ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ  ΤΜΗΜΑ ΦΥΣΙΚΗΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΙΑΤΡΙΚΗ ΦΥΣΙΚΗ

ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΕΚΦΥΛΙΣΜΟΥ
ΜΕΣΩΣΠΟΝΔΥΛΙΟΥ ΔΙΣΚΟΥ ΜΕ ΧΡΗΣΗ
ΑΛΓΟΡΙΘΜΩΝ ΑΝΑΛΥΣΗΣ ΥΦΗΣ ΨΗΦΙΑΚΗΣ
ΕΙΚΟΝΑΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ

ΣΟΦΙΑ ΜΙΧΟΠΟΥΛΟΥ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΤΡΑ 2007
UNIVERSITY OF PATRAS

SCHOOL OF MEDICINE DEPARTMENT OF PHYSICS

INTERDEPARTMENTAL PROGRAMME OF POSTGRADUATE STUDIES IN MEDICAL PHYSICS

COMPUTER AIDED CHARACTERIZATION OF DEGENERATIVE DISK DISEASE EMPLOYING DIGITAL IMAGE TEXTURE ANALYSIS AND PATTERN RECOGNITION ALGORITHMS

SOFIA MICHOPoulos

MASTER THESIS

PATRAS 2007
ACKNOWLEDGEMENTS

The deepest and most sincere gratitude is to my supervisor, Professor George Panayiotakis for his trust, continuous support and priceless guidelines, during this Master Thesis.

I would like to thank Professor Elias Panagiotopoulos (MD) for his participation, valuable contribution and co-operation during this work.

In addition I would like to thank Professor Dionysis Cavouras for his scientific support and useful advices in all the stages of this work.

I am also indebted to Assistant Professor Lena Costaridou for her precious help and support during this work.

I am truly grateful to Dr. Ioannis Boniatis, for helping and guiding me in every step from the beginning of this work.

My sincerest appreciation to my colleagues and friends in the Department of Medical Physics for their precious help and discussions.

But the biggest and deepest thanks and appreciations deserve to my loving family and friends. Nothing I could say could possibly reveal the gratitude I own to them for everything they have done for me.
### CHAPTER A

**INTRODUCTION**

A.1.1. THE PROBLEM ...................................................... 2  
A.1.2. THESIS ORIGINALITY ........................................... 2  
A.1.3. PUBLICATIONS .................................................. 3  
A.1.4. THESIS LAYOUT ................................................... 4

### CHAPTER B

**THEORETICAL PART**

B.1. THE CERVICAL SPINE  
B.1.1. ANATOMY OF THE CERVICAL SPINE  
B.1.2. DEGENERATIVE DISC DISEASE OF THE CERVICAL SPINE  
B.1.3. IMAGING OF THE CERVICAL SPINE AND DIAGNOSIS OF DEGENERATIVE DISC DISEASE

B.2. INTRODUCTION TO MAGNETIC RESONANCE IMAGING  
B.2.1. INTRODUCTION TO MRI .......................................... 15  
B.2.2. NUCLEAR MAGNETIC RESONANCE PHYSICS .................... 15  
B.2.3. IMAGING PRINCIPLES ........................................... 16  
B.2.4. IMAGING HARDWARE ............................................ 17  
B.2.5. T1 AND T2 WEIGHTED IMAGES .................................. 18  
B.2.6. MRI OF THE CERVICAL SPINE .................................. 19

B.3. INTRODUCTION TO IMAGE ANALYSIS  
B.3.1. INTRODUCTION TO IMAGE ANALYSIS AND PATTERN RECOGNITION METHODS ........................................... 21  
B.3.2. IMAGE PREPROCESSING AND ROI SELECTION ............... 21  
B.3.3. GENERATION OF FEATURES .................................... 22  
B.3.4. FEATURES DISCRIMINATING POWER INVESTIGATION .......... 29  
B.3.5. CLASSIFICATION .................................................. 30  
B.3.6. FEATURES SELECTION ........................................... 35  
B.3.7. SYSTEM DESIGN AND PERFORMANCE EVALUATION ........ 36
# TABLE OF CONTENTS

## CHAPTER C

### MATERIALS AND METHODS

C.1. DATA ACQUISITION AND INTERVERTEBRAL DISC CHARACTERIZATION ................................................................. 39  
C.2. DESIGN OF THE IMAGE ANALYSIS SYSTEM ........................................ 40  
C.3. SYSTEM PERFORMANCE EVALUATION METHOD ................... 43  

## CHAPTER D

### RESULTS AND DISCUSSION

D.1. STATISTICAL ANALYSIS RESULTS .............................................. 45  
D.2. EVALUATION OF THE DESIGNED SYSTEM AND COMMENTS ON THE RESULTS ........................................... 46  

## CHAPTER E

### CONCLUSIONS & FUTURE WORK

E.1. CONCLUSIONS ........................................................................... 50  
E.2. FUTURE WORK ........................................................................ 50  

## REFERENCES

## APPENDIX I

ABBREVIATIONS .............................................................................. 51  

## APPENDIX II

LIST OF FIGURES ............................................................................ 52  

## APPENDIX III

LIST OF TABLES ............................................................................. 53  

SUMMARY ....................................................................................... 54  
ΠΕΡΙΛΗΨΗ ..................................................................................... 55
CHAPTER A

INTRODUCTION

A.1.1. THE PROBLEM.................................................................2
A.1.2. THESIS ORIGINALITY..................................................2
A.1.3. PUBLICATIONS..............................................................3
A.1.4. THESIS LAYOUT............................................................4
A.1.1. THE PROBLEM

The Degenerative Disc Disease (DDD) is considered to be a major cause of pain and disability in the adult working population\textsuperscript{[1]}. It is an age-associated condition that is caused due to the degeneration of intervertebral discs and the narrowing of intervertebral disc spaces. In the case of cervical DDD, the symptoms are mainly neck pain and stiffness. If the degeneration is severe then back pain and muscle weakness occur as well and cervical spine surgery is considered\textsuperscript{[5]}.

To evaluate DDD different imaging modalities are used, but MRI is the preferred examination since it provides excellent soft tissue contrast, it does not involve ionizing radiation and it is not invasive. Using sagittal images of the cervical spine, an orthopedist measures the intervertebral disc spaces and detects the degenerated discs\textsuperscript{[3,4]}.

Computerized approaches for the detection of degenerated discs based on disc morphology have been proposed in the past\textsuperscript{[9]}. In this study a different approach which exploits the differentiation of the disc texture is used for classification purposes. When a disc is degenerated, alterations on its biochemical composition occur. These alterations affect the texture of MR images\textsuperscript{[23]}. More specifically, the loss of hydration results in speckled or darker appearance of the disc in T2-weighted images. The proposed image analysis system uses textural features calculated from the intervertebral discs, which comprise the regions of interest, in order to classify them as normal or degenerated (narrowed).

A.1.2. THESIS ORIGINALITY

The originality of this Master Thesis relies on:

- The adaptation of image texture as a factor of differentiation between normal and degenerated cervical intervertebral discs.

- The design of an image analysis system for the classification of cervical intervertebral discs using textural features calculated from MR images.
A.1.3. PUBLICATIONS

This work has been accepted for publication in international conference proceedings and for presentation in a national conference.

Publications in international conference proceedings


Presentation in national conference

A.1.4. THESIS LAYOUT

The layout if this thesis is presented as following:

Chapter B analyzes the theoretical part. Section B.1 is an overview to the anatomy of the cervical spine and the degenerative disc disease. Section B.2 is an introduction to the magnetic resonance imaging. Information concerning image analysis is found in Section B.3.

Chapter C refers to the materials and methods. It begins with Section C.1, where the data acquisition and characterization method is presented. Section C.2 describes the steps for the design of the image analysis system. Finally, Section C.3 presents the methods that are used for the evaluation of the performance of this system.

Chapter D presents and analyzes the results of the present study. Section D.1 is about the statistical analysis and features reduction results and it is followed by Section D.2 which presents the results and comments on the system design process.

Finally, some conclusions concerning the results produced during the course of this study are being presented in Section E.1. In Section E.2 suggestions for future work are made, targeting to the improvement of the results and the enrichment of this study.
CHAPTER B
THEORETICAL PART

B.1. THE CERVICAL SPINE

B.1.1 ANATOMY OF THE CERVICAL SPINE.................................6
B.1.2 DEGENERATIVE DISC DISEASE OF THE CERVICAL SPINE........9
B.1.3 IMAGING OF THE CERVICAL SPINE AND DIAGNOSIS OF
       DEGENERATIVE DISC DISEASE........................................11
B.1. ANATOMY OF THE CERVICAL SPINE

The cervical spine is made up of the first seven vertebrae in the spine (C1-C7) as it is shown in Figure 1. It starts just below the skull and ends at the top of the thoracic spine. The cervical spine has a backward "C" shape (called the lordotic curve) and is much more mobile than either of the thoracic or lumbar regions of the spine. The purpose of the cervical spine is to contain and protect the spinal cord, support the skull, and enable diverse head movement (i.e. rotate side to side, bend forward and backward)\(^{33}\).

Between each vertebra in the cervical spine lie the intervertebral discs which act as cushions or shock absorbers and also permit some movement between the vertebral bodies. The entire spinal column is joined together by ligaments that allow the spine to bend and twist carrying the weight of the human body with just the right balance of strength and flexibility. In addition to the intervertebral discs, special joints between each of the vertebral bodies, called facet joints, allow the individual bones of the spine to move and rotate with respect to each other. These joints are important because they can be a source of pain if they become arthritic. Many muscle groups, that move the trunk and the limbs, also attach to the spinal column. The muscles, that closely surround the bones of the spine, are important for maintaining posture and helping the spine to carry the loads created during normal activities, work, and play. Strengthening these muscles can be an important part of physical therapy and rehabilitation.

Each vertebra is shaped in a special way so that when they are stacked together, the spinal cord is protected from damage by the bones of the entire spinal column. The spinal cord is part of the central nervous system and a direct extension of the brain. It is made up of a large collection of nerves and carries messages from the brain to the entire body\(^{33}\).
Figure 1. Anatomy of the cervical spine

Figure 2. Transverse view of the cervical spine components
B.1.1.1 Vertebrae

The vertebrae support the majority of the weight imposed on the spine. The body of each vertebra is attached to a bony ring that consists of several parts. A bony projection on either side of the vertebral body supports the arch that protects the spinal canal. The laminae are the parts of the vertebrae that form the back of the bony arch, which surrounds and covers the spinal canal. There is a transverse process on either side of the arch where some of the muscles of the spinal column attach to the vertebrae. The spinous process is the bony portion of the vertebral body that can be felt as a series of bumps in the center of a person's neck and back (see Figures 1 and 2)[32].

B.1.1.2 Intervertebral Disc

The discs located in between each vertebrae function as shock absorbers and as joints. They are designed to absorb the stresses carried by the spine while allowing the vertebral bodies to move with respect to each other. They made up of a strong outer ring of fibers called the annulus fibrosis, and a soft center called the nucleus pulposus. The outer layer (annulus) helps keep the disc's inner layer intact. The nucleus of the disc has a very high water content making it very moist (see Figures 2 and 3)[32].

B.1.1.3 Facet Joint

The facets connect the bony arches of each of the vertebral bodies. There are two facet joints between each pair of vertebrae one on each side. Facet joints connect each vertebra with the next vertebra above and below. They are primarily designed to allow the vertebral bodies to rotate with respect to each other[32].

B.1.1.4 Neural Foramen

The neural foramen is the opening where the nerve roots exit the spine and travel to the rest of the body. There are two neural foramen located between each pair of vertebrae. The foramen creates a protective passageway for the nerves that carry signals between the spinal cord and the rest of the body (see Figure 2)[32].
B.1.1.5 Spinal Cord and Nerve Roots

The spinal cord extends from the base of the brain to the bottom of the first lumbar vertebra. It is enclosed in a protective membrane called the dura mater which forms a watertight sack around the spinal cord and nerves. The spinal cord is surrounded by spinal fluid inside this sack. The nerves in each area of the spinal cord connect to specific parts of the body (see Figures 2 and 3). The nerves of the cervical spine go to the upper chest and arms. The nerves also carry electrical signals back to the brain creating sensations. Damage to the nerves, nerve roots, or spinal cord can lead to symptoms such as pain, tingling, numbness and weakness.\[32\].

Figure 3. Anatomical diagram of the spinal cord and nerve roots

B.1.2 Degenerative Disc Disease of the Cervical Spine

Degenerative disc disease (DDD) is part of the natural process of growing older. With age, the intervertebral discs lose their flexibility, elasticity, and shock absorbing characteristics. The ligaments that surround the disc (annulus fibrosis), become brittle and they are more easily torn. At the same time, the soft gel-like center of the disc (nucleus pulposus), starts to dry out and shrink (see Figure 4). Degenerative disc disease is as certain as death and taxes, and to a certain degree this
process happens to everyone. However, not everyone develops symptoms as a result of degenerative disc disease. Many people who have "normal" necks have MRI's that show disc herniations, degenerative changes, and narrowed spinal canals\textsuperscript{[2,34]}. 

When degenerative disc disease becomes painful or symptomatic, it can cause several different symptoms, including neck pain, nerve root pathology, and spinal cord compression. These symptoms are caused by the fact that worn out discs are a source of pain because they do not function as well as they once did, and as they shrink, the space available for the nerve roots and the spinal cord also shrinks\textsuperscript{[1]}. As the discs between the intervertebral bodies start to wear out, the entire cervical spine becomes less flexible. This results to neck pain and stiffness, especially towards the end of the day.

The common symptoms that suggest that degenerative disc disease may be responsible for a person's pain include neck pain, pain that radiates down to the back of the shoulder blades or into the arms, numbness and tingling, and sometimes even difficulties with hand dexterity or walking. Muscle weakness occurs at a later stage in the degenerative process than pain does, and it is a sign that disease is relatively more serious. In severe cases of cervical DDD, where there is evidence of spinal cord compression, individuals may experience symptoms of sciatica and back pain, and lower extremity weakness\textsuperscript{[34,35]}. 

![Examples of Disc Problems](image)

Figure 4. Image showing examples of different intervertebral disc problems
B.1.3 IMAGING OF THE CERVICAL SPINE AND DIAGNOSIS OF DEGENERATIVE DISC DISEASE

B.1.3.1 Clinical examination

The diagnosis of degenerative disc disease begins with a complete physical examination of the neck, arms and lower extremities. The doctor examines patient’s neck for flexibility, range of motion, and the presence of certain signs that suggest that the nerve roots or spinal cord are affected by degenerative changes. This often involves testing the strength of muscles to make sure that they are still working normally. Patients are often asked to fill out a questionnaire that refers to the occurrence and frequency of symptoms of pain, numbness, tingling and weakness\textsuperscript{[5,9,31]}.

B.1.3.2 Imaging the DDD

A set of x-rays is also usually ordered when a patient with neck pain goes to see a doctor. If degenerative disc disease is present, the x-rays will often show a narrowing of the spaces between the vertebral bodies, which indicates that the disc has become very thin or has collapsed (see Figure 5). Bone spurs begin to form around the edges of the vertebral bodies and the edges of the facet joints in the spine. These bone spurs can be seen on an x-ray and they are called osteophytes. As the discs collapse and bone spurs form, the space available for the nerve roots and the spinal cord starts to shrink. The nerve roots are especially vulnerable to compression at the exit of the spinal canal, the neuroforamen (see Figure 4)\textsuperscript{[2,3,9,33]}.

![Figure 5. X-ray views of the cervical spine](image-url)
A CT scan can be used in order to evaluate the bony anatomy in the cervical spine. The method is non-invasive, it is quick, it provides excellent visualization of bone in the axial projection and shows the root-canals and paraspinal area. It assists to detect bony elements narrowing the spinal canal, and thus it can show how much space is available for the nerve roots and spinal cord within the spinal canal. The main draw-backs are that it cannot image soft tissues (discs and nerves), and if a large area is imaged, there will be significant irradiation of the patient\[3\].

A magnetic resonance imaging scan (MRI) is a useful tool in the diagnosis of DDD. It is able to cut through multiple layers of the spine and show any abnormality of soft tissues, such as nerves and ligaments. The test also can be used to verify: loss of water in a disc, facet joint hypertrophy (enlargement), stenosis (narrowing of spinal canal), or a herniated disc (protrusion or rupture of the intervertebral disc). It helps determining where the nerve roots or spinal cord are being compressed. The MRI scan has become the most common test, to look at the cervical spine\[3,4,7,8\]. Figure 6 shows one transverse and one sagittal T2-weighted MR image of the cervical spine. In can be noticed that the disc between C4-C5 vertebrae is narrowed and bulges backwards (herniation) resulting in nerve root compression. The principles of MRI are explained in more detail at the second part of this chapter.

![Figure 6. Transverse (on the left) and Saggital (on the right) views of a cervical spine T2-weighted MRI scan](image)

A myelogram is an invasive method used to assess DDD. It involves the injection of radiographic contrast media (dye) into the sac (dura) surrounding the spinal cord and nerves. After the injection x-rays, or CT scans are performed. The dye
creates contrast in the images and helps a doctor see if there is a herniated disc, or pressure on the spinal cord or spinal nerves (see Figure 7)[8]. Before the MRI scan was developed, the myelogram was the only test that doctors had, to look for a herniated disc. It is still used today, but not nearly as often, since it is replaced by MRI which can accurately assess disc herniation and is non invasive[7].

Figure 7. Transverse (on the left) and Saggital (on the right) views of the cervical spine imaged employing T2-weighted MRI scan

**B.1.3.2 DDD treatment**

The treatment for DDD, for patients who do not have evidence of nerve root compression, includes non-steroidal anti-inflammatory drugs and physical therapy or even a soft cervical collar that allows the neck to rest. If there is evidence of muscle weakness, caused by nerve root or spinal cord compression, surgery is offered in order to relieve the pressure on the nerves is more of an urgent priority[5,32].
CHAPTER B

THEORETICAL PART

B.2. INTRODUCTION TO MAGNETIC RESONANCE IMAGING

B.2.1. INTRODUCTION TO MRI
B.2.2. NUCLEAR MAGNETIC RESONANCE PHYSICS
B.2.3. IMAGING PRINCIPLES
B.2.4. IMAGING HARDWARE
B.2.5. T1 AND T2 WEIGHTED IMAGES
B.2.6. MRI OF THE CERVICAL SPINE
B.2.1. INTRODUCTION TO MRI

Magnetic resonance imaging (MRI) is an imaging technique that provides high quality and high contrast images of the inside of the human body. It is based on the principles of nuclear magnetic resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules\[31\]. MRI is a tomographic modality. It produces NMR images of slices through the human body. Each slice is composed of volume elements (voxels), with a volume of 3 mm\(^3\) per voxel. The intensity of each picture element (pixel) in the MR image is proportional to the signal intensity of the contents of the corresponding voxel of the object that is being imaged. In the human body MRI primarily images the NMR signal from the hydrogen nuclei, which comprise approximately 63% of the human body\[30\].

MRI is based on the absorption and emission of energy in the radio frequency range (RF pulses) of the electromagnetic spectrum. In order to image objects smaller than the wavelength of the energy that is used (RF pulse) it has to measure the spatial variations in the phase and frequency of the RF energy that is being absorbed and emitted by the imaged objects.

Using different RF sequences image contrast can be changed. The main advantage of magnetic resonance imaging over computed tomography (CT) is that it provides high soft tissue contrast. In contrast to CT, no ionizing radiation is being used by MRI scanners\[8\].

B.2.2. NUCLEAR MAGNETIC RESONANCE PHYSICS

Nuclear magnetic resonance (NMR) is a physical phenomenon based upon the magnetic properties of the nuclei of the atoms. It occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to a second oscillating magnetic field (electromagnetic RF pulse). The nuclei that experience this phenomenon are called magnetic nuclei, they contain single protons or neutrons and thus they have an intrinsic magnetic moment and angular momentum. NMR most commonly studies hydrogen nuclei which are abundant in nature\[31\].
When a magnetic nucleus is placed in an external magnetic field, the spin vector aligns itself with the field, just like a magnet would. There is a low energy configuration state (N-S-N-S alignment for the magnetic poles) and a high energy state (N-N-S-S). This nucleus can undergo a transition between the two energy states by the absorption of a photon. The energy of this photon must exactly match the energy difference between the two states. For this to occur the frequency (ν) of the photon has to be in the RF range. In clinical MRI, ν is typically between 15 and 80 MHz for hydrogen imaging.

When a group of magnetic nuclei is placed in a magnetic field, each one of them aligns in one of the two possible orientations. The NMR signal results from the difference between the energy absorbed by the nuclei which make a transition from the lower energy state to the higher energy state, and the energy emitted by those which simultaneously make a transition from the higher energy state to the lower energy state. Consequently the signal in NMR is proportional to the population difference between the states\(^{31}\).

### B.2.3. IMAGING PRINCIPLES

MRI is primarily used to construct pictures of the NMR signal from the hydrogen atoms in an object. In medical MRI, radiologists are most interested in looking at the NMR signal from water and fat, the major hydrogen containing components of the human body.

In order to image the positions of the spins (hydrogen nuclei) magnetic field gradients (ie. variations in the magnetic field with respect to position) are used. The amplitude of the signal is proportional to the number of spins in a plane perpendicular to each gradient. This principle forms the basis behind all magnetic resonance imaging.

Three magnetic field gradients are used in MRI. These are the slice selection, frequency encoding and phase encoding gradient. The slice selection is a one-dimensional, linear magnetic field gradient applied during the period the RF pulse is applied. This field allows selecting the slice that will be imaged. In addition the frequency encoding and phase encoding gradients are used to give information about the coordinates of each voxel within the slice. Finally, Fourier transform techniques
are used, to decode the signal and retrieve spatial information in order to construct the MR image$^{[30]}$.

**B.2.4. IMAGING HARDWARE**

The major components of an MRI scanner are: a static magnetic field, an RF transmitter and receiver, three orthogonal, controllable magnetic gradients and the control computer.

The static magnetic field is produced by a (usually cylindrical) magnet which is the largest and most expensive component of the scanner. The MRI magnets are usually superconducting electromagnets which can provide static magnetic fields with strength of 1 to 3 Tesla. This type of magnets have coils made of niobium-titanium alloy that is immersed in liquid helium. In this way the coil is cooled to a temperature close to absolute zero (0 K) and therefore its resistance to flow of electric current is practically zero. Thus it can create and maintain very high strength, stable magnetic fields.

The three orthogonal gradient coils lie within the body magnet. These coils are room temperature coils and they produce low strength gradients (20-100mT/m) in the static field in the X, Y, and Z directions. In addition the RF coil is also embedded the magnet. This coil transmits the pulse sequences which rotate the spin of the nuclei and at the same time it serves as a receiver that detects the signal from the spins within the body. In some cases surface RF receiver coils are used as well in order to obtain better images (images with a higher signal to noise ratio).

During the examination the patient is positioned within the magnet by a computer controlled patient table which has a positioning accuracy of 1 mm. The scan room is surrounded by an RF shield that serves in preventing the high power RF pulses from radiating out through the hospital and at the same time it prevents the various RF signals from television and radio stations from being detected by the imager. Finally the computer is the control component of the scanner. It programs the radiofrequency and pulse sequence of the RF coil and also collects the signal and composes the MR images which can be seen on a video display and/or printed on film$^{[30]}$. 
B.2.5. T1 AND T2 WEIGHTED IMAGES

Several imaging techniques (referred as imaging sequences) such as spin-echo gradient-echo and inverse recovery are used in MR imaging. In addition different time weighting is used to create contrast in MR images. The most commonly used MR images are T1 and T2 weighted spin echo images.

T1 refers to the time required for a certain percentage of the tissue's nuclei to realign with the magnetic field (longitudinal relaxation) and it is typically about 1 second. T2 refers to the time required for local dephasing of nuclei following the application of the transverse energy pulse (transverse relaxation).

Image contrast is created by using a selection of image acquisition parameters that weights signal by T1 or T2. In the spine, T1-weighting causes the spinal cord and vertebrae to appear light gray, while the cerebrospinal fluid and healthy intervertebral discs appear dark. When T2 imaging is used the contrast is reversed so that the spinal cord and vertebrae appear dark, whereas the cerebrospinal cord and healthy intervertebral discs appear white (see Figure 8)[3,7].

Figure 8. T1 weighted (on the left) and T2 weighted (on the right) sagittal MR images of the cervical spine
B.2.6. MRI OF THE CERVICAL SPINE

MRI of the spine is used to evaluate different types of tissue, including the spinal cord, vertebral disks and spaces between the vertebrae through which the nerves travel, as well as distinguish healthy tissue from diseased tissue (see Figure 8). By using a combination of different imaging planes and pulse sequence parameters the abovementioned structures can be accurately depicted. From a morphological aspect MRI may be the most accurate means of evaluating the cervical spine and intervertebral discs.

Unlike Computed Tomography and conventional X-Rays in which the image depends on information related to electron density, MRI signals are influenced by the T1 and T2 relaxation times and proton density, providing greater tissue contrast. Thus the role of MRI may go beyond gross anatomical appraisal to actual characterization of pathologic and biochemical changes within tissue\textsuperscript{[8]}.
CHAPTER B
THEORETICAL PART

B.3. INTRODUCTION TO IMAGE ANALYSIS

B.3.1. INTRODUCTION TO IMAGE ANALYSIS AND PATTERN RECOGNITION METHODS ................................................................. 21
B.3.2. IMAGE PREPROCESSING AND ROI SELECTION ................................. 21
B.3.3. GENERATION OF FEATURES ................................................................. 22
B.3.4. FEATURES DISCRIMINATING POWER INVESTIGATION ................. 29
B.3.5. CLASSIFICATION ........................................................................... 30
B.3.6. FEATURES SELECTION .................................................................. 35
B.3.7. SYSTEM DESIGN AND PERFORMANCE EVALUATION ............... 36
B.3.1. INTRODUCTION TO IMAGE ANALYSIS AND PATTERN RECOGNITION METHODS

Image analysis and pattern recognition systems are used as diagnostic tools that can extract useful information from medical images and make a decision on the pathology (diagnosis). The design of a Medical Image Analysis System (MIAS) comprises 3 steps: features generation, classification and system design and performance evaluation. Two more steps are necessary to prepare the input for the system. These are image acquisition and preprocessing and the selection of Regions of Interest (ROIs). Once the ROIs are selected the system extracts features from each one of them and classifies the images to provide a diagnosis. These systems are used as training tools to help the inexperienced physicians or as second opinion diagnostic tools by the experienced ones\(^\text{[25]}\).

B.3.2. IMAGE PREPROCESSING AND ROI SELECTION

Image preprocessing targets to equalize images and enhance their contrast. Image equalization provides that all images have approximately the same histogram. This is important when the system has to compare images acquired using different imaging setup.

Histogram Equalization filters are used to enhance image contrast by modifying the image histogram. These filters rescale the grey levels of the original image and force the histogram of the image to follow the desired distribution. This is achieved by using a mapping function based on the probability density models for the desired distribution of the output histogram.

Adaptive Histogram Equalization (AHE) is a technique that calculates a different mapping function for each pixel in the image based on a local histogram. This technique provides improved local area contrast enhancement in comparison to the simple histogram equalization algorithms and thus it is often used as a prepossessing step\(^\text{[29]}\).

After enhancing the images, Regions of Interest (ROIs) are selected from each image. These ROIs are those areas in the image which provide useful diagnostic information. ROI selection can be performed either by manual or by automatic
segmentation. Automatic methods are preferred over manual segmentation methods because they are repeatable and user invariant and thus they are considered to be more trustworthy. However there are cases were automatic segmentation is not possible and thus manual segmentation is used.

B.3.3. GENERATION OF FEATURES

A digital image is a set of discrete pixels, each carrying a gray-tone value, and they are arranged in such a way in the image that they form anatomical regions, such as bone, muscles etc. These anatomical regions have different texture and shape which are qualitatively evaluated by the physicians in order to discriminate among normal and abnormal regions\textsuperscript{[16,23]}. An image analysis system uses textural and morphological features in order to evaluate the images. These features quantify the texture and shape of structures within the image and thus they allow the system to discriminate among normal and abnormal regions\textsuperscript{[11]}. A set of features is extracted from each ROI and used by the image analysis system to classify this ROI. The remaining of this chapter offers a brief description of different textural features.

B.3.3.1 Features from 1st order statistics.

These features are extracted from the histogram of the ROI and thus they provide information about the frequency of appearance of each gray-level within the ROI\textsuperscript{[25,29]}. These features are:

1. Mean value

\[
m = \frac{\sum_i \sum_j g(i,j)}{N}
\]  (1)

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

2. Standard deviation

\[
std = \sqrt{\frac{\sum_i \sum_j (g(i,j) - m)^2}{N}}
\]  (2)

The \(std\) describes the variation of the ROI gray levels from the mean value.
3. Skewness

\[ s_k = \frac{1}{N} \sum_i \sum_j \left( \frac{g(i,j) - m}{\text{std}} \right)^3 \] (3)

The sk describes the distribution asymmetry around the mean.

4. Kurtosis

\[ k = \frac{1}{N} \sum_i \sum_j \left( \frac{g(i,j) - m}{\text{std}} \right)^4 \] (4)

The k is indicative of the histogram shape and describes the distribution sharpness compared to the normal distribution.

**B3.3.2 Features from 2st order statistics.**

Textural features calculated from 2nd order statistics provide information about the spatial distribution of the gray levels within the ROI\(^{[16,18,28]}\). This type of information can be extracted from the co-occurrence and run-length matrixes.

**B.3.3.2.1 Features extracted from co-occurrence matrices.**

The co-occurrence matrix describes the frequency that two gray-levels of the ROI appear together (two neighboring pixels) in a particular dimension\(^{[18]}\). A different co-occurrence matrix is calculated for each one of the basic directions (horizontal-0°, diagonal-45°, vertical-90°, and antidiagonal-135°). The following example illustrates the calculation of the four co-occurrence matrixes for a specific ROI. Table 1 shows the array \(g(i,j)\) of a ROI with four gray-levels. In addition Table 2 shows the four co-occurrence matrices (\(p0(i,j)\), \(p45(i,j)\), \(p90(i,j)\) and \(p135(i,j)\)) calculated from the ROI for each one of the four directions.

<table>
<thead>
<tr>
<th>g(i,j)=</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1. A four gray-levels ROI
where, \( P_{0}(0,0) = 4 \) is the number of times that neighboring pixels \([0,0]\) appear in the \( g(i,j) \) matrix both in the \( 0^0 \) and \( 180^0 \) directions. The rest of the co-occurrence matrices are constructed in a similar way.

In some cases, the four matrices are combined into one by addition in order to save computational time. This addition of the matrices results in loss of directionality but this usually is not a concern in medical image texture analysis\(^{17}\). 11 of the 14 textural features that can be computed from the co-occurrence matrix of the ROI are demonstrated below. These features describe qualitative characteristics of texture such as homogeneity, contrast, or the presence of organized structures within the ROI.

1. Angular Second Moment

\[
ASM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \left( p(i,j) \right)^2 
\]

where \( N_g \) is the number of gray levels in ROI, \( i,j = 1,\ldots,N_g \), and \( p(i,j) \) is the co-occurrence matrix.

2. Contrast

\[
CON = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \left( p(i,j) \right)^2 \right\}_{i-j} = n
\]

3. Inverse Different Moment

\[
IDM = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{p(i,j)}{1+(i-j)^2}
\]

4. Entropy

\[
ENT = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \log(p(i,j))
\]
5. Correlation

\[
COR = \frac{\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} (ij) p(i, j) - m_x m_y}{\sigma_x \sigma_y}
\]  

(9)

where \(m_x, m_y, \sigma_x\) and \(\sigma_y\) the respective mean values and standard deviations of \(p_x\) and \(p_y\), as described in equations (1) and (2), and where

\[
p_x(i) = \sum_{j=1}^{N_{max}} p(i, j)
\]

(10) and

\[
p_y(j) = \sum_{i=1}^{N_{max}} p(i, j)
\]

(11)

6. Sum of Squares

\[
SSQ = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} (1 - m)^2 p(i, j)
\]

(12)

7. Sum Average

\[
SAVE = \sum_{i=2}^{2N_g} ip_{x+y}(i)
\]

(13)

where

\[
p_{x+y}(k) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} p(i, j), i + j = k, k = 2, 3, ..., 2N_g
\]

(14)

8. Sum Entropy

\[
SENT = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log p_{x+y}(i)
\]

(15)

9. Sum Variance

\[
SVAR = -\sum_{i=2}^{2N_g} (i - SAVE)^2 p_{x+y}(i)
\]

(16)

10. Difference Variance

\[
DVAR = \sum_{i=2}^{2N_g} (i - SAVE)^2 p_{x-y}(i)
\]

(17)

11. Difference Entropy

\[
DENT = -\sum_{i=0}^{N_x-1} p_{x-y}(i) \log p_{x-y}(i)
\]

(18)

where

\[
p_{x-y}(k) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} p(i, j), |i - j| = k, k = 2, 3, ..., N_x - 1
\]

(19)
26

B.3.3.2.2 Features extracted from run length matrices.

A run length is a set of consecutive, pixels that have the same gray-level value. The length of the run is the number of pixels in the run considered. The "gray-level run length matrix" (R-L matrix) shows the frequency of each run length of each gray-level in the image, in a given direction. The lines of the R-L matrix correspond to different gray-levels while the columns correspond to different lengths of the run length in the ROI\(^{[18]}\). Just like the co-occurrence matrix, there are four directions of run lengths (horizontal-0°, diagonal-45°, vertical-90°, and antidiagonal-135°). The following example illustrates the calculation of the four R-L matrixes for the ROI shown in Table 1. Table 3 shows the four R-L matrices (RL0(i,j), RL45(i,j), RL90(i,j) and RL135(i,j)) calculated from that ROI for each one of the four directions.

<table>
<thead>
<tr>
<th>R(_{0}(i,j)=)</th>
<th>R(_{45}(i,j)=)</th>
<th>R(_{90}(i,j)=)</th>
<th>R(_{135}(i,j)=)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 0 0</td>
<td>1 2 0 0</td>
<td>0 1 1 0</td>
<td>3 1 0 0</td>
</tr>
<tr>
<td>0 2 0 0</td>
<td>2 1 0 0</td>
<td>0 2 0 0</td>
<td>2 1 0 0</td>
</tr>
<tr>
<td>0 1 1 0</td>
<td>1 2 0 0</td>
<td>3 1 0 0</td>
<td>5 0 0 0</td>
</tr>
<tr>
<td>0 1 0 0</td>
<td>2 0 0 0</td>
<td>2 0 0 0</td>
<td>2 0 0 0</td>
</tr>
</tbody>
</table>

Table 3. The four R-L matrices calculated from the ROI in table 1.

where R\(_{45}(0,0) = 1\) is the number of times that run length = 1 (i.e. 1 point of gray level 0) appears in the image g(i,j) along the 45° direction. Again, R\(_{45}(3,2) = 2\), i.e. in g(i,j) there are two combinations of two consecutive points of gray-level 3 along 45° direction. The rest of the R-L matrices are constructed in a similar way.

In some cases, the four matrices are combined to one just like it is done with the co-occurrence matrices in order to save computational time. The 5 textural features that can be computed from the R-L matrix of the ROI are demonstrated below. These features describe the length, distribution and homogeneity of run lengths within the ROI.

1. Short Run Emphasis

\[
SRE = \frac{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} r(i, j)}{N_x N_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} r(i, j)^2}
\] (20)

26
where \( r(i,j) \) is the run length matrix, \( N_g \) is the number of gray values in the image, \( N_r \) is the largest possible run, \( i=1,\ldots,N_g, j=1,\ldots,N_r \).

2. Long Run Emphasis
\[
LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 r(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} r(i,j)}
\]

3. Gray Level Non Uniformity
\[
GLNU = \frac{\sum_{i=1}^{N_g} \left( \sum_{j=1}^{N_r} r(i,j) \right)^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} r(i,j)}
\]

4. Run Length Non Uniformity
\[
RLNU = \frac{\sum_{j=1}^{N_r} \left( \sum_{i=1}^{N_g} r(i,j) \right)^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} r(i,j)}
\]

5. Run Percentage
\[
RP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} r(i,j)}{P}
\]

where \( P \) is the total possible number of runs in the ROI

B.3.3.2.3 Features extracted from Laws Texture Energy Measures.

The Texture Energy Measures are computed by first applying small convolution kernels to a digital image, and then performing a nonlinear windowing operation\[^{19,10}\]. The 2-D convolution kernels that are typically used for texture discrimination are generated from a set of five one-dimensional convolution kernels of length five each (see Table 4). The letters L,E,S,W and R stand for Level, Edge, Spot, Wave, and Ripple.
L5 = [ 1 4 6 4 1 ]
E5 = [ -1 -2 0 2 1 ]
S5 = [ -1 0 2 0 -1 ]
W5 = [ -1 2 0 -2 1 ]
R5 = [ 1 -4 6 -4 1 ]

Table 4. Laws one-dimensional convolution kernels

From these one-dimensional convolution kernels, 25 different two-dimensional convolution kernels can be calculated by convolving a vertical 1-D kernel with a horizontal 1-D kernel. As an example, the L5E5 kernel is found by convolving a vertical L5 kernel with a horizontal E5 kernel. In order to calculate the Texture Energy Images (TEM images), the ROI is convolved with each one of the 25 two-dimensional kernels and a windowing operation is performed. Finally the 25 TEM images are normalized and combined to create 14 TEM images that are rotationally invariant (see Table 5). From each one of the 14 TEM images 4 first order statistics textural features (see section B.3.3.1) are calculated and used to encode the textural information of the TEM images. In total 56 textural features are calculated from a single ROI. The TEMs are associated to the inhered microstructure of the image.

Table 5. Laws Texture Energy Images

<table>
<thead>
<tr>
<th>E5L5TR = E5L5T + L5E5T</th>
<th>W5S5TR = W5S5T + S5W5T</th>
</tr>
</thead>
<tbody>
<tr>
<td>S5L5TR = S5L5T + L5S5T</td>
<td>W5L5TR = W5L5T + L5W5T</td>
</tr>
<tr>
<td>R5S5TR = R5S5T + S5R5T</td>
<td>R5W5TR = R5W5T + W5R5T</td>
</tr>
<tr>
<td>R5L5TR = R5L5T + L5R5T</td>
<td>W5W5TR = W5W5T * 2</td>
</tr>
<tr>
<td>S5E5TR = S5E5T + E5S5T</td>
<td>E5E5TR = E5E5T * 2</td>
</tr>
<tr>
<td>R5E5TR = R5E5T + E5R5T</td>
<td>S5S5TR = S5S5T * 2</td>
</tr>
<tr>
<td>W5E5TR = W5E5T + E5W5T</td>
<td>R5R5TR = R5R5T * 2</td>
</tr>
</tbody>
</table>

B.3.3.3 Features Normalization.

The textural features that have been presented so far can have very different ranges of values. This might result on an imbalanced influence of features with larger and lower values. To avoid this textural features have to be normalized to have similar mean value and range. Usually zero mean value and unit standard deviation are used for normalization\cite{29}. The following equation performs normalization of a feature with
a value \( x_i \) by subtracting its mean value and dividing the result with the standard deviation (25).

\[
\bar{x}_i = \frac{x_i - \text{mean}}{\text{std}}
\]  

(25)

**B.3.4 FEATURES DISCRIMINATING POWER INVESTIGATION**

In order to verify the discriminating ability of the abovementioned textural features, a null hypothesis test has to be performed. Such a test is used to examine the hypothesis that the means of two normally distributed populations are equal. If the null hypothesis is rejected then the populations are considered to be statistically independent within a predetermined confidence level \( p \) (where usually \( p < 0.05 \) is used).

Student’s t-test is most commonly used for this purpose. It assumes that the values of both populations are normally distributed and that their variances are equal. Student’s t-test can be either paired or un-paired depending on the data. The following equations demonstrate the calculation of the t-value for paired and unpaired data. After calculating the t value the p value is found from matrices. If the null hypothesis is rejected the textural feature which is tested is considered to be able to demonstrate statistically significant differences between the two populations\(^{27}\).

The populations are considered to be normally distributed but this is not always the case. In order to provide that both populations have a normal distribution of data the Lilliefors test is used. It is an adaptation of the Kolmogorov-Smirnov test. It is used to test the null hypothesis that data come from a normally distributed population\(^{15}\). If the null hypothesis is rejected this indicates that the populations are normally distributed and thus the t-test could be used.

**paired t-test**

\[
t = \frac{\bar{X}_D}{s_D} \sqrt{N}
\]  

(26)

For this equation, the differences D between all pairs must be calculated. \( \bar{X}_D \) is the mean value of the population of differences and \( s_D \) is the standard deviation.
unpaired t-test
\[ t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_{\bar{X}_1-\bar{X}_2}}} \] (27)

where \( s_{\bar{X}_1-\bar{X}_2} = \sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \) (28)

where \( \bar{X}_1 \) and \( \bar{X}_2 \) are the mean values of the two populations, \( s_1 \) and \( s_2 \) are their standard deviations and \( n_1, n_2 \) are the number of participants in each population.

The features that are proven to have statistically significant differences are used to design the system whereas those features that are not statistically independent are not used. This process is called Features Reduction by means of statistical analysis methods\(^{[26]}\).

**B.3.5. CLASSIFICATION**

A supervised classification system is used to classify inputs (feature vectors) into defined output categories (classes). The heart of this system is the classifier which is a decision-making algorithm that has been designed to perform the abovementioned task. In order to classify unknown data, the classifier has to be trained to learn the differences between classes using feature vectors of known class. Based on the input feature vectors, the classifier creates a ‘unique’ description of each predefined classification category. The input feature vectors from each class are referred to as ‘training sets’. When training is completed, the system is ready to decide on the classification of new data that haven’t been used for training\(^{[24]}\). Some of the classification algorithms that are commonly used are demonstrated below.

**B.3.5.1 Minimum Distance Classifier**

This is one of the simplest classifiers. It calculates the Euclidian distance of the input feature vector from the center of each class (center is the mean value vector)
and classifies the input vector to the class that minimizes this distance\cite{25}. The discriminant function of this classifier is given by equation (29).

\[
d_j(x) = x^T m_j - \frac{1}{2} (m_j m_j^T)
\]  

(29)

where \(m_j\) is the mean vector of class \(j\) and \(x\) is the input feature vector

**B.3.5.2 Nearest Neighbor Classifier**

This classifier calculates the distances between the input feature vector and each one of the training vectors and classifies the input vector to the class of the vector with the minimum distance (the nearest neighbor)\cite{25}. A variation of the nearest neighbor classifier is the k-NN (k-nearest neighbor) which classifies the input vector to the class where most on the k nearest neighbors belong.

**B.3.5.3 Least Squares Minimum Distance Classifier**

The Least Squares Minimum Distance Classifier (LSMD) is a variation of the minimum distance classifier\cite{24}. To overcome the problem of overlapping classes where the minimum distance classifier fails, the LSMD forces the data of each class to cluster around a pre-selected point in the augmented feature space. In this case even if the data of each class are not sufficiently represented by the mean value of the class they are forced to do so at the augmented feature space. Then the augmented feature space is transformed to the classifier space following the mean square error minimization. Finally the transformed data are classified employing the minimum distance rule. Equation (30) gives the transformation from the augmented feature space to the classifier space under the limitation of equation (31) which is the mean square error minimization rule. The discriminant function is given in equation (32).

\[
A = \left[ \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} V_i (x_{i,j}^T) \right]^{-1} \left[ \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} (x_{i,j} x_{i,j}^T) \right]^{-1}
\]  

(30)

\[
\mathbf{e}_j = \mathbf{x}_{i,j} - \mathbf{V}_i = \mathbf{A} x_{i,j} - \mathbf{V}_i
\]  

(31)

\[
d_j(x) = \sum_{i=1}^{M} a_{c_i} x_i - b_c
\]  

(32)
where \( V \) is the pre-selected point for each class around which data are forced to cluster, \( i=1,\ldots,k \) is the number of the class, \( x_{i,j} \) is the input feature vector and \( x \) is the transformed feature vector. In addition \( e_i \) is the calculated error, while \( M \) is the number of features, \( a_c \) and \( b_c \) are the weight coefficients that derive from the transformation function \( A \).

A variation of the LSMD classifier is the Quadratic classifier (QLSMD) which uses the second order discriminant function given by equation (33). This classifier can draw a second order multidimensional surface between the data and thus it can classify more effectively data that belong to overlapping classes\(^{20,24}\).

\[
d_i(x) = \sum_{j=1}^{M} a_{ci,j} x_i + \sum_{j=1}^{M-1} \sum_{k=j+1}^{M} a_{ci,j} x_j x_k + \sum_{j=1}^{M} a_{ci,j} x_j - b_c \tag{33}
\]

**B.3.5.4 Bayesian Classifier**

The Bayesian is probabilistic classifier and based on Bayes theorem. It is designed to classify the data providing the minimum probability error\(^{22,25}\). It’s main drawback is that it assumes a Gaussian distribution of data. However it is an effective and simple algorithm that can provide accurate results even when a small amount of training data is used. The discriminant function of the Bayesian classifier is given by equation (34).

\[
d_i = x^T C^{-1} m_i - \frac{1}{2} m_i^T C^{-1} m_i \tag{34}
\]

where \( x \) is the input feature vector \( i=1,2,\ldots,k \) is the number of the class, \( m_i \) is the mean feature vector given by equation (35) and \( C \) is the covariance matrix for class \( i \) given by equation (36).

\[
m_i = \frac{1}{N_i} \sum_{k=1}^{N_i} z_k \tag{35}
\]

\[
C_i = \left( \frac{1}{N_i} \sum z z^T \right) - m_i m_i^T \tag{36}
\]
where \( z_k \) is the mean feature vector of class \( k \) and \( N_i \) is the number of feature vectors that belong to class \( i \).

A variation of the Bayesian classifier is the Quadratic Bayesian whose discriminant function is given by equation (37).

\[
d_i = \ln(P_i) - \frac{1}{2} \ln|C_i| - \frac{1}{2} \left( (x - m_i)^T C_i^{-1} (x - m_i) \right) \tag{37}
\]

where again \( x \) is the input feature vector \( i=1,2,\ldots,k \) is the number of the class, \( m_i \) is the mean feature vector, \( C \) is the covariance matrix for class \( i \) and \( P_i \) is the a priori probability of an input pattern vector to belong in class \( i \).

**B.3.5.4 Probabilistic Neural Network**

The Probabilistic Neural Network Classifier (PNN) classifier is primarily based on Bayes theorem for conditional probability and Parzen’s method for estimating probability density function of random variables\(^{[21]}\).

Its architecture consists of four layers (see Figure 9). An input layer, used to input the new feature vector. A pattern layer, in which all the distances from the input feature vector to the training feature vectors are computed, and a summation layer in which the outputs from the pattern layer are connected to the summation units. Each output is connected on the summation unit of the class it represents. The last layer is the output layer in which the decision class for the unknown pattern vector is chosen. An important advantage of the PNN classifier is that it minimizes the expected risk of classifying patterns in the wrong category. Moreover it is guaranteed to converge to an optimal classifier as the size of the representative training set increases. The disadvantages of this classifier are that the entire training set must be stored and used during testing and that the amount of computation needed to classify an unknown pattern depends on the size of the training set thus it has large memory requirements and results in slow execution for a large set of data.
The general form of the discriminant function of the PNN classifier for class $j$ is given by equation (38) and the input feature vector is classified to the class with the larger value of the discriminant function.

$$g_j(x) = \frac{1}{(2\pi)^{p/2} \sigma^p N_j} \sum_{i=1}^{N_j} w(y)$$

(38)

where $w(y)$ is the kernel function and $y$ is a function of the input feature vector $x$ as given by equation (39). In addition, $x_i$ is the $i$-th training pattern vector, $N_j$ is the number of patterns in class $j$, $\sigma$ is a smoothing parameter, and $p$ is the number of features employed in the feature vector.

$$y_i = \frac{1}{\sigma} \sqrt{\|x - x_i\|^2}$$
Equations (40), (41) and (42) demonstrate different functions that can be used as kernels for the PNN classifier.

Gaussian:

\[
w(y) = e^{-\frac{y^2}{2}}
\]  

\[ (40) \]

Exponential:

\[
w(y) = e^{-|y|}
\]  

\[ (41) \]

Reciprocal:

\[
w(y) = \frac{1}{1 + y^2}
\]  

\[ (42) \]

**B.3.6. FEATURES SELECTION**

In this chapter, so far several feature extraction and classification methods have been presented. An essential step in the design of an image analysis system is the selection of the optimum features combination. Although it seems reasonable that the more features one uses the higher the classification accuracy that is achieved this is not usually correct. When the number of features employed by a system increases, so does the complexity of the classifier, and thus the computational power that is needed increases tremendously\[25\]. In addition it has been shown that the overall error increases when too many features are employed by the system.

The optimum combination of features is considered to be that combination which provides the highest classification accuracy employing the minimum number of features. Several methods can be used to determine the optimum combination of features, such as Exhaustive search, Sequential Forward and Sequential Backward Selection, Principal and Independent Component Analysis and t-test statistical methods. The Exhaustive Search (ES) method is the most accurate but it is computationally intensive\[26\]. This method is calculating all possible features combinations and tests the performance of the classifier for each one of these combinations in order to find the optimum one. Another method employs Student’s
t-test in order to detect which features have the highest discriminating power. When the number of features is large a combination of Exhaustive Search and Student’s t-test is usually preferred.

**B.3.7. SYSTEM DESIGN AND PERFORMANCE EVALUATION**

In order to design the image analysis system all the abovementioned methods have to be combined. The design procedure follows 7 steps.

1. Image Preprocessing
2. ROIs Selection
3. Features Generation
4. Features Reduction employing statistical independence analysis
5. Features Selection
6. Classifier Selection
7. Performance Evaluation

So far in this chapter steps one to five of the system design procedures have been covered. In step six the best classifier has to be selected for the given set of data. The optimum combination of features that has been selected in step six is used to design all the classifiers. Then the classification accuracy of each one is calculated using equation (43) and the classifier that demonstrated the highest classification accuracy is chosen to be used by the system. Step seven of the design procedure refers to the performance evaluation of the image analysis system. In medical image analysis problems the amount of data available for training the system is usually limited. This is the reason that methods like Self Consistency (SC) and Leave One Out (LOO) are used to estimate the classification error of the system\textsuperscript{[25]}.

SC and LOO methods exploit the given set of data by using them alternatively for training and design in order to estimate the classification accuracy. The SC method estimates the reliability of the classifier using all available training data, first for training and then for testing. The result of self-consistency is an optimistically biased estimate of the classifier performance, since the same dataset is used for both training and testing. However, it is an important step for assessing the reliability of the classifier in correctly classifying the data that are used for its design. In contrast to the SC, the LOO method is not biased since it uses different set of data for training and
testing the classifier. In this method all but one feature vectors are used for training the classifier and the feature vector that was left out is used as the unknown vector to be classified in order to test the classification performance. This process is repeated as many times as the number of feature vectors, by leaving out a different one each time\(^{26}\).

In order to evaluate the system’s performance the confusion matrix shown in Table 6 is used. The following equations demonstrate the calculation of the classification accuracy (43), sensitivity (44), and specificity (45) of the system.

<table>
<thead>
<tr>
<th>Decision Class</th>
<th>True Class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 2</td>
</tr>
<tr>
<td>Class 1</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
</tr>
<tr>
<td>Class 2</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
</tr>
</tbody>
</table>

Table 6. Truth table demonstrating then confusion matrix used to evaluate the classification performance

where, \( n_{11} \) and \( n_{22} \) are the number of correctly classified feature vectors, and \( n_{21} \) and \( n_{12} \) are the numbers of misclassified vectors.

\[
\text{Overall classification accuracy} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100 \quad (43)
\]

\[
\text{Sensitivity} = \frac{n_{11}}{n_{11} + n_{12}} \times 100 \quad (44)
\]

\[
\text{Specificity} = \frac{n_{22}}{n_{21} + n_{22}} \times 100 \quad (45)
\]

where equations (44) and (45) refer to the calculation of sensitivity and specificity of the system in the case that the problem is to characterize the data belonging to Class 2.
CHAPTER C
MATERIALS & METHODS

C.1. DATA ACQUISITION AND INTERVERTEBRAL DISC CHARACTERIZATION .............................................39
C.2. DESIGN OF THE IMAGE ANALYSIS SYSTEM .............................................40
C.3. SYSTEM PERFORMANCE EVALUATION METHOD .....................................43
C.1. DATA ACQUISITION AND INTERVERTEBRAL DISC CHARACTERIZATION

The sample comprised 14 saggital images of the cervical spine. These images were obtained employing a 1.0 Tesla MRI scanner (Philips Intera) with a surface neck coil. The images were printed on film and reviewed by an experienced orthopedist. Employing the classification scheme proposed by Matsumoto, the orthopedist characterized the intervertebral discs in reference to the narrowing of the disc space (see Table 7)\(^6\). The selection of the specific classification scheme was based on the fact that it is specifically designed for cervical intervertebral disc characterization in contrast to most classification schemes that are designed for the characterization of intervertebral disc of the lumbar spine\(^9\). In addition this scheme is widely accepted and used for this specific purpose.

From the 70 intervertebral discs (intervertebral disc spaces) examined, 54 were found to be normal since they had no disc space narrowing or less than 25\% loss of height in comparison to the most adjacent normal disc space (M=0). The remaining 16 discs were characterized as degenerated (narrowed). 14 of them were slightly narrowed since they had 25\% to 50\% loss of height (M=1) and 2 discs were severely narrowed since they had more than 50\% loss of height (M=2).

<table>
<thead>
<tr>
<th>Grade(^a)</th>
<th>Binary notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc degeneration</td>
<td>0 Bright as or slightly less bright than CSF</td>
</tr>
<tr>
<td></td>
<td>1 Dark and/or speckled</td>
</tr>
<tr>
<td></td>
<td>2 Almost black</td>
</tr>
<tr>
<td>Posterior disc protrusion</td>
<td>0 Disc material confined within the posterior margin of the VB</td>
</tr>
<tr>
<td></td>
<td>1 Disc material protruding beyond the posterior margin of the VB without cord compression</td>
</tr>
<tr>
<td></td>
<td>2 Beyond VB with cord compression</td>
</tr>
<tr>
<td>Anterior disc protrusion</td>
<td>0 Disc material confined within the anterior margin of the VB</td>
</tr>
<tr>
<td></td>
<td>1 Disc material protruding beyond the anterior margin of the VB</td>
</tr>
<tr>
<td>Narrowing of the disc space</td>
<td>0 No narrowing or less than 25% loss in height compared with the most adjacent normal disc space</td>
</tr>
<tr>
<td></td>
<td>1 25% to 50% loss of height</td>
</tr>
<tr>
<td></td>
<td>2 More than 50% loss of height</td>
</tr>
<tr>
<td>Foraminal stenosis</td>
<td>0 No obliteration of intraforaminal fat</td>
</tr>
<tr>
<td></td>
<td>1 Disc material or bony spurs obliterating intraforaminal fat unilaterally or bilaterally</td>
</tr>
</tbody>
</table>

\(^a\) CSF, cerebrospinal fluid; VB, vertebral body

Table 7. The cervical intervertebral disc classification scheme proposed by Matsumoto et al
The disc space measurements for the characterization of disc space narrowing were manually performed and a computerized example is shown in Figure 10. In this image the first three cervical discs were characterized as normal whereas the fourth and fifth discs were characterized as degenerated (narrowed with M=1).

Figure 10. T2-weighted sagittal images of the cervical spine: (a) original image demonstrating the disc space measurement technique, (b) the corresponding image processed using wavelet denoising and adaptive histogram equalization.

C.2. DESIGN OF THE IMAGE ANALYSIS SYSTEM

The design procedure follows the 7 steps that have been analyzed in Section B.3.7. The implementation of each one of these steps is presented in the following sections.

C.2.1. Image Preprocessing

Each one of the 14 MR Images was preprocessed using the wavelet multiscale denoising (4 decomposition levels) and adaptive contrast enhancement techniques (adaptive histogram equalization). For this purpose dedicated digital image processing
software was used (Medical Visualization Tool)\textsuperscript{[12,13,14]}. Figure 11 is an example of a MR image before and after enhancement. Contrast enhancement allows for better visualization of the structures and facilitates the segmentation of regions of interest.

**C.2.2. ROIs Selection**

On each enhanced MR image, 5 ROIs, corresponding to the patient’s five cervical intervertebral discs were determined by manual delineation. This task was accomplished employing the same software that was used for image enhancement\textsuperscript{[12,13,14]}. Figure 11 shows an enhanced image in which the discs have been delineated along with the ROIs corresponding to these discs.

![Figure 11. Enhanced image demonstrating the cervical intervertebral discs delineation and five ROIs corresponding to the delineated intervertebral discs.](image-url)
C.2.3. Features Generation

A total of 72 textural features were extracted from each segmented disc ROI, utilizing custom developed algorithms. In particular, (i) four textural features were computed from the ROI’s grey level histogram (first order statistics textural features\textsuperscript{[25]}), (ii) seven from the ROI’s grey level co-occurrence matrix (co-occurrence features\textsuperscript{[28]}), (iii) five using the ROI’s grey level run-length matrix (run-length features\textsuperscript{[18]}) and (iv) fifty six from the TEM images (Laws textural features\textsuperscript{[19]}).

C.2.4. Features Reduction employing statistical analysis

In order to investigate the discriminating ability of the abovementioned textural features, unpaired Student’s t-test was used\textsuperscript{[27]}. In addition the Lilliefors test was performed for each set of features, to verify that both normal and degenerated populations had Gaussian distributions of data\textsuperscript{[15]}. All statistical processing was performed utilizing the “Matlab Statistics Toolbox”.

Those features that were proven to have statistically significant differences (p<0.01), were retained and used to design the image analysis system, whereas the ones that were not found to be statistically independent, were rejected. In this way the total number of features used to design the system was reduced, since only the most discriminating features were retained\textsuperscript{[26]}.

C.2.5. Features Selection

The next step was to determine the optimum combination of features (i.e. the features combination providing the highest classification accuracy employing the minimum number of feature)\textsuperscript{[26]}. For this purpose the Exhaustive Search method was used to examine all possible combinations of 2 to 5 features.

C.2.6. Classifier Selection

In order to improve the performance of the system eight different classification algorithms were developed and tested using Matlab\textsuperscript{®} (see Table 8). The classification accuracy of each algorithm was tested using the LOO method. The
classifier that provided the highest accuracy along with the optimum features combination was chosen as the basis of the image analysis system.

<table>
<thead>
<tr>
<th>Minimum Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nearest Neighbor</td>
</tr>
<tr>
<td>k- Nearest Neighbor (with k=5)</td>
</tr>
<tr>
<td>Least Squares Minimum Distance</td>
</tr>
<tr>
<td>Quadratic Least Squares Minimum Distance</td>
</tr>
<tr>
<td>Bayesian</td>
</tr>
<tr>
<td>Quadratic Bayesian</td>
</tr>
<tr>
<td>Probabilistic Neural Networks</td>
</tr>
</tbody>
</table>

Table 8. The classification algorithms that were tested

C.3. SYSTEM PERFORMANCE EVALUATION METHOD

The performance of the proposed image analysis system was evaluated using the LOO method in order to calculate the following three parameters.

(a) the classification accuracy of the system
(b) the sensitivity of the system in detecting a degenerated disc
(c) the specificity of the system in the same task
CHAPTER D

RESULTS AND DISCUSSION

D.1 STATISTICAL ANALYSIS RESULTS...............................................45
D.2. EVALUATION OF THE DESIGNED SYSTEM AND COMMENTS
      ON THE RESULTS.................................................................46
RESULTS AND DISCUSSION

D.1. STATISTICAL ANALYSIS RESULTS

The Lilliefors test verified that textural features of both normal and degenerated populations had Gaussian distributions of data, thus t-test analysis could be used. Statistical analysis revealed the existence of statistically significant differences (p<0.05) between normal and degenerated discs for the generated feature values. The specific finding can be considered as indicative of differentiations of intervertebral disc image texture due to DDD. It has been shown that disc texture changes in MR images correlate with structural, histological and biochemical alterations in the intervertebral disc which are indicative of the DDD. Thus, texture differentiation seems reasonable.

In order to reduce the number of features that would be used for the design of the system only the features that provided high statistical independence between the normal and degenerated populations were used. 10 out of the 72 textural features were found to be most discriminating (p<0.01). The names and types of these textural features are shown in Table 9. At this point it should be noticed that 8 out of 10 textural features that are considered to be the most discriminating are calculated from Texture Energy Measures. Thus it can be said that TEMs are pretty sensitive to texture differentiation between normal and degenerated intervertebral discs.

<table>
<thead>
<tr>
<th>Textural Features</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>1st order statistics</td>
</tr>
<tr>
<td>Sum average</td>
<td>Co-occurrence matrix</td>
</tr>
<tr>
<td>Standard deviation (from E5L5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Kurtosis (from S5L5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Kurtosis (from W5L5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Kurtosis (from R5L5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Standard deviation (from S5E5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Standard deviation (from W5E5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Standard deviation (from R5E5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Mean (from R5S5TR)</td>
<td>Texture Energy Measures</td>
</tr>
</tbody>
</table>

Table 9. The most discriminating textural features
D.2. EVALUATION OF THE DESIGNED SYSTEM AND COMMENTS ON THE RESULTS

D.2.1 BEST FEATURES COMBINATION RESULTS

The optimum combination of features was selected from the reduced set of features using the Exhaustive Search method. All combinations of 2 to 5 features were tested to find the optimum one which comprised 4 textural features. Three of these features were calculated from TEMs while one was calculated from 1st order statistics (i.e. the histogram of the image). The optimum features combination is given in Table 10.

<table>
<thead>
<tr>
<th>Textural Features</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>1st order statistics</td>
</tr>
<tr>
<td>Kurtosis (from R5L5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Standard deviation (from S5E5TR)</td>
<td>Texture Energy Measures</td>
</tr>
<tr>
<td>Standard deviation (from W5E5TR)</td>
<td>Texture Energy Measures</td>
</tr>
</tbody>
</table>

Table 10. The features that comprise the optimum combination

D.2.2. CLASSIFICATION ALGORITHMS RESULTS

In order to find the algorithm that would provide the highest classification accuracy each one of the 8 classifiers was designed using the best features combination. Using the LOO method the classification accuracies of these algorithms were calculated and are shown in Table 11.

<table>
<thead>
<tr>
<th>Classification Algorithm</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Distance</td>
<td>87.5</td>
</tr>
<tr>
<td>Nearest Neighbor</td>
<td>81.3</td>
</tr>
<tr>
<td>k- Nearest Neighbor (with k=5)</td>
<td>81.3</td>
</tr>
<tr>
<td>Least Squares Minimum Distance</td>
<td>93.8</td>
</tr>
<tr>
<td>Quadratic Least Squares Minimum Distance</td>
<td>78.1</td>
</tr>
<tr>
<td>Bayesian</td>
<td>93.8</td>
</tr>
<tr>
<td>Quadratic Bayesian</td>
<td>78.1</td>
</tr>
<tr>
<td>Probabilistic Neural Networks</td>
<td>84.4</td>
</tr>
</tbody>
</table>

Table 11. The classification accuracies of different algorithms
Both the Bayesian and Least Squares Minimum Distance algorithms achieved the highest classification accuracy. The LSMD classifier was the one chosen to be used by the system because in contrast to the Bayesian it does not assume normal distributions of data. The LSMD has the advantage to translate the data from the feature space to the classifier’s space and thus it provides better dissociation between classes with overlapping data.

**D.2.3 SYSTEM PERFORMANCE EVALUATION**

The system was designed using the LSMD classifier along with the four textural features that comprised the best features combination. To evaluate the performance of the system, the accuracy, sensitivity and specificity parameters were calculated and are presented in Table 12.

<table>
<thead>
<tr>
<th>Disc characterization</th>
<th>Normal</th>
<th>Degenerated</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>15</td>
<td>1</td>
<td>93.8%</td>
</tr>
<tr>
<td>Degenerated</td>
<td>1</td>
<td>15</td>
<td>93.8%</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td>93.8%</td>
</tr>
</tbody>
</table>

Table 12. Truth table demonstrating classification results concerning the discrimination between normal and degenerated discs.

The classification accuracy, sensitivity and specificity of the system had all reached 93.8%, since only one normal and one degenerated disc were misclassified by the system.

For comparison reasons the classification accuracy for a system designed using only features calculated from first and second order statistics was also calculated. Employing again the LOO and Exhaustive search methods the best classification algorithm and best combination of features were selected. The LSMD classifier designed with the standard deviation, run length emphasis, run percentage, inverse different moment and difference variance textural features provided the highest classification accuracy. In addition the sensitivity and specificity were 81.3% and 100% respectively (see Table 13).
By comparing the two tables it should be noticed that the classification accuracy was improved by 3.2%. Thus the system that employed all textural features (TEMs, 1\textsuperscript{st} and 2\textsuperscript{nd} order statistics) performed better in the task of classification in comparison to that which did not use the TEMs. Moreover the sensitivity of the system was improved by 12.3% whereas the specificity decreased by 6.2%. Thus it can be stated that TEMs contributed significantly on the task of discriminating between normal and degenerated disc. But the most important achievement is that they increased significantly the system’s sensitivity.

<table>
<thead>
<tr>
<th>Disc characterization</th>
<th>Normal</th>
<th>Degenerated</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>16</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Degenerated</td>
<td>3</td>
<td>13</td>
<td>81.3%</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td>90.6%</td>
</tr>
</tbody>
</table>

Table 13. Truth table demonstrating classification results for comparison reasons

Finally it should be noticed that the results presented so far have a major limitation which can greatly affect their accuracy. This limitation refers to the size of the sample used for the design of the system, which is quite small. Due to the limited amount of data used to design the system it is possible that the results of this study could greatly vary if different sets of data were used for the same purpose. Thus to have consistent results that would be used in clinical conditions the system has to be trained using more data to achieve both optimization and consistency for the classification accuracy.
CHAPTER E

CONCLUSIONS AND FUTURE WORK

E.1. CONCLUSIONS ......................................................................................50
E.2. FUTURE WORK ...................................................................................50
E.1. CONCLUSIONS

The proposed computer-based image analysis system discriminated successfully normal from degenerated cervical intervertebral discs. The system was designed on the basis of textural features, which were found sensitive enough to capture structural alterations of the intervertebral cervical disc, associated to the Degenerative Disc Disease. It consisted of the LSMD classifier designed using with the best combination of features which comprised four textural features and achieved managed to successfully characterize 30 out of 32 intervertebral discs. The classification accuracy was 93.8% and the sensitivity and specificity of the system had exactly the same value. It was also shown that TEMs had high discriminating power and through a comparison study they were proved to significantly increase the system’s sensitivity. Further work and validation of the system employing a larger sample is needed to make the system a trustworthy and useful tool to the physicians as a decision-support tool.

E.2. FUTURE WORK

Future work will focus on improving the classification accuracy by employing different classification algorithms and textural features. In addition morphological features could be tested for the characterization of DDD. In reference to the method of segmentation, an algorithm for the automatic segmentation of the intervertebral disc that would replace the manual delineation, would make the system more used friendly and more importantly, the results will be user independent. Moreover it would be useful to test the system on a larger sample and adapt it to discriminate between slightly and severely degenerated discs (ie discs with M=1 and discs with M=2).

The proposed system might also be useful for the characterization of intervertebral discs of the lumbar spine. A future study could go further from the characterization of the discs as normal or narrowed (degenerated) to the detection of different characteristics of DDD. For example disc herniation (protrusion) and intraforaminal fat obliteration are two more criteria used for the evaluation of DDD which are also included in the classification scale proposed by Matsumoto et al.

Finally a different approach on the problem would be to quantitatively estimate the intervertebral disc degeneration, and calculate a percentage for it.
REFERENCES

Scientific Papers:


References

Books:


Web Pages:


[34] Congress of Neurological Surgeons (2007), http://www.neurosurgeon.org/

## APPENDIX I

### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHE</td>
<td>Adaptive Histogram Equalization</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>DDD</td>
<td>Degenerative Disc Disease</td>
</tr>
<tr>
<td>ES</td>
<td>Exhaustive Search</td>
</tr>
<tr>
<td>K-NN</td>
<td>K-Nearest Neighbor</td>
</tr>
<tr>
<td>LOO</td>
<td>Leave One Out</td>
</tr>
<tr>
<td>LSMD</td>
<td>Least Squares Minimum Distance</td>
</tr>
<tr>
<td>M</td>
<td>Matsumoto’s scale</td>
</tr>
<tr>
<td>MD</td>
<td>Minimum Distance</td>
</tr>
<tr>
<td>MIAS</td>
<td>Medical Image Analysis System</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic Resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NN</td>
<td>Nearest Neighbor</td>
</tr>
<tr>
<td>PNN</td>
<td>Probabilistic Neural Network</td>
</tr>
<tr>
<td>QLSMD</td>
<td>Quadratic Least Squares Minimum Distance</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>SC</td>
<td>Self Consistency</td>
</tr>
<tr>
<td>TEM</td>
<td>Texture Energy Measures</td>
</tr>
</tbody>
</table>
APPENDIX II

LIST OF FIGURES

Figure 1. Anatomy of the cervical spine

Figure 2. Transverse view of the cervical spine components

Figure 3. Anatomical diagram of the spinal cord and nerve roots

Figure 4. Image showing examples of different intervertebral disc problems

Figure 5. X-ray views of the cervical spine

Figure 6. Transverse (on the left) and Sagittal (on the right) views of a cervical spine T2-weighted MRI scan

Figure 7. Transverse (on the left) and Sagittal (on the right) views of the cervical spine imaged employing T2-weighted MRI scan

Figure 8. T1 weighted (on the left) and T2 weighted (on the right) sagittal MR images of the cervical spine

Figure 9. Schematic diagram of the PNN classifier

Figure 10. T2-weighted sagittal images of the cervical spine

Figure 11. Enhanced image demonstrating the cervical intervertebral discs delineation
APPENDIX III

LIST OF TABLES

Table 1.  A four gray-levels ROI

Table 2.  The four co-occurrence matrices calculated from the ROI in table 1

Table 3.  The four R-L matrices calculated from the ROI in table 1

Table 4.  Laws one-dimensional convolution kernels

Table 5.  Laws Texture Energy Images

Table 6.  Truth table demonstrating then confusion matrix used to evaluate the classification performance

Table 7.  The cervical intervertebral disc classification scheme proposed by Matsumoto et al

Table 8.  The classification algorithms that were tested

Table 9.  The most discriminating textural features

Table 10.  The features that comprise the optimum combination

Table 11.  The classification accuracies of different algorithms

Table 12.  Truth table demonstrating classification results concerning the discrimination between normal and degenerated discs.

Table 13.  Truth table demonstrating classification results for comparison reasons
ABSTRACT

Introduction: A computer-based classification system is proposed for the characterization of cervical intervertebral disc degeneration from saggital magnetic resonance images.

Materials and methods: Cervical intervertebral discs from saggital magnetic resonance images were assessed by an experienced orthopaedist as normal or degenerated (narrowed) employing Matsumoto’s classification scheme. The digital images were enhanced and the intervertebral discs which comprised the regions of interest were segmented. First and second order statistics textural features extracted from thirty-four discs (16 normal and 16 degenerated) were used in order to design and test the classification system. In addition textural features were calculated employing Laws TEM images. The existence of statistically significant differences between the textural features values that were generated from normal and degenerated discs was verified employing the Student’s paired t-test. A subset with the most discriminating features (p<0.01) was selected and the Exhaustive Search and Leave-One-Out methods were used to find the best features combination and validate the classification accuracy of the system. The proposed system used the Least Squares Minimum Distance Classifier in combination with four textural features with comprised the best features combination in order to classify the discs as normal or degenerated.

Results: The overall classification accuracy was 93.8% misdiagnosing 2 discs. In addition the system’s sensitivity in detecting a narrow disc was 93.8% and its specificity was also 93.8%.

Conclusion: Further investigation and the use of a larger sample for validation could make the proposed system a trustworthy and useful tool to the physicians for the evaluation of degenerative disc disease in the cervical spine.
ΠΕΡΙΛΗΨΗ

Σκοπός: Η στένωση των μεσοσπονδύλων δίσκων της αυχενικής μοίρας, ως κύρια έκφραση εκφυλιστικής νόσου, είναι μια από τις σημαντικότερες αιτίες πρόκλησης πόνου στην περιοχή του αυχένα. Στην κλινική πράξη η αξιολόγηση της στένωσης γίνεται μέσω μέτρησης του μεσοσπονδύλου διαστήματος, σε διάφορες απεικονίσεις της αυχενικής μοίρας του ασθενούς. Στην παρούσα εργασία προτείνεται μια υπολογιστική μέθοδος ανάλυσης εικόνας, για την αυτοματοποιημένη εκτίμηση της στένωσης από εικόνες μαγνητικής τομογραφίας.

Υλικό και Μέθοδος: Μελετήθηκαν 34 μεσοσπονδύλιοι δίσκοι από οβελιαίες τομές μαγνητικής τομογραφίας της αυχενικής μοίρας, οι οποίες ελήφθησαν με χρήση Τ2 ακολουθίας. Η στένωση των μεσοσπονδύλων δίσκων αξιολογήθηκε από έμπειρο ορθοπαιδικό βάσει της κλίμακας Matsumoto. Οι δίσκοι χωρίστηκαν σε δύο κατευγορίες: (α) 16 φυσιολογικοί και (β) 16 δίσκοι που παρουσίαζαν στένωση. Με χρήση διαδραστικού περιβάλλοντος επεξεργασίας εικόνες καθορίστηκε το περιγραμμα των μεσοσπονδύλων δίσκων οι οποίοι αποτελούν τις προς ανάλυση περιοχές ενδιαφέροντος (Π.Ε.). Σε κάθε Π.Ε. εφαρμόστηκαν αλγόριθμοι εξαγωγής χαρακτηριστικών υψής. Συγκεκριμένα υπολογιστικά χαρακτηριστικά υψής από στατιστικά πρώτης και δεύτερης τάξης καθώς και χαρακτηριστικά από τα μέτρα ενέργειας υψής κατα Law's. Τα παραπάνω χαρακτηριστικά, ποσοτικοποιούν διαγνωστικές πληροφορίες της έντασης του σήματος της Π.Ε. και συσχετίζονται με τη βιοχημική σύσταση των απεικονιζόμενων δομών. Τα εξαγόμενα χαρακτηριστικά υψής αξιοποιήθηκαν για τη σχεδίαση του ταξινόμησης ελάχιστης απόστασης ελαχίστων τετραγώνων, ο οποίος χρησιμοποιήθηκε για το διαχωρισμό μεταξύ φυσιολογικών δίσκων και δίσκων που παρουσίαζαν στένωση (εκφυλισμένων).

Αποτελέσματα: Η ακρίβεια της ταξινόμησης φυσιολογικών και εκφυλισμένων μεσοσπονδύλων δίσκων ανήλθε σε 93.8%. Η ευαισθησία καθώς και η ειδικότητα της μεθόδου, σε ότι αφορά την ανίχνευση εκφυλισμένων δίσκων, είναι επίσης 93.8%.

Συμπέρασμα: Με δεδομένο το μικρό μέγεθος του δείγματος που χρησιμοποιήθηκε για το σχεδιασμό της μεθόδου, απαιτούνται περαιτέρω εργασίες πιστοποίησης της ακρίβειας ταξινόμησης, προκειμένου η μέθοδος αυτή να αξιοποιηθεί από ακτινολόγους και ορθοπαιδικούς, ως βοηθητικό διαγνωστικό εργαλείο.