Development of new MRI-based analysis methods for improved diagnosis of low back pain patients

Christian Waldenberg

Department of Medical Radiation Sciences
Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2023

Cover illustration: Artwork in the shape of the lumbosacral spine created by Christian Waldenberg and partly generated with http://www.wordclouds.com/. The words reflect the professions, sciences, and methods applied in this thesis.

Development of new MRI-based analysis methods for improved diagnosis of low back pain patients

christian.waldenberg@gu.se
christian.waldenberg@vgregion.se

ISBN 978-91-8069-193-2 (PRINT)
ISBN 978-91-8069-194-9 (PDF)

Printed in Borås, Sweden 2023
Printed by Stema Specialtryck AB

"… medicine is so broad and complex that it is difficult, if not impossible, to capture the relevant information in rules."

– William B. Schwartz 1987

Development of new MRI-based analysis methods for improved diagnosis of low back pain patients

Christian Waldenberg

Department of Medical Radiation Sciences, Institute of Clinical Sciences
Sahlgrenska Academy, University of Gothenburg,
Gothenburg, Sweden

ABSTRACT

Background. Low back pain (LBP) is the leading cause of disability worldwide, where three of four individuals experience back pain at some point in their lives. The pathophysiological background of LBP is probably multifactorial, where bone marrow damage, tissue changes in the vertebral endplates, and intervertebral disc (IVD) degeneration have been recognized as tissue changes linked to pain. Annular fissures within the IVD are of particular interest, as they may be associated with vascular and nerve ingrowth. However, these tissue changes are also common in asymptomatic individuals, making it difficult to correctly identify the cause of pain in the individual patient.
Aim. This thesis aims to develop data-driven MRI-based analysis methods to improve the understanding of spinal tissue changes and their association with LBP with the ultimate purpose of improving diagnostics.

Paper I. MR images of 49 IVDs in 10 LBP patients were analyzed with unsupervised clustering methods to objectively and continuously classify the IVD heterogeneity related to degenerative changes. IVD degeneration could successfully be quantified with the proposed method.
Paper II. The lumbar IVDs of 25 LBP patients and 12 matched controls were examined with T2-mapping to quantify possible differences in IVD signal behaviors between patients and controls. The cohorts differed significantly in nucleus pulposus signal. A sub-analysis revealed that this signal difference was related to IVD fissures, visible as high-intensity zones at the outer part of the annulus fibrosus.
Paper III. Radiomics features were extracted from 123 IVDs (n=43 LBP patients) and examined with conventional MRI followed by discography and computed tomography. The features were further analyzed using artificial neural networks and a radiomics-based attention mapping technique to identify the presence and position of possible annular fissures. The method showed

great potential and was found to classify the presence of fissures with 100% sensitivity and 97% specificity. The method also identified the position of fissures in 87% of the analyzed IVDs.
Paper IV. Radiomic features from 61 LBP patients examined with conventional MR imaging were extracted and then analyzed using machine-learning techniques to explore possible associations between annular fissures and vertebral lesions. The findings suggest that radiomics can objectively detect vertebral tissue changes associated with adjacent annular fissures.

Conclusion. With data-driven methods, such as radiomics and attention mapping, tissue changes both within the IVD and the vertebra were well revealed in LBP patients. Further, the methods could be used to find associations between different types of tissue changes and were sensitive to subtle and imperceptible changes associated with disc degeneration and annular fissuring. These analysis methods could contribute to improved MRI diagnostics for LBP patients.

Keywords: Annular fissure, bone marrow lesion, image analysis, intervertebral disc, intervertebral disc degeneration, low back pain, machine learning, radiomics, texture analysis.

ISBN 978-91-8069-193-2 (PRINT)
ISBN 978-91-8069-194-9 (PDF)

SAMMANFATTNING PÅ SVENSKA

Ländryggssmärta är den vanligaste orsaken till funktionshinder. Cirka tre av fyra individer kommer någon gång i livet uppleva ryggsmärta. Utöver det stora personliga lidandet, medför en kronisk ländryggssmärta dessutom kostnader för samhället. Orsaken till ospecifik ländryggssmärta är inte helt känd då den sannolikt beror av flera saker. Bland annat har forskning visat att patologiska benmärgsförändringar i ryggens kotor är kopplade till smärta. Vidare anses vävnadsdegeneration av ryggens intervertebrala diskar, såsom fissurer (sprickor), vara väl kopplat till smärta. Denna avhandling syftar till att, med hjälp av magnetresonans (MR) -bilder och data-drivna analysmetoder, förbättra förståelsen av förändringar i ryggradsvävnad och deras samband med ländryggssmärta med mål att förbättra diagnostiken för denna patientgrupp.

I **studie I** undersöktes vävnadsheterogeniteten inuti diskar med hjälp av en dataklustringsmetod. Med metoden kunde vi visa att diskar, avbildade med MR, uppvisar ett distinkt histogrammönster som ändrades på ett förutsägbart sätt då disken degenererar. Detta beteende kunde modelleras för att systematiskt klassificera diskdegeneration på ett objektivt sätt.
I **studie II** undersöktes om heterogeniteten inuti ländryggsdiskar skiljer mellan patienter med ländryggssmärta och friska individer. Modellen, som utvecklades i studie I, tillsammans med statistiska analysmetoder, användes för att påvisa eventuella skillnader. Det påvisades en skillnad i de centrala delarna i diskarna som var främst kopplade till förekomst av fissurer i diskarna.
I **studie III** utvecklades en metod för att detektera diskfissurer i de diagnostiska MR-bilder som används dagligen i klinik. MR-bilder analyserades med neurala nätverk i kombination med metoder som extraherar bildmarkörer, även kallat för radiomics. Denna metod uppnådde en god förmåga att urskilja även fissurer som normalt inte är synliga i MR-bilder.
I **studie IV** undersöktes om det finnas en association mellan diskfissurer och vävnadsförändring i intilliggande kotor. Diskar och kotor, som avbildats med MR, analyserades med maskininlärningsmodeller kombinerat med bildmarkörer som kan spegla en vävnadens egenskaper. Analysen visade att kotor uppvisar vävnadsförändringar som är associerade med intilliggande fissurer och att bildmarkörerna i detalj kan påvisa dessa.

Sammanfattningsvis erbjuder de nyutvecklade metoderna en hög känslighet för subtila och, för ögat osynliga, vävnadsförändringar inom både diskar och kotor hos patienter med ländryggssmärta. Vidare kan metoderna användas för att hitta samband mellan olika typer av vävnadsförändringar associerade till ländryggssmärta. Tillsammans bör de nya metoderna kunna bidra till en förbättrad diagnostik för patientgruppen med ospecifik ländryggssmärta.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Waldenberg C**, Hebelka H, Brisby H, Lagerstrand KM.
MRI histogram analysis enables objective and continuous classification of intervertebral disc degeneration.
European Spine Journal.
doi.org/10.1007/s00586-017-5264-7 (2017)

II. **Waldenberg C**, Hebelka H, Brisby H, Lagerstrand KM.
Differences in IVD characteristics between low back pain patients and controls associated with HIZ as revealed with quantitative MRI.
PLoS ONE 14(8): e0220952.
doi.org/10.1371/journal.pone.0220952 (2019)

III. **Waldenberg C**, Eriksson S, Brisby H, Hebelka H, Lagerstrand KM.
Detection of Imperceptible Intervertebral Disc Fissures in Conventional MRI – An AI Strategy for Improved Diagnostics.
Journal of Clinical Medicine, 12(1), 11.
doi.org/10.3390/jcm12010011 (2023)

IV. **Waldenberg C**, Brisby H, Hebelka H, Lagerstrand KM.
Associations between annular fissures and adjacent vertebral body changes – confirmed by CT discography and MRI-based radiomics.
Manuscript

CONTENT

ABBREVIATIONS v
1 INTRODUCTION 7
2 SPINE ANATOMY AND FUNCTION 9
2.1 Pathophysiology and its Association with low back pain 10
3 SPINE IMAGING TECHNIQUES 13
3.1 MRI physics related to image intensity and contrast 14
3.2 Conventional MRI in the spine 15
3.3 T2 mapping 16
4 ANALYSIS TECHNIQUES 19
4.1 Radiological analysis 19
4.1.1 Pfirrmann classification 19
4.1.2 Modic changes 21
4.1.3 High-intensity zones 22
4.1.4 Dallas discogram description 23
4.2 Machine learning 24
4.2.1 Texture analysis 27
4.2.2 Neural networks and deep learning 30
5 AIMS 33
6 SUMMERY OF STUDIES 35
6.1 Overview of study cohorts 35
6.2 Paper I 36
6.3 Paper II 37
6.4 Paper III 38
6.5 Paper IV 39
7 CONCLUSION 43
8 DISCUSSION 45
8.1 Can MR post-processing techniques improve spine diagnostics? 45
8.2 How much data is enough? 48

8.3 What model is the most appropriate? 50
9 FUTURE PERSPECTIVES 53
ACKNOWLEDGEMENT 55
REFERENCES 56

ABBREVIATIONS

AF	Annulus Fibrosus
AI	Artificial Intelligence
ANN	Artificial Neural Network
alMRI	axial loaded MRI
CNN	Convolutional Neural Networks
CSF	Cerebrospinal Fluid
CT	Computed Tomography
DDD	Dallas Discogram Description
DL	Deep Learning
HIZ	High-intensity zone
LBP	Low Back Pain
IVD	Intervertebral Disc
MC	Modic Change
ML	Machine Learning
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NP	Nucleus Pulposus
T1W	T1-weighted
T2	Transverse relaxation time
T2W	T2-weighted
XAI	eXplainable Artificial Intelligence

1 INTRODUCTION

Low back pain (LBP) is considered the most frequent source of chronic disability for both genders in the working years [6, 7], with about three of four individuals experiencing back pain at some point in their lives [8]. Although the etiology of the disease is unclear, it is estimated that approximately 10% of the patients will develop chronic LBP [9] and, according to a recent "call for action" viewpoint published by *The Lancet*, the burden of LBP is estimated to continue to grow with the aging population [10].

Several spinal tissue changes are associated with LBP, including changes in the intervertebral disc's (IVD) biochemical properties, structural properties, and metabolic function, as well as inflammation and infection in the endplates/vertebra e [11-15]. Disc degeneration is suggested to be one of the main causes of LBP. It involves changes in the IVD matrix, including fragmentation of the nucleus pulposus followed by dehydration and decreased viable cells. This may weaken the annulus fibrosus, making it susceptible to annular fissures and ingrowth of nerve endings that can reach deep into the IVD, which altogether may contribute to LBP, sometimes referred to as discogenic pain [16, 17]. In addition, the densification of pain-inducing vertebral nociceptors has been associated with IVD degeneration and fissuring [18]. Tissue changes to the vertebra or cartilaginous and bony vertebral endplates have been reported to be risk factors for LBP [19, 20]. However, all the tissue changes mentioned above have also been shown to be common in asymptomatic individuals [21], making it difficult to find the cause of LBP, which limits the possibilities for adequate help and treatment.

Chronic LBP is usually defined as pain that persists or recurs for longer than three [22]. The evaluation of such patients includes the patient's pain description, clinical examination, and imaging before treatment advice is given. To date, the best imaging tool available to determine and visualize possible pathology within the spinal column for LBP patients is magnetic resonance imaging (MRI). Conventional T1-weighted (T1W) and T2-weighted (T2W) sequences are commonly used to image the spine. When studied subjectively, the images only enable a rough overview of the morphology. Such qualitative interpretation is not specific enough to derive the source of the pain. Perhaps this is because the degenerative changes visualized with conventional MRI are also common among asymptomatic individuals and naturally increase with age [23]. Further, with no indications of severe underlying conditions, conventional spine MRI has not been shown to assist in treatment decisions that improve clinical outcomes [24]. You et al. found that

90% of all MRI scans of back pain patients did include detectable tissue changes, but the clinical significance of such changes was unclear [25]. This indistinct correlation between spinal tissue changes on MRI and clinical symptoms may result from the insensitive and unspecific categorization schemes used to evaluate the MR images. As such, changes in tissue structure may remain undetected even though they are present in the images. Therefore, conventional MRI and the currently available analysis tools are usually insufficient to link changes directly to LBP and pinpoint the source of the pain. The lack of precision in the diagnosis makes it difficult to select those patients who will benefit from targeted treatment. Hence, we need to increase the knowledge of LBP by developing reliable diagnostic tools for an improved understanding of the etiology and treatment of the disease.

Also, by utilizing unconventional MRI techniques, a deeper understanding of the relation between spinal tissue changes and LBP might be reached. For example, mapping techniques, e.g., T2-mapping, can offer quantitative representations of tissue structure and function. Further, machine learning (ML) techniques can convert conventional MR images into quantitative features [2], automate cognitive tasks, and aid in medical decision-making to improve consistency. Within other diagnostic areas, authors have shown that phenotypic tissue patterns in diagnostic imaging can reflect tissue characteristics on a cellular and genetic level [26]. On this note, utilizing ML, we have successfully exploited inter-pixel relationships in MR images to detect tissue changes that cannot be identified by subjective visual interpretation alone and related those findings to symptoms and clinical outcomes for patients with LBP [1, 2, 4]. However, within spine research, ML's potential is not fully utilized as it is today mainly used to automate cognitive tasks such as segmenting and classifying tissue and bone structures [27]. Also, texture analysis of MR images has proven promising. It has the potential to objectively describe the tissue shape and structure via quantitative features, i.e., "radiomics," calculated from qualitative conventional MR images. Combined with classification algorithms, the features may have the ability to characterize tissue and relate it to possible pathology, as shown here in paper IV [4, 28].

2 SPINE ANATOMY AND FUNCTION

The human spine column consists of 24 individual vertebrae and two sections of naturally fused vertebrae, the sacrum (S1-S5) and the coccyx. The most cranial part of the column is composed of the seven cervical vertebrae (C1-C7), followed by 12 thoracic (T1-T12), five lumbar (L1-L5) vertebrae, the sacrum, and finally, the coccyx. The posterior portion of the vertebra forms the vertebral arch, while the anterior part consists of the vertebral body, which is responsible for the support of the spine. The vertebral body is divided into mainly two parts, the outer rim, consisting of cortical bone, and the inner region, composed of trabecular bone, which serves as structural support and contains deposits of phosphate and calcium [29]. Besides porous trabecular bone, the inner part of a young and healthy vertebra is mainly composed of red, haematopoietically active bone marrow. However, the tissue composition is known to be dynamic and varies with sex, age, and activity levels. Specifically, increased age is associated with decreased bone density and trabeculae bone [30]. Further, red bone marrow is known to be replaced by fatty components, where an age-dependent linear increase in fat content has been observed [31].

