intensity from midsagittal T2-weighted MRI. As discussed in section 2.5.2 this mean signal intensity reflects the water and proteoglycan content of the intervertebral disc [TERT1991] and is a sensitive measure of degeneration severity [NIEM2008]. However, although it is a good measure of biochemical changes within the disc, it does not account for morphological and structural disc changes.

Alternative quantification methods include the measurement of T1, T1ρ and T2 relaxation times, as well as the ADC. As discussed in section 2.5.2 these methods provide additional information about the disc tissue, but their application is limited by the requirement for specific acquisition protocols and additional imaging time.

The present study aims in developing new measures for the quantification of disc degeneration from conventional T2-weighted MR images. In this context, texture and shape descriptors are utilized aiming to capture additional image information. Texture descriptors are exploited for assessing degenerative changes resulting from biochemical tissue alterations (water and proteoglycan content), while shape descriptors were used for assessing gross anatomical and structural changes of the disc tissue. Moreover, the spinal level of each disc (ranging from L1-L2 to L5-S1) is used as an additional feature encoding prior knowledge of the disc's location representing anatomical information. The motivation behind using a conventional imaging protocol (T2-weighted MRI), is to produce a method that could be applicable in widely available clinical protocols for spine imaging, without the constrains of specific imaging protocols.

A preliminary study on disc degeneration quantification was performed with the 1.5 Tesla dataset and the corresponding results were presented in UKRC2009. The full texture and shape analysis was conducted using the 3.0 Tesla dataset (because this had a more reliable severity grading ground truth). Texture analysis results have been accepted for publication in Acta Radiologica.

The disc quantification methods and results presented here were applied on the 255 intervertebral discs of the 3.0 Tesla dataset. This section summarizes the quantitative features calculated from each segmented disc ROI, and describes the feature selection process used for identifying the most descriptive features with respect to disc degeneration. Next the statistical analysis methods used for testing the descriptive ability of the texture and shape descriptors and comparing their performance to the standard method (based on mean disc intensity) are presented.

5.2.1 Methods: Quantitative Features Calculation

All 255 discs of the 3.0 Tesla dataset were segmented by the Atlas-RFCM method, and a set of 89 textural and 32 shape descriptors together with a location descriptor (the disc level) was extracted from each segmented disc region. The background of texture and shape analysis is given in Section 5.1, while features definitions can be found in Appendix 1. The set of descriptors extracted from each segmented disc ROI comprised:

- 4 first order features - the mean, standard deviation, skewness and kurtosis
- 24 co-occurrence features - first co-occurrence matrices were calculated for four different directions(0°, 45°, 90° and 135°) with a pixel distance d=2, and from these matrices the mean values and ranges of the 12 co-occurrence features over the four directions where calculated
- 5 run length features - the short run emphasis, long run emphasis, grey-level non uniformity, run length non uniformity and run percentage
- 56 Laws texture energy features - the mean, standard deviation, skewness and kurtosis of each one of the 14 Laws Texture Energy Images calculated from the disc-ROI
- 7 region descriptors - the area, perimeter, eccentricity, major axis length, minor axis length, convex area and solidity
- 8 central moments - the μ_{00}, μ_{11}, μ_{02}, μ_{20}, μ_{12}, μ_{21}, μ_{03}, μ_{30} were calculated from the binary template of the disc-ROI

- 7 Hu's invariant moments - the φ1-φ7 measures where calculated from the binary template of the disc-ROI
- 10 Fourier descriptors – the first ten Fourier coefficients were calculated from the disc-ROI boundary
- 1 location descriptor – the disc level, taking values in the range [1-5] corresponding to lumbar disc levels (L1-L2, L2-L3, L3-L4, L4-L5,L5-S1).

5.2.2. Methods: Optimal Feature Selection

In the disc quantification process, a set of 122 texture and shape features have been calculated from each segmented disc-ROI. These features contain a lot of redundant information, which would unnecessarily complicate the classification process, and could adversely affect the outcome. The objective of dimensionality reduction is to obtain an accurate data representation while eliminating redundant components. Thus, the dimensionality reduction methods need to identify or construct uncorrelated features with high information content. Dimensionality reduction methods can help improve classification performance and alleviate the curse of dimensionality in cases where the number of training samples is limited. In addition, they can reduce computational complexity and enable better understanding of the underlying process [JAIN2000, THEO2003].

Here two methods are tested and compared to help identify the most descriptive feature set with respect to disc degeneration severity, which is the classification task of the present study. These methods are Principal Component Analysis (PCA) and Stepwise Discriminant Analysis (SDA) which are briefly outlined below. The reduced feature sets produced by the two methods were used for the classification of disc degeneration severity and their performance was assessed using ROC analysis as described in section 6.3.4.

Stepwise Discriminant Analysis

SDA is a well established method, which is highly popular with CAD applications [TOUR2001]. This method aims to reduce the number of features utilized by a classification system while preserving as much of the class

discriminatory information as possible. In this context, only the most important features of a given dataset are selected aiming to maximize class separability, while features bearing no additional information are removed. SDA has been utilized in a variety of classification problems such as the characterization of masses and microcalcifications in X-ray mammography, the characterization of pulmonary nodules in lung CT and the identification of subcellular patterns from human cell arrays [CHAN1998, MCNI1999, CONR2004, KARA2007].

SDA works by identifying the features that maximize the interclass distance, while keeping the classes as tightly clustered as possible (minimize within class variance). This is achieved by minimizing the Wilks' ratio Λ (Equation 5.1).

$$\Lambda = \frac{\det(S_W(\bar{x}))}{\det(S_T(\bar{x}))} \tag{5.1}$$

where $\bar{x}$ is the feature vector, S_w is the within class scatter matrix and S_T is the total scatter matrix.

Λ equals zero for perfect discrimination while it equals one in case of no discrimination between classes. The SDA procedure starts with the full feature set and it sequentially selects to remove and add features (dimensions) from the model in order to minimize Λ and thus increase the model's discrimination ability. Features are removed from the model if the F-to-remove value is smaller than the corresponding F-to-remove-threshold, while they are added if the F-to-enter value is larger than a predefined F-to-enter threshold. These F-to-enter and F-to-remove values give the statistical significance of change in Λ resulting from the feature addition or removal according to the F-statistic (Equations 5.2, 5.3) [LINE2007]. Here, the probability to enter and probability to remove features were set to p_{enter}=0.05 and p_{remove}=0.10.

$$F_{enter} = \frac{(n-p-q)}{(q-1)} \frac{1-\Delta\Lambda(p+1)}{\Delta\Lambda(P+1)} \tag{5.2}$$

$$F_{remove} = \frac{(n-p-q+1)}{(q-1)} (\Delta\Lambda(p-1)-1) \tag{5.3}$$

$$\Delta\Lambda(p+1) = \frac{\Lambda(p+1)}{\Lambda(p)} \tag{5.4}$$

where n is the number of data points, q is the number of classes, p is the number of features and ΔΛ is the partial Λ statistic given by equation 5.4.

Principal Component Analysis

This is the best know dimensionality reduction method. It is widely used in pattern recognition applications, while a variety of non-linear techniques based on PCA have also been developed [JAIN2000]. However, PCA still remains the golden standard as the more complex non-linear techniques have not yet been shown to outperform it [MAAT2007].

PCA performs dimensionality reduction while retaining as much variance in the feature space as possible. Since high variance is related to high information content, this method can "compact information" to fewer dimensions by identifying the dimensions of maximum variance. PCA has been used in a variety of CAD applications such as breast cancer diagnosis from X-ray mammography, pulmonary nodules diagnosis from CT images and the diagnosis of Alzheimer's from PET and SPECT [AYRE2005, LIU2009, WAY2009, LOPE2009].

PCA performs an orthogonal transformation from the initial feature space to a new set of uncorrelated variables which are called principal components. A complete description is beyond the scope of this thesis, a good tutorial is given in [SHLE2005]. Equation 5.5 outlines the PCA transformation with X being the original dataset, Y the transformed dataset and P the principal components matrix. In order to maximize variance, while minimizing redundancy the principal components (p_1,p_2,…p_n) which compose the transformation matrix P must be the eigenvectors of the covariance matrix given by equation 5.6.

$$Y = P \cdot X \tag{5.5}$$

$$C_X = \frac{1}{n-1} X \cdot X^T \tag{5.6}$$

The first principal component lies in the direction of maximum variance, the second in the next highest variance direction and so on. As a consequence most information is retained within the first few dimensions of the transformed space and thus dimensionality reduction can be performed with small information loss.

5.2.3. Methods: Disc Quantification Comparison to the Standard Method

The texture and shape features performance in evaluating intervertebral disc degeneration were compared to those of the adjusted mean disc signal, which is sought as the standard method for quantifying degeneration. For this purpose, the adjusted disc signal intensity was calculated as the ratio of the mean disc over the mean CSF signal intensities. The mean CSF signal was calculated by selecting a rectangular region of size 3x5pixels within the CSF in the anterior part of the dural sac adjacent to the disc as shown in Figure 5.1. In cases where the CSF signal intensity measurement was considered unreliable, due to flow artefacts or narrowing of the dural sack, the observation was recorded as missing and was removed from the case sample analyzed [LUOM2001, NEIM2008].

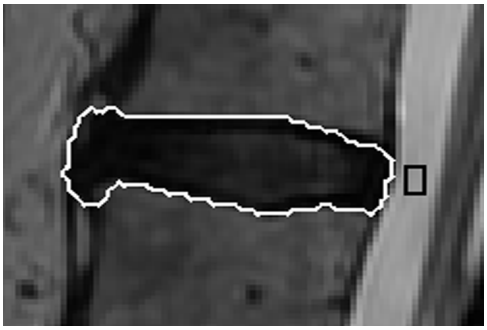


Figure 5.1. The segmented disc region (delineated by the white line) along with the manually defined CSF-ROI (black rectangle) used for adjusting disc signal intensity.

5.2.4. Methods: Statistical Analysis for Quantification Results Interpretation

Statistical analysis was used to determine the ability of SDA selected features to quantify disc degeneration severity. The radiologists' consensus grading presented in Chapter 3 served as ground truth for testing the features discriminating ability.

Spearman's ρ correlation coefficient was employed for measuring the association between the ground truth and quantitative features values. This correlation is used for assessing the validity of texture and shape features for disc degeneration severity quantification. Moreover, The Kruskal-Wallis one way

analysis of variance was used for statistically testing the overall ability of these features to distinguish between discs belonging to different grades of degeneration severity. The statistical significance level was set to $p<0.05$. In addition, the same statistical comparison with a Bonferroni correction (adjusting for multiple tests) was employed for testing the separability of neighbouring degeneration severity classes. This would provide information with respect to descriptors ability to evaluate early or advanced degenerative changes. Moreover, the texture and shape features' effectiveness in quantifying disc degeneration was compared to that of the standard quantification method based on disc's adjusted signal intensity.

5.2.5. Methods: Testing Quantitative Measurements Repeatability

The repeatability of quantitative measurements was tested using Intraclass correlation coefficients. The quantitative feature values are subject to small variations in the segmentation and CSF-ROI selection processes. Specifically, the atlas-based disc segmentation method requires manual selection of 2 initialization landmarks. This landmark selection can affect the segmentation result and thus impose small changes in the texture and shape feature values calculated from the segmented disc region. Additionally, the CSF-ROI selection requires manual input of a single landmark corresponding to the centre of the ROI. This could result in a small variation of the calculated mean CSF value and consequently affect the value of the adjusted mean intensity descriptor. To test the effect of manual landmark selection on the quantitative feature values, the manual positioning of disc and CSF-ROI landmarks was performed twice on a subset of 50 randomly selected intervertebral discs. Consequently, two measurements of these quantitative values were acquired and intraclass correlation coefficients were calculated between these measurements in order to test their repeatability.

5.3. Results and Discussion: Disc Quantification and Association to Pathology

As discussed in Section 2.5.2 the quantification of disc characteristics can provide an objective and reproducible evaluation of degeneration severity, which is sensitive to small changes. This is particularly important for evaluating the disease progress and monitoring the response to treatment [AUER2006].

In this study the quantification of degeneration severity was based on conventional T2-weighted magnetic resonance images of the lumbar spine. These images encode both biochemical and morphological information about the disc tissue allowing the detection of early degenerative changes [MODI2007, NEIM2008]. Here, texture and shape features were extracted from the image, while the validity and reliability of these features in disc degeneration quantification was evaluated by: (i) testing their association to clinical grading of disc degeneration severity, (ii) comparing correlation results to those of the standard quantification method based on adjusted disc mean signal intensity, (iii) evaluating features ability to distinguish between discs from different degeneration severity grades and finally (iv) testing measurements repeatability.

Both PCA and SDA methods were tested for dimensionality reduction, aiming to find the most descriptive feature set with respect to disc degeneration severity. Both methods yielded very similar results (see Appendix 3), and SDA was selected here to form the basis of the proposed CAD system. This selection was based on the requirement for meaningful quantitative measures [TOUR2000]. Specifically, the principal components calculated by PCA are linear combinations of the initially calculated feature set, and thus do not convey a clear physical meaning. On the other hand SDA simply selects a subset of these initially calculated features, which is straight forward to interpret.

Specifically, SDA yielded a set of 7 features which are sought as the most appropriate for characterizing disc degeneration severity. The selected feature set is presented in Section 5.3.1., while features association to clinical grading of degeneration is discussed in Section 5.3.2. Moreover, Section 5.3.3. presents the ANOVA and Multicomparison results commenting on features ability to

distinguish between early and advanced stages of degeneration, while Section 5.3.4. discusses on measurements repeatability.

5.3.1. Results and Discussion: The selected feature set

This section provides a brief description of the 7 quantitative features selected by the SDA method:

1. Sum of Squares: this is a co-occurrence derived feature that quantifies the variance of grey-level spatial dependencies within the image and its value decreases for increasing image homogeneity.
2. Information Measure of Correlation 1: this feature is also derived from co-occurrence matrices and it describes linear dependencies between the grey-levels within the ROI and its value increases with increasing image homogeneity.
3. Range Sum of Squares: another co-occurrence derived feature that quantifies the changes in variance across different image directions.
4. Range of Entropy: a feature quantifying changes in grey-level randomness across the different image directions [HARA1979].
5. Grey Level Non Uniformity: this is a run-length feature that quantifies the uniformity of grey level run lengths within the image. Its value increases for non-isotropically distributed runs along the different grey levels [GALL1975, KARA2009].
6. Solidity: this is a shape feature that represents the ratio of the disc-ROI area over the area of the smallest polygon than can inscribe the ROI, and thus is can be seen as a measure of image compactness [MATL2010].
7. Disc Level: this feature introduces useful anatomical information, as the lower lumbar levels are more prone to degeneration, due to spinal anatomy and the increased compressive load applied to the corresponding discs [BATT2004].

5.3.2 Results and Discussion: Quantitative Features Association to Clinical Grading Ground Truth

The suitability of the selected textural and shape features for quantifying disc degeneration severity was investigated through their association to radiologists' consensus grading. As described in Section 5.2. Spearman's correlation coefficients were utilized for this task, while the correlation to mean disc signal intensity, which is sought as a reference method was also calculated for comparison reasons. Table 5.2. summarizes the results of this correlation analysis, while the key points are analyzed in the bullet points list below.

Table 5.2. Features Correlation to Clinical Grading

Quantitative Features		**Association to Clinical Grading**	
Feature Class	Feature Name	Spearman ρ	$p < 0.001$
Method of Reference			
	Adjusted Mean Intensity	-0.5567	✓
New Approach			
	Sum of Squares	-0.812	✓
Co-occurrence	Information Measure of Correlation 1	0.799	✓
	Range Sum of Squares	-0.397	✓
	Range Entropy	-0.730	✓
Run-Length	Grey Level Non Uniformity	0.439	✓
Region	Solidity	-0.450	✓
Location	Disc Level	0.224	X

Looking at the method of reference:

- Mean Disc Signal Intensity displays a negative correlation to disc degeneration severity. This does not come as a surprise, since progressive degeneration is related to disc dehydration resulting in lower disc signal in T2-weighted images. The strength of correlation is statistically significant ($p<0.001$).

And at the selected feature set:

- The correlation signs between textural features values and clinical grading of degeneration indicate that the disc image becomes more homogeneous with increasing severity of degeneration. The strength of correlation is statistically significant, and these features appear to provide good measures of degeneration severity.
- Three of the textural features analysed (Sum of Squares, Information Measure of Correlation 1 and Range Entropy) display stronger correlation to degeneration severity than Mean Disc Signal Intensity does.
- The negative association of the Solidity feature to degeneration severity indicates a decrease in disc shape compactness with progressing degeneration.
- The positive correlation between the Disc Level and clinical grading indicates that the lower disc levels tend to be more severely degenerated. However, this correlation is not statistically significant.

5.3.2.1 Disc Degeneration and Image Homogeneity – The role of texture

The correlation analysis results indicate an increase in disc image homogeneity with progressive degeneration. Here, the connection between image homogeneity and the degenerative alterations of the disc is analysed and the role of texture in evaluating degeneration severity is discussed.

As outlined in chapter 2, the degenerative process results in nucleus dehydration (due to proteoglycan fragmentation) and the disc progressively becomes more fibrous [AN2004]. This is why degenerated discs appear darker in T2 weighted MR images, while the nucleus-annulus distinction progressively decreases [PFIR2001]. These degenerative alterations result in a decrease of grey levels variation with progressive degeneration, and the disc image appears more homogeneous. An example for this is provided in Figure 5.2, where grey-level histograms for discs assigned to Grades II, III, IV and V are illustrated. Moving from Grade II to Grade V we can see how the histogram shifts towards lower

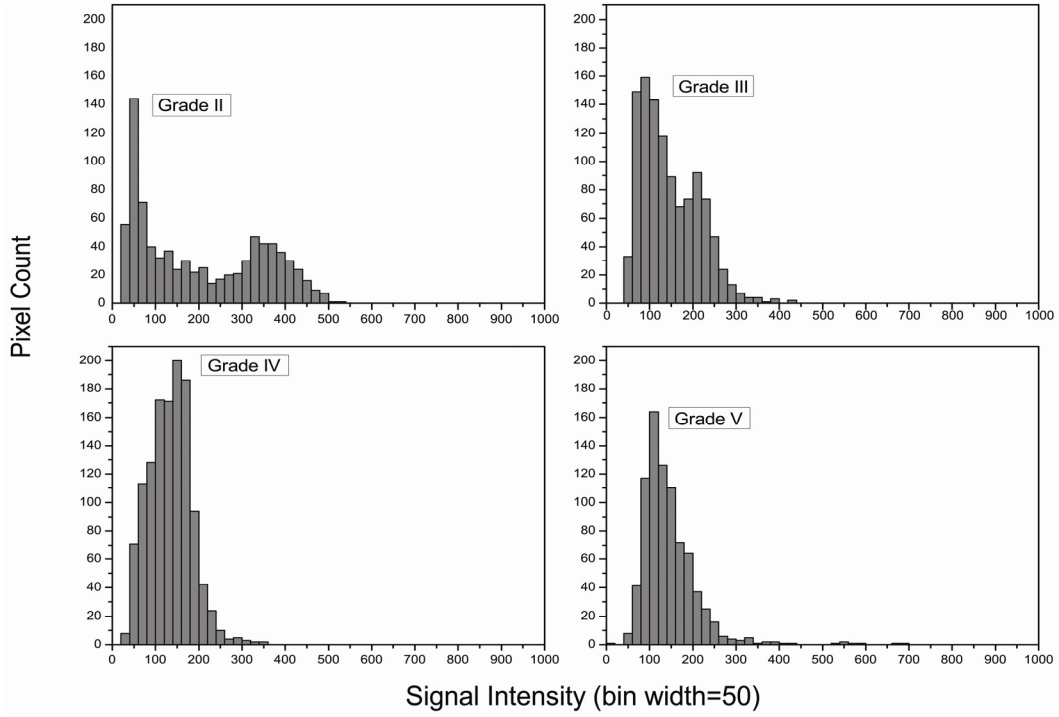


Figure 5.2. Representative histograms of discs for the different grades of degeneration severity. The more degenerated the disc the lower the mean value and the narrower the histogram (fewer frequent intensities are dominant).

signal intensity values indicating disc dehydration. Grades II and III are characterized by bimodal distributions, with the peaks corresponding to the mean annulus (left) and nucleus (right) grey-levels. The histograms of Grades IV and V are unimodal distributions indicating the complete loss of nucleus annulus distinction. When moving from Grade II to Grade III, the peak corresponding to the disc's nucleus gradually shifts towards lower grey-level values, while in Grade IV the two peaks merge indicating the progressive loss of distinction between disc nucleus and annulus. This loss of distinction results in narrowing of the histogram when moving to higher grades of degeneration indicating a decrease in image grey level range (i.e. the disc image becomes more homogeneous).

Textural features have the ability to measure image homogeneity and thus quantify the progressive loss of nucleus annulus distinction, which is an important factor in the degenerative process. As shown in Table 5.2, the strength of correlation between three of the five textural features (namely the Sum of Squares, Information Measure of Correlation 1 and Range Entropy) and clinical diagnosis of degeneration severity is actually higher than the corresponding one for discs'

adjusted mean signal intensity. This can be attributed to the ability of these features to represent additional image information regarding the grey-level variation.

The effectiveness of texture quantification may be related to biochemical and structural tissue alterations. In the degenerative process, proteoglycan fragmentation results in reduced water binding capacity of the nucleus. In addition, a slow substitution of the fine type II collagen fibres of the inner nucleus by coarser type I fibres is taking place. These alterations lead to nucleus dehydration and annulus encroaching on the nucleus, which macroscopically appear as a loss of nucleus annulus distinction [AN2004]. In MRI this results in an increase of image homogeneity, which can be captured by textural features. Thus, these features can indirectly evaluate disc alterations resulting in loss of distinction between nucleus and annulus by measuring the grey-level variations within the image. By this we do not imply that the use of textural features should substitute the adjusted mean intensity quantification method, but they should rather be used together to obtain the maximum amount of information regarding disc degeneration severity.

5.3.3. Results and Discussion: Features discriminating ability in early versus severe degeneration

The properties of quantitative features were further investigated by analyzing their ability to differentiate between the different degeneration severity grades. Table 5.3. summarizes the results of the Kruskal-Wallis ANOVA, as well as the comparison between neighbouring grades. The key points of this analysis are outlined in the following list.

Looking at the method of reference:

- Mean Disc Signal Intensity demonstrates statistically significant differences along the range of different degeneration severity grades. However, when looking into neighbouring grades, this method can distinguish normal from mild and moderate degeneration (Grades II, III and IV), but not between the moderate and severe stages (Grades IV and V).

Table 5.3. ANOVA and Multicomparison Results

Quantitative Features		**Kruskal –Wallis ANOVA**		**Comparison between neighbouring Grades** (Bonferroni and p<0.05)		
Feature Class	Feature Name	Chi square	p<0.05	Grade II Grade III	Grade III Grade IV	Grade IV Grade V
Reference Method						
	Adjusted Mean Intensity	64	✓	✓	✓	X
New Approach						
Co-occurrence	Sum of Squares	137	✓	✓	✓	X
	Information Measure of Correlation 1	132	✓	✓	✓	X
	Range Sum of Squares	68	✓	✓	X	X
	Range Entropy	112	✓	✓	✓	X
Run-Length	Grey Level Non Uniformity	47	✓	X	✓	X
Region	Solidity	48	✓	X	✓	✓
Location	Disc Level	24	✓	X	X	✓

Looking at the selected feature set:

- All seven features comprising the selected feature set demonstrated statistically significant differences along the range of different degeneration grades.
- Looking into neighbouring grades, textural features can generally distinguish between the early stages of disease, but not between moderate and severe degeneration (Grades IV and V).
- In contrast, the Solidity and Disc Level features cannot effectively distinguish between the early stages of degeneration, but they display statistically significant differences between the moderate and severe stages (Grades IV and V).

5.3.3.1 Degeneration severity stages - the role of texture and shape analysis

The multicomparison results in Table 5.3. indicate that mean disc signal intensity and textural features offer good measures for early and intermediate degeneration. On the other hand looking at the separability of Grades IV and V these measures cannot distinguish between moderately and severely degenerated discs. This was somewhat expected for the textural features, since Grade IV and V discs are fibrous and have completely lost their nucleus-annulus distinction appearing homogeneous in MRI.

Although textural information does not help distinguishing among the more advanced stages of degeneration, shape quantification can assist in this task. Solidity effectively distinguishes between Grade IV and Grade V discs ($p<0.05$). This can be explained by the morphological differences between these grades as defined by Pfirrmann [PFIR2001]. Grade IV corresponds to discs with normal height or a moderate height decrease, with a rather smooth elliptical outline. On the other hand, Grade V represents a collapsed disc space where endplate damage yields an irregular and less compact outline for the disc. This change in disc shape compactness is reflected in Solidity values, which is thus providing a good measure for advanced degeneration stages.

5.3.4. Results and Discussion: Quantitative measurements repeatability

Intra-class correlation coefficients values used for testing measurements' repeatability were found to be over 0.97 for both adjusted mean signal intensity and the selected textural and shape descriptors values, indicating very high repeatability.

5.4. Discussion on Quantification Methods and Results

The correlation analysis and ANOVA results presented in this chapter indicated that the proposed texture and shape features provide good measures of degeneration severity. An advantage of the proposed texture quantification in comparison to the reference method, is that the textural features used are independent of grey level shifting and thus do not require an intra-body reference for grey-level adjustment. Consequently, texture quantification can be used for evaluating disc degeneration severity, even in cases where the adjusted mean intensity method cannot be calculated due to the lack of a CSF reference. The lack of such a reference can be due to flow artefacts or narrowing of the dural sac [NIEM2008], which in our data sample affected 19% of all cases. However, it should be pointed out that the use of inhomogeneity correction techniques, such as those described in section 4.4.4., could potentially alleviate the CSF reference problem. In addition, looking into shape quantification, Solidity is a scale invariant feature and thus the corresponding results are independent of patient size.

In addition, since the proposed quantification approach utilizes conventional T2-weighted MR images it could be easily applicable in clinical practice, in contrast with quantification approaches, such as the measurement of ADCs or T1 and T2 relaxation times that require specific imaging protocols. A limitation of this quantification study is the lack of Grade I discs (according to Pfirmann's scale), which represent the normal young population.

The proposed quantification method was directly compared to the adjusted mean signal intensity method, but a similar comparison with other quantification approaches is not possible with our data set. However, some correlation coefficient values from other studies, which also used Pfirrmann's method as their grading standard, will be given here as an indication. In a 2006 study by Auerbach *et al.* on a set of 50 lumbar discs (10 patients), the T1ρ quantification measure was found to be statistically significantly correlated to the clinical grading of degeneration, yielding a correlation coefficient value equal to -0.51 [AUER2006]. In addition, in a 2010 study by Blumenkrantz *et al.* conducted on a set of 80

lumbar discs (16 patients), both T1ρ and T2 relaxation times were found to be significantly associated to clinical grading of degeneration. Specifically, correlation coefficient values for the T2 time were equal to -0.61 while for the T1ρ time equal to -0.84 [BLUM2010]. In the present study, on a data set of 255 discs (51 patients) multiple features resulted in statistically significant correlation to clinical grading, while the strongest overall association was observed for the Sum of Squares feature which yielding a correlation coefficient equal to -0.81.

5.5. Summary and Analysis of Research Objectives

This section demonstrated the strong potential of disc texture and shape features in quantifying intervertebral disc degeneration from conventional T2-weighted MR images. Textural features are particularly sensitive to early stages of degeneration and thus could potentially be useful in evaluating the outcome of new treatment methods, such as growth factor therapy. On the other hand the Solidity shape feature is a good measure for distinguishing between moderate and severe degeneration. The proposed approach provides simple, fast and highly repeatable quantification of disc degeneration, from conventional MR images. Overall, disc image texture and shape quantification could be a valuable tool for tracking the evolution of disc degeneration and monitoring the response to treatment.

Table 5.4. summarizes the quantification results by analysing the research objectives for disc quantification, while Table 5.5. discusses on the quantification hypotheses tested.

Table 5.4. Research Objectives Summary and Discussion

Disc Degeneration Quantification
5. To calculate image features describing textural and shape properties of the disc region in order to quantify disc degeneration severity *First and second order textural features, along with shape features describing the disc region properties, statistical moments and Fourier descriptors were extracted from each segmented disc region aiming to quantify disc degeneration severity. Stepwise discriminant analysis was used for selecting the most descriptive feature set.*
6. To investigate the ability of quantitative features in evaluating disc degeneration by calculating their association to degeneration severity as defined by radiologists consensus grading *The quantification methods were applied on 255 discs of the 3.0 Tesla*

dataset which were accompanied by a ground truth based on three radiologists' consensus gradings. Correlation coefficients were employed for testing the validity of quantitative features for evaluating disc degeneration. Textural features measuring image homogeneity were highly associated to degeneration severity, and could distinguish among the early disease stages. In addition, solidity a feature measuring disc shape compactness was found to be sensitive in distinguishing among the more advanced stages of degeneration.

7. To compare textural and shape features' performance to the performance of the mean signal intensity quantification method in the task of disc degeneration quantification.

   A subset of textural features were found to be more strongly associated to disc degeneration than the mean signal intensity, which is sought as the method of reference. These features appear to quantify the progressive loss of nucleus annulus distinction, which is an important feature of the degenerative process. On the other hand shape features correlation to the clinical ground truth was lower than that of the reference method.

8. To test the repeatability of quantitative features measurements

   Measurements repeatability was tested utilizing intraclass correlation coefficients. Results indicated that quantitative features measurements are highly repeatable (ICC>0.97).

Table 5.5. Research Hypotheses Summary and Discussion

Disc Degeneration Quantification

3. Textural and shape image features are suitable for describing intervertebral disc degeneration severity. This would be confirmed if these features values are statistically significantly correlated to the clinical diagnosis (grading) of disease.

Confirmed: *as shown in Table 5.2. statistically significant associations were found between textural and shape features values, and the clinical grading ground truth. This indicates the strong potential of these features in evaluating degeneration severity.*

4. Textural features are more suitable than shape features for evaluating the early stages of degeneration. This is because in the early stages of disease only the internal disc structure is affected but its outline remains almost intact. Moreover, shape features are more suitable that texture features for assessing the more advanced stages of disease, when internally the disc has become fibrotic.

   Confirmed: *as shown in Table 5.3., textural features can discriminate between early stages of degeneration (Grades II-IV), while shape features are more effective for distinguishing between moderate and severe degeneration (Grades IV-V)*

Chapter 6. Computer Aided Diagnosis (CAD)

This chapter presents the computer aided diagnosis methods and results. It starts with an introduction to CAD systems in section 6.1, while section 6.2 gives an overview of CAD systems development for spine diagnosis. The methods developed in the present study for CAD of disc degeneration are analyzed in Section 6.3, while the corresponding results are presented and discussed in Section 6.4. Finally, Section 6.5 offers a conclusion and analyzes the research objectives and hypotheses.

6.1. Introduction to CAD

This Section introduces Computer Aided Diagnosis methods in 6.1.1., presents the design and main components of CAD applications for tissue characterization in 6.1.2. and provides a brief review on supervised classification algorithms in 6.1.3.

6.1.1. CAD Introduction or what is it good for?

Computer Aided Diagnosis systems utilize pattern recognition methods for quantitatively analyzing medical images. These systems aim to detect and characterize abnormal findings based on image appearance. The precursor of computer aided diagnosis is automated computer diagnosis an idea that emerged in the 60s. At that time it was assumed that computerized methods would replace the physicians in the task of diagnostic image interpretation. However, the current belief is that computerized methods should be utilized by the physicians to help them in the decision making process without replacing them, thus the name computer aided diagnosis [DOI2007]. Specifically, nowadays CAD is intended to be used as "a second reader", influencing the radiologist's opinion. The use of CAD aims at improving diagnostic accuracy by increasing the sensitivity in detection and characterization of lesions and also improving agreement between observers [DOI1999].

CAD systems are designed for detecting and characterizing abnormal findings in medical images and they are divided in two categories. Systems that detect suspicious regions in medical images are described by the acronym CADe. Moreover, the term CADx describes systems that quantitatively analyze the properties of these suspicious regions and characterize them as normal or abnormal [GIGE2008].

CAD systems have been developed for a variety of diagnostic tasks, including breast cancer [CHEN1999, KARA2007], lung cancer and interstitial disease [AWAI2004, KORF2009], colon cancer [YOSH2002], liver disease

[KATO2007], as well as skeletal disease such as hip and knee osteoarthritis [BONI2007a, SHAM2009].

Professor Kunio Doi of the University of Chicago, stated that "CAD has become on of the major research subjects in medical imaging and diagnostic radiology" [DOI2007]. Nowadays, CAD research is rapidly expanding to other imaging modalities and clinical tasks. It is expected that in the future CAD systems will be incorporated into diagnostic workstations and used as support tools for image interpretation in a variety of tasks in daily clinical work [DOI2007, GIGE2008].

It should be noticed that the benefits of using CAD as a second reader in clinical practice are controversial. Taking as an example breast cancer detection which is the most widely used CAD application, it has been shown that breast CAD systems can detect cancers missed by the clinicians but they also produce a number of false positives [BIRD2001]. Looking into the influence of breast CAD on the decision making process, results are disappointing. Specifically, CAD has been shown to yield an increase of patient recall rates without improving the cancer detection rate [TAYL2008]. Although the detection and classification accuracy of a CAD system is important, it is actually the system's influence on radiologist's opinion that matters the most. The following section describes the main components of a CAD system, from the point of view of applications developed for tissue characterization (CADx).

6.1.2. CAD System Design: The main CADx components

The layout of a typical CADx system is given in Figure 6.1. In CADx it is assumed that the location of the region of interest is predefined by either a CADe system or the user himself [COST2010]. The CAD main components are: (a) preprocessing, (b) segmentation of the region of interest (ROI), (c) quantification of ROI properties and (d) classification of the ROI as normal or abnormal or as belonging to one of n classes of pathology.

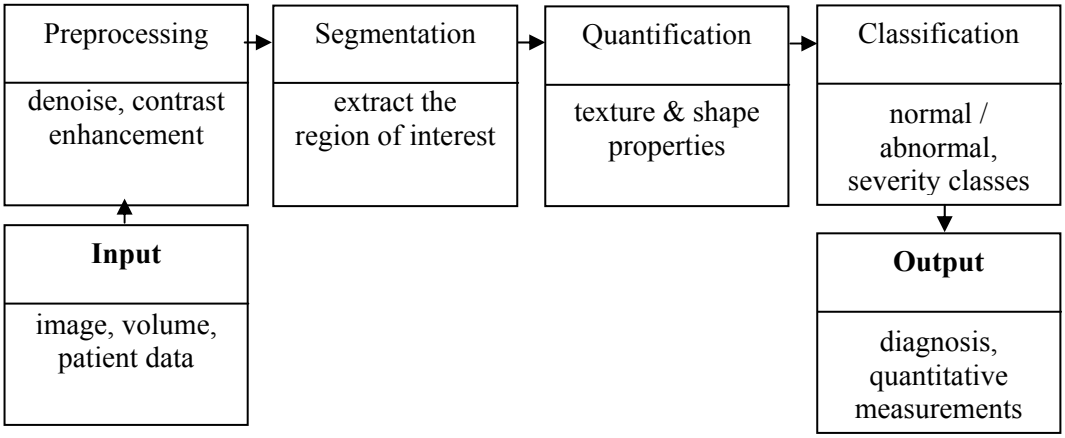


Figure 6.1. CAD layout.

The preprocessing step in CAD aims into enhancing the region of interest within the image. This can be achieved by the application of image processing techniques, such as denoising and contrast enhancement, as well as techniques aiming to the reduction of artefacts and correction of field inhomogeneities. It should be noticed that preprocessing in CAD aims into facilitating the computer in the task of ROI segmentation. This differs from the traditional application of image enhancement, which aims into improving the image for the human observer [DOI1999]. An example for this is the case of MRI field intensity inhomogeneities, which have little impact on visual diagnosis of disease but can greatly affect the performance of automatic segmentation algorithms and thus need to be corrected prior to segmentation [SLED1998]. Moving on to the segmentation step, its accuracy can affect the outcome of the CAD system and thus the segmentation performance needs to be validated.

The next step in CAD is the quantification of ROI properties, which is also known as the "feature extraction process". In this process a set of features describing size, shape and/or textural properties of the ROI are calculated. Feature extraction aims at capturing image properties reflecting the underlying pathology. These features values are then exploited for differentiating between healthy and diseased tissue [COST2010].

In the design stage of a CAD system, a large number of quantitative features is extracted from each ROI. Dimensionality reduction methods are then exploited in order to select features maximizing class separability and reduce redundant information. This process facilitates training and testing of the CAD

system by finding which features can effectively differentiate between the healthy and diseased state [COST2010].

The last step of a CAD system is classification. In this step the ROI is assigned to the appropriate class according to the output of a classification algorithm. The classes represent various stages of health and disease, such as benign and malignant lesions as they appear in X-ray mammography, normal, ground glass or reticular patterns of interstitial pneumonia in lung CT, or the various stages of osteoarthritis in hip X-rays [CHEN1999, KORF2009, BONI2007a].

6.1.3. Brief review on classification algorithms

A classification algorithm is used for assigning an input into one of *n* defined output categories called classes. The classifier input is called the "feature vector", and it is a set of quantitative feature values describing the properties of the item to be classified. The classification algorithm is trained to distinguish among the different classes. In the training process feature vectors of known class (the training data) are employed in order to calculate the classifier's discriminant function. This function encodes the knowledge of the system on how to distinguish among the different classes on the basis of feature values. The training process is also known as "classifier design". Following this design process, the trained classifier is now ready to be used for assigning an output class to every new feature vector of unknown class [THEO2003].

In case of image based CAD systems the classifier input (feature vector) encodes quantitative information about a given ROI within the medical image, such as its texture, shape or size. The classification algorithm utilizes this quantitative information in order to assign an output class to the ROI (e.g. normal/abnormal, benign/malignant) [GIGE2008].

There is a great variety of classification algorithms. Simple classification methods include the Minimum Distance, the Bayesian and the Nearest Neighbour classifiers, while more elaborate approaches are based on Neural Networks and Support Vector Machines [JAIN2000]. The following sections provide short

descriptions of some simple classification algorithms, which are used later in this study.

6.1.3.1. Minimum Distance Classifier

The Minimum Distance (MD) classifier assigns the input feature vector to the class whose centre is at the minimum distance from the feature vector. The class centres are calculated as the means of the training data belonging to each class. Equation 6.1 gives the discriminant function of the minimum distance classifier, based on the Euclidean distance calculation. A limitation of the simple minimum distance classifier is that the resulting classification boundary is always linear. However, depending on the problem at hand, a more complex solution might be better. A simple variation of this classifier is the Mahalanobis minimum distance classifier which can give a quadratic classification boundary [DUDA2001, GONZ2001]. The corresponding discriminant function is given by equation 6.2.

$$f_{j\,Eucl}(x) = x_{in}{}^{T} m_j - \frac{1}{2}(m_j m_j^T) \tag{6.1}$$

$$f_{j\,Mahal}(x) = \sqrt{(x_{in} - m_j)^T C^{-1} (x_{in} - m_j)} \tag{6.2}$$

$$C = \left(\frac{1}{N_j} \sum x_j x_j{}^{T} \right) - m_j m_j^T \tag{6.3}$$

where m_j is the mean vector of class j, x_j is the *j-th* training feature vector of this class, N_j is the number of training vectors for class j, C_j is the covariance matrix of class j and x_{in} is the input feature vector with T representing the transpose matrix.

6.1.3.2. Nearest Neighbour Classifier

The Nearest Neighbour (NN) is a very simple classification algorithm which makes a decision based on the neighbours of the input vector. The neighbours are the training samples situated closer to the input vector within the feature space. In its simplest form this classifier assigns the input vector to the class of its nearest neighbour. The k-NN classifier, finds the k nearest neighbours and assigns the input vector to the class that represents the majority of these neighbours. A variation of the k-NN classifier is the distance weighted k-NN described by Dudani [DUDA1976]. In this approach the contribution of each one

of the k nearest neighbours is weighted by a factor that is inversely proportional to its distance form the input features vector. The distance weighted k-NN has been shown to provide improved classification accuracy in comparison to the simple k-NN and especially in cases of small or moderate size training samples. Just as with minimum distance classifiers, various distance measures can be used by the k-NN for finding the nearest neighbours, such as the Euclidean, the L1 norm and the Mahalanobis distances (Equations 6.4.-6.6.). A drawback of the k-NN is that classes with more frequent examples tend to dominate the outcome [NIXO2002].

$$d_{Euclidean} = \sqrt{(x - x_j)^2} \tag{6.4}$$

$$d_{L1norm} = \left| x - x_j \right| \tag{6.5}$$

$$d_{Mahalanobis} = \sqrt{(x - x_j)^T C^{-1} (x - x_j)} \tag{6.6}$$

where C_j is the covariance matrix of class j given by Equation 6.3, x_j is the *j-th* training feature vector of this class and x is the input feature vector.

6.1.3.3. Bayesian Classifier

It is a simple probabilistic classification algorithm based on Bayes' theorem, which gives the posterior probability (also know as conditional probability) of an event when the prior probabilities and its likelihood are known. In case of classification, the posterior probability denotes the probability that the input vector belongs to a certain class. The classifier's discriminant function is given in Equation 6.7. The Bayesian classifier provides classification with the minimum probability error. It only requires a small amount of training data in order to estimate the parameters (means and variances of the features) and this makes it a simple and efficient classification algorithm. Its main limitation is the requirement for normal distributions of data for each class [THEO2003, GONZ2001].

$$f_j(x) = x^T C_j^{-1} m_j - \frac{1}{2} m_j^T C^{-1} m_j \tag{6.7}$$

where m_j is the mean vector of class j, C is mean covariance matrix of all classes and x is the input feature vector.

6.1.3.4. Probabilistic Neural Network Classifier

The Probabilistic Neural Network (PNN) classifier combines statistical theory with neural networks. Artificial neural networks are highly non-linear computational systems that can approximate the behaviour of complex phenomena and systems which are difficult to model explicitly. The PNN classifier architecture for a two class classification problem is outlined in Figure 6.2. It is a four layer neural network, starting with an "input layer" and moving on to the "pattern layer", the "summation layer" and finally the "output". The pattern layer has a different neuron for each one of the training feature vectors, and it functions by computing the distances between the input vector and each one of the training feature vectors. The summation layer has one neuron for each class, which provides a weighted summary of distance measure calculated by the pattern layer. The outcome layer acts as a threshold providing as classification result the class which yielded the highest sum in the summation level.

Equation 6.8 gives the discriminant function of the PNN classifier. A great advantage of this classifier is that it makes no assumptions about data distributions. Moreover, it converges to an optimal classifier as the number of the training data increases. Finally, being many times faster than traditional neural networks, it can be easily retrained every time additional data become available [SPEC1990, HAJM2002].

$$f_j(x) = \frac{1}{(2\pi)^{p/2}\sigma^p N_j} \sum e^{\frac{\|x-x_j\|^2}{2\sigma}} \tag{6.8}$$

where x is the unknown feature vector, x_j is the *j-th* training feature vector, N_j is the number of patterns in class j, σ is a smoothing parameter, and p is the number of features employed in the feature vector

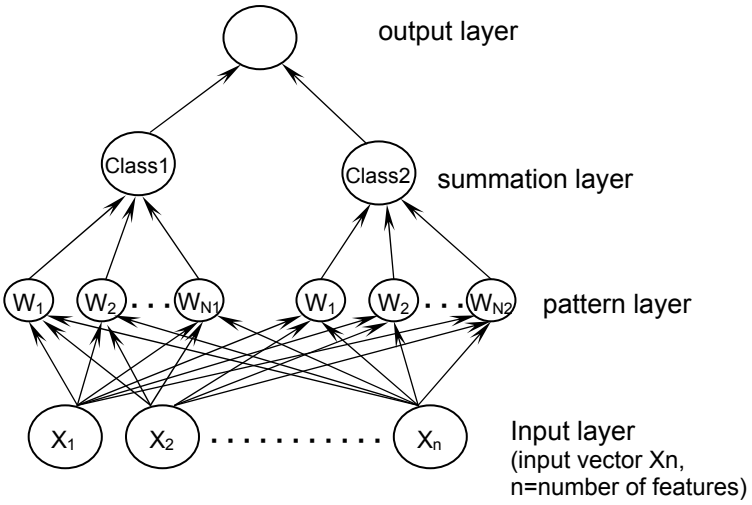


Figure 6.2. Outline of the PNN classifier architecture.

.

6.2. Literature Review on Spine CAD

So far, only a small number of studies work towards computer aided diagnosis of spine pathology. These systems investigate various abnormalities such as disc space narrowing and herniation, vertebral osteophytes and spinal stenosis. This section offers an overview to these spine CAD methods.

The very first study dates back to 1983 and is entitled "Computer Aided Diagnosis of Lumbar Disc Herniation". It presents an automated system, named the "LOWBACK", which takes as input data the physical findings, clinical results and pain history of lumbar spine patients and gives as output a suggestion on the most appropriate treatment (conservative or surgery). This system does not perform image analysis, but is rather one of the early approaches in computer-assisted medical decision making [HUDG1983].

Moving on to studies analyzing medical images, Tsai *et al*. presented a method for evaluating the shape and position of lumbar disc herniation [TSAI2002]. This method is applied on axial images of the intervertebral disc and the authors suggest it could be used with either CT or MRI data. In this study, a B-spline function is used for approximating a "circular" disc boundary, leaving out the herniation area. Next, geometrical features describing the distance of this herniation area from the disc centre, as well as its orientation are extracted. These features are used for classifying the herniation as bulging, protrusion or extrusion. Moreover, the orientation features are used for deciding if the herniation is central or lateral. The diagnostic results of this CAD system were confirmed by operative findings in 16 patients' lumbar spines with 18 herniations, reporting that all hernias were accurately evaluated. The authors conclude that the proposed method can be used for accurate classification of lumbar disc herniation and assist surgical planning.

Charmathy *et al*. [CHAR2004] presented a computer aided diagnosis method for grading disc space narrowing from lateral X-ray images of the cervical spine. Their method utilized b-spline interpolation between manually defined landmarks for delineating the vertebral bodies. Next quantitative features measuring height and area properties of the intervertebral space are calculated in

order to characterize the narrowing between adjacent vertebrae. Finally, disc space narrowing is classified on the basis of the distance features values using k-means clustering together with self organizing maps techniques. A four grade system is utilized for characterizing the severity of narrowing (normal to significant narrowing). The method was tested on 294 discs and results indicated classification accuracy of over 82% for all grades, allowing an error of ± one grade. Given the similarity of neighbouring grades, the proposed method appears to work well on a complex task and could thus serve computer aided evaluation of disc space narrowing.

A computer aided diagnosis method for the characterization of vertebrae as normal or abnormal with respect to the presence of anterior osteophytes from lateral X-rays is presented by Cherukuri *et al*. [CHER2004]. The vertebrae were delineated using b-spline interpolation of manually defined landmarks in the same way as in [CHAR2004]. Next, area features were calculated from the convex hull image of each segmented vertebra, in order to quantify the vertebra's shape variation from its ideal rectangular shape. Finally, a multilayer perceptron was used for vertebrae classification. The authors report that results indicated that the convex hull features can be used for successfully discriminating between vertebrae with and without osteophytes.

A study on the computer aided diagnosis of spinal stenosis from lateral X-rays was presented in [KOOM2006]. In this study, morphological features describing the vertebral bodies and intervertebral discs height and width as well as the width of the spinal canal are exploited for characterizing a variety of conditions related to spinal stenosis. These conditions are posterior osteophytes, posterior apophyseal arthropathy, disc space narrowing, spondylolisthesis and spinal stenosis. A Bayesian framework was employed for classification purposes. A set of 16 images was used for testing the system. Classification accuracies ranged between 70% and 80%, which the authors suggest would be suitable for screening purposes.

Finally, a recent study by Alomari *et al*. from the University of Buffalo, presents a method for computer aided diagnosis of lumbar disc pathology from lower spine MRI [ALOM2010]. The authors developed a model of disc appearance

by putting together information regarding the grey level values of the disc's neighbourhood, as well as its location and relevant distance from adjacent discs. A dataset of 80 cases including discs with various types of pathologies was used for testing the model's ability to distinguish between normal and abnormal discs. The average accuracy of the binary classification system was reported to be 91.3%.

6.3. CAD for disc degeneration diagnosis

The third and last aim of the present study was to design a computer aided diagnosis system exploiting MR image information for evaluating disc degeneration severity. The CAD system presented here combines texture and shape information extracted from segmented disc regions in order to assign each intervertebral disc to the appropriate grade of degeneration severity. This system is designed for supporting the clinicians in diagnosing intervertebral disc degeneration. It is intended to be used as a second reader aiming to improve diagnostic accuracy and reduce disagreement between clinical experts. The proposed system is designed to work with conventional T2-weighted MR images and thus could be directly applicable in clinical practice.

This chapter presents the design of the CAD system, along with the methods used for validating its performance. The proposed CAD system works by classifying the discs to degeneration severity grades, according to Pfirmann's scheme, which is outlined in Table 4.2. The system was trained and tested utilizing the 255 discs of the 3 Tesla dataset. Ground truth grading for disc degeneration severity was obtained by the three radiologists' consensus.

Before designing the current system, a preliminary study on computer aided diagnosis of lumbar disc degeneration was conducted using the 170 discs of the 1.5 Tesla dataset with the three grades evaluation scheme outlined in section 3.2. This study was presented in the Computer Assisted Radiology and Surgery 2009 Conference.

6.3.1. Design of the Proposed CAD System

Figure 6.3. presents an outline of the CAD system. The major components of this system are the disc segmentation method, the feature extraction process, the feature selection method and finally the classification scheme. The disc segmentation method was presented in detail in Chapter 4, while Chapter 5 outlined the feature extraction and selection methods used for disc degeneration

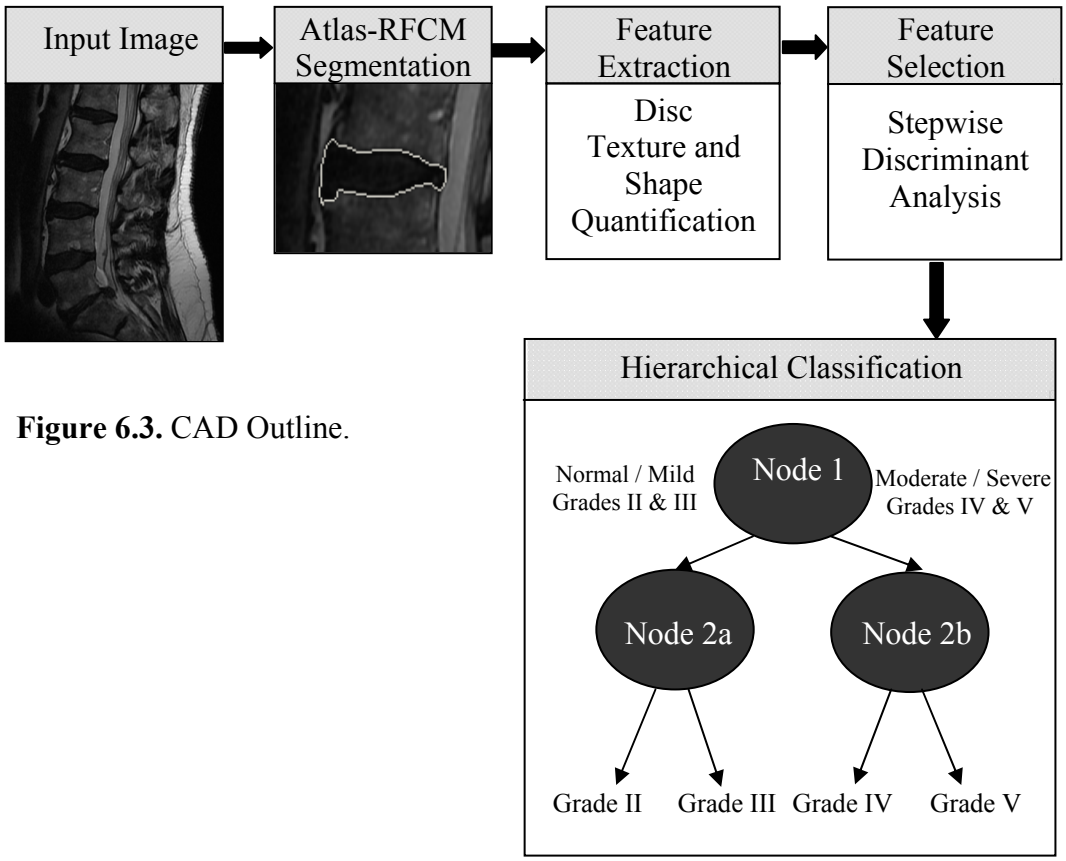


Figure 6.3. CAD Outline.

quantification. This chapter describes the classification scheme together with methods used for validating the system's performance.

Disc classification is based on the hierarchical scheme depicted in Figure 6.1, while four different classification algorithms were tested and compared aiming to optimize system performance. Finally, the Receiver Operating Characteristic Curve Analysis (ROC-Analysis) was employed for assessing the CAD performance.

6.3.2 Classification Scheme

The classification task of this study is to distinguish between four different grades of degeneration severity (II-V, according to the five grade Pfirmann's scheme). A hierarchical classification scheme was designed for this purpose and

its outline is illustrated in Figure 6.3. The first level of this classification scheme has a single node used for distinguishing light from severe pathology. Thus it is classifying normal and mildly degenerated discs (Grades II and III) in one class, and moderately and severely degenerated discs (Grades IV and V) in the other. The second level has two nodes, one for distinguishing between normal and mildly degenerated discs (Grade II from Grade III in node 2a), and the second for distinguishing between moderately and severely degenerated discs (Grades IV and V in node 2b).

Four different classification methods were employed and their performances were compared in order to identify the most suitable one for disc degeneration severity classification. These methods are the Minimum distance, Bayesian, Probabilistic Neural Networks and the distance weighted k-Nearest Neighbour classifier, which were briefly reviewed in Section 6.1.3. All four classification methods were trained and tested on the task of classifying the 255 intervertebral discs to degeneration severity classes using the hierarchical classification scheme. The classifiers' performance was evaluated using Receiver Operator Characteristic Curve Analysis as described in section 6.3.4.

6.3.3. Training and Testing the CAD System

The CAD accuracy was validated using a modified version of the leave-one-out method. This method is commonly used for testing CAD systems' performance [YOSH2002, BONI2006, DOI2007, KARA2007]. In leave-one-out, all but one of the data samples (here the intervertebral discs) are used for training the classification scheme, and the left-out sample is used for testing. This process is repeated recursively until all data samples have been used for testing, and the classification accuracy is calculated as the number of correctly classified left-out samples over their sum. The leave-one-out is a computationally intensive method, but offers the advantage that training is achieved with basically all samples, while maintaining the independence of training and testing sets. Thus, it provides an unbiased estimate of the classification error probability [JAIN2000, THEO2003].

Here, the leave-one-out method was modified by keeping out all five discs of each individual patient in each round. This modification was used for eliminating any bias that could be introduced in the classification process by using data from the same person for training and testing. This is because there might be some intra-patient dependence of discs' appearance and degeneration severity, although different image regions are used. The modified version is called leave-one-patient-out method. According to this method, the 5 lumbar discs of each individual patient were left-out of the training sample and the remaining 250 discs were used for training the CAD system. The left-out discs were then used for testing the systems accuracy and this process was recursively repeated for all 51 patients in the data sample.

6.3.4. CAD Performance Evaluation

A simple way to evaluate the performance of a CAD system is by measuring its classification accuracy, which is the fraction of cases that are correctly classified. Other simple measures, commonly used for performance evaluation are the sensitivity and specificity of the classification system. These measures values depend on the selection of a classification threshold. However, in most problems the data samples do not fall into one of two clearly defined categories, but there is an overlap between the distributions of the two data classes. In such cases, there is a trade off between the sensitivity and specificity of the classification system depending on the selection of its threshold value. The main limitation of the accuracy, sensitivity and specificity measures is that their values correspond to an arbitrary selected threshold value and cannot provide a global description of classification performance [METZ1988].

On the other hand, a different classification performance evaluation method called Receiver Operating Characteristic Curve Analysis (ROC-Analysis) intentionally forces the decision threshold to vary. The ROC curve is a plot of the True Positive Fraction (TPF) versus the False Positive Fraction (FPF). The TPF and FPF measures are defined in Equations 6.9-6.10. Figure 6.4. depicts the

$$TPF = \frac{tp}{tp + fn}\% \tag{6.9}$$

$$FPF = \frac{fp}{fp + tn}\% \tag{6.10}$$

where *tp* represents the number of true positive events, *fp* the number of false positives, *tn* the number of true negatives and *fn* the number of false negatives.

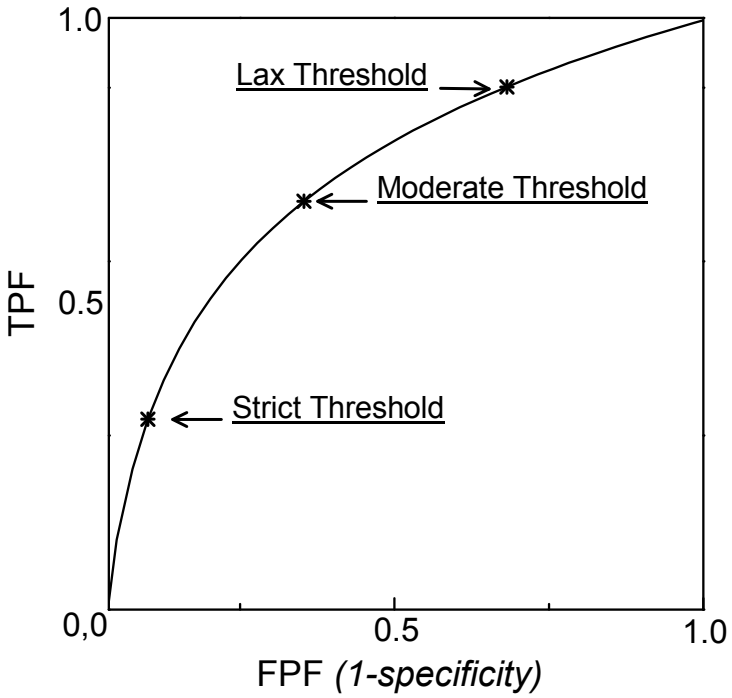


Figure 6.4. Example of a ROC curve and with points in the curve representing strict, moderated and lax thresholds.

ROC-space with a representative example of a ROC curve. Each point in the curve corresponds to a different decision threshold. The curve is produced by varying the decision threshold of the classification scheme and adding a new point pair of TPF and FPF values in the ROC space for each threshold value. The advantage of ROC analysis is that it provides a global description of classification performance, which is independent of the threshold value [METZ1988]. It provides richer information than scalar measures (such as accuracy, sensitivity and specificity) allowing more accurate comparison between different classification systems [FAWC2006].

Classification algorithms' performances can be compared by means of the Area Under the ROC curve (known as AUROC or A_z) measure. The AUROC

measure gives the probability of classifying a randomly chosen positive instance higher than a randomly chosen negative instance. Its values range between 0 and 1, with 0.5 representing random guessing [FAWC2006].

In the present study, ROC Analysis was used for evaluating the classification performance of the CAD system. First the performances of four different classification algorithms were tested, and the optimal one was selected to serve as basis for the CAD as described in section 6.3.2. System validation was performed by plotting the ROC curves and calculating the AUROC values for each node of the hierarchical classification scheme, together with their confidence intervals. In addition, Cohen's Kappa was used for testing the CAD agreement to clinical grading truth provided by radiologists' consensus, as well as each individual reader separately.

6.4. Results and discussion: CAD for Disc Degeneration severity classification

The CAD system presented here was trained and tested on a data sample of 255 intervertebral discs acquired by a 3 Tesla scanner. Radiologists' consensus served as ground truth, in an effort to incorporate multiple observers' knowledge in the classification process. The CAD system exploited disc textural and shape information for discriminating between different grades of disc degeneration. This information was given as input to the hierarchical classification scheme which classified the intervertebral discs to one of four degeneration severity grades.

Section 6.4.1 presents the classification results comparing different algorithms and justifying the classifier selection approach. Next, Section 6.4.2 presents the CAD agreement to clinical diagnosis, while Section 6.4.3. discusses further the CAD results.

6.4.1 Classification Algorithms Performance and Optimal Classifier Selection

As described in section 6.3.2., four different classification algorithms (MD, Bayesian, PNN and k-NN) were tested as part of the hierarchical classification scheme in order to identify the algorithm that would optimize the performance of the CAD system.

Figure 6.5 gives the ROC curves of these classification algorithms for node 1 of the hierarchical classification scheme (Figure 6.3), while Table 6.1 summarizes the corresponding Area Under ROC values for all four classifiers tested. These data represent the classifiers' ability to distinguish normal and mild (Grades II and III) form moderate and severe (Grades IV and V) disc degeneration. It is worth noting that there are very small differences between the ROC curves of the four classifiers. The k-NN appears to perform marginally better and this is also reflected by its slightly higher AUROC value as shown in Table 6.1.

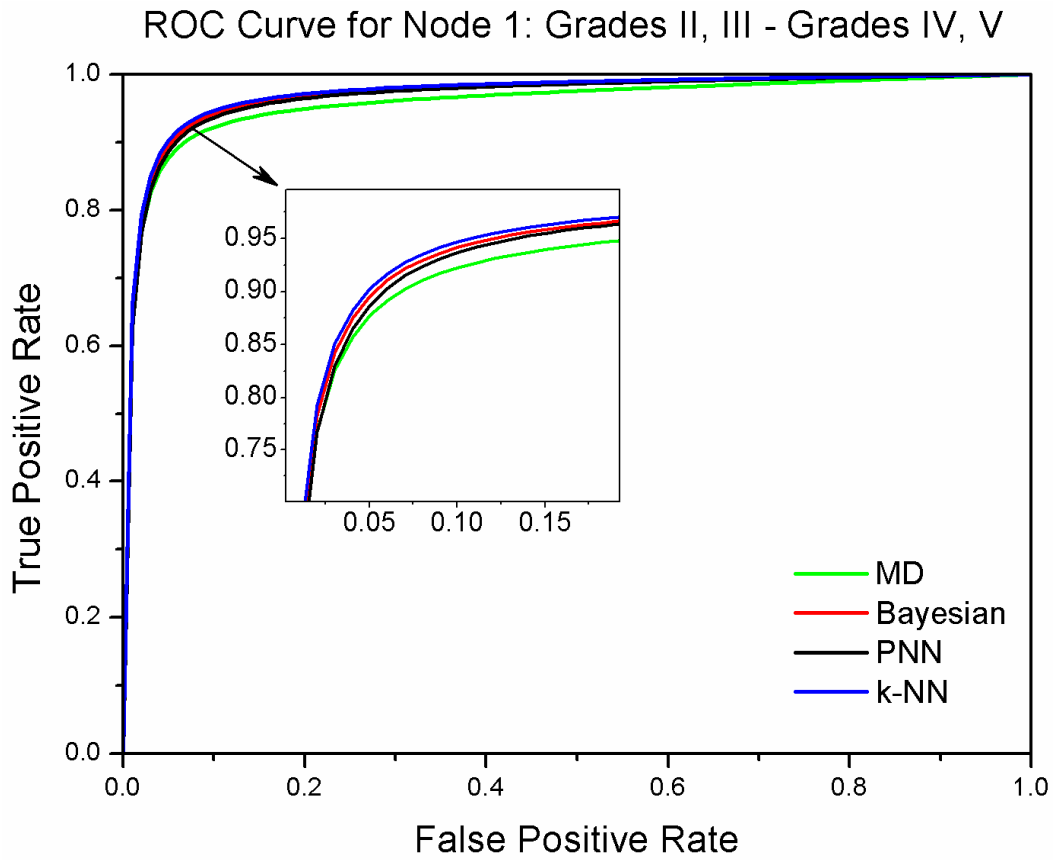


Figure 6.5. The four classifiers' ROC curves for Node 1 which distinguishes normal and mild discs from moderately and severely degenerated discs (Grades II and III from Grades IV and V). The zoom-in box highlights subtle differences between classifiers' performances.

Table 6.1. Classification performance for Node 1 by means of AUROC

Classifier	AUROC ± Standard Error	95% Confidence interval [Lower, upper]
MD	0.957 ± 0.014	[0.930 , 0.985]
Bayesian	0.969 ± 0.012	[0.946 , 0.992]
PNN	0.967 ± 0.012	[0.943 , 0.990]
k-NN	0.971 ± 0.013	[0.949 , 0.993]

Figure 6.6 gives the ROC curves of the classification algorithms for node 2a of the hierarchical classification scheme. This node is distinguishing normal (Grade II) from mild (Grade III) disc degeneration. Table 6.2 summarizes the corresponding Area Under ROC values for all four classifiers tested.

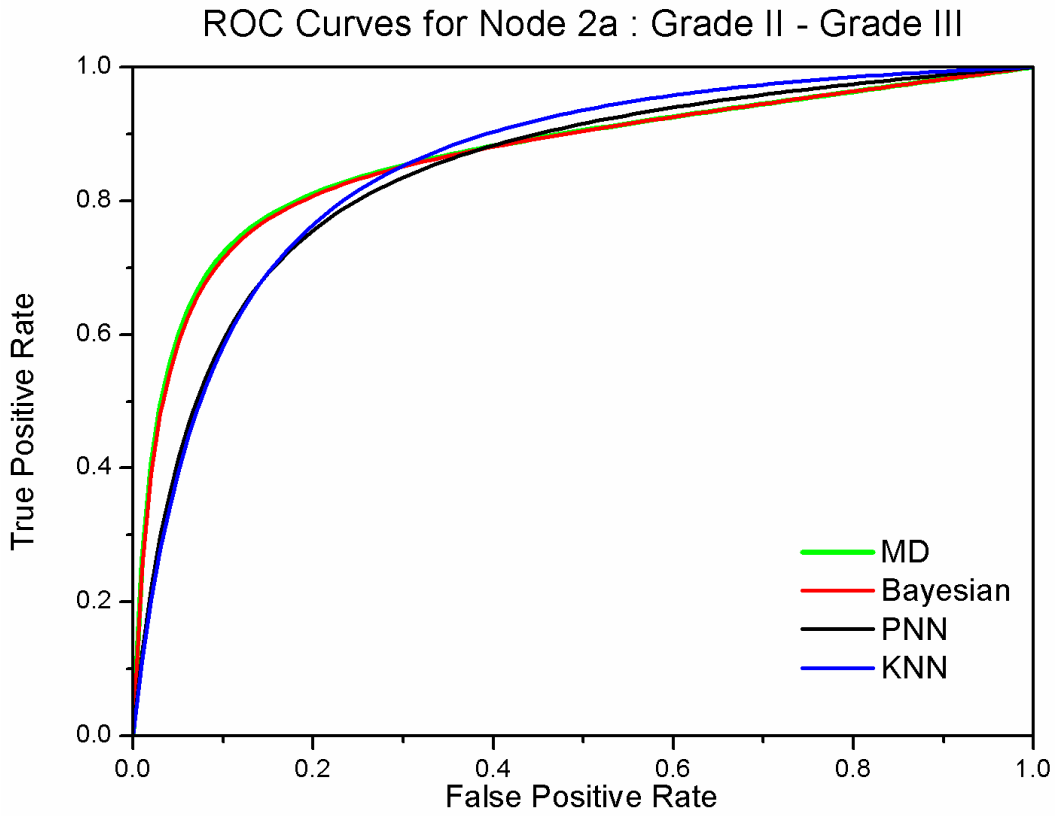


Figure 6.6. The classifiers' ROC curves for Node 2a which distinguishes between normal discs and discs with mild degeneration (Grade II from Grade III).

Table 6.2. Classification performance for Node 2a by means of AUROC

Classifier	AUROC ± Standard Error	95% Confidence interval [Lower, upper]
MD	0.867 ± 0.029	[0.809 , 0.925]
Bayesian	0.865 ± 0.030	[0.807 , 0.923]
PNN	0.844 ± 0.032	[0.781 , 0.907]
k-NN	0.854 ± 0.031	[0.793 , 0.915]

Finally, Figure 6.7 gives the ROC curves of the same classification algorithms for node 2b of the hierarchical classification scheme. These curves represent the classifiers' ability to distinguish between moderate and severe disc degeneration, Grade IV from Grade V. Finally, Table 6.3 summarizes the corresponding Area Under ROC values for all four classifiers tested. Again the differences between the ROC curves of the four classifiers are very small in both node 2a and node 2b. In both these cases, the Minimum Distance classifier appears to perform marginally better resulting in a slightly higher AUROC value that the other three classification algorithms.

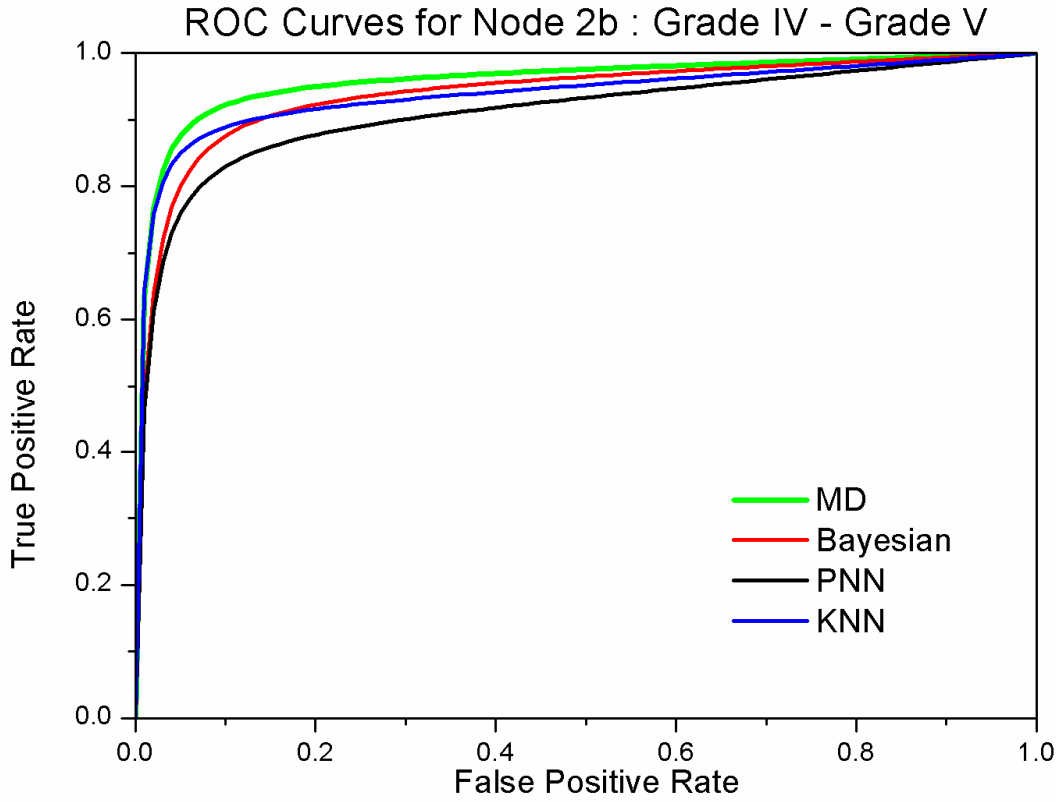


Figure 6.7. The classifiers' ROC curves for Node 2b which distinguishes between moderate and severe disc degeneration (Grade IV and Grade V).

Table 6.3. Classification performance for Node 2b by means of AUROC

Classifier	AUROC ± Standard Error	95% Confidence interval [Lower, upper]
MD	0.958 ± 0.017	[0.924 , 0.992]
Bayesian	0.939 ± 0.021	[0.897 , 0.981]
PNN	0.910 ± 0.027	[0.857 , 0.963]
k-NN	0.937 ± 0.022	[0.894 , 0.980]

The performance differences between the four classification algorithms are very small, as indicated by the ROC curves and AUROC values. Specifically, the k-NN classifier had the highest overall performance for distinguishing normal and mild from moderately and severely degenerated discs (Node 1) with the corresponding increase in its AUROC value being as small as 0.01. The results are similar for Nodes 2a and 2b with the Minimum Distance classifier displaying slightly higher AUROC values. However, the differences in AUROC values are extremely small with each classifier's AUROC value lying well within the others' confidence intervals. Consequently, none of the four classifiers appears to perform

better than the other, or equivalently all classifiers seem to provide equally good solutions.

For the final design of the CAD system, a single classification algorithm will be selected. According to Ockham's razor "entities must not be multiplied beyond necessity", or in other words "the simplest among equally good solutions is usually the optimal one" [William of Ockham]. In this case the simplest solution from a computational point of view is given by the Minimum Distance classifier. Thus, this algorithm was chosen here to serve the current classification task on the merit of simplicity but with adequate performance.

.

6.4.2. CAD Agreement to Clinical Diagnosis

Moving on to further evaluation of the CAD outcome, Table 6.4. gives the contingency matrix between CAD classification results and the ground truth provided by radiologists' consensus. The overall classification performance was 78.8%. The most challenging task was distinguishing normal (Grade II) from mildly degenerated discs (Grade III) in node 2a. These grades also yielded the greatest proportion of disagreement between the three radiologists' gradings as shown in Section 3.3.

Looking into the misclassified data, most cases represent a one grade difference. There is also one disc case with an error of two degrees, where a Grade II disc has been erroneously classified as Grade IV. Misclassified data are equally distributed over and under the principal diagonal of the contingency matrix indicating that the CAD does not yield any systematic over- or under-estimation.

Table 6.4. Contingency matrix between the CAD output and clinical grading

Consensus / CAD output	**Grade II**	**Grade III**	**Grade IV**	**Grade V**
Grade II	43	10	0	0
Grade III	8	68	12	0
Grade IV	1	12	66	1
Grade V	0	0	10	24

Table 6.5. provides Cohen's Kappa statistics evaluating the CAD system's agreement to the severity grading ground truth, as well as to each individual reader's grading separately. The average kappa value is equal to 0.87 representing almost perfect agreement between the CAD and radiologists' consensus, according to Landis and Koch's interpretation method summarized in Table 3.3 [LAND1977].

In addition, the kappa values for CAD agreement to individual radiologists' grading ranges from 0.69 to 0.88. It is noteworthy that kappa is higher for the most experienced radiologist (Reviewer 2), and lower for the least experienced one. Moreover, these values are higher than the interobserver agreements presented in Table 3.8. Thus the CAD system being trained with radiologists' consensus provides an outcome with higher agreement to the individual observers, than their agreement with each other. This is promising with respect to the CAD system's ability to improve interobserver agreement. However, the influence of the system on the radiologist's diagnostic accuracy (i.e. when the system acts as a second reader) needs to be evaluated.

Table 6.5. CAD Agreement to Clinical Grading

Observers Test	Kappa	95% Confidence Interval
CAD- Consensus Grading	0.8719	[0.7398 , 1.0000]
CAD- Reviewer 1	0.8092	[0.6457 , 0.9727]
CAD- Reviewer 2	0.8832	[0.7583 , 1.0000]
CAD- Reviewer 3	0.6867	[0.4892 , 0.8842]

6.5 Conclusion and Analysis of Research Objectives

Up to date, there is only a single other study reporting on CAD for disc pathologies from MRI [ALOM2010]. Alomari *et al*. presented a binary classification system distinguishing between normal and abnormal discs, as described in Section 6.2. That system is not specific to disc degeneration, since disc abnormality as defined by the authors includes various pathologies (degeneration, desiccation, herniation). Although the aims of the two systems differ, the reported performance appears to be similar. Specifically, Alomari *et al*. report average classification accuracy equal to 91.3%. The corresponding accuracy of the study presented in this thesis is 90.2%, when translated to a binary classification scheme (separating Grades II, III from Grades IV, V). The study by Alomari together with the study presented here, work towards establishing a new field for CAD applications, the diagnosis of intervertebral disc pathology from MRI.

Concluding, the CAD system described in the present study exploits textural, shape and disc context information for classifying the intervertebral discs to classes of degeneration severity. The system resulted in almost perfect agreement to the clinical grading provided by expert radiologists' consensus. This CAD system could be a valuable tool in hands of physicians to support clinical diagnosis of disc degeneration, but first its effect when used as a second reader needs to be tested.

Tables 6.6 and 6.7 summarize the objectives and hypotheses of the current chapter discussing the extent to which they were tested.

Table 6.6. Research Objectives Summary and Discussion : CAD

9. To design a computer aided diagnosis system putting together the segmentation and quantification steps with a classification scheme for categorizing the discs to disc degeneration severity grades
A CAD system was designed exploiting the segmentation and quantification steps for disc degeneration diagnosis. A hierarchical classification scheme

was developed for classifying the discs as normal, mildly, moderately, or severely degenerated utilizing texture and shape information extracted from the segmented disc region.

10. To develop methods for training and testing the CAD system using radiologists consensus as grading truth, and evaluating its classification performance.

    The CAD system was trained and tested using the "leave-one-patient-out" method, while ROC analysis was employed for evaluating classification performance. Results indicate good classification performance and an almost perfect agreement between the CAD output and radiologists' consensus grading (κ=0.87).

11. To apply the CAD system in a new dataset and test its influence on the radiologist's opinion when the system is used as a second reader

    This objective has not been met but the CAD effect in clinical diagnosis is currently under investigation. A new dataset comprising 120 intervertebral discs from 24 patients' lumbar scans has been collected. The three radiologists are currently on the process of clinically grading disc degeneration using the CAD output as a second opinion. The effect of the proposed CAD system on clinical diagnosis will be evaluated though the resulting changes in observers' agreement.

Table 6.7. Research Hypotheses Summary and Discussion: CAD

5. The CAD system can help improve diagnostic accuracy when used as a second reader. This would be supported if the disagreement between different clinical experts decreases when using the CAD.

   Not confirmed: *this hypothesis is currently being tested by evaluating the influence of CAD on radiologists' decision making using a new dataset comprising 120 discs.*

Chapter 7. Conclusion

This chapter summarizes the results of this thesis and outlines its contributions in Section 7.1. The limitations of this work are presented in Section 7.2, while Section 7.3 discusses the open questions and offers a summary of ongoing work together with suggestions for future research.

7.1. Thesis Contributions

The present study had three main aims: the segmentation of intervertebral discs, the quantification of disc properties for evaluating disc degeneration and the development of a computer aided diagnostic system for grading degeneration severity. The contributions of this thesis with respect to each task are summarized in Table 10.3 and discussed below.

Table 7.1. Contributions of this Study

Intervertebral Disc Segmentation
A novel atlas-based segmentation method was developed and tested: • this method can segment both normal and degenerated discs • it provides accurate segmentation meeting clinical requirements • it offers important interaction time savings in comparison to manual disc segmentation
Disc Degeneration Quantification
An alternative approach to disc quantification was introduced for the evaluation of degeneration severity: • this method exploits textural and shape features providing additional information to the standard method based on mean disc signal intensity • textural features measuring disc image homogeneity where found to be particularly sensitive in evaluating early degenerative changes • disc shape information is be able to distinguish among the advanced stages of degeneration • the proposed method analyses conventional T2 weighted images, and thus can be directly applicable in clinical practice
Computer Aided Diagnosis
A CAD system was developed for grading degeneration severity: • providing accurate classification into four severity grades • resulting in high agreement to the clinical ground truth, yet further testing is required to assess the system's influence on radiologists' opinion

With respect to disc segmentation, three different approaches were tested. The optimal one is Atlas-RFCM, a method that incorporates prior anatomical knowledge in the segmentation process by means of a probabilistic disc atlas, combined with fuzzy clustering techniques and smoothness constraints. This method resulted in accurate segmentation of both normal and degenerated intervertebral discs, meeting the clinical requirements. Moreover, the segmentation process is computationally inexpensive offering important time savings over manual segmentation. To the best of our knowledge, this is the first study working towards the segmentation of degenerated discs and reporting on the resulting accuracy. Overall, the proposed method was shown to be a useful tool for segmenting the intervertebral discs prior to the quantification and computer-aided diagnosis tasks.

Moving on to quantification, an alternative approach to the standard method of mean disc intensity quantification was introduced here. Specifically, the proposed system exploited texture and shape information extracted from conventional T2-weighted magnetic resonance images of the lumbar spine for the quantification of degeneration severity. Texture and shape analysis has been shown to be a valuable tool in a variety of diagnostic problems, and the present study exploited these methods in a new application which showed promising results. Specifically, textural features quantifying disc signal intensity homogeneity (in this case co-occurrence derived features such as the sum of squares and information measure of correlation) were strongly associated to intervertebral disc degeneration. These features capture degenerative alterations related to the progressive loss of nucleus-annulus distinction, and are particularly sensitive to distinguishing among the early stages of disease. Moreover disc shape information (expressed by the solidity feature) can help distinguishing between the moderate and severe degeneration states, which affect the disc's outline. The proposed approach can support clinical diagnosis of disc degeneration from conventional T2-weighted images in a precise and repeatable manner. Disc texture and shape quantification could offer a valuable tool for tracking the evolution of the disease. Moreover, they could be used for monitoring the response to

treatment, a particularly important task for the development of new treatment methods.

Finally, a novel CAD system was designed exploiting the previously extracted quantitative information for disc degeneration diagnosis. This is the first study reporting on CAD for degeneration severity grading. The CAD system presented here provided accurate classification of intervertebral discs to classes representing normal, mild, moderate and severe degeneration, performing in high agreement to clinical grading provided by expert radiologists. The CAD is intended to be used as a second reader aiming to increase diagnostic accuracy and improve interobserver agreement, but further investigation is required to test the system's effectiveness in clinical practice.

7.2. Thesis Limitations

Table 7.2 presents the limitations of this study. In my opinion the two most significant limitations are the severity grading ground truth and CAD testing issues, which are further discussed below.

Table 7.2. Limitations of this Study

Data Collection and Clinical Grading
• The MR data analysed in this study are 2D, since conventional clinical imaging protocols were utilized. However volumetric data obtained by 3D protocols would be desirable, as further described in the limitations for disc quantification. Although 3D imaging of the spine is possible the application of 3D protocols in clinical practice is limited. • The ground truth for degeneration severity relies on qualitative evaluation based on the disc appearance in MRI, yet no histological confirmation is available. It is neither ethical nor practicable to obtain disc samples for histological examination from low back pain patients, as in most cases they follow conservative therapy. • There are no Grade I discs, since no young patients data were collected in this study.
Intervertebral Disc Segmentation
• The segmentation method is working in 2D. However, its extension to 3D is straightforward as demonstrated by the pseudo-3D approach • The probabilistic atlas is designed on the basis of normal discs only, and this could introduce shape bias particularly in cases of degenerated discs where the boundary shape differs.
Disc Degeneration Quantification
• The texture and shape quantification relies on 2D analysis, due to the limitations imposed by data. Such limitations would be of particular importance in the case of shape analysis for the evaluation of disc space narrowing and herniation, as no volumetric information could be extracted.

Computer Aided Diagnosis

- The CAD influence in clinical diagnosis when the system is used as a second reader remains to be tested

The severity grading ground truth used for evaluating the quantification results and training the CAD system, is based on the visual evaluation of MR images. A histological confirmation of degenerative findings would be desirable in order to establish a more objective ground truth. However, obtaining histological samples is difficult and would only be possible in cases of patients undergoing spine surgery and only for the levels being treated. Here, in an effort to establish a more reliable ground truth three radiologists were asked to review the data independently and their consensus grading was utilized.

The proposed CAD system has not been adequately tested. Specifically, the system yielded good classification accuracy and almost perfect agreement to clinical grading, but its effectiveness as a decision making aid remains to be tested. It is highly important to evaluate how the CAD system influences the radiologist and affects the decision making process when used as a second reader, but this was not possible so far due to limited clinical time availability. Currently, the three radiologists who carried out the initial gradings of the 3.0 Tesla dataset are evaluating a new dataset comprising 120 discs. They review the images first without and then with the help of CAD and record both gradings. This evaluation aims into investigating the effect of CAD on interobserver agreement.

7.3. Open Questions and Future Research

This section summarizes some future research steps in the fields of disc segmentation, quantification and computer aided diagnosis, linked to the methods and results of this study. Moreover, it offers an outline of the tasks currently being investigated in continuation of the work presented in this thesis.

Further work on **intervertebral disc segmentation** could include:

- modification of the Atlas-Based segmentation method by incorporating shape information in the FCM framework aiming to control the influence of this shape prior on the segmentation process and reduce the corresponding bias inherent to atlas-based methods [BAZI2007].
- testing the proposed atlas-based segmentation on volumetric data, and designing a true-3D approach.
- expanding the current methods for the segmentation of vertebral bodies and/or other spinal structures.

For **disc quantification**:

- further quantitative features could be tested for quantifying disc degeneration severity.
- the texture and shape information quantified in this study could be exploited for evaluating other specific disc pathologies such as narrowing and herniation.
- this quantitative analysis could be applied on 3D data in order to obtain volumetric measurements of the disc shape
- longitudinal studies could be used for further investigating the ability of disc quantitative analysis in monitoring the disease progress
- texture and shape features could be exploited as biomarkers for evaluating the effect of new therapeutic methods

With respect to **CAD** for intervertebral disc degeneration:

- more elaborate classification schemes could be tested

- a larger dataset can be used for training the system (also including Grade I discs)
- and just as with the quantification techniques, the current system could be adapted for evaluating other specific disc pathologies such as herniation and narrowing

Currently, the tools presented in this thesis are being exploited for further evaluating disc pathology. Specifically, in collaboration with the Magnetic Resonance and Image Analysis Research Centre (MARIARC) of the University of Liverpool the quantification tools presented here are exploited for evaluating the effect of Alkaptonuria on the intervertebral discs. Alkaptonuria (AKU) is a metabolic disorder, which results in cartilaginous tissue damage [FIND2010]. It is a rare condition and little is known about the disease pathways, but our collaborators in Liverpool believe that spine is the first structure to be affected [personal communication with Dr Lakshminarayan Ranganath, University of Liverpool]. Our role in this project is to quantify the disc alterations from MR images of AKU patients and provide biomarkers for assessing the disease progress. This analysis could be exploited in the future for testing the effect of AKU treatment (treatment is currently being tested on an animal model, in Liverpool).

Moreover, the quantification and computer aided diagnosis of disc herniation is being tested. Specifically, a preliminary study on hernia segmentation and quantification has been conducted on a small dataset (100 discs). Results indicated that shape analysis methods can provide precise and repeatable quantification of the herniation. This study was presented by Professor Andrew Todd-Pokropek in UKRC 2010. In addition, the output of this work was used for designing a CAD system for lumbar disc hernia characterization. Preliminary results were presented in the Computer Assisted Radiology and surgery Conference (CARS2010). These studies demonstrated the potential of disc shape analysis for hernia quantification and computer aided diagnosis. However, 3D data are required for obtaining volumetric measurements of the disc

and hernia size and shape, and thus getting a more precise quantitative evaluation of this condition.

Finally, the data collection continues, while as described in the previous section a new set of 120 discs is currently being reviewed by the three radiologists who reviewed the initial 3.0 Tesla data, in order to test the influence of the CAD system on radiologists' opinions when it is used as a second reader.

References

ADAM2006 Adams, M.A., Roughley, P.J., 2006. What is intervertebral disc degeneration, and what causes it? Spine 31, 2151-2161.

ALOM2010 Alomari, R., Corso, J., Chaudhary, V., Dhillon, G., 2010. Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI. International Journal of Computer Assisted Radiology and Surgery 5, 287-293.

AN2004 An, H.S., Anderson, P.A., Haughton, V.M., Iatridis, J.C., Kang, J.D., Lotz, J.C., Natarajan, R.N., Oegema Jr, T.R., Roughley, P., Setton, L.A., Urban, J.P., Videman, T., Andersson, G.B.J., Weinstein, J.N., 2004. Introduction. Disc degeneration: Summary. Spine 29, 2677-2678.

ANDE1999 Andersson, G.B.J., 1999. Epidemiological features of chronic low-back pain. Lancet 354, 581-585.

ARAN2010 Arana, E., Royuela, A., Kovacs, F.M., Estremera, A., Sarasbar, H., Amengual, G., Galarraga, I., MartÃ-nez, C., Muriel, A., Abraira, V., Del Real, M.T.G., Zamora, J., Campillo, C., Lumbar spine: Agreement in the interpretation of 1.5-T MR images by using the nordic modic consensus group classification form. Radiology 254, 809-817.

AUER2006 Auerbach, J.D., Johannessen, W., Borthakur, A., Wheaton, A.J., Dolinskas, C.A., Balderston, R.A., Reddy, R., Elliott, D.M., 2006. In vivo quantification of human lumbar disc degeneration using T1ρ-weighted magnetic resonance imaging. European Spine Journal 15.

AWAI2004 Awai, K., Murao, K., Ozawa, A., Komi, M., Hayakawa, H., Hori, S., Nishimura, Y., 2004. Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on Radiologists Detection Performance1. Radiology 230, 347-352.

AYRE2005 Ayres, F., Rangayyan, R., 2005. Characterization of architectural distortion in mammograms. IEEE engineering in medicine and biology 24, 59-67.

BANK2000 Bankman, I., Spisz, T., Pavlopoulos, S., 2000. Two-Dimensional Shape and Texture Quantification. In: Bankman, I. (Ed.), Handbook of medical imaging. Academic Press, USA, pp. 215-230.

BART1998 Bartlett, R.J.V., Hill, C.R., Gardiner, E., 1998. A comparison of T2 and gadolinium enhanced MRI with CT myelography in cervical radiculopathy. British Journal of Radiology 71, 11-19.

BATT1995 Battie, M.C., Haynor, D.R., Fisher, L.D., Gill, S.K., Gibbons, L.E., Videman, T., 1995. Similarities in degenerative findings on magnetic resonance images of the lumbar spines of identical twins. Journal of Bone and Joint Surgery - Series A 77, 1662-1670.

BATT2004 Battie, M.C., Videman, T., Parent, E., 2004. Lumbar disc degeneration: Epidemiology and genetic influences. Spine 29, 2679-2690.

BAZI2007 Bazin, P.L., Pham, D., Statistical and topological atlas based brain image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007, 94-101.

BELK2007 Belkoff, S., 2007. Biomechanics of the Spine. In: Van Goethem, J., Van den Hauwe, L., Parizel, P. (Eds.), Spinal Imaging Diagnostic Imaging of the Spine and Spinal Cord. Springer-Verlag Berlin Heidelberg

BENN2005 Benneker, L.M., Heini, P.F., Anderson, S.E., Alini, M., Ito, K., 2005. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. European Spine Journal 14, 27-35.

BEZD1984 Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences 10, 191-203.

BEZD1993 Bezdek, J.C., Hall, L.O., Clarke, L.P., 1993. Review of MR image segmentation techniques using pattern recognition. Medical Physics 20, 1033-1048.

BIRD2001 Birdwell, R.L., Ikeda, D.M., Shaughnessy, K.F., Sickles, E.A., 2001. Mammographic Characteristics of 115 Missed Cancers Later Detected with Screening Mammography and the Potential Utility of Computer-aided Detection1. Radiology 219, 192-202

BLUM2010 Blumenkrantz, G., Zuo, J., Li, X., Kornak, J., Link, T., Majumdar, S., In vivo 3.0-tesla magnetic resonance T2 relaxation mapping in subjects with intervertebral disc degeneration and clinical symptoms. Magnetic Resonance in Medicine 63, 1193-1200.

BODE1996 Boden, S.D., 1996. The use of radiographic imaging studies in the evaluation of patients who have degenerative disorders of the

lumbar spine. Journal of Bone and Joint Surgery - Series A 78, 114-124.

BOGD1997 Bogduk, N., 1997. Clinical Anatomy of the Lumbar Spine and Sacrum. Churchill Livingstone.

BONI2006 Boniatis, I.S., Costaridou, L.I., Cavouras, D.A., Panagiotopoulos, E.C., Panayiotakis, G.S., 2006. Quantitative assessment of hip osteoarthritis based on image texture analysis. British Journal of Radiology 79, 232-238.

BONI2007a Boniatis, I., Cavouras, D., Costaridou, L., Kalatzis, I., Panagiotopoulos, E., Panayiotakis, G., 2007. Computer-aided grading and quantification of hip osteoarthritis severity employing shape descriptors of radiographic hip joint space. Computers in biology and medicine 37, 1786-1795.

BONI2007b Boniatis, I., Costaridou, L., Cavouras, D., Kalatzis, I., Panagiotopoulos, E., Panayiotakis, G., 2007. Assessing hip osteoarthritis severity utilizing a probabilistic neural network based classification scheme. Medical engineering & physics 29, 227-237.

BRAI1998 Braithwaite, I., White, J., Saifuddin, A., Renton, P., Taylor, B.A., 1998. Vertebral end-plate (Modic) changes on lumbar spine MRI: Correlation with pain reproduction at lumbar discography. European Spine Journal 7, 363-368.

BRIN1998 Brinkmann, B.H., Manduca, A., Robb, R.A., 1998. Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. IEEE Transactions on Medical Imaging 17, 161-171.

BROW1992 Brown, L.G., 1992. Survey of image registration techniques. ACM Computing Surveys 24, 325-376.

CARR2009a Carragee, E.J., Don, A.S., Hurwitz, E.L., Cuellar, J.M., Carrino, J., Herzog, R., 2009. Does Discography Cause Accelerated Progression of Degeneration Changes in the Lumbar Disc. Spine 34, 2338–2345.

CARR2009b Carrino, J.A., Lurie, J.D., Tosteson, A.N.A., Tosteson, T.D., Carragee, E.J., Kaiser, J., Grove, M.R., Blood, E., Pearson, L.H., Weinstein, N.J., Herzog, R., 2009. Lumbar spine: Reliability of MR imaging findings. Radiology 250, 161-170.

CAST2004 Castellano, G., Bonilha, L., Li, L.M., Cendes, F., 2004. Texture analysis of medical images. Clinical Radiology 59, 1061-1069.

CHAL1997 Chalana, V., Kim, Y., 1997. A methodology for evaluation of boundary detection algorithms on medical images. IEEE Transactions on Medical Imaging 16, 642-652.

CHAR2004 Chamarthy, P., Stanley, R.J., Cizek, G., Long, R., Antani, S., Thoma, G., 2004. Image analysis techniques for characterizing disc space narrowing in cervical vertebrae interfaces. Computerized Medical Imaging and Graphics 28, 39-50.

CHAN1998 Chan, H.P., Sahiner, B., Lam, K.L., Petrick, N., Helvie, M.A., Goodsitt, M.M., Adler, D.D., 1998. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. Medical Physics 25, 2007-2018.

CHEN1999 Chen, D.R., Chang, R.F., Huang, Y.L., 1999. Computer-aided diagnosis applied to US of solid breast nodules by using neural networks. Radiology 213, 407-412.

CHEN2007 Chen, W., Giger, M.L., Li, H., Bick, U., Newstead, G.M., 2007. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. Magnetic Resonance in Medicine 58, 562-571.

CHER2004 Cherukuri, M., Stanley, R.J., Long, R., Antani, S., Thoma, G., 2004. Anterior osteophyte discrimination in lumbar vertebrae using size-invariant features. Computerized Medical Imaging and Graphics 28, 99-108.

CHEV2007 Chevrefils, C., ChEriet, F., Grimard, G., Aubin, C.E., 2007. Watershed segmentation of intervertebral disk and spinal canal from MRI images. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 1017-1027.

CHEV2009 Chevrefils, C., Cheriet, F., Aubin, C.Ã., Grimard, G., 2009. Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from MR images. IEEE Transactions on Information Technology in Biomedicine 13, 608-620.

COLE1979 Coleman, G.B., Andrews, H.C., 1979. Image Segmentation by Clustering. Proc IEEE 67, 773-785.

CONR2004 Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lörch, T., Ellenberg, J., Pepperkok, R., Eils, R., 2004. Automatic identification of subcellular phenotypes on human cell arrays. Genome Research 14, 1130.

COOT1992 Cootes, T., 1992. Active shape models. Springer-Verlag.

COOT2000 Cootes, T., 2000. 'Model-Based Methods in Analysis of Biomedical Images' Image Processing and Analysis. Oxford University Press, Oxford, USA.

COST2005 Costaridou, L. (Ed.), 2005. Medical Image Analysis Methods. Taylor and Francis Group, USA.

COST2010 Costaridou, L., 2010. CADx in Mammography. Biomedical Image Processing - Methods and Applications. Deserno, T. (Ed.), Springer-Verlag.

DAM2007 Dam, E.B., 2007. Quantitative Automated Musculoskeletal Analysis. Academic radiology 14, 1153-1155.

DOI1999 Doi, K., MacMahon, H., Katsuragawa, S., Nishikawa, R.M., Jiang, Y.J., 1999. Computer-aided diagnosis in radiology: potential and pitfalls. European Journal of Radiology 31, 97-109.

DOI2007 Doi, K., 2007. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 31, 198-211.

DRAK2004 Drake, R., Vogl, W., Mitchell, A., 2004. Gray's anatomy for students. Churchill Livingstone.

DRAK2008 Drake, R., Vogl, W., Mitchell, A., Tibbitts, R., Richardson, P., 2008. Gray's Atlas of Anatomy. Churchill Livingstone.

DUDA1976 Dudani, S.A., 1976. Distance-Weighted K-Nearest-Neighbor Rule. IEEE Transactions on Systems, Man and Cybernetics SMC-6, 325-327.

DUDA2001 Duda, R., Hart, P., Stork, G., 2001. Pattern Classification. John Wiley & Sons, USA.

DURG1996 Durgin, F.H., Proffitt, D.R., 1996. Visual learning in the perception of texture: Simple and contingent aftereffects of texture density. Spatial Vision 9, 423-474.

FARD2001 Fardon, D.F., Milette, P.C., 2001. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Spine 26.

FAWC2006 Fawcett, T., 2006. An introduction to ROC analysis. Pattern recognition letters 27, 861-874.

FIND2010 Findakure, 2010. Find Alkaptonuria Research Project.

FUJI1999 Fujiwara, A., Tamai, K., Yamato, M., An, H.S., Yoshida, H., Saotome, K., Kurihashi, A., 1999. The relationship between facet joint osteoarthritis and disc degeneration of the lumbar spine: An MRI study. European Spine Journal 8, 396-401.

GALL1975 Galloway, M.M., 1975. Texture analysis using gray level run lengths. Computer Graphics and Image Processing 4, 172-179.

GIGE2008 Giger, M.L., Chan, H.P., Boone, J., 2008. Anniversary paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. Medical Physics 35, 5799-5820.

GONZ2002 Gonzalez, R.C., Woods, R.E., 2002. Digital Image Processing. Prentice Hall.

GRAY1918 Gray, H., 1918. Anatomy of the Human Body. In: Febiger, L. (Ed.). bartleby.com.

HAJM2003 Hajmeer, M., Basheer, I., 2003. Comparison of logistic regression and neural network-based classifiers for bacterial growth. Food Microbiology 20, 43-55.

HARA1973 Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics smc 3, 610-621.

HARA1979 Haralick, R.M., 1979. Statistical And Structural Approaches To Texture. Proc IEEE 67, 786-804.

HARA1985 Haralick, R.M., Shapiro, L.G., 1985. Image segmentation techniques. Computer Vision, Graphics, and Image Processing 29, 100-132.

HAUG2006 Haughton, V., 2006. Imaging intervertebral disc degeneration. Journal of Bone and Joint Surgery - Series A 88, 15-20.

HILI2004 Hilibrand, A.S., Robbins, M., 2004. Adjacent segment degeneration and adjacent segment disease: The consequences of spinal fusion? Spine Journal 4.

HU1962 Hu, M.-K., 1962. Visual pattern recognition by moment invariants. Information Theory, IRE Transactions on 8, 179-187.

HUDG1983 Hudgins, W.R., 1983. Computer-aided diagnosis of lumbar disc herniation. Spine 8, 604-615.

JAIN2000 Jain, A.K., Duin, R.P.W., Mao, J., 2000. Statistical Pattern Recognition: A Review. IEEE Trans. Pattern Anal. Mach. Intell. 22, 4-37.

JENS1994 Jensen, M.C., Brant-Zawadzki, M.N., Obuchowski, N., Modic, M.T., Malkasian, D., Ross, J.S., 1994. Magnetic resonance imaging of the lumbar spine in people without back pain. New England Journal of Medicine 331, 69-73.

JINK2003 Jinkins, J.R., Dworkin, J.S., Green, C.A., Greenhalgh, J.F., Gianni, M., Gelbien, M., Wolf, R.B., Damadian, J., Damadian, R.V., 2003. Upright, weight-bearing, dynamic-kinetic magnetic resonance imaging of the spine - Review of the first clinical results. Journal of the Hong Kong College of Radiologists 6, 55-74.

JOHA2006 Johannessen, W., Auerbach, J.D., Wheaton, A.J., Kurji, A., Borthakur, A., Reddy, R., Elliott, D.M., 2006. Assessment of human disc degeneration and proteoglycan content using T1ρ weighted magnetic resonance imaging. Spine 31, 1253-1257.

KARA2007 Karahaliou, A., Skiadopoulos, S., Boniatis, I., Sakellaropoulos, P., Likaki, E., Panayiotakis, G., Costaridou, L., 2007. Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis. Br J Radiol 80, 648-656.

KARA2009 Karahaliou, A., 2009. Texture Analysis of Mammographic Images for Computer-Aided Breast Cancer Diagnosis. Department of Medical Physics. University of Patras, Patras.

KASS1988 Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. International Journal of Computer Vision 1, 321-331.

KATO2007 Kato, H., Kanematsu, M., Zhang, X., Saio, M., Kondo, H., Goshima, S., Fujita, H., 2007. Computer-Aided Diagnosis of Hepatic Fibrosis: Preliminary Evaluation of MRI Texture Analysis Using the Finite Difference Method and an Artificial Neural Network. Am. J. Roentgenol. 189, 117-122.

KATZ2006 Katz, J.N., 2006. Lumbar disc disorders and low-back pain: Socioeconomic factors and consequences. Journal of Bone and Joint Surgery - Series A 88, 21-24.

KERT2001a Kerttula, L., 2001. Magnetic Resonance Imaging of the Intervertebral Disc: Post-traumatic findings and the value of diffusion-weighted MR imaging. Department of Diagnostic Radiology. University of Oulu, Oulu.

KERT2001b Kerttula, L., Kurunlahti, M., Jauhiainen, J., Koivula, A., Oikarinen, J., Tervonen, O., 2001. Apparent diffusion coefficients and T2 relaxation time measurements to evaluate disc degeneration : A quantitative MR study of young patients with previous vertebral fracture. Acta Radiologica 42, 585-591.

KETT2006 Kettler, A., Wilke, H.J., 2006. Review of existing grading systems for cervical or lumbar disc and facet joint degeneration. European Spine Journal 15, 705-718.

KIM2003 Kim, D.Y., Kim, J.H., Noh, S.M., Park, J.W., 2003. Pulmonary nodule detection using chest CT images. Acta Radiologica 44, 252-257.

KJAE1995 Kjaer, L., Ring, P., Thomsen, C., Henriksen, O., 1995. Texture analysis in quantitative MR imaging tissue characterisation of normal brain and intracranial tumours at 1.5 T. Acta Radiologica 36, 127-135.

KJAE2005 Kjaer, P., Leboeuf-Yde, C., Korsholm, L., Sorensen, J.S., Bendix, T., 2005. Magnetic resonance imaging and low back pain in adults: A diagnostic imaging study of 40-year-old men and women. Spine 30, 1173-1180.

KJAE2006 Kjaer, P., Korsholm, L., Bendix, T., Sorensen, J.S., Leboeuf-Yde, C., 2006. Modic changes and their associations with clinical findings. European Spine Journal 15, 1312-1319.

KOES2001 Koes, B.W., Van Tulder, M.W., Ostelo, R., Kim Burton, A., Waddell, G., 2001. Clinical guidelines for the management of low back pain in primary care: An international comparison. Spine 26, 2504-2513.

KOOM2006 Koompairojn, S., Hua, K.A., Bhadrakom, C., 2006. Automatic classification system for lumbar spine X-ray images. Proceedings - IEEE Symposium on Computer-Based Medical Systems, pp. 213-218.

KORF2007 Korfiatis, P., Skiadopoulos, S., Sakellaropoulos, P., Kalogeropoulou, C., Costaridou, L., 2007. Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT. British Journal of Radiology 80, 996-1005.

KORF2009 Korfiatis, P., Karahaliou, A., Kalogeropoulou, C., Lazamtzo, A., Costaridou, L., 2009. Texture Based Identification and Characterization of Interstitial Pneumonia Patterns in Lung Multidetector CT. Information Technology in Biomedicine, IEEE Transactions on PP, 1-1.

KOVE2000 Kovesi, P., 2000. MATLAB and Octave functions for computer vision and image processing. School Comput. Sci. Softw. Eng., Univ. Western Australia, Perth, W.A., Australia p. http://www.csse.uwa.edu.au/pk/research/matlabfns/.

KUND2003 Kundel, H.L., Polansky, M., 2003. Measurement of Observer Agreement1. Radiology 228, 303-308.

LAND1977 Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics, 159-174.

LAWS1980 Laws, K.I., 1980. Textured image segmentation. University of Southern California, IPI Report 940.

LERS1993 Lerski, R.A., Straughan, K., Schad, L.R., Boyce, D., Bluml, S., Zuna, I., 1993. VIII. MR image texture analysis - An approach to tissue characterization. Magnetic Resonance Imaging 11, 873-887.

LIKA2001 Likar, B., Viergever, M.A., PernuÅ¡, F., 2001. Retrospective correction of MR intensity inhomogeneity by information minimization. IEEE Transactions on Medical Imaging 20, 1398-1410.

LINE2007 Linek, M., Jungmann, M., Berlage, T., Pechnig, R., Clauser, C., 2007. Rock classification based on resistivity patterns. Journal of Geophysics and Engineering 4, 171-183.

LIU2007 Liu, J., Ma, W., 2007. An effective recognition method of breast cancer based on PCA and SVM algorithm. Medical Biometrics, Lecture Notes in Computer Science 4901, 57-64.

LOPE2009 Lopez, M., Ramirez, J., Gorriz, J., Salas-Gonzalez, D., Alvarez, I., Segovia, F., Puntonet, C., 2009. Automatic tool for alzheimer's disease diagnosis using PCA and bayesian classification rules. Electronics letters 45, 389—391.

LUOM2000 Luoma, K., RiihimAki, H., Luukkonen, R., Raininko, R., Viikari-Juntura, E., Lamminen, A., 2000. Low back pain in relation to lumbar disc degeneration. Spine 25, 487-492.

LUOM2001 Luoma, K., Vehmas, T., RiihimAki, H., Raininko, R., 2001. Disc height and signal intensity of the nucleus pulposus on magnetic resonance imaging as indicators of lumbar disc degeneration. Spine 26, 680-686.

MAJU2006 Majumdar, S., 2006. Magnetic resonance imaging and spectroscopy of the intervertebral disc. NMR in Biomedicine 19, 894-903.

MANI2000 Maniadakis, N., Gray, A., 2000. The economic burden of back pain in the UK. Pain 84, 95-103.

MARI2009 Marinelli, N.L., Haughton, V.M., MuÃ±oz, A., Anderson, P.A., 2009. T2 relaxation times of intervertebral disc tissue correlated with water content and proteoglycan content. Spine 34, 520-524.

MAAT2007 Van der Maaten, L., Postma, E., Van den Herik, H., 2007. Dimensionality reduction: A comparative review. Computer and Information Science 10, 1-35, Published Online.

MATL2010 Matlab, 2010. Measure properties of image regions. http://www.mathworks.com/access/helpdesk/help/toolbox/images/regionprops.html

MCIN1996 McInerney, T., Terzopoulos, D., 1996. Deformable models in medical image analysis: A survey. Medical Image Analysis 1, 91-108.

MCIN2000a McInerney, T., Terzopoulos, D., 2000. Deformable Models. In: Bankman, I. (Ed.), Handbook of medical imaging. Academic Press, USA, pp. 127-145.

MCIN2000b McInerney, T., Terzopoulos, D., 2000. T-snakes: Topology adaptive snakes. Medical Image Analysis 4, 73-91.

MCNI1999 McNitt-Gray, M.F., Hart, E.M., Wyckoff, N., Sayre, J.W., Goldin, J.G., Aberle, D.R., 1999. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. Medical Physics 26, 880.

METZ1978 Metz, C.E., 1978. Basic principles of ROC analysis+. Elsevier, pp. 283-298.

MILE1997 Milette, P.C., 1997. The proper terminology for reporting lumbar intervertebral disk disorders. American Journal of Neuroradiology 18, 1859-1866.

MODI1988 Modic, M.T., Masaryk, T.J., Ross, J.S., Carter, J.R., 1988. Imaging of degenerative disk disease. Radiology 168, 177-186.

MODI2007 Modic, M.T., Ross, J.S., 2007. Lumbar degenerative disk disease. Radiology 245, 43-61.

MWAL2008 Mwale, F., Iatridis, J.C., Antoniou, J., 2008. Quantitative MRI as a diagnostic tool of intervertebral disc matrix composition and integrity. European Spine Journal 17.

NIEM2008 Niemelainen, R., Videman, T., Dhillon, S.S., Battie, M.C., 2008. Quantitative measurement of intervertebral disc signal using MRI. Clinical Radiology 63, 252-255.

NIXO2002 Nixon, M., Aguado, A., 2002. Feature Extraction and Image Processing Sons Inc.

OCKH2010 Ockham, 2010. Ockham's Razor. Encyclopaedia Britannica. http://www.britannica.com/EBchecked/topic/424706/Ockhams-razor

OTSU1979 Otsu, N., 1979. Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern SMC-9, 62-66.

PAAJ1997 Paajanen, H., Erkintalo, M., Parkkola, R., Salminen, J., Kormano, M., 1997. Age-dependent correlation of low-back pain and lumbar disc degeneration. Archives of Orthopaedic and Trauma Surgery 116, 106-107.

PAL1993 Pal, N.R., Pal, S.K., 1993. A review on image segmentation techniques. Pattern Recognition 26, 1277-1294.

PARI2007 Parizel, P., Van Goethem, J., Van den Hauwe, L., Voormolen, M., 2007. Degenerative Disc Disease. In: Van Goethem, J., Van den Hauwe, L., Parizel, P. (Eds.), Spinal Imaging Diagnostic Imaging of the Spine and Spinal Cord. Springer-Verlag Berlin Heidelberg

PARK2003 Park, H., Bland, P.H., Meyer, C.R., 2003. Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Transactions on Medical Imaging 22, 483-492.

PARV2006 Parveen, R., Ruff, C., Todd-Pokropek, A., 2006. Three Dimensional Tissue Classifications in MR Brain Images. Computer Vision Approaches to Medical Image Analysis, pp. 236-247.

PERI2003 Periaswamy, S., Farid, H., 2003. Elastic registration in the presence of intensity variations. IEEE Transactions on Medical Imaging 22, 865-874.

PERR2006 Perry, J., Haughton, V., Anderson, P.A., Wu, Y., Fine, J., Mistretta, C., 2006. The Value of T2 Relaxation Times to Characterize Lumbar Intervertebral Disks: Preliminary Results. AJNR Am J Neuroradiol 27, 337-342.

PFIR2001 Pfirrmann, C.W.A., Metzdorf, A., Zanetti, M., Hodler, J., Boos, N., 2001. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine 26, 1873-1878.

PHAM1999 Pham, D.L., Prince, J.L., 1999. Adaptive fuzzy segmentation of magnetic resonance images. IEEE Transactions on Medical Imaging 18, 737-752.

PHAM2000 Pham, D.L., Xu, C., Prince, J.L., 2000. Current methods in medical image segmentation. Annual Review of Biomedical Engineering, pp. 315-337.

PHAM2001 Pham, D.L., 2001. Spatial models for fuzzy clustering. Comput. Vis. Image Underst. 84, 285-297.

PRAT2001 Pratt, W., 2001. Digital Image Processing. John Wiley & Sons, New York, USA.

QAZI2007 Qazi, A.A., Dam, E.B., Olsen, O.F., Nielsen, M., Christiansen, C., 2007. Automatic detection of diseased regions in knee cartilage. Progress in Biomedical Optics and Imaging - Proceedings of SPIE.

RADI2010 radiologyinfo, 2010. Spine CT Dose. http://www.radiologyinfo.org/en/info.cfm?pg=spinect

RAIN1995 Raininko, R., Manninen, H., Battié, M.C., Gibbons, L.E., Gill, K., Fisher, L.D., 1995. Observer variability in the assessment of disc degeneration on magnetic resonance images of the lumbar and thoracic spine. Spine 20, 1029.

RANG2005 Rangayyan, R., 2005. Biomedical Image Analysis. CRC Press, Boca Raton, Florida, USA.

REED1993 Reed, T.R., du Buf, J.M.H., 1993. A Review of Recent Texture Segmentation and Feature Extraction Techniques. Computer Vision and Image Understanding 57, 359-372.

ROBE1997 Roberts, N., Gratin, C., Whitehouse, G.H., 1997. MRI analysis of lumbar intervertebral disc height in young and older populations. Journal of Magnetic Resonance Imaging 7, 880-886.

ROGO2000 Rogowska, J, 2000. Overview and Fundamentals of Medical Image Segmentation. In: Bankman, I. (Ed.), Handbook of medical imaging. Academic Press, USA, pp. 55-85.

ROHF2005 Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D.B., Maurer, J.C.R., Suri, J., Wilson, D.L., Laxminarayan, S., 2005. Quo Vadis, Atlas-Based Segmentation? , The Handbook of Medical Image Analysis -- Volume III: Registration Models. Kluwer Academic / Plenum Publishers, pp. 435-486.

ROSS2010 Ross, J.S., 2010. Babel 2.01. Radiology 254, 640-641.

SCHA1993 Schad, L.R., Bluml, S., Zuna, I., 1993. IX. MR tissue characterization of intracranial tumors by means of texture analysis. Magnetic Resonance Imaging 11, 889-896.

SEIF2006 Seifert, S., WAchter, I., Dillmann, R., 2006. Segmentation of intervertebral discs, trachea and spinal cord from MRI images. International Journal of Computer Assisted Radiology and Surgery 1, 80-82.

SEIF2009 Seifert, S., WÃ¤chter, I., Schmelzle, G., Dillmann, R., 2009. A knowledge-based approach to soft tissue reconstruction of the cervical spine. IEEE Transactions on Medical Imaging 28, 494-507.

SHAF1999 Shafaie, F.F., Wippold Ii, F.J., Gado, M., Pilgram, T.K., Riew, K.D., 1999. Comparison of computed tomography myelography and magnetic resonance imaging in the evaluation of cervical spondylotic myelopathy and radiculopathy. Spine 24, 1781-1785.

SHAM2009 Shamir, L., Ling, S.M., Scott Jr, W.W., Bos, A., Orlov, N., MacUra, T.J., Eckley, D.M., Ferrucci, L., Goldberg, I.G., 2009. Knee X-ray image analysis method for automated detection of osteoarthritis. IEEE Transactions on Biomedical Engineering 56, 407-415.

SHAP2001 Shapiro, L.G., Stockman, G.C., 2001. Computer Vision. Prentice-Hall, New Jersey, USA

SHI2007 Shi, R., Sun, D., Qiu, Z., Weiss, K.L., 2007. An efficient method for segmentation of MRI spine images. 2007 IEEE/ICME International Conference on Complex Medical Engineering, CME 2007, pp. 713-717.

SHLE2005 Shlens, J., 2005. A tutorial on principal component analysis. Systems Neurobiology Laboratory University of California at San Diego, Published Online.

SLED1998 Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in mri data. IEEE Transactions on Medical Imaging 17, 87-97.

SLUI2005 Sluimer, I., Prokop, M., Van Ginneken, B., 2005. Toward automated segmentation of the pathological lung in CT. IEEE Transactions on Medical Imaging 24, 1025-1038.

SPEC1990 Specht, D.F., 1990. Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification. IEEE Transactions on Neural Networks 1, 111-121.

TALL1998 Tallroth, K., 1998. Plain CT of the degenerative lumbar spine. European Journal of Radiology 27, 206-213.

TAYL2008 Taylor, P., Potts, H.W.W., 2008. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. European Journal of Cancer 44, 798-807.

TERT1991 Tertti, M., Paajanen, H., Laato, M., Aho, H., Komu, M., Kormano, M., 1991. Disc degeneration in magnetic resonance imaging: A comparative biochemical, histologic, and radiologic study in cadaver spines. Spine 16, 629-634.

THEO2003 Theodoridis, S., Koutroubas, K., 2003. Pattern Recognition. Academic Press.

TOUR1999 Tourassi, G.D., 1999. Journey toward computer-aided diagnosis: Role of image texture analysis. Radiology 213, 317-320.

TOUR2001 Tourassi, G.D., Frederick, E.D., Markey, M.K., Floyd Jr, C.E., 2001. Application of the mutual information criterion for feature selection in computer-aided diagnosis. Medical Physics 28, 2394.

TSAI2002 Tsai, M.D., Jou, S.B., Hsieh, M.S., 2002. A new method for lumbar herniated inter-vertebral disc diagnosis based on image analysis of transverse sections. Computerized Medical Imaging and Graphics 26, 369-380.

UDUP1996 Udupa, J., Samarasekera, S., 1996. Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation. Graphical Models and Image Processing 58, 246-261.

URBA2003 Urban, J., Roberts, S., 2003. Degeneration of the intervertebral disc. Arthritis Res Ther 5, 120 - 130.

VADA2007 Vadala , G., Sowa, G.A., Kang, J.D., 2007. Gene therapy for disc degeneration. Expert Opinion on Biological Therapy 7, 185-196.

VANT2007 Van Tulder, M., Koes, B.W., 2007. Evidence-Based Medicine for Low Back Pain. In: Van Goethem, J., Van den Hauwe, L., Parizel, P. (Eds.), Spinal Imaging Diagnostic Imaging of the Spine and Spinal Cord. Springer-Verlag Berlin Heidelberg

VIDE2003 Videman, T., Battie, M.C., Gibbons, L.E., Maravilla, K., Manninen, H., Kaprio, J., 2003. Associations between back pain history and lumbar MRI findings. Spine 28, 582-588.

VOVK2007 Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of intensity inhomogeneity in MRI. IEEE Transactions on Medical Imaging 26, 405-421.

WAY2009 Way, T., Sahiner, B., Chan, H., Hadjiiski, L., Cascade, P., Chughtai, A., Bogot, N., Kazerooni, E., 2009. Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features. Medical Physics 36, 3086-3098.

WIKI2010 Wikimedia Commons by Wikipedia. http://commons.wikimedia.org/wiki/Main_Page

WOLA2005 Wolansky, L.J., Parikh, D.D., Shah, K.J., Yalamanchili, R., Farkas, J., 2005. Magnetic resonance imaging protocols for cervical disc disease: what is your neighbor up to? Journal of Neuroimaging 15, 183-187.

XU1998 Xu, C., Prince, J.L., 1998. Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7, 359-369.

YOSH2002 Yoshida, H., Nappi, J., MacEneaney, P., Rubin, D.T., Dachman, A.H., 2002. Computer-aided Diagnosis Scheme for Detection of Polyps at CT Colonography1. Radiographics 22, 963-979.

ZIJD1994 Zijdenbos, A.P., Dawant, B.M., 1994. Brain segmentation and white matter lesion detection in MR images. Critical Reviews in Biomedical Engineering 22, 401-465.

Appendix 1. Texture and Shape Features Equations

First order features

Equations A.1.-A.4. give the definitions of first order features [THEO2003].

Mean
$$m = \frac{\sum_i \sum_j g(i,j)}{N} \tag{A.1}$$

Standard Deviation
$$std = \sqrt{\frac{\sum_i \sum_j (g(i,j)-m)^2}{N}} \tag{A.2}$$

Skewness
$$sk = \frac{1}{N} \frac{\sum_i \sum_j (g(i,j)-m)^3}{std^3} \tag{A.3}$$

Kurtosis
$$k = \frac{1}{N} \frac{\sum_i \sum_j (g(i,j)-m)^4}{std^4} \tag{A.4}$$

where $g(i,j)$ is the pixel intensity in position (i,j) and N is the total number of pixels within the ROI.

Co-occurrence features

Equations A.5.-A.16. give the definitions of co-occurrence features as they appear in [HARA1979], while the variables used for calculating these features are defined by equations A.17.-A.30.

Angular Second Moment
$$ASM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i,j))^2 \tag{A.5}$$

Contrast
$$CON = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}, \quad |i-j| = n \tag{A.6}$$

Inverse Difference Moment
$$IDM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i,j) \tag{A.7}$$

Entropy
$$ENT = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j)) \tag{A.8}$$

Correlation
$$COR = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{A.9}$$

Sum of Squares $$SSQ = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(1-\mu)^2 p(i,j) \quad \text{(A.10)}$$

Sum Average $$SAVE = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad \text{(A.11)}$$

Sum Entropy $$SENT = -\sum_{i=2}^{2N_g} p_{x+y}(i)\log(p_{x+y}(i)) \quad \text{(A.12)}$$

Sum Variance $$SVAR = -\sum_{i=2}^{2N_g}(i-SENT)^2 p_{x+y}(i) \quad \text{(A.13)}$$

Difference Entropy $$DENT = -\sum_{i=1}^{N_g} p_{x-y}(i)\log(p_{x-y}(i)) \quad \text{(A.14)}$$

Information Measure of Correlation 1 $$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad \text{(A.15)}$$

Information Measure of Correlation 2 $$IMC2 = (1 - e^{-2(HXY2-HXY)})^{1/2} \quad \text{(A.16)}$$

Notation and variables definitions for co-occurrence features calculation:

$p(i,j)$ denotes the $(i,j)^{th}$ entry of a co-occurrence matrix P(i,j) normalized by the number of neighbouring resolution cell pairs R: $p(i,j) = P(i,j)/R$ (A.17)

N_g is the number of grey-levels in the ROI (following quantization)

$p_x(i)$ the i^{th} entry in the marginal probability matrix obtained by summing the rows of $p(i,j)$: $$p_x(i) = \sum_{j=1}^{N_g} p(i,j) \quad \text{(A.18)}$$

$p_y(j)$ the j^{th} entry in the marginal probability matrix obtained by summing the columns of $p(i,j)$: $$p_y(j) = \sum_{i=1}^{N_g} p(i,j) \quad \text{(A.19)}$$

μ_x, μ_y are the mean values of p_x and p_y : $$\mu_x = \sum_{i=1}^{N_g} i p_x(i) \quad \text{(A.20)}$$

$$\mu_y = \sum_{y=1}^{N_g} j p_y(j) \quad \text{(A.21)}$$

σ_x, σ_y are the standard deviations of p_x and p_y: $$\sigma_x = \sum_{i=1}^{N_g}(i-\mu_x)^2 p_x(i) \quad \text{(A.22)}$$

$$\sigma_y = \sum_{j=1}^{N_g}(j-\mu_y)^2 p_y(i) \quad \text{(A.23)}$$

$$p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \ where \ i+j=k, \qquad k = 2,3,...,2N_g \tag{A.24}$$

$$p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \ where \ |i-j|=k, \qquad k = 0,1,...,N_g - 1 \tag{A.25}$$

$$HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j)) \tag{A.26}$$

$$HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j)) \tag{A.27}$$

$$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)p_y(j)\log(p_x(i)p_y(j)) \tag{A.28}$$

$$HX = -\sum_{i=1}^{N_g} p_x(i)\log(p_x(i)) \tag{A.29}$$

$$HY = -\sum_{j=1}^{N_g} p_y(j)\log(p_y(j)) \tag{A.30}$$

Run-length matrix features

Equations A.31-A.35 provide the 5 textural features that can be computed from run-lengths matrices as they appear in [GALL1975].

Short Run Emphasis

$$SRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} \frac{r(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)} \tag{A.31}$$

Long Run Emphasis

$$LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} j^2 r(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)} \tag{A.32}$$

Grey Level Non Uniformity

$$GLNU = \frac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_r} r(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)} \tag{A.33}$$

Run Length Non Uniformity

$$RLNU = \frac{\sum_{j=1}^{N_r}\left(\sum_{i=1}^{Ng} r(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)} \quad \text{(A.34)}$$

Run Percentage

$$RP = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}{P} \quad \text{(A.35)}$$

where $r(i,j)$ is the run length matrix, N_g is the total number of grey values in the image, N_r is the largest possible run and P is the total possible number of runs in the image.

Fourier Descriptors

Fourier descriptors are calculated by the complex sequence *z(n)*, which is constructed from the contour by expressing the ROI boundary coordinates in complex format as shown in Equation A.36. The sequence *z(n)* is periodic and its Fourier series is given as one period of the DFT as defined in Equation A.37. The Fourier descriptors are given by the DFT coefficients *Z(k)* [BANK2000].

$$z(n) = x(n) + jy(n) \quad \text{(A.36)}$$

$$Z(k) = \frac{1}{N}\sum_{n=0}^{N-1} z(n) e^{-j\frac{2\pi nk}{N}} \quad \text{(A.37)}$$

with *x(n), y(n)* being the coordinates of the n_{th} contour point of the ROI *with {n / [0,N-1]}*, and *Z(k)* is the k_{th} Fourier coefficient of this object.

Statistical Moments

Central moments are defined by Equation A.37. When moments are computed for binary images only shape information is quantified, whereas when computed for grey-scale images information about the grey-levels intensity distribution is also retained [BANK2000, RANG2005].

$$\mu_{pq} = \sum_{x=1}^{M-1}\sum_{y=1}^{N=1}(x-\bar{x})^p(y-\bar{y})^q f(x,y) \tag{A.37}$$

where $\bar{x}$ and $\bar{y}$ are the image centroids, m_{pq} is the order *(p+q)* moment of the ROI *f(x,y)* and μ_{pq} is the corresponding central moment.

Moreover, Equation A.38 gives scale invariant moments. While finally Equations A.39-A.45 give Hu's moments, which are characterized by translation, scaling and rotation invariance [HU1962].

$$\eta_{pq} = \mu_{pq} \big/ \mu_{00}^{(p+q+2)/2} \tag{A.38}$$

$$\phi_1 = \eta_{20} + \eta_{02} \tag{A.39}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{A.40}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{A.41}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{A.42}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{12} + \eta_{30})[(\eta_{12} + \eta_{30})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{A.43}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{12} + \eta_{30})(\eta_{21} + \eta_{02})^2\end{aligned} \tag{A.44}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{12} + \eta_{30})[(\eta_{12} + \eta_{30})^2 - 3(\eta_{21} + \eta_{30})^2] \\ & + (3\eta_{12} - \eta_{03})(\eta_{12} + \eta_{30})[3(\eta_{12} + \eta_{30})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{A.45}$$

where η_{pq} is the order *(p+q)* scale invariant central moment of the image, and ϕ_x representing Hu's moments.

Appendix 2. Abbreviations

2D	Two Dimensional
3D	Three Dimensional
ADC	Apparent Diffusion Coefficient
ANOVA	Analysis of Variance
Atlas-FCM	Atlas Combined with Fuzzy Clustering Method
Atlas-RFCM	Atlas Combined with Robust Fuzzy Clustering
AUROC	Area Under the ROC curve
CAD	Computer Aided Diagnosis
CADe	CAD for abnormality Detection
CADx	CAD for characterization
CJV	Coefficient of Joint Variation
CT	Computed Tomography
CSF	Cerebrospinal Fluid
CV	Coefficient of Variation
DFT	Discrete Fourier Transform
$\mathbf{D_{mean}}$	Mean Boundary Distance
DSI	Dice Similarity Index
FCM	Fuzzy C Means
ICC	Intraclass Correlation Coefficient
k-NN	k- Nearest Neighbours Classifier
MD	Minimum Distance Classifier
MEAN	Mean Value
MEDIC	Multi Echo Data Image Combination
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
N3	Nonparametric Nonuniform intensity Normalization
NN	Nearest Neighbour Classifier
PD	Proton Density
PNN	Probabilistic Neural Networks Classifier
PX	Pixel

RF	Radio Frequency
RFCM	Robust Fuzzy C Means
ROC	Receiver Operator Characteristic curve
S	Second
SDA	Stepwise Discriminant Analysis
STD	Standard Deviation
T1	Longitudinal Relaxation Time (spin – lattice)
T2	Transverse Relaxation Time (spin – spin)
TE	Echo Time
TR	Repetition Time
VB	Vertebral Bodies

Appendix 3. ROC Results for SDA and PCA

Table A3. ROC analysis results for SDA and PCA using the MD classifier

	SDA		PCA	
	AUROC ± Standard Error	95% Conf. Int. [Lower, upper]	AUROC ± Standard Error	95% Conf. Int. [Lower, upper]
Node 1	0.957 ± 0.014	[0.930 , 0.985]	0.956 ± 0.013	[0.931, 0.981]
Node 2a	0.867 ± 0.029	[0.809 , 0.925]	0.832 ± 0.038	[0.757 , 0.907]
Node 2b	0.958 ± 0.017	[0.924 , 0.992]	0.943 ± 0.029	[0.903 , 0.983]

Appendix 4. Data processing Tools Details

All algorithms where implemented in Matlab (versions 7.04 and 2009b).

The Medical Image Processing Analysis and Visualization (MIPAV) software was used for inhomogeneity correction.

Statistical analysis was performed with combined use of Microcal Origin (version 8.0) and Matlab (Statistics Toolbox version 7).