Between the vertebral bodies and IVDs lie the endplates. The endplates are of biomechanical importance as they transfer the mechanical load and distribute the intradiscal pressure onto the adjacent vertebral body, preventing the nucleus pulposus (NP) in the adjacent IVD from bulging into the porous trabecular bone of the vertebrae [32]. These typically sub-millimeter thin layers of hyaline cartilage consist of a hydrated proteoglycan gel reinforced by collagen fibers [33]. The cartilaginous endplates persist throughout normal skeletal maturation and support the adjacent IVDs by transporting fluid and essential solutes into and out of the IVD [34]. Mainly the proteoglycan molecules within the collagen matrix of the endplate regulate the solute transport, and a loss of these molecules has been associated with the degeneration of the adjacent IVD by accelerating the loss of proteoglycans from NP of the IVD [34]. Thus, the endplates play a critical role in maintaining the health status of the adjacent IVD. The tissue structure of the cartilaginous endplates is similar to that of the articular cartilage of the synovial joints. However, unlike articular cartilage, the cartilaginous endplates lack a direct connection to the adjacent vertebral bone but are instead connected to the IVD through the lamellae of the medial (inner) annulus fibrosus (AF) [33]. During skeletal maturity, however, the cartilaginous endplate undergoes graduate mineralization, ultimately leading to calcification of the endplate cartilage and the replacement by bone tissue [35] that fuse with the vertebrae, forming a bony endplate. This transformation likely decreases the important diffusion

and nutrient exchange between the vertebrae and adjacent IVD [36].

The vertebrae are separated by the IVDs, which enable flexibility of the spine (i.e., flexion, extension, and rotation). The IVD is the largest avascular structure in the human body [33] and consists of mainly two tissues, the inner NP surrounded by the outer annulus AF [37]. The AF comprises 10 to 15 concentric layers of firm yet flexible collagenous lamellae. The lamella, which contains fibroblastic cells and parallel collagen type I fibers interconnected with perpendicular collagen fibers, gives rise to the tensile strength capable of withstanding high compressive forces that can reach over 3.000 N during active lifting [38, 39]. The NP comprises a gelatinous core consisting of mainly randomly oriented collagen type II fibers and an organized network of radially oriented elastin fibers [39-41]. The fibers are all fixed in aggrecan, a hydrated proteoglycan gel responsible for the attraction of water through osmotic pressure [39, 42]. The main function of the NP is to evenly distribute mechanical pressure across the adjacent endplate and vertebral body, while the AF encloses the NP, preventing it from herniating or leaking out from the IVD.

2.1 PATHOPHYSIOLOGY AND ITS ASSOCIATION WITH LOW BACK PAIN

The natural history of the degenerative process of the spine is an ongoing research topic and involves a complex interaction between environmental and genetic factors. The pathophysiological background of LBP is probably multifactorial, where a few spinal tissue changes have been shown to be closely related to LBP. These include IVD degeneration [11, 12], tissue changes in the vertebra [15], and damaged cartilaginous and bony endplates [19, 20].

Besides the age-dependent transformation of the vertebral tissue, several typical tissue changes can be found linked to LBP and degeneration. Foremost, bone marrow lesions (Modic changes [43]) such as edema, occurring as a consequence of inflammation, non-age-induced fatty replacements, and sclerosis, are associated with LBP. Degeneration of the IVD can occur due to trauma but does also naturally occur with increasing age. This involves matrix changes in the NP, including tissue fragmentation followed by dehydration and a decrease in viable cells, ultimately weakening the AF and making it susceptible to annular fissures. If present, the fissures can further accelerate the degenerative process [44], induce tissue stress concentrations [45, 46] and serve as a point of entrance for vascularized tissue accompanied by nociceptive nerve endings [11, 12, 16, 17]. In addition, it has been suggested that regions

of damaged endplates that are associated with IVD degeneration are subjected to the densification of nociceptors capable of indirectly transmitting pain [32].

Degenerated IVDs may also possibly negatively influence adjacent endplates and vertebrae. Although the etiology is unclear, bone marrow lesions have been suggested to result from inflammatory mediators diffusing from adjacent IVDs at sites with endplate damage [47]. Such pro-inflammatory factors are more abundant within the NP of degenerated IVDs [48], which may further drive the development of vertebral bone marrow lesions. On the other hand, bone marrow lesions and endplate damage have been shown to be associated with decreased or locally impaired nutrient supply of the IVD. This, in turn, is thought to contribute to IVD degeneration [34, 49], leading to altered extracellular matrix composition and compromised IVD tissue integrity [50]. Consequently, the tissue may be susceptible to structural damage and degeneration, further propagating the degenerative progress in the IVD and the adjacent vertebrae [51]. In addition, vertebral bodies and endplates are highly innervated by nociceptors [52, 53], which are densified in regions with damaged endplates associated with increased IVD degeneration and fissuring [18].

3 SPINE IMAGING TECHNIQUES

Radiological imaging aims to diagnose, monitor, and treat disease, preferably using minimal ionizing radiation and non-invasive techniques [54]. There are several possible imaging techniques available to evaluate the status of the spine. Computed tomography (CT) enables a rapid assessment of traumatized patients [55] and offers superior sensitivity and specificity in identifying skeletal injuries, e.g., fractures, compared to plain radiographs [56]. Although the bony structure of the vertebral bodies is well depicted with MRI, subtle fractures are better visualized in CT due to the dense, and thus highly attenuating, outer cortical bone [57]. However, CT does not provide high-contrast images in soft tissue and is, therefore, unsuitable for screening ligamentous injury and spinal cord lesions. Even for IVD herniations, it provides limited visibility [58]. In this regard, MRI is undoubtedly the superior imaging technique for imaging soft tissue. With its inherently superior image contrast, MRI is the best-suited modality to evaluate degeneration of the spinal soft tissues [59], such as IVD degeneration and vertebral composition [43]. In addition to being sensitive to the presence and properties of hydrogen-dense tissue, MRI has the potential to highlight certain tissue properties, e.g., fat involvement, and to evaluate the function of tissues by advanced MR techniques, e.g., diffusion/perfusion-weighted MRI. Based on such measurements, quantitative parameter maps displaying the tissue of interest can be reconstructed, which enables advanced analysis and objective interpretation [60].

Discography, first introduced by Lindblom and Hirsch in 1948 [61, 62], was originally used as a method for diagnosing herniated discs but soon became a method used to investigate LBP patients with assumed pain signaling from the IVD. During discography, a contrast agent is injected into the NP during fluoroscopy, followed by a CT to visualize disc morphology, shape, and possible tissue changes of the IVD, e.g., fissuring. The purpose of discography is not only to obtain a detailed characterization of the IVD but also to investigate if an IVD is painful, i.e., provocative discography. Possible pain signaling from the IVD is evaluated by pressurizing the IVD to elucidate if a familiar pain response is induced. Today, however, the procedure is rarely performed in clinical practice due to its questionable reliability/reproducibility, invasive nature, and concomitant side effects [63, 64]. Using advanced analysis techniques, we have developed a novel technique for classifying and visualizing annular fissures in conventional MR images. The technique is based on a noninvasive approach and offers increased availability without the concomitant side effects of discography [4] (paper III).

3.1 MRI PHYSICS RELATED TO IMAGE INTENSITY AND CONTRAST

In most medical applications, the images generated by MRI scanners are created by sampling the signal from hydrogen protons, which are abundant in the human body. Each proton has a quantum-mechanical property called spin, which can collectively be described with classical mechanics as the angular momentum of a sphere rotating around its axis. The proton is positively charged, and as such, its rotation induces a local magnetic field. A strong external magnetic field (MR scanner) causes the net sum of the normally randomly oriented spins to align toward the direction of the external field, creating a net magnetization that can be probed by the scanner.

The magnetic components must be separated to distinguish the faint magnetic field generated by the protons (range of μT) from the external field (T). This is achieved by exciting the protons, i.e., transmitting a radiofrequency (RF) pulse whose energy is absorbed by the protons, causing the angular momentum of the protons to be "flipped" into the transversal plane, perpendicular to the external magnetic field. The electromagnetic signal from the protons is subsequently detected as they induce electric currents in external receiver coils placed near the tissue of interest.

As soon as the RF pulse is delivered, the protons begin to relax, i.e., revert to their original state. This relaxation occurs in two distinct ways called T2-relaxation (spin-spin relaxation) and T1-relaxation (spin-lattice relaxation). As the protons move around in the tissue, they interact with each other by inflicting a change in the local magnetic field and proton precision speed. This causes the protons to lose unity in their phasing, decreasing the signal sum of all protons. This decaying signal process is referred to as T2-relaxation. As this motion is random, the signal decreases exponentially with time. The T2-relaxation is tissue-specific, and the rate of signal decay depends on how free the protons are within the tissue. A relatively free molecule, e.g., water, that rapidly interacts with other atoms or molecules, causing it to rotate and vibrate (i.e., tumble), will see a local magnetic field that fluctuates rapidly, which effectively averages out over the course of a few milliseconds. Such a molecule is less affected by loss in phasing unity compared to a molecule where the protons are tightly bound, e.g., in large molecules. As a result, molecules with relatively free protons will have a long T2 time. Inhomogeneities in the main magnetic field cause a similar loss in phasing unity. Contrary to true T2-relaxation, this process can be reverted.

T1-relaxations involve the process of protons dispersing the energy they gained from the RF-pulse to the surrounding tissue. This causes the net magnetization to gradually return and be parallel to the external magnetic field. The speed of this process differs between tissues and largely depends on how tightly the proton is bounded to its molecule. Specifically, to induce such energy dispersion, the protons need to tumble against neighboring protons, nuclei, or molecules at a rate that causes magnetic moment fluctuations at the Larmor frequency. Fluctuations of these frequencies result in a rapid T1-relaxation and are most common among protons with intermediate binding, e.g., fat tissue [65]. The Larmor frequency is specific to the external magnetic field strength, and, as a result, the T1-value of tissue is specific to the field strength of the MRI scanner.

3.2 CONVENTIONAL MRI IN THE SPINE

Conventional MRI (i.e., T1W, T2W, and fat/fluid saturated images) is indispensable in the diagnostic routine and is the preferred method for spine imaging to evaluate spine morphology and possible tissue changes [43, 59]. The scan parameters can be tuned to capture the different relaxation characteristics of the tissues. By carefully selecting the time between repeated RF-pulse excitations and the time between the excitation and signal sampling, the MRI can be sensitized to the T1 or T2-relaxation characteristics of the tissue. As such, the choice of scan protocol directly determines the contrast appearance of the tissues in the image [66].

As indicated in the previous section, T2W imaging is sensitive to fluids that will appear bright in the images. As such, due to the richly hydrated proteoglycans, a healthy NP in the IVDs will appear bright. During degeneration, however, the loss of proteoglycans and hydration is reflected by a decreased signal. As such, T2W imaging is commonly used to evaluate the degeneration of IVDs [67]. Similarly, the cerebrospinal fluid (CSF) and edema caused by local inflammations in the vertebrae and IVDs also appear bright on T2W images. Contrary to T2W imaging, where tissues with long T2-relaxation time appear bright on the images, tissues with short T1-relaxation time, such as fat, appear bright on a T1W image. The bright appearance of fat in T1W images makes T1W imaging useful for identifying fatty replacements in the vertebra. Commonly, T1W imaging is used to visualize detailed tissue morphology and is selected due to the generally high SNR, short scan time, and high contrast, especially for paramagnetic contrast agents, e.g., gadolinium.

Even though conventional MRI techniques are the core of clinical scan protocols, the techniques come with limitations. Conventional T1W and T2W imaging are designed to emphasize the difference in the nuclear magnetic relaxation between the tissues by selecting appropriate scan parameters to gain optimal tissue-to-tissue contrast. The acquired signal intensity is only qualitative and is dependent on the scan parameters and the signal amplification by the scanner hardware. As such, the absolute signal intensity in T1W and T2W images might not be comparable between different scans even if the same MRI scanner and protocol are used and is most probably not comparable between different MR scanners at different sites. However, post-processing can, in some instances, be conducted to compute semi-quantitative measures [68]. Many methods are proposed to deal with the signal variability between conventional MRI scans, and several approaches have proven useful, including intensity normalization using histogram matching [69] and normalization of the signal intensity using a reference tissue [70]. In papers I and II, special care was taken to use the same hardware with fixed signal amplification and scan parameters across all examinations. As such, the pixel values across all MRI scans could be compared within the same study.

3.3 T2 MAPPING

The MRI-based T2-mapping technique is not commonly utilized in clinical practice, as the scanning procedure is time-consuming. Also, the technique is challenging as the measurements can be influenced by several factors, including magnetic field inhomogeneities and stimulated echo formation from multi-echo trains, etc. [71-73]. Unlike conventional MRI techniques (e.g., T1W and T2W imaging), however, T2-mapping is an objective and quantitative measure that enables direct quantification of the tissue spin-spin relaxation properties, which can be used to compare results from different sites and studies and to follow-up patients longitudinally.

The spin-spin relaxation time (T2 value) has been shown to be a sensitive metric for studying the structural integrity of the collagen matrix, tissue anisotropy, and changes in cartilage water content [74]. Also, the T2 value in the IVD has been reported to be influenced by the pathology of the IVD, such as herniations and annular fissures [75]. The T2-mapping technique is also a proven marker for reflecting pathophysiological changes in the spine. For example, we have demonstrated regional differences in the T2 value over the IVD and between IVDs with different degrees of degeneration [2] (paper II), and further that axial loading of the spine during MRI (alMRI) momentarily

induces significant T2 changes in the disc and aggravates symptoms in LBP patients [76].

The T2 value is determined by measuring the signal intensity, S(t), at multiple time points, t, after the excitation (i.e., different echo times). A mono-exponential function, $S(t) = S0 \exp(-t/T2)$, is then fitted to sampled signal where S0 is the signal value at the excitation, i.e., t=0, and T2 is the tissue-specific relaxation time. This model fitting is completed for each pixel in the image to form a T2-map with calculated T2 values pixel-by-pixel [66] (Figure 1).

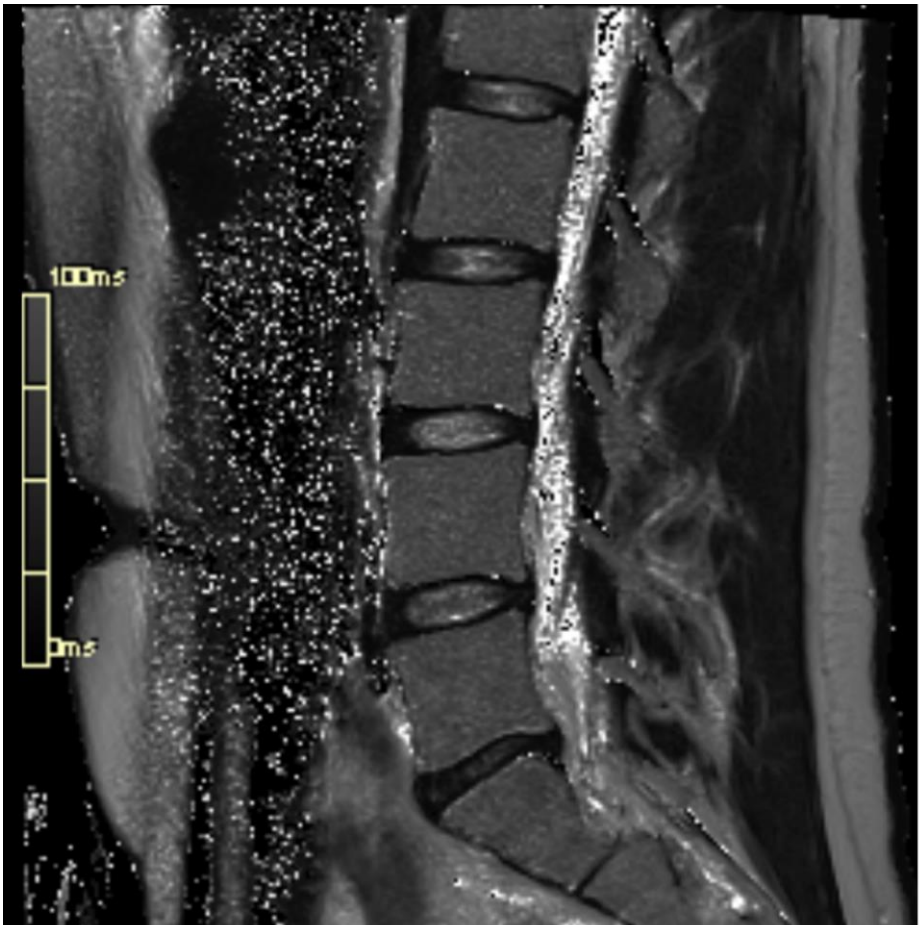

Figure 1. Example of a T2-map in a sagittal view of the lower spine displaying differences in the disc structure.

4 ANALYSIS TECHNIQUES

The continuously increasing availability of medical imaging technology and increased expectations by both patients and physicians drive the rising use of diagnostic imaging. The vast amount of data generated each day offers an opportunity for a more precise evaluation of characteristics of patient pathology, with the potential to personalize medicine and tailor patient treatment. Fundamentally, medical image analysis is divided into two separate steps; inspection of the image followed by interpretation [77]. In this chapter, both radiological and data-driven analysis methods will be discussed.

4.1 RADIOLOGICAL ANALYSIS

With over 4.2 billion radiological imaging exams performed worldwide each year [78], medical images are one of the most commonly used tools to aid in diagnostic and clinical decisions, risk stratification, and decision support related to surgery. For most organs and anatomical structures, classification schemes exist that are clinically used to guide the interpretation. These schemes allow systematic evaluation of the medical images and enable comparison of results between studies and sites. Related to the evaluation of the spine, a few characterization schemes used in this thesis are described below.

4.1.1 PFIRRMANN CLASSIFICATION

A common classification scheme to determine the grade of degeneration of the IVD in MR images is the Pfirrmann grading scheme [67]. Based on midsagittal conventional T2W images, IVD degeneration is classified according to a five-grade scale. This scale includes an evaluation of the IVD shape, the IVD height, IVD signal homogeneity, the relative signal intensity of the NP, and the distinction between the NP and AF (Figure 2). As described in the original article authored by Pfirrmann et al. [67], the IVDs are graded as follows:

Grade I – The IVD structure is homogeneous, with a bright hyperintense signal intensity and a normal disc height.

Grade II – The IVD structure is inhomogeneous, with a hyperintense signal. The distinction between the NP and AF is clear, and the disc height is normal, with or without horizontal gray bands.

Grade III – The IVD structure is inhomogeneous, with an intermediate gray signal intensity. The distinction between NP and AF is unclear, and the disc height is normal or slightly decreased.

Grade IV – The IVD structure is inhomogeneous, with a hypointense dark gray signal intensity. The distinction between NP and AF is lost, and the disc height is normal or moderately decreased.

Grade V – The IVD structure is inhomogeneous, with a hypointense black signal intensity. The distinction between NP and AF is lost, and the disc space is collapsed.

Even though this classification scheme is widely used and has been proven useful in the clinic, for example, to relate IVD degeneration to biomechanical IVD properties [79], to IVD diffusion properties [80], and to endplate perfusion [81], the classification scheme is rough with limited discriminatory ability [82]. To enable a quantitative and continuous marker of degeneration for evaluation of small differences in degeneration between IVDs and over time, we have proposed a method based on Gaussian mixture models and histogram statistics [2] (paper I). Using both conventional MR images and T2 maps, the method has demonstrated excellent linear correlation with IVD degeneration, both in our work and others [83].

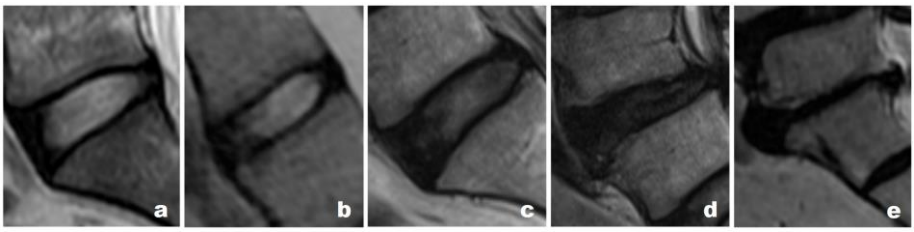

Figure 2. Sagittal view of L5-S1 intervertebral discs of different degeneration grades examined with T2-weighted MRI. According to the Pfirrmann classification, the intervertebral discs are graded as follows: (a) Grade I, (b) Grade II, (c) Grade III, (d) Grade IV, (e) Grade V.

4.1.2 MODIC CHANGES

Modic changes (MCs) are signs of pathological bone marrow lesions localized in the vertebral body and adjacent to the endplates. The signs are visible in conventional T1W and T2W images and have been shown to involve changes in the bone marrow and endplate which may include inflammatory and fibrotic changes and increased fat content (Figure 3). The classification of these bone marrow lesions was credited to Modic et al. [43], who recommended a categorical three-type scale based on typical signs in the MR images. The MC types are categorized as follows:

Type 1 – Lesions that appear dark on T1W images and bright on T2W images to classify the presence of bone marrow edema
Type 2 – Lesions that appear semi-bright on both T1W and T2W images to classify fatty replacements of the red bone marrow
Type 3 – Lesions that appear dark on both T1W and T2W images to classify bone sclerosis.

The exact underlying causes of MC are unclear, but as previously discussed, mechanical and biochemical stress to both the IVD and the vertebra seems to trigger inflammatory cascades leading to an induction of the pathology [84]. The correlation between MCs and LBP varies greatly between studies [13]. This may be attributed to the relatively rough classification scheme with limited possibility to detect small and diffuse lesions or lesion types with mixed tissue content. Though, it has been shown that the MC prevalence is frequently associated with the presence of back pain [85]. Also, the prevalence is higher among men and increases with age [84, 86].

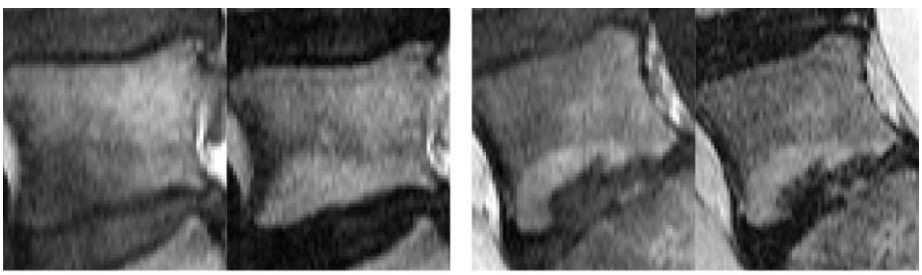

Figure 3. Example of type 1 and type 2 Modic Change (MC) lesions visible on T1-weighted (left) and T2-weighted (right) sagittal MR images of vertebrae.

4.1.3 HIGH-INTENSITY ZONES

High-intensity zone (HIZ), an MRI-based classification first introduced by April et al. [87], is defined as a hyperintense region in the outer AF of the lumbar IVD corresponding to an annular fissure. It is accepted that an HIZ lesion reflects a fissure in the AF lamellar structure (Figure 4). Such fissures can be inflamed, causing local edema that can appear as a bright region on a T2W image. According to the original article [87], the lesion should be as intense as the CSF and separated from the NP. Although it is well known that some high signaling fissures could still be invisible or only semi-bright on T2W images due to the limited spatial resolution in relation to the fissure size. Such so-called partial volume effects reduce the sensitivity of identifying such fissures [4, 88]. This is in line with paper III, where we showed that HIZ as a predictor for an outer annular fissure resulted in a low sensitivity of less than 63%. Even though the prevalence of HIZ is higher among symptomatic patients compared to asymptomatic [1, 89] and the occurrence of an HIZ has been reported to be associated with LBP [12, 17, 87], HIZ does not reliably indicate symptomatic disc disruption, making the diagnostic role of HIZ limited [89].

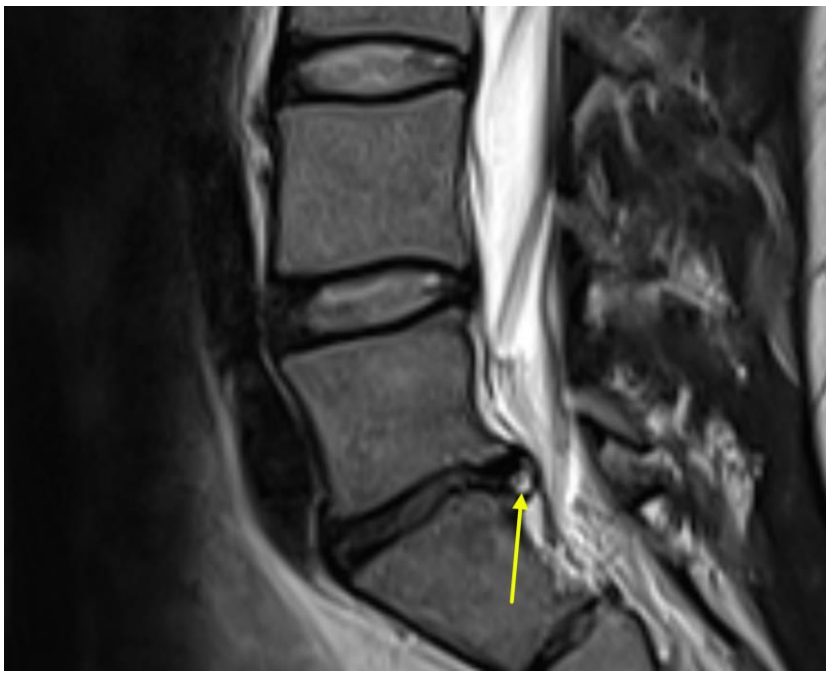

Figure 4. A T2-weighted image in the sagittal view of the lower spine displaying a bright high-intensity zone lesion at the dorsal intervertebral disc (arrow).

4.1.4 DALLAS DISCOGRAM DESCRIPTION

Dallas Discogram Description (DDD) is a radiological classification scheme first defined by Sachs et al. [90] that can be used to categorize the extent and severity of annular fissures with CT in an IVD after an invasive discography procedure. The classification system is graded 0-3 as follows (Figures 5 and 6):

Grade 0 – Intact NP with no contrast leaking into the AF
Grade 1 – Contrast leaks out of the NP into the inner AF
Grade 2 – Contrast leaks out of the NP into the outer AF
Grade 3 – Contrast leaks out of the NP into and beyond the outer AF.

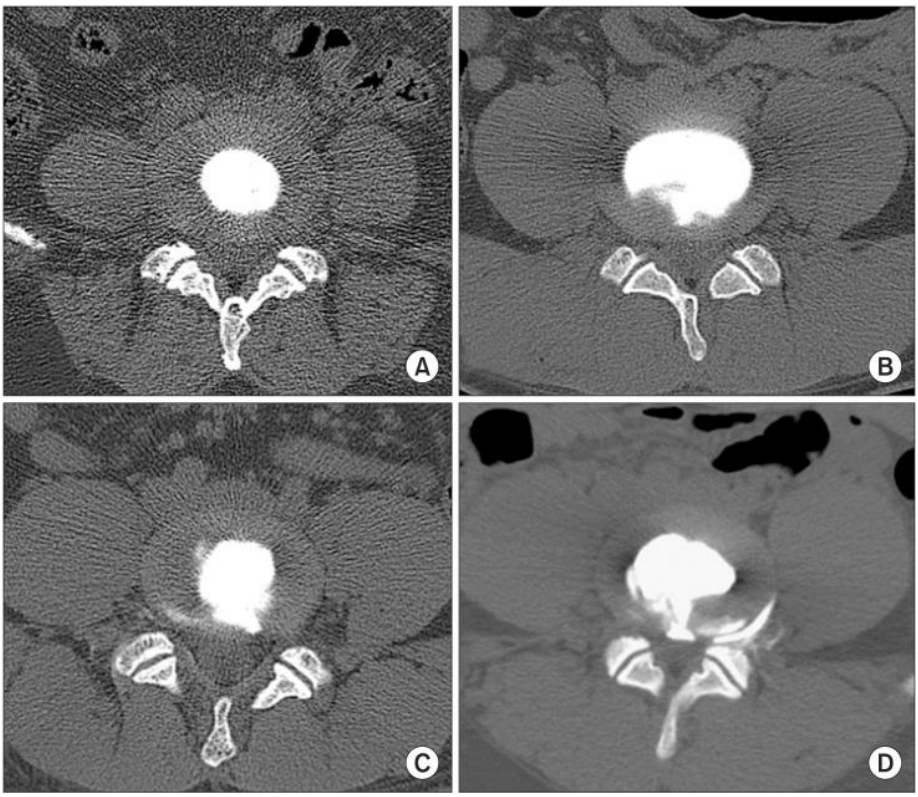

Figure 5. Axial view of IVDs seen on CT discogram. The contrast agent is visible in white within the intervertebral disc (A-D) and also outside the peripheral intervertebral disc (D), displaying the distribution of injected contrast agent and, thus, the extent of fissuring. The annular disruption is categorized according to Dallas discogram description as follows; (A) Grade 0, (B) Grade 1, (C) Grade 2, (D) Grade 3.

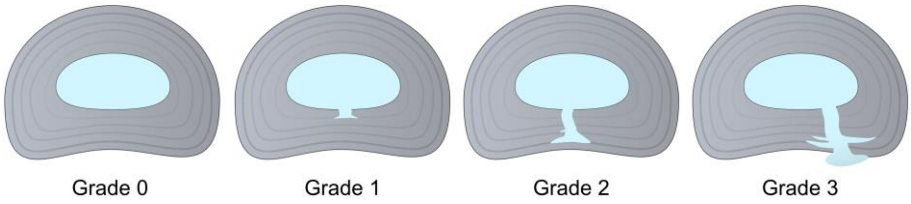

Figure 6. Schematic illustration of the Dallas Discogram Description classification system visualizing the extent of fissuring in the annulus fibrosus for each of the four classification grades 0-3.

4.2 MACHINE LEARNING

Within diagnostic principles, the application of ML to patient data has shown encouraging results [91-93] and is playing an increasingly important role in medical research and clinical practice [94]. For instance, ML has been applied to aid in screening, diagnosis, prognosis, prediction of response to therapy, and even to classify molecular disease subtypes [95]. The number of published articles applying ML related to LBP has rapidly increased in the past years, with a notable upsurge starting in 2020 (Figure 7). Since then, ML related to LBP has been utilized to, for example, recognize degenerated IVDs [96, 97], identify annular fissures [4, 98] and vertebral fractures [99], analyze scoliotic curves [100], and more [101, 102].

Traditionally, engineers have created software-based models that depend on predefined logical rules to generate an outcome, such as classifying pathology or suggesting treatment. However, as described already in 1987 by Schwartz et al., the medical field is "so broad and complex that it is difficult, if not impossible, to capture the relevant information in rules" [103]; why the need for another approach to process and utilize the data was recognized.

Seen as a part of artificial intelligence, ML differs from the traditional models as it leverages data algorithms to "learn," i.e., gradually improve the performance for the given task by studying a set of observations rather than being programmed with rules. This approach to processing data has the fundamental key advantage of being able to process vast amounts of data, potentially millions of input features and examples, to learn patterns related to a particular output, such as a disease or classification decision.

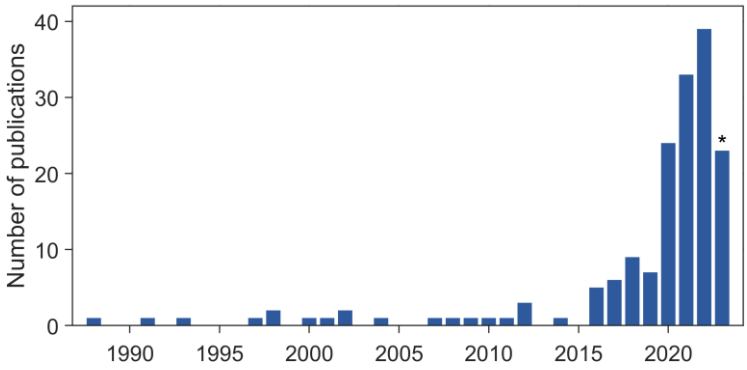

*Figure 7. The number of machine learning papers related to low back pain indexed yearly in PubMed since its first appearance in 1988. * Note that the number of papers indexed in 2023 includes papers listed up to and including April 5.*

The data was retrieved by the following query search on PubMed: ("artificial intelligence"[Title/Abstract] OR "feature extraction"[Title/Abstract] OR "computer vision"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "radiomics"[Title/Abstract] OR "texture analysis"[Title/Abstract] OR "neural network"[Title/Abstract]) AND "low back pain"[Title/Abstract].

Although the ML technique mimics human behavior by learning from examples, one caveat of the technique is its general requirement of access to large amounts of data to learn a new concept, as opposed to human learning, where only one or a few examples might suffice [104]. For this reason, care must be taken when choosing the model for the given task. A complex model with many degrees of freedom trained on too few data samples might fit well to the training data but run the risk of being biased and not generalizing well to new, unseen data. That is, the model might have a low ability to capture the characteristics of the population. Although no exact sample size requirements exist for ML models, models with a higher degree of freedom will generally require more samples during training. Despite rarely being performed [105], there are methodologies available to estimate the sample size requirements of a given ML model. Most often, a “post hoc” curve-fitting approach is applied as a part of an exploratory analysis. The method relies on empirical testing where the model performance is extrapolated as a function of sample size. This method has been shown to be accurate in determining ample sample size to achieve the expected model accuracy [105, 106]. However, this approach will not describe how well the model generalizes to other datasets, as this is partly determined by the diversity in the dataset used during the training and validation of the model. It must be noted that the sample size required is not

only limited to the model but also to the specific problem at hand. An ML model applied to a specific targeted application or population may require a relatively small sample size, perhaps in the range of hundreds, while a model applied to heterogeneous populations, or when differences between imaging phenotypes are subtle, requires much larger datasets [107]. Equally important is to consider the demographic and geographic biases, including age distribution in the study cohort, proportions of race and sex, differences in imaging machines (vendors types, acquisition protocols), differences in the geographic area, etc. [108]. Although small datasets often are adequate for proof-of-concept studies during research, generally larger datasets with unbiased images and labels, confirmation of its clinical utility followed by clinical trials and observational outcome studies are required to develop and verify models to be widely implemented into clinical practices [109]. Fortunately, research groups are gaining increased access to large open-source medical imaging datasets [108], which might aid researchers in gathering diverse data to develop such generalizable models.

The ML domain is divided into two subfields, conventional ML and deep learning (DL). Both subfields share the general concepts of supervised and unsupervised learning and can be used to both predict and infer outcomes. Conventional ML leverage handcrafted features tailored to reflect the problem at hand, while DL models, which typically consist of several layers of artificial neural networks (ANNs), offer end-to-end learning from images with no need to craft features manually. Instead, DL algorithms can learn features in hierarchical data representations directly from images, thus minimizing the need for prior knowledge in the field. Given a large enough dataset, DL models commonly outperform conventional ML models; this is especially true for object recognition problems [110]. Further, DL models offer new advanced analysis, such as generating synthetic CT images [111] and quantitative MRI maps from conventional MRI [112], so-called image-to-image translation [113]. However, DL models are generally more complex and require larger datasets. Perhaps more importantly, conventional ML is often much more interpretable and thus trusted, which is particularly important in the medical field [114-116].

In classification problems, supervised learning is the method of choice [94], which is defined by its use of labeled datasets, i.e., data used to train the model is coupled to their corresponding class or label of interest. This labeled data is used to train (or "supervise") the model to map the input data to the desired output. In medicine, the input data can consist of homogeneous or diverse data such as medical images and electronic health records, while the output can be, for example, a classification decision. Linear classifiers, support vector

machines, decision trees, and random forests are all common classification algorithms within conventional ML. Supervised learning is also common in regression problems which are useful for understanding the relationship between input and output data by estimating the relationships between dependent variables' "labels" and independent variables' "features." Linear regression and logistic regression are popular regression models used to calculate the linear combination of input data that best fits the labels. In papers III and IV, we have implemented classification algorithms and logistic regression models trained with supervised learning and combined those with complementing analysis techniques in order to extract subtle information in MR images related to spinal tissue changes.

As opposed to supervised learning, in unsupervised learning, the ML models are not trained on data related to predefined output labels; instead, the model aims to detect patterns, similarities, or differences in data samples under given characteristics without human labeling, also referred to as data mining [117]. K-means clustering and hierarchical clustering are two common algorithms used for groping (clustering) data by its characteristics rather than labels. Unsupervised learning models are also used for dimensionality reduction. This involves reducing the number of features or data points in a given dataset to a manageable size without losing data integrity (more on this under headline 5.3.1). This is often used during feature preprocessing, as in paper IV.

There are several excellent review articles, original articles, and conference papers that, in detail, explain the most common ML models and how to optimize them. Please refer to the following references for in-depth information [118, 119].

4.2.1 TEXTURE ANALYSIS

Texture analysis is an intensively discussed topic in the field of pattern recognition. The technique aims to extract quantitative and ideally reproducible representations from images. These include both visually apparent structures and complex patterns that are not easily recognized with human interpretation [4, 120]. Within medicine, these features are often called "radiomics" and are particularly well-established within the field of oncology [26]. Several studies have shown that radiomic features correlate with tissue changes at the cellular level [26, 121] and have, as such, been shown to be associated with tumor aggressiveness and to predict response to treatment. The radiomics technique offers high-throughput quantitative image analysis that can be used to discover new markers or patterns that are associated with tissue change and disease [122].

There are hundreds of features available, all of which can be categorized into three main groups. A brief description of the feature categories is given here:

- *Shape-based features;* describe the size and shape of the structure of interest and do not analyze the grey level distribution inside the region of interest.
- *Histogram-based features (first-order statistical features);* describe the distribution of voxel intensities within the image
- *Texture-based features;* describe the relationships between image voxels and can further be sub-divided into mainly five sub-groups:
    - *gray-level co-occurrence matrix (GLCM);* captures spatial relationships of pairs of voxels
    - *gray-level run length matrix (GLRLM);* relies on the spatial distribution of runs of consecutive voxels with the same gray level
    - *gray-level size zone matrix (GLSZM);* quantifies areas (also called zones) of interconnected neighboring voxels with the same gray-level
    - *neighborhood gray-tone difference matrix (NGTDM);* relies on differences between the gray level of a voxel relative to the mean gray level of its neighboring voxels
    - *neighborhood gray-level dependence matrix (NGLDM);* similar to NGTDM, NGLDM also relies on differences between the gray level of a voxel relative to neighboring voxels. However, here, a neighboring voxel within a predefined distance is considered. If these voxels are within a predefined range of grey level differences relative to the central voxel, the pixels or voxels are considered connected.

The number of features can be further extended by filtering the image before being subjected to further analysis and feature extraction. However, doing so might obstruct the interpretability of individual feature as the features will no longer be calculated directly from the medical image that visualizes the tissue of interest. Instead, the features will be calculated based on an image often heavily altered by filtration. However, if the interpretability is not of top priority, radiomic features calculated on filtered images can, in some cases, offer an improvement in an association. Particularly features calculated on wavelet-filtered images seem to be useful [123]. Also, wavelet-filtered MR images have shown excellent repeatability and reproducibility [124]. The most common filters to be applied before feature calculation are [125]:

- *Wavelet* filtering
- *Laplacian of Gaussian* filter (edge enhancement filter)
- Filter by taking the *square*, *square root, logarithm,* or *exponential* of the image grey level intensities
- Filter by calculating the magnitude of the local *gradient* in the image
- Filter by calculating the *Local Binary Pattern*

Although individual radiomic features might correlate with specific tissue changes (as in paper IV) or clinical outcomes, the large number of features extracted during analysis, typically in the range of hundreds, benefit from being processed with ML techniques to establish feature combinations that can create additive or synergetic effect to the specific association. This is clearly visible in paper IV, where the best predictive ability was achieved using 5-7 carefully selected features of high importance combined with ML techniques.

As is also evident in paper IV, the commonly vast number of features often needs to be reduced to select the most important features that reflect the problem at hand. Reducing the number of features has the benefit of:

1) *reducing the risk of the model overfitting,* resulting in low model generalizability to other datasets. Minimizing redundant data can reduce noise, and thus the risk of finding false patterns leading to model overfitting. Also, some features might be completely irrelevant to the target and have no correlation with the problem at hand. Such features add noise to the model and increase the risk of model overfitting.
2) *increase model performance*. The common event of only a few training examples paired with many more features can lead to the phenomenon known as the "Curse of dimensionality": In a high-dimensional space, the training data can become so sparse that no meaningful patterns can be learned by the model, resulting in poor performance. Similarly, the Hughes phenomenon [126] and peaking phenomenon [127] states that, for a fixed sample size, the predictive power of a classifier first increases and then decreases as the number of dimensions or features used in the model is increased. As such, reducing the number of features might be beneficial. This phenomenon is very visible in paper IV, Figure 1.
3) *increase the interpretability of the model* by reducing model complexity.
4) *reducing the computational cost* of training the model.

There are different approaches to reducing the number of features used as input to the model, and they can be grouped into mainly three families of methods, one unsupervised and two supervised. Typically, unsupervised methods are applied first to reduce the bulk of features.

Unsupervised feature reduction. As previously described, unsupervised methods do not rely on labels but instead, search for patterns in the calculated features themselves. Unsupervised feature selection can be applied to fill in missing data, remove features with high multicollinearity (redundant data), and remove features with low variance (low ability to explain target label).

Supervised feature reduction. Supervised methods are mainly divided into filters and wrapper methods. Wrapper methods incrementally add or remove features that are used to train an ML model. The performance is recorded for each subset of features, and the top-performing subset is then selected. This is a very common feature selection method, as the method tends to select a high-performing combination of features for the chosen model. Filter methods, on the other hand, have the advantage of not being model specific and being less computationally expensive. They analyze each feature's statistical relation with its corresponding target using correlation and mutual information measures. However, these methods only evaluate each feature in isolation, but as described before, a combination of features often has a greater predictive value. The supervised filter methods use Euclidean distance, Pearson correlation, and information measures to analyze relevance and redundancy between features and between features and targets [128].

Although not discussed here, also semi-supervised feature reduction methods exist and mainly belong to the family of filter methods. For an in-depth explanation of the concepts and a more elaborate description of the topic of feature selection, please refer to the following paper [128] and book chapter [129].

4.2.2 NEURAL NETWORKS AND DEEP LEARNING

The first influential work in the study of the mechanism of vision was published in 1959 by Hubel and Wiesel [130], which concluded that the simple brain cells of the visual cortex respond to oriented edges and that visual processing starts with the recognition of simple structures. As the information advances through the visual processing pathway, increasingly complex information is composed until the brain can finally recognize the outside visual world. Today, 60 years later, this model still inspires researchers and is the core theory behind neural networks.

A feedforward artificial neural network, sometimes referred to as a multilayer perceptron, is a class of neural networks that consists of fully interconnected neurons, or nodes, that are capable of nonlinear mapping between input and output data [131].

In contrast to conventional fully-connected feedforward networks and conventional machine learning techniques, convolutional neural networks (CNNs) retain the spatial information in the input images. This is achieved by processing input data in a hierarchical order where one neuron responds to a restricted region of the input data, known as the receptive field. Different neurons map to different receptive fields, which together cover the whole visual field and compose a feature map for a convolutional layer. By stacking these convolutional layers, increasingly complex structures can be recognized, exploiting the spatial information of the original image.

CNNs are commonly used for image segmentation and classification tasks and are currently outperforming other ML techniques in the realm of object detection and classification [110, 132]. Further, CNNs are the model of choice for segmentation purposes. For example, U-Net is a widely used CNN for segmentation tasks [133]. For example, a variant of the U-net has been used to segment IVDs accurately [134]. Another use of CNNs is the ability to create attention maps. An attention mapping technique, Grad-Cam [135], construct attention maps by computing the gradient of the given classification score with respect to the feature maps of late convolutional layers. The attention maps highlight pixels in the input image that are important to achieve the given classification score [135]. This powerful tool is commonly used to verify that the CNN is focusing on the correct regions of the input image. However, the technique can also be used to visualize correlations and pinpoint important structures in diagnostic images, as was done using a method based on radiomic features in paper III.

5 AIMS

This thesis aims to develop data-driven MRI-based analysis methods to improve the understanding of spinal tissue changes and their association with LBP with the ultimate purpose of improving diagnostics.

The specific aims of studies I to IV were:

I. To investigate the feasibility of histogram analysis to quantify IVD heterogeneity to obtain a tool for objective and continuous grading of IVD degeneration.

II. To explore possible regional differences in IVD tissue composition between patients with chronic LBP and controls by analyzing quantitative T2 maps and correlating possible differences with the radiological marker HIZ.

III. To develop a method to detect the presence and position of annular fissures in conventional MR images using a combination of analysis techniques, including texture analysis, ANNs, and attention mapping.

IV. To determine possible associations between annular fissures and adjacent vertebra in a clinical setting using radiomics.

6 SUMMERY OF STUDIES

6.1 OVERVIEW OF STUDY COHORTS

Paper I. The explorative study included 10 patients with chronic LBP (n=49 IVDs, age 25-69 years, mean age 41 years, 6 male); see Figure 8. The inclusion criteria was LBP > 6 months.
Paper II. The case-control study included 25 patients with chronic LBP (n = 124 IVDs, age 25-69 years, mean age 38 years, 11 male) and 12 matched controls (n = 59 IVDs, age 25-59 years, mean 38 years, 7 male). The inclusion criteria was LBP > 6 months.
Paper III. This study cohort included 43 patients with chronic LBP (n = 123 IVDs, mean age 45 years, 18 male), which was initially designed to study the impact of spinal loading and disc degeneration on provocative discography and were prospectively included between April 2007 and March 2010. The inclusion criteria were LBP > 6 months with failed conservative therapy. Individuals with low image quality (low signal-to-noise ratio, motion artifact) or failed examination(s) were excluded from the study.
Paper IV. The study included the same patients as in Paper III and additionally 16 patients. In total, 61 patients with chronic LBP were included (n = 177 IVDs, age 24–63 years, mean age 45 years, 29 male), prospectively included between April 2007 and March 2010. Inclusion criteria were LBP > 6 months with failed conservative therapy. Low-quality images (low signal-to-noise ratio, motion artifact) were excluded from the study.

Besides the overlap in the cohort between the studies in papers III and IV, there was no overlap between the papers (Figure 8).

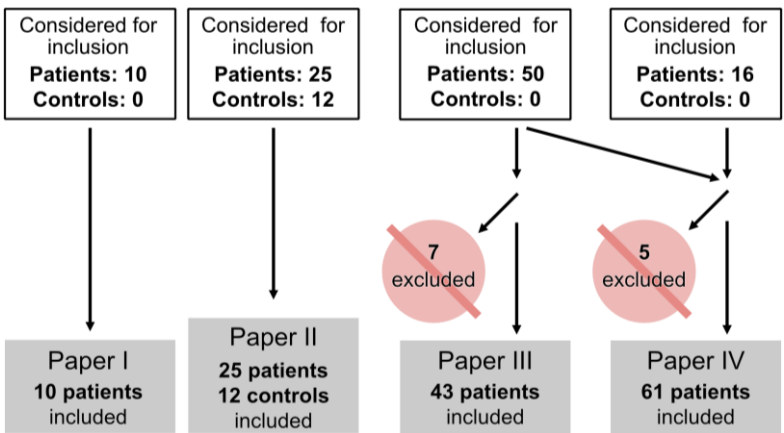

Figure 8. Flowchart displaying subject inclusion and exclusion for paper I to IV.

6.2 PAPER I

Motivation. Radiological IVD degeneration classification is based on visual and qualitative interpretation of conventional MR images. As such, the qualitative grading of IVD degeneration is hampered by subjective interpretation and limited to discrete classification grades.

Methods. The IVDs in the lumber spine of the patients were examined with conventional MRI and quantitative T2-mapping. Each IVD was semi-automatically segmented on three mid-sagittal slices, of which histogram features of the IVD heterogeneity were extracted using an unsupervised Gaussian mixture clustering model (Figure 9). The calculated histogram feature values were correlated to IVD degeneration in terms of Pfirrmann grades to assess its usability.

Results. Features calculated on both T2 maps and conventional MR images displayed similar results where two well-separated and distinct peaks were visible in histograms of well-hydrated IVDs. The topology of the histograms was altered with increased IVD degeneration. Specifically, the distance between peaks decreased, representing a reduced distinction between the AF and NP (Figure 10). The histogram features were shown to correlate with Pfirrmann grade ($p<0.02$).

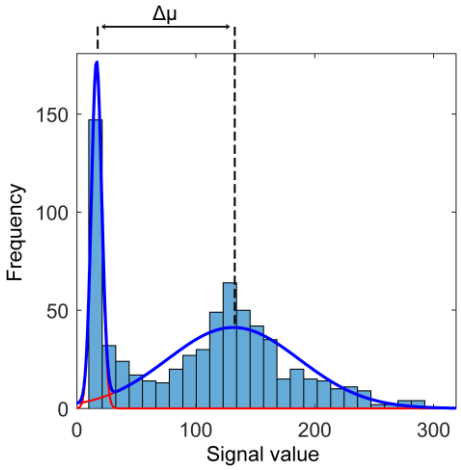

Figure 9. Example of Gaussian mixture models clustering, where the separation between high and low grey scale values, Δμ, was calculated.

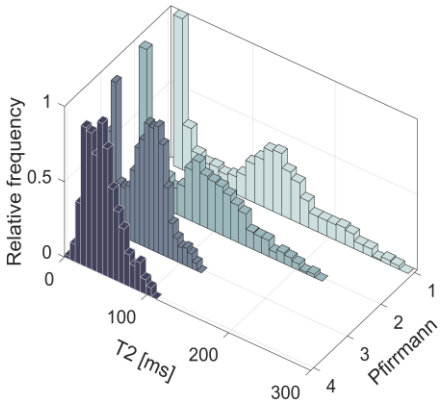

Figure 10. Examples of histograms displaying the distribution of voxel-values within IVDs of different degeneration measured with Pfirrmann grading. The distributions were derived from T2-weighted images.

6.3 PAPER II

Motivation. Quantitative MRI can provide objective measures of IVD tissue relaxation characteristics. However, there are limited studies comparing regional quantitative IVD measures of symptomatic and asymptomatic individuals.

Methods. The IVDs in the lumber spine of patients and age-matched controls were examined with T2-mapping. Each IVD was semi-automatically segmented into three mid-sagittal slices, of which the IVD heterogeneity was investigated by dividing the IVDs into five sub-regions in the posterior-anterior orientation. From each sub-region, the mean T2 value and standard deviation were calculated. Further, the distribution of T2 values in the IVDs was investigated using histogram analysis by Gaussian mixture models clustering. In addition, the analysis was conducted for IVDs of different degeneration grades, as described by Pfirrmann. To investigate possible IVD heterogeneity caused by IVD fissuring, an additional analysis, using the same analysis procedure was conducted after excluding IVDs affected by fissuring visible as HIZ.

Results. Significantly different T2 values between patients and controls were found in sub-regions representing the NP and the border zone between the NP and the posterior AF (p = 0.047–0.050) (Figure 11). After excluding all IVDs with HIZ, no significant differences between the two cohorts were found. Neither for the IVDs globally (p = 0.054–0.995) nor for individual sub-regions (p = 0.053–0.869).

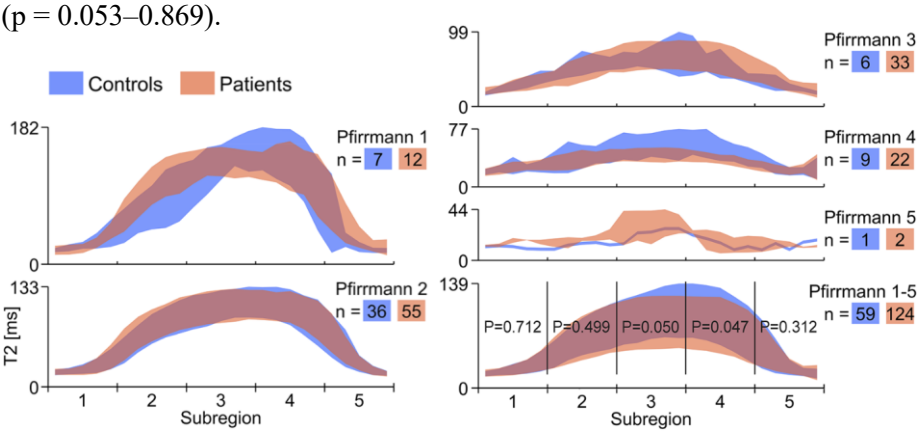

Figure 11. The distribution of T2-map values for intervertebral discs (IVDs) of different Pfirrmann grades presented as two overlapping patches ranging from anterior IVD (sub-region 1) to posterior IVD (sub-region 5). The height of each patch represents the mean T2 value ± one standard deviation.

6.4 PAPER III

Motivation. Annular fissures in IVDs are associated with LBP. Although clearly visible at CT discograms, no sensitive non-invasive method exists to detect annular fissures.

Methods. IVDs in the lumbar spine of LBP patients were examined with conventional MRI and discography followed by a CT. Based on the CT images, the occurrence and extension of the annular fissures were graded according to DDD by a senior radiologist. Each IVD was semi-automatically segmented on five mid-sagittal slices, after which radiomic features describing the IVD heterogeneity and intrinsic tissue properties were calculated using the T2W images. To determine the location of annular fissures, 22 features that were mainly sensitive to changes in the AF were selected and used to train a classification model based on shallow ANNs. During training, 10-fold stratified cross-validation was performed to reduce the risk of model overfitting. An attention mapping technique based on the radiomic features was used to predict the location of the fissures (Figure 12).

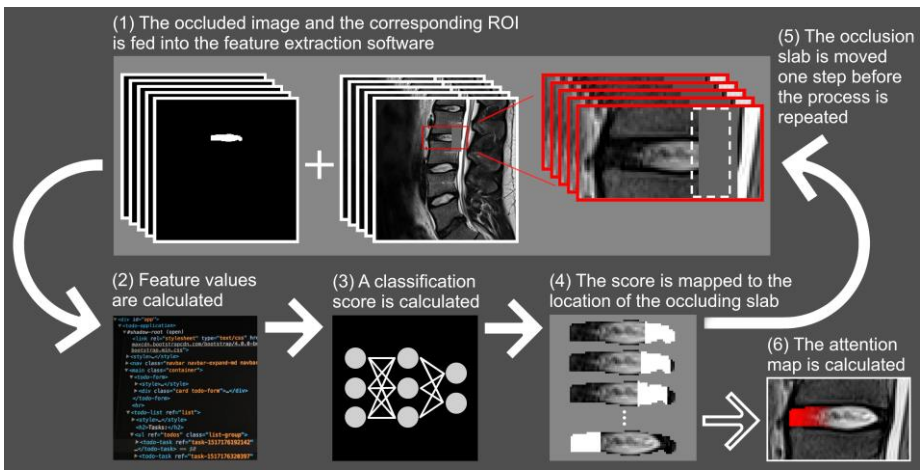

*Figure 12. Flowchart displaying the method used to calculate the attention maps. The region of interest (ROI) and the corresponding image of the intervertebral disc overlaid with an occluding slab are fed into feature extraction software (**1**). The feature values are calculated (**2**) and fed into the artificial neural network to calculate a classification score (**3**). This classification score is mapped back to the position of the occluding slab (**4**). Next, the position of the slab is moved one pixel in the anterior direction (dashed line in (**1**)) and the process is repeated (**5**). When the occluding slab has populated all positions of the intervertebral disc, the attention map is constructed (**6**).*

Results. Twenty (out of 123) IVDs were found to be unsuitable for the attention mapping technique as they exhibited severely disrupted AF (<50% continuously intact outer third AF) and lacked delimitable fissuring pathology. As such, they were excluded from further analysis. The attention maps accurately displayed the true position of the fissures in 90 (87%) of the analyzed discs (Figure 13).

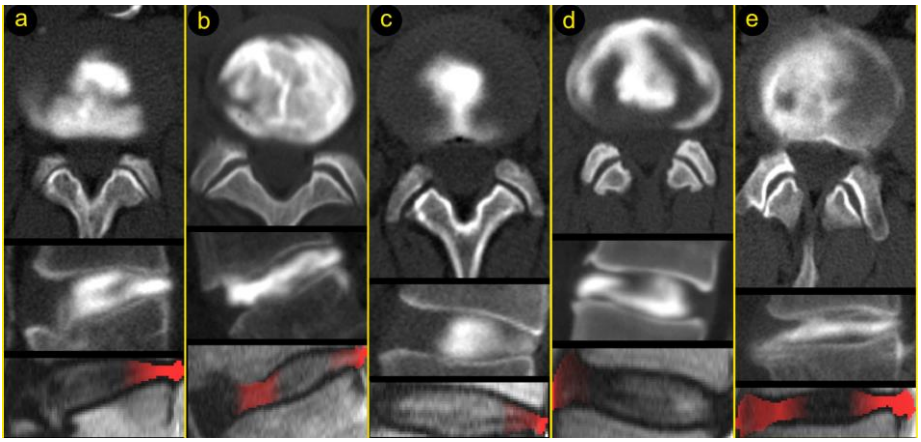

Figure 13. Representative examples of computed tomography discograms (top and middle) and the corresponding attention maps overlaid on a sagittal T2-weighted MR image (bottom). The contrast media, injected into the nucleus pulposus (visible in white), spreads into annular fissures, revealing their position and extent. The images (a-e) display instances where the proposed method correctly identifies the outer annular fissures in the sagittal T2-weighted images visible in red.

6.5 PAPER IV

Motivation. It has been shown that vertebral tissue changes are closely related to increased IVD degeneration and fissuring. However, the etiology is unclear, and radiological markers are not specific enough to enable precision diagnostics, which is why more competent methods to study associations between tissue changes are required.

Methods. 177 IVDs and adjacent vertebrae in the lumbar spine of the patients were examined with conventional MRI and discography followed by a CT. Based on the CT images, the occurrence and extension of the annular fissures were graded according to DDD by a senior radiologist. Since mainly IVD fissures reaching the outer AF are believed to induce LBP, IVDs with fissures reaching the outer AF were separated from the rest. Five to seven mid-sagittal slices of each vertebra were manually segmented, after which radiomic features describing the vertebral heterogeneity and intrinsic tissue properties were calculated on the T1W and T2W images. Before calculating the radiomics, the following preprocessing of the MR images was conducted to ensure reproducible features:

- MR images were interpolated to isotropic voxels of size 1x1x1 mm^3.
- MR image volumes were normalized to the mean volume signal intensity ± 3 standard deviations.
- MR image intensity discretization was performed using a fixed bin width of three.

To generate generalizable results, the number of features was reduced by considering (1) the variability of ROI segmentation, (2) the robustness of image acquisition and reconstruction, and (3) the redundancy of features related to their collinearity (4). To further select the most meaningful features best associated with IVD fissures, the remaining features were reduced further using ML algorithms combined with backward feature selection and logistic regression, finally selecting three features (Figure 14).

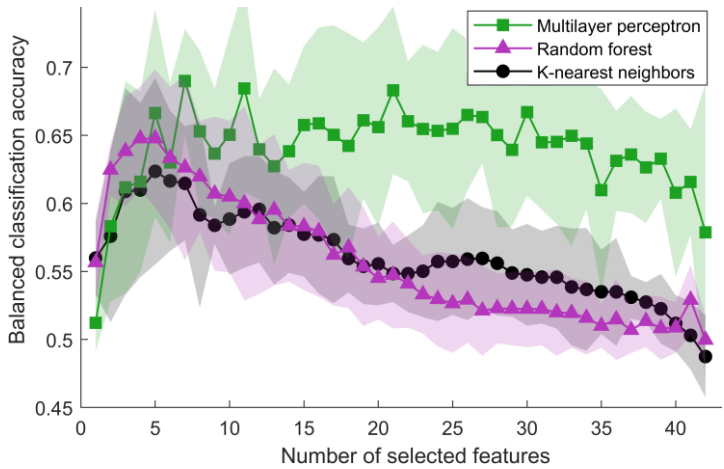

Figure 14. Model performance in identifying an annular fissure as a function of the number of included features calculated from the adjacent vertebra. The figure displays the mean accuracy ± one standard deviation (transparent patch) calculated from 5-fold cross-validation.

Results. By using three features that were associated with crosstalk between IVD with fissures and adjacent vertebrae, the logistic regression accurately classified 82% of the IVDs, reaching a sensitivity of 96.5% and a specificity of 28.7%. All three selected features were derived from only the T1W images suggesting that features extracted from the T2W image contributed with no additional information when features from the T1W image were included in the analysis. It was found that an inhomogeneous vertebra with few or small bright areas on a T1W image was likely to be adjacent to an IVD with a fissure (Figure 15).

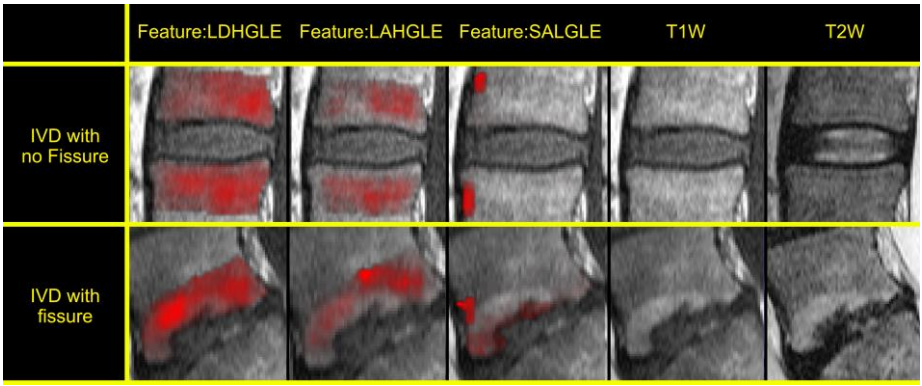

Figure 15. Examples of feature maps overlaid onto T1W images of the vertebra. In the examples, the logistic regression correctly classified the adjacent IVD. T1-weighted (T1W) and T2-weighted (T2W) images with no overlay are added for reference.

7 CONCLUSION

With data-driven methods, such as radiomics and attention mapping, tissue changes both within the IVD and the vertebra were well revealed in LBP patients. Further, the methods can be used to find associations between different types of tissue changes and were sensitive to subtle and imperceptible changes associated with disc degeneration and annular fissuring. These analysis methods could contribute to improved MRI diagnostics for LBP patients.

The specific conclusions for the paper I to IV were:

I. Histogram features correlated well with IVD degeneration, suggesting that histogram analysis is a suitable tool for objective classification, here applied to IVD degeneration.
II. Significant differences in T2 values, primarily in the NP region, were found between the patient and control cohorts, mainly related to IVD fissuring visible as HIZ, which may reflect altered IVD structure and function associated with HIZ.
III. The proposed method, utilizing radiomic features and ML models, was able to identify annular fissures in conventional MRI, allowing for new non-invasive research related to the presence and position of individual fissures.
IV. Radiomic features can objectively describe tissue heterogeneity and detect vertebral tissue changes associated with adjacent annular fissures.

8 DISCUSSION

This thesis aims to move the current research front forward within LBP diagnostics by developing and exploring non-conventional post-processing techniques based on artificial intelligence (AI), such as ML, attention mapping, and texture analysis. With these methods, not only quantitative relaxation mapping but also conventional relaxation-weighted MR images can be exploited to create objective features that provide the ability to predict pathology and potentially follow-up treatment longitudinally. This is of great advantage as it reduces the need for lengthy parametric scanning protocols, resulting in increased patient comfort and lower acquisition costs. Most importantly, it offers the possibility to extract new information which has been "hidden" in the big data that has been collected over decades. This may open a window of possibilities to further improved diagnostics and an increased understanding of causal relationships. Thus, this thesis represents an important stepping-stone toward better understanding the multifactorial pathology related to LBP and may potentially, in its extension, differentiate between different types of painful spine conditions from what is otherwise categorized under the same heading; non-specific LBP.

8.1 CAN MR POST-PROCESSING TECHNIQUES IMPROVE SPINE DIAGNOSTICS?

MRI is considered to be the most suitable imaging method to evaluate IVD degeneration [136], and the images are usually classified using the radiological marker Pfirrmann grading. However, radiological classification markers are subjective and limited to only a few discrete and nonlinear stages. The development of an objective and continuous classification method could generate unbiased classification results and aid physicians in following patients longitudinally. In Paper I, we exploited the heterogeneity in the IVD tissue using an unsupervised clustering algorithm to develop an objective and continuous classification method. We chose this specific method based on prior knowledge that the distinction between the IVD AF and NP diminishes during degeneration which is clearly reflected in the histograms of the IVDs. Further, the model evaluates the whole disc simultaneously, not individual voxels, which makes it insensitive to noise and local variance. The work was hypothesis-driven and based on the idea that both quantitative and weighted T2 contrast would reflect the same disc heterogeneity pattern. That is, we investigated if both quantitative MRI and conventional MRI could be

converted into quantitative handcrafted features that correlated with IVD degeneration. The features that were calculated from both the quantitative and conventional MR images had a similar discriminative ability, indicating that conventional MRI, in a proper setting, can have a similar diagnostic value to that of quantitative T2 maps. Recently, a paper similar to paper I used texture analysis to classify IVD degeneration [137]. The authors utilized fractal dimension, a method used to measure complexity in a pattern [138], to measure tissue heterogeneity. They concluded that fractal analysis was a suitable tool for objective and continuous classification of IVD degeneration. In paper III, we also investigated the value of texture analysis for IVD characterization. However, we did not use texture analysis to classify general IVD degeneration but to phenotype potentially painful fissures in the IVDs. Since paper I was a feasibility study that included only patients with chronic LBP, we hypothesized that the different histogram topologies in IVDs of the same Pfirrmann grade resulted from fissures in the IVDs. This hypothesis was based on the knowledge that fissures may contain clusters of nuclear cells originating from the NP, granulation tissue, and ingrowth of nerve endings and, as such, likely increase IVD heterogeneity and thus the topology of the IVDs histograms [17].

In paper II, this hypothesis was tested. We investigated the heterogeneity and the regional behavior of IVDs among patients with chronic LBP and matched controls using, among other things, the clustering method developed in paper I. Changes in the regional behavior were correlated to the occurrence of annular fissures visible as HIZ. The study concluded that annular fissures are not only correlated to a signal increase at the outer part of the annulus but also with a decrease in hydration of the NP which could result from altered IVD micro-structure due to IVD fissuring. From a pathophysiological view, the altered IVD micro-structure most likely reflects reduced proteoglycan concentration and elevated fibrosis and matrix cell density, as indicated by decreased T2 values inside the NP of IVDs with fissures [17]. The finding that the signal changes not only at the position of the fissure but also at the center of the disc when a fissure is present in the IVD is an important aspect to consider in the study design of paper III, where the IVD tissue characteristics were probed to identify annular fissures. In this paper (paper II), quantitative T2 maps were utilized. One main advantage of using quantitative T2 maps as opposed to conventional qualitative MRI is the ability to directly compare T2 values between different studies and sites, enabling good grounds for a longitudinal assessment. However, as discussed in paper II and under 4.3, parametric T2-mapping suffers from potentially low accuracy and reproducibility, even when evaluating parametric maps from different vendors and scanners at the same magnetic field strengths [54]. One often implemented method to ensure rigor evaluation is the use of phantoms [139], as was done to

evaluate the acquired T2 maps in Paper II. It was found that the T2 values representative for IVDs of Pfirrmann grade 3 had an error of less than 7%. This, being a systematic error, should not influence the conclusions of the current study. However, the error should be considered in direct relation to other studies.

In paper III, where we had access to a unique dataset from both high-resolution CT-discography and T2W MRI, we utilized texture analysis combined with ML models and complementing analysis techniques to visualize the location of individual fissures in IVDs through attention maps, also known as saliency maps. Such maps highlight parts of the input image that is important for the classifier to make its decision for the target class. However, other popular techniques are available that also produce similar saliency maps based on DL techniques, such as the Grad-CAM technique [135], which, according to Google Scholar, has 13635 citations as of April 10th, 2023. The Grad-CAM technique considers the whole input image to create the attention map. However, as Paper II indicated, annular fissures correlated with changes also at the center of the IVD (the NP). Therefore, another method was utilized that visualized fissures in the AF but, at the same time, did not visualize tissue changes in the NP that were a consequence of the fissures. For that purpose, we designed an approach based on texture analysis, where only a small subset of 22 features that were mainly sensitive to changes in the outer IVD tissue (the AF) was selected.

The discriminative ability of texture analysis to detect annular fissures was proven in Paper III, and the idea was further explored in Paper IV. Previous studies have reported cross-talk between the IVDs and vertebra where, for instance, pro-inflammatory gene expression in the IVDs correlated with fibrogenic changes in adjacent vertebrae [140]. However, radiological markers, such as HIZ and MC, have not yet reached conclusive associations. To overcome this, we explored the association directly in the MR images by applying texture analysis and ML models, as opposed to radiological markers, with the aim of analyzing vertebral heterogeneity patterns in greater detail. Radiomic features were calculated from conventional T1W and T2W images, where only three remained after extensive feature reduction. Unexpectedly, the remaining features were all calculated from the T1W images, suggesting that the T2W images had little or only similar explaining ability to what T1W could offer. One can speculate that this is because the distribution av important tissue patterns within the vertebra are the most important for the association and not only the ability to separate water and fat. The findings could also be explained by the fact that T1W images have a higher signal-to-contrast ratio compared to T2W images. It should be mentioned that when all radiomic features based on

T1W images were removed from the analysis, only a slightly reduced association was archived (data not shown). This implies that higher image quality is most important for diagnostic precision. It would be interesting to investigate the association between fissures and vertebrae using features calculated from also functional imaging, such as diffusion and perfusion. It is known that vertebral perfusion normally diminishes after skeletal maturity but that the perfusion can be re-generated in the event of IVD damage. Perfusion and diffusion analysis combined with the current features might enable a deeper insight into the association between pathology and tissue changes. Comparisons between LBP patients and matched controls might elucidate possible associations of vertebral and IVD tissue changes and tissue function to LBP, further adding to the collective knowledge of painful associations.

8.2 HOW MUCH DATA IS ENOUGH?

The encouraging results in object recognition and classification problems, where ML algorithms outperform the human observer, spread hope that similar progress will occur in the medical field. The rapid advancement in computer vision tasks may result from the very large and labeled datasets available, sometimes containing millions of images. In the medical field, however, much smaller datasets are often collected. Publicly available datasets considered "large" typically include subjects in the order of hundreds to thousands [108]. Although a one-to-one comparison of dataset sizes would not be fair, as there are generally fewer classes with less variation in medical datasets compared to datasets in computer vision problems [141]. The datasets used in the ML studies in this thesis included relatively few subjects, ranging from 10 to 61. This, of course, might impact the generalizability of the applied models and methods. By generalizability, I here refer to the model's ability to be successfully applied to a dataset reflecting the true general public. Although relatively few subjects were used in the current studies, an effort was made to choose models and techniques suitable for the amount of data available to generate meaningful results. For example, in paper I, a clustering method based on a mixture of Gaussians was applied. In this case, the model was not used to suggest probabilistic predictions. Rather, each analyzed IVD and subject stood separated from the rest. As such, the study sample size did not affect clustering interpretability.

In paper IV, where radiomic features extracted from conventional MR images were used to investigate the association between annular fissures and vertebral tissue changes, the radiomics features were greatly reduced before being

analyzed with ML techniques. By applying an ensemble of ML models, which also has been shown to increase generalizability [142], a few most important radiomic features were selected to describe possible associations through logistic regression. The end product was a rather simple logistic regression model with few parameters. As was previously described, models with few degrees of freedom are easier to interpret and are less prone to overfit and thus do a better job at generalizing to other datasets.

In paper III, where radiomics were extracted from 43 patients and 123 IVDs to predict the presence of annular fissure, cross-validation, a standard solution to validate a model [143], was implemented to reduce the risk of model overfit and to imply model generalizability. As the hyperparameters, i.e., parameters that control the model learning process, were not tuned during training, cross-validation is a recognized method to estimate model performance [144]. Of course, also other techniques could have been implemented to possibly increase model generalizability, such as data augmentation, i.e., a technique where several slightly modified copies of input data are used to train ML to reduce over-fitting. Although the true generalizability of the model to the general population remains unknown, more subjects by itself would not necessarily increase model generalizability. Instead, if additional subjects were to be included in the study, the subjects should be referred from other sites to form a more heterogeneous cohort. As described before, most importantly, model generalizability in the medical field comes from including subjects with diverse backgrounds. Factors such as sex and age distribution, geographic location, etc., can play an equally important role in obtaining a model that generalizes well to other datasets and the general population. For example, it has been shown that an individual's exercise activity can influence the hydration levels in IVDs [145]. Although easily overlooked, this can probably influence the model used in paper III as the model probes the IVD to make its prediction, and increased IVD hydration can change the IVD appearance on T2W images and quantitative values in T2 maps.

In a recent meta-analysis covering more than 500 publications leveraging ML techniques to classify Alzheimer's disease, the authors underline the need for a diverse dataset. The study found that larger datasets did not improve the accuracy of predicting Alzheimer's disease. But rather, the accuracy mostly decreased with an increased number of subjects [141]. The author hypothesizes that this might come from the fact that larger sample sizes might closer reflect real-life populations. Although relatively small datasets might suffice for proof-of-concept studies, the results from the meta-analysis, again, stress the importance of including a diverse cohort when planning studies and creating models that aim to be widely implemented into clinical practice [109].

One plausible method to investigate model generalizability is to evaluate the model or method using a public benchmark [146]. This public benchmark should also decide suitable performance metrics which could lead to unbiased conclusions. However, the scarce data available in the medical field often do not allow this approach. This is very true for the data used in papers III and IV, where we analyzed discography images, a procedure that is rarely performed today. It must be noted, though, that the subject data size should satisfy the intended use. Larger datasets or increased generalizability would probably not have changed the conclusions of papers III or IV. That being said, to quote Michael W. Browne [143], "Every model we are likely to consider will be wrong to some extent." This is likely true as a model is only a plausible approximation to reality.

8.3 WHAT MODEL IS THE MOST APPROPRIATE?

AI has experienced considerable advancement in the last decade and has been largely driven by the increase in data availability and computational power [147]. Within medicine, AI algorithms are now often performing on par [148] or even better than physicians [149, 150] in identifying tissue changes and pathology. In most cases, DL-based models outperform others. However, despite the promising results, few DL models have reached clinical implementation [151]. Perhaps because the high predictive accuracy is often achieved through increased model complexity, turning such systems into a "black box," leading to a lack of model interpretability and trust [114-116].

In contrast to conventional ML using handcrafted radiomic features, DL approaches using CNNs automatically extract features and hierarchical data representations from input images without human intervention [152]. However, handcrafted features are derived through mathematical equations [153]. As such, handcrafted features offer increased interpretability where the observer can evaluate model results to explain important characteristics in the image data. Another important aspect that, in some cases, might be in favor of conventional ML models is its need for fewer data. Although higher predictive performance can usually be reached using DL models, conventional ML models often require fewer data samples than DL equivalents to achieve comparable performance [123].

Inspired by Zhou, B. et al. [154], who first introduced the concept of discriminative localization maps, i.e., an attention map that highlights regions of an image used by a CNN to identify the category of the image, we developed

a method for attention mapping using radiomic features instead of CNNs (paper III). In this study, specific features were chosen that were able to identify the location of an annular fissure. Using this approach allowed the use of a simple model with relatively few parameters, making the method and model derivable. In study IV, we investigated the possible association between an annular fissure and adjacent vertebrae using radiomic features. Although unexplored, classification performance could probably have increased using a DL approach, but the association results, which were the aim of the study, would not have been easily achievable.

The need for trustworthy models, especially in critical domains such as healthcare and self-driving cars, has formed the field of explainable Artificial Intelligence (XAI) [155], which focuses on the understanding and interpretation of the behavior of AI systems. Although increasingly many contributions are made to this field, most methods for increasing the interpretability of DL models target classification tasks and are often explained by the use of varieties of attention maps-techniques to highlight the most important areas of an image related to the target class [114], e.g., CAM, Grad-CAM [135], LIME [156] and SHARP[157], to name a few. The field of XAI is evolving at a high pace, and although not yet mature [114, 158], the continued commitment might benefit complex ML models to be better understood in the future.

9 FUTURE PERSPECTIVES

The field of ML is rapidly evolving, with over 50.000 published papers in 2022. The surge of new ML models and techniques requires researchers to continually monitor the ML field to benefit from recent advances. The rapid evolution might, of course, make the current models used in this thesis soon obsolete. However, hopefully, more capable models will emerge and benefit the medical field to further advance the understanding of pathology related to LBP.

All included papers in this thesis are performed systematically to tackle some of the shortcomings in the field of spine research by offering new tools that display not only visible global changes of general degeneration. However, the thesis still does not aim to offer a diagnostic tool that overcomes the clinical challenges for this group of patients. Of course, it cannot be overlooked that patients with LBP have a higher prevalence of spinal tissue changes compared to individuals without pain and that MRI can depict specific behaviors that may reflect the cause of the pain. However, there is still a long way to go to establish evidence, and maybe we will never find the true cause of pain. A recent study showed that educating patients on their LBP significantly decreased their anxiety, stress, depression, and pain intensity [159], indicating that pain sensation might not only be related to tissue changes. Perhaps patients initially do experience pain related to pathology, but this pathology might heal or change character, leaving only the sensation of pain and, as such, will not be displayed by MRI. It has been shown that prior LBP is a predictor for future LBP [160], where more than 40% of the patients will suffer a recurrence of pain within one year [161]. As such, potential tissue changes related to LBP might be better recognized during longitudinal studies where changes are closely followed over time.

Numerous studies have evaluated spinal tissue changes and their associations related to LBP using conventional MRI-based radiological markers with the aim of better understanding the pathology and, ultimately, improving LBP diagnostics; however, without yet reached a conclusion. New image analysis methods, such as those described in this thesis, may have the feasibility to push the research forward. Although progress in the field is undeniably occurring, I believe that conventional MR images, with or without advanced analysis methods, have a limited explaining ability and that we need to introduce unexplored functional methods into the research field. Functional MRI, for example, diffusion and perfusion, can directly assess tissue function, which relaxation weighting and mapping methods do not offer. Combined with

present analysis methods, functional MRI may depict new signal behaviors important for LBP. We have exemplified that our explanatory analysis methods can display subtle, even to the eye, invisible tissue changes. Perhaps these methods can also explain similar subtle changes in data derived from functional MRI. I believe combining morphological images, functional methods, and new advanced image analysis is a plausible way forward. Additionally, combining functional MRI with longitudinal studies can hopefully capture subtle tissue changes that evolve over time. Such an approach may further increase our understanding of causality and ultimately bring us closer to understanding the source of LBP and thereby improving the treatment of this large patient cohort.

Going forward, a longitudinal study that includes a large cohort of LBP patients and matched controls might possibly advance the understanding of the occurrence of tissue changes associated with LBP. Specifically, clinical and functional MRI evaluated by explanatory analysis techniques developed in this thesis, such as handcrafted clustering features used to investigate the IVD heterogeneity, pixel-wise classification algorithms of vertebral pathology, and models using radiomics features evaluated in papers III and IV might have the potential to uncover yet unknown associations to continue to grow our understanding of what causes LBP.

ACKNOWLEDGEMENT

First and foremost, I want to express my deepest gratitude to my supervisor **Kerstin Lagerstrand**. Thank you for believing in me and allowing me to embark on this journey together with you. Your vision, enthusiasm, and many bright ideas have inspired me to explore new territory. Thank you for letting me shape my own projects, and even though some have been unsuccessful, you still encouraged me to continue to challenge myself. Your sturdy support and continued guidance have been of monument value. Although it was, at times, hard work (for probably both of us), I am now thankful for your determination and perseverance in wanting to pass on your knowledge, and I hope I can continue to learn and grow from you in the future. Thank you!

Secondly, I want to thank my co-supervisors:

Hanna Hebelka, thank you for your guidance. I have stumbled, and I have strayed, but you always gently guide me back in the right direction. Thank you for always finding the time to help me. Even under very short notice, you have always been available. Your perspective as a researcher and clinician has been invaluable to me.

Helena Brisby, our discussions have always been of the greatest value to me. Your broad knowledge and experience as a researcher and clinician are undoubtedly exceptional. Thank you for all the very important perspectives and input you have given me throughout my Ph.D. studies.

Thank you to my co-authors, **Stefanie Eriksson**, **Leif Torén, Nikolaos Papadimitriou**, and **Anna Grimby Ekman.** It has been a pleasure to work with you.

I want to thank the **MR physics group** at Sahlgrenska University Hospital. Thank you for your support and wisdom; your collective knowledge is unparallel. Thank you for the inspiring coffee breaks and many cheerful nights during conferences. I hope there are many more to come.

Thank you to everyone at the **Department of Medical Radiation Sciences**, **DSF**, and **MFT** for supporting me throughout my time as a Ph.D. student.

Lastly, I want to send a big thank you to my **family** and wife, **Emmy**. Thank you for always believing in me and have encouraged me during my Ph.D. studies.

REFERENCES

1. Waldenberg, C., et al., *Differences in IVD characteristics between low back pain patients and controls associated with HIZ as revealed with quantitative MRI.* PLoS One, 2019. **14**(8): p. e0220952.
2. Waldenberg, C., et al., *MRI histogram analysis enables objective and continuous classification of intervertebral disc degeneration.* European Spine Journal, 2017.
3. Song, J., et al., *Prevalence of Lumbosacral Transition Vertebrae in Symptomatic Adults and the Levels of Degeneration in the Suprajacent Disc.* Surgeries, 2023. **4**(1): p. 120-126.
4. Waldenberg, C., et al., *Detection of Imperceptible Intervertebral Disc Fissures in Conventional MRI—An AI Strategy for Improved Diagnostics.* Journal of Clinical Medicine, 2023. **12**(1): p. 11.
5. Kim, S.M., et al., *Analysis of the Correlation Among Age, Disc Morphology, Positive Discography and Prognosis in Patients With Chronic Low Back Pain.* Ann Rehabil Med, 2015. **39**(3): p. 340-6.
6. Andersson, G.B., *Epidemiological features of chronic low-back pain.* Lancet, 1999. **354**(9178): p. 581-5.
7. Modic, M.T. and J.S. Ross, *Lumbar Degenerative Disk Disease.* Radiology, 2007. **245**(1): p. 43-61.
8. Andersson, G.B.J., *Epidemiological features of chronic low-back pain.* The Lancet, 1999. **354**(9178): p. 581-585.
9. Urban, J.P.G. and S. Roberts, *Degeneration of the intervertebral disc.* Arthritis Res Ther, 2003. **5**(3): p. 120.
10. Buchbinder, R., et al., *Low back pain: a call for action.* The Lancet, 2018. **391**(10137): p. 2384-2388.
11. Videman, T. and M. Nurminen, *The occurrence of anular tears and their relation to lifetime back pain history: a cadaveric study using barium sulfate discography.* Spine (Phila Pa 1976), 2004. **29**(23): p. 2668-76.
12. Lim, C.-H., et al., *Discogenic lumbar pain: association with MR imaging and CT discography.* European Journal of Radiology, 2005. **54**(3): p. 431-437.
13. Jensen, T.S., et al., *Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain.* Eur Spine J, 2008. **17**(11): p. 1407-22.
14. Liu, C., et al., *Quantitative estimation of the high-intensity zone in the lumbar spine: comparison between the symptomatic and asymptomatic population.* Spine J, 2014. **14**(3): p. 391-6.
15. Mok, F.P.S., et al., *Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort.* The Spine Journal, 2016. **16**(1): p. 32-41.

16. Ross, J.S., M.T. Modic, and T.J. Masaryk, *Tears of the anulus fibrosus: assessment with Gd-DTPA-enhanced MR imaging.* AJNR Am J Neuroradiol, 1989. **10**(6): p. 1251-4.
17. Peng, B., et al., *The pathogenesis and clinical significance of a high-intensity zone (HIZ) of lumbar intervertebral disc on MR imaging in the patient with discogenic low back pain.* European Spine Journal, 2006. **15**(5): p. 583-587.
18. Fields, A.J., E.C. Liebenberg, and J.C. Lotz, *Innervation of pathologies in the lumbar vertebral end plate and intervertebral disc.* The Spine Journal, 2014. **14**(3): p. 513-521.
19. Urban, J.P.G., S. Smith, and J.C.T. Fairbank, *Nutrition of the Intervertebral Disc.* Spine, 2004. **29**(23): p. 2700-2709.
20. Conger, A., et al., *Vertebrogenic Pain: A Paradigm Shift in Diagnosis and Treatment of Axial Low Back Pain.* Pain Medicine, 2022. **23**(Supplement_2): p. S63-S71.
21. Kjaer, P., et al., *Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women.* Spine (Phila Pa 1976), 2005. **30**(10): p. 1173-80.
22. Treede, R.-D., et al., *Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11).* pain, 2019. **160**(1): p. 19-27.
23. Brinjikji, W., et al., *Systematic Literature Review of Imaging Features of Spinal Degeneration in Asymptomatic Populations.* American Journal of Neuroradiology, 2015. **36**(4): p. 811-816.
24. Chou, R., et al., *Imaging strategies for low-back pain: systematic review and meta-analysis.* Lancet, 2009. **373**(9662): p. 463-72.
25. You, J.J., et al., *Indications for and results of outpatient computed tomography and magnetic resonance imaging in Ontario.* Can Assoc Radiol J, 2008. **59**(3): p. 135-43.
26. Yip, S.S. and H.J. Aerts, *Applications and limitations of radiomics.* Phys Med Biol, 2016. **61**(13): p. R150-66.
27. Huber, F.A. and R. Guggenberger, *AI MSK clinical applications: spine imaging.* Skeletal Radiology, 2022. **51**(2): p. 279-291.
28. Eriksson, S., et al., *Texture Analysis of Magnetic Resonance Images Enables Phenotyping of Potentially Painful Annular Fissures.* Spine (Phila Pa 1976), 2022. **47**(5): p. 430-437.
29. Grant, L.A. and N. Griffin, *Grainger & Allison's Diagnostic Radiology Essentials E-Book*. 2018: Elsevier Health Sciences.
30. Twomey, L. and J. Taylor, *Age changes in lumbar vertebrae and intervertebral discs.* Clinical orthopaedics and related research, 1987(224): p. 97-104.
31. Schellinger, D., et al., *Normal lumbar vertebrae: anatomic, age, and sex variance in subjects at proton MR spectroscopy—initial experience.* Radiology, 2000. **215**(3): p. 910-916.

32. Lotz, J., A. Fields, and E. Liebenberg, *The role of the vertebral end plate in low back pain.* Global spine journal, 2013. **3**(3): p. 153-163.
33. Moore, R.J., *The vertebral end-plate: what do we know?* European Spine Journal, 2000. **9**(2): p. 92-96.
34. Roberts, S., et al., *Transport properties of the human cartilage endplate in relation to its composition and calcification.* Spine, 1996. **21**(4): p. 415-420.
35. ODA, J., H. TANAKA, and N. TSUZUKI, *Intervertebral Disc Changes with Aging of Human Cervical Vertebra: From the Neonate to the Eighties.* Spine, 1988. **13**(11): p. 1205-1211.
36. Moore, R.J., *The vertebral endplate: disc degeneration, disc regeneration.* European Spine Journal, 2006. **15**(3): p. 333-337.
37. Walker, J., 3rd, et al., *Discography in practice: a clinical and historical review.* Current reviews in musculoskeletal medicine, 2008. **1**(2): p. 69-83.
38. Arjmand, N., et al., *Predictive equations for lumbar spine loads in load-dependent asymmetric one- and two-handed lifting activities.* Clinical Biomechanics, 2012. **27**(6): p. 537-544.
39. Dowdell, J., et al., *Intervertebral disk degeneration and repair.* Neurosurgery, 2017. **80**(3 Suppl): p. S46.
40. Inoue, H. and T. Takeda, *Three-dimensional observation of collagen framework of lumbar intervertebral discs.* Acta Orthop Scand, 1975. **46**(6): p. 949-56.
41. Yu, J., et al., *Elastic fibre organization in the intervertebral discs of the bovine tail.* Journal of anatomy, 2002. **201**(6): p. 465-475.
42. Glaser, C., *New techniques for cartilage imaging: T2 relaxation time and diffusion-weighted MR imaging.* Radiol Clin North Am, 2005. **43**(4): p. 641-53, vii.
43. Modic, M.T., et al., *Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging.* Radiology, 1988. **166**(1 Pt 1): p. 193-9.
44. Sharma, A., T. Pilgram, and F.J. Wippold, *Association between Annular Tears and Disk Degeneration: A Longitudinal Study.* American Journal of Neuroradiology, 2009. **30**(3): p. 500-506.
45. Adams, M.A., *Biomechanics of back pain.* Acupunct Med, 2004. **22**(4): p. 178-88.
46. Deneuville, J.-P., et al., *Quantitative MRI to Characterize the Nucleus Pulposus Morphological and Biomechanical Variation According to Sagittal Bending Load and Radial Fissure, an ex vivo Ovine Specimen Proof-of-Concept Study.* Frontiers in Bioengineering and Biotechnology, 2021. **9**(489).
47. Dudli, S., et al., *Pathobiology of Modic changes.* European Spine Journal, 2016. **25**: p. 3723-3734.

48. Risbud, M.V. and I.M. Shapiro, *Role of cytokines in intervertebral disc degeneration: pain and disc content.* Nat Rev Rheumatol, 2014. **10**(1): p. 44-56.
49. Aoki, J., et al., *End plate of the discovertebral joint: degenerative change in the elderly adult.* Radiology, 1987. **164**(2): p. 411-414.
50. Benneker, L.M., et al., *2004 Young Investigator Award Winner: vertebral endplate marrow contact channel occlusions and intervertebral disc degeneration.* Spine, 2005. **30**(2): p. 167-173.
51. Holm, S., et al., *Experimental disc degeneration due to endplate injury.* J Spinal Disord Tech, 2004. **17**(1): p. 64-71.
52. Bailey, J.F., et al., *Innervation patterns of PGP 9.5-positive nerve fibers within the human lumbar vertebra.* Journal of anatomy, 2011. **218**(3): p. 263-270.
53. Antonacci, M.D., D.R. Mody, and M.H. Heggeness, *Innervation of the human vertebral body: a histologic study.* Clinical Spine Surgery, 1998. **11**(6): p. 526-531.
54. Gulani, V. and N. Seiberlich, *Quantitative MRI: Rationale and Challenges*, in *Advances in Magnetic Resonance Technology and Applications*, N. Seiberlich, et al., Editors. 2020, Academic Press. p. xxxvii-li.
55. Linsenmaier, U., et al., *Whole-body computed tomography in polytrauma: techniques and management.* European Radiology, 2002. **12**(7): p. 1728-1740.
56. Platzer, P., et al., *Clearing the cervical spine in critically injured patients: a comprehensive C-spine protocol to avoid unnecessary delays in diagnosis.* European Spine Journal, 2006. **15**(12): p. 1801-1810.
57. Wilmink, J., *MR imaging of the spine: trauma and degenerative disease.* European Radiology, 1999. **9**: p. 1259-1266.
58. Parizel, P.M., et al., *Trauma of the spine and spinal cord: imaging strategies.* European Spine Journal, 2010. **19**(1): p. 8-17.
59. Wilmink, J.T., *MR imaging of the spine: trauma and degenerative disease.* European Radiology, 1999. **9**(7): p. 1259-1266.
60. Li, X. and S. Majumdar, *Quantitative MRI of articular cartilage and its clinical applications.* Journal of Magnetic Resonance Imaging, 2013. **38**(5): p. 991-1008.
61. Lindblom, K., *Diagnostic puncture of intervertebral disks in sciatica.* Acta orthopaedica scandinavica, 1948. **17**(1-4): p. 231-239.
62. Hirsch, C., *An attempt to diagnose the level of a disc lesion clinically by disc puncture.* Acta orthopaedica scandinavica, 1949. **18**(1-4): p. 132-140.
63. Carragee, E.J., et al., *2009 ISSLS Prize Winner: Does discography cause accelerated progression of degeneration changes in the lumbar disc: a ten-year matched cohort study.* Spine (Phila Pa 1976), 2009. **34**(21): p. 2338-45.

64. Hebelka, H., A. Nilsson, and T. Hansson, *Pressure Increase in Adjacent Discs During Clinical Discography Questions the Methods Validity.* Spine, 2014. **39**(11): p. 893-899.
65. Bojorquez, J.Z., et al., *What are normal relaxation times of tissues at 3 T?* Magnetic Resonance Imaging, 2017. **35**: p. 69-80.
66. McRobbie, D.W., *MRI from picture to proton*. 3. ed. ed, ed. E.A. Moore and M.J. Graves. 2017: Cambridge : Cambridge University Press.
67. Pfirrmann, C.W., et al., *Magnetic resonance classification of lumbar intervertebral disc degeneration.* Spine (Phila Pa 1976), 2001. **26**(17): p. 1873-8.
68. Shah, M., et al., *Evaluating intensity normalization on MRIs of human brain with multiple sclerosis.* Medical Image Analysis, 2011. **15**(2): p. 267-282.
69. Nyúl, L.G. and J.K. Udupa, *On standardizing the MR image intensity scale.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 1999. **42**(6): p. 1072-1081.
70. Shinohara, R.T., et al., *Statistical normalization techniques for magnetic resonance imaging.* Neuroimage Clin, 2014. **6**: p. 9-19.
71. Prasloski, T., et al., *Applications of stimulated echo correction to multicomponent T2 analysis.* Magn Reson Med, 2012. **67**(6): p. 1803-14.
72. Lukzen, N.N., et al., *The generating functions formalism for the analysis of spin response to the periodic trains of RF pulses. Echo sequences with arbitrary refocusing angles and resonance offsets.* Journal of Magnetic Resonance, 2009. **196**(2): p. 164-169.
73. Lebel, R.M. and A.H. Wilman, *Transverse relaxometry with stimulated echo compensation.* Magnetic Resonance in Medicine, 2010. **64**(4): p. 1005-1014.
74. Mosher, T.J. and B.J. Dardzinski, *Cartilage MRI T2 relaxation time mapping: overview and applications.* Semin Musculoskelet Radiol, 2004. **8**(4): p. 355-68.
75. Trattnig, S., et al., *Lumbar intervertebral disc abnormalities: comparison of quantitative T2 mapping with conventional MR at 3.0 T.* Eur Radiol, 2010. **20**(11): p. 2715-22.
76. Nilsson, M., et al., *Axial loading during MRI influences T2-mapping values of lumbar discs: a feasibility study on patients with low back pain.* European Spine Journal, 2016. **25**(9): p. 2856-2863.
77. Krupinski, E.A., *Current perspectives in medical image perception.* Attention, Perception, & Psychophysics, 2010. **72**(5): p. 1205-1217.
78. Mahesh, M., A.J. Ansari, and J. Fred A. Mettler, *Patient Exposure from Radiologic and Nuclear Medicine Procedures in the United States and Worldwide: 2009–2018.* Radiology. **0**(0): p. 221263.

79. Shan, Z., et al., *Correlation between biomechanical properties of the annulus fibrosus and magnetic resonance imaging (MRI) findings.* European Spine Journal, 2015. **24**(9): p. 1909-1916.
80. Yu, H.J., et al., *In vivo quantification of lumbar disc degeneration: assessment of ADC value using a degenerative scoring system based on Pfirrmann framework.* European Spine Journal, 2015. **24**(11): p. 2442-2448.
81. Muftuler, L.T., et al., *Association between intervertebral disc degeneration and endplate perfusion studied by DCE-MRI.* European Spine Journal, 2015. **24**(4): p. 679-685.
82. Griffith, J.F., et al., *Modified Pfirrmann Grading System for Lumbar Intervertebral Disc Degeneration.* Spine, 2007. **32**(24): p. E708-E712.
83. Zheng, H.-D., et al., *Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI.* Nature Communications, 2022. **13**(1): p. 841.
84. Viswanathan, V.K., A.P. Shetty, and S. Rajasekaran, *Modic changes - An evidence-based, narrative review on its patho-physiology, clinical significance and role in chronic low back pain.* J Clin Orthop Trauma, 2020. **11**(5): p. 761-769.
85. Herlin, C., et al., *Modic changes—their associations with low back pain and activity limitation: a systematic literature review and meta-analysis.* PloS one, 2018. **13**(8): p. e0200677.
86. Jensen, T.S., et al., *Characteristics and natural course of vertebral endplate signal (Modic) changes in the Danish general population.* BMC Musculoskelet Disord, 2009. **10**: p. 81.
87. Aprill, C. and N. Bogduk, *High-intensity zone: a diagnostic sign of painful lumbar disc on magnetic resonance imaging.* Br J Radiol, 1992. **65**(773): p. 361-9.
88. Lei, D., et al., *Painful Disc Lesion: Can Modern Biplanar Magnetic Resonance Imaging Replace Discography?* Clinical Spine Surgery, 2008. **21**(6): p. 430-435.
89. Carragee, E.J., S.J. Paragioudakis, and S. Khurana, *Lumbar High-Intensity Zone and Discography in Subjects Without Low Back Problems.* Spine, 2000. **25**(23): p. 2987-2992.
90. Sachs, B.L., et al., *Dallas discogram description. A new classification of CT/discography in low-back disorders.* Spine (Phila Pa 1976), 1987. **12**(3): p. 287-94.
91. Yu, K.-H., A.L. Beam, and I.S. Kohane, *Artificial intelligence in healthcare.* Nature biomedical engineering, 2018. **2**(10): p. 719-731.
92. Hamet, P. and J. Tremblay, *Artificial intelligence in medicine.* Metabolism, 2017. **69**: p. S36-S40.
93. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Medical image analysis, 2017. **42**: p. 60-88.
94. Castiglioni, I., et al., *AI applications to medical images: From machine learning to deep learning.* Physica Medica, 2021. **83**: p. 9-24.

95. Rajkomar, A., J. Dean, and I. Kohane, *Machine Learning in Medicine.* New England Journal of Medicine, 2019. **380**(14): p. 1347-1358.
96. Yang, T., et al., *The application of key feature extraction algorithm based on Gabor wavelet transformation in the diagnosis of lumbar intervertebral disc degenerative changes.* PLoS One, 2020. **15**(2): p. e0227894.
97. Gao, F., et al., *Automated Grading of Lumbar Disc Degeneration Using a Push-Pull Regularization Network Based on MRI.* Journal of Magnetic Resonance Imaging, 2021. **53**(3): p. 799-806.
98. Eriksson, S., et al., *Texture Analysis of Magnetic Resonance Images Enables Phenotyping of Potentially Painful Annular Fissures.* Spine (Phila Pa 1976), 2021.
99. Murata, K., et al., *Artificial intelligence for the detection of vertebral fractures on plain spinal radiography.* Scientific Reports, 2020. **10**(1): p. 1-8.
100. Thong, W., et al., *Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models.* European Spine Journal, 2016. **25**(10): p. 3104-3113.
101. D'Antoni, F., et al., *Artificial Intelligence and Computer Vision in Low Back Pain: A Systematic Review.* International Journal of Environmental Research and Public Health, 2021. **18**(20): p. 10909.
102. D'Antoni, F., et al., *Artificial Intelligence and Computer Aided Diagnosis in Chronic Low Back Pain: A Systematic Review.* International Journal of Environmental Research and Public Health, 2022. **19**(10): p. 5971.
103. Schwartz, W.B., R.S. Patil, and P. Szolovits, *Artificial Intelligence in Medicine.* New England Journal of Medicine, 1987. **316**(11): p. 685-688.
104. Fei-Fei, L., R. Fergus, and P. Perona, *One-shot learning of object categories.* IEEE transactions on pattern analysis and machine intelligence, 2006. **28**(4): p. 594-611.
105. Balki, I., et al., *Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review.* Can Assoc Radiol J, 2019. **70**(4): p. 344-353.
106. Cho, J., et al., *How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?* arXiv preprint arXiv:1511.06348, 2015.
107. Chang, K., et al., *Distributed deep learning networks among institutions for medical imaging.* Journal of the American Medical Informatics Association, 2018. **25**(8): p. 945-954.
108. Willemink, M.J., et al., *Preparing Medical Imaging Data for Machine Learning.* Radiology, 2020. **295**(1): p. 4-15.
109. Park, S.H. and K. Han, *Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for*

Medical Diagnosis and Prediction. Radiology, 2018. **286**(3): p. 800-809.
110. Russakovsky, O., et al., *ImageNet Large Scale Visual Recognition Challenge.* International Journal of Computer Vision, 2015. **115**(3): p. 211-252.
111. Lei, Y., et al., *MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks.* Medical Physics, 2019. **46**(8): p. 3565-3581.
112. Sveinsson, B., et al., *Synthesizing Quantitative T2 Maps in Right Lateral Knee Femoral Condyles from Multicontrast Anatomic Data with a Conditional Generative Adversarial Network.* Radiology: Artificial Intelligence, 2021. **3**(5): p. e200122.
113. Alotaibi, A., *Deep generative adversarial networks for image-to-image translation: A review.* Symmetry, 2020. **12**(10): p. 1705.
114. Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis, *Explainable AI: A Review of Machine Learning Interpretability Methods.* Entropy, 2021. **23**(1): p. 18.
115. Wang, L., *Heterogeneous data and big data analytics.* Automatic Control and Information Sciences, 2017. **3**(1): p. 8-15.
116. Falla, D., et al., *Machine learning approaches applied in spinal pain research.* Journal of Electromyography and Kinesiology, 2021. **61**: p. 102599.
117. Jiménez-Carvelo, A.M., et al., *Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review.* Food Research International, 2019. **122**: p. 25-39.
118. Ray, S. *A quick review of machine learning algorithms*. in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. 2019. IEEE.
119. Yang, L. and A. Shami, *On hyperparameter optimization of machine learning algorithms: Theory and practice.* Neurocomputing, 2020. **415**: p. 295-316.
120. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data.* Radiology, 2016. **278**(2): p. 563-77.
121. Moon, S.H., et al., *Correlations between metabolic texture features, genetic heterogeneity, and mutation burden in patients with lung cancer.* European Journal of Nuclear Medicine and Molecular Imaging, 2019. **46**(2): p. 446-454.
122. Mayerhoefer, M.E., et al., *Introduction to Radiomics.* Journal of Nuclear Medicine, 2020. **61**(4): p. 488-495.
123. Zhou, J., et al., *Predicting the response to neoadjuvant chemotherapy for breast cancer: wavelet transforming radiomics in MRI.* BMC Cancer, 2020. **20**(1): p. 100.

124. Bianchini, L., et al., *A multicenter study on radiomic features from T2-weighted images of a customized MR pelvic phantom setting the basis for robust radiomic models in clinics.* Magnetic Resonance in Medicine, 2021. **85**(3): p. 1713-1726.
125. Van Griethuysen, J.J., et al., *Computational radiomics system to decode the radiographic phenotype.* Cancer research, 2017. **77**(21): p. e104-e107.
126. Hughes, G., *On the mean accuracy of statistical pattern recognizers.* IEEE Transactions on Information Theory, 1968. **14**(1): p. 55-63.
127. Sima, C. and E.R. Dougherty, *The peaking phenomenon in the presence of feature-selection.* Pattern Recognition Letters, 2008. **29**(11): p. 1667-1674.
128. Cai, J., et al., *Feature selection in machine learning: A new perspective.* Neurocomputing, 2018. **300**: p. 70-79.
129. Tang, J., S. Alelyani, and H. Liu, *Feature selection for classification: A review.* Data classification: Algorithms and applications, 2014: p. 37.
130. Hubel, D.H. and T.N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex.* The Journal of physiology, 1959. **148**(3): p. 574-591.
131. Gardner, M.W. and S.R. Dorling, *Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences.* Atmospheric Environment, 1998. **32**(14): p. 2627-2636.
132. Alzubaidi, L., et al., *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.* Journal of Big Data, 2021. **8**(1): p. 53.
133. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 2015. Springer.
134. Dolz, J., C. Desrosiers, and I. Ben Ayed. *IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet*. in *Computational Methods and Clinical Applications for Spine Imaging: 5th International Workshop and Challenge, CSI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers*. 2019. Springer.
135. Selvaraju, R.R., et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
136. Jarvik, J.G. and R.A. Deyo, *Diagnostic evaluation of low back pain with emphasis on imaging.* Ann Intern Med, 2002. **137**(7): p. 586-97.
137. Ma, J., et al., *Is fractal dimension a reliable imaging biomarker for the quantitative classification of an intervertebral disk?* European Spine Journal, 2020. **29**: p. 1175-1180.

138. Bisoi, A.K. and J. Mishra, *On calculation of fractal dimension of images.* Pattern Recognition Letters, 2001. **22**(6): p. 631-637.
139. Ogier, A.C., et al., *The road toward reproducibility of parametric mapping of the heart: a technical review.* Frontiers in Cardiovascular Medicine, 2022. **9**.
140. Dudli, S., et al. *INTERVERTEBRAL DISC/BONE MARROW CROSS-TALK WITH MODIC CHANGES (ISSLS Prize-Basic Science): 55*. in *Spine Journal Meeting Abstracts*. 2017. LWW.
141. Varoquaux, G. and V. Cheplygina, *Machine learning for medical imaging: methodological failures and recommendations for the future.* npj Digital Medicine, 2022. **5**(1): p. 48.
142. Hansen, L.K. and P. Salamon, *Neural network ensembles.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990. **12**(10): p. 993-1001.
143. Browne, M.W., *Cross-validation methods.* Journal of mathematical psychology, 2000. **44**(1): p. 108-132.
144. Berrar, D., *Cross-Validation*. 2019.
145. Belavý, D.L., et al., *Running exercise strengthens the intervertebral disc.* Sci Rep, 2017. **7**: p. 45975.
146. *Combining Shape and Learning for Medical Image Analysis: Robust, Scalable and Generalizable Registration and Segmentation*. 2019: Chalmers University of Technology.
147. Esteva, A., et al., *A guide to deep learning in healthcare.* Nature Medicine, 2019. **25**(1): p. 24-29.
148. Jamaludin, A., et al., *ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist.* Eur Spine J, 2017. **26**(5): p. 1374-1383.
149. Zhou, W., et al., *Ensembled deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images.* Nature Communications, 2021. **12**(1): p. 1259.
150. Golany, T., et al., *Physicians and Machine-Learning Algorithm Performance in Predicting Left-Ventricular Systolic Dysfunction from a Standard 12-Lead-Electrocardiogram.* Journal of Clinical Medicine, 2022. **11**(22): p. 6767.
151. van der Laak, J., G. Litjens, and F. Ciompi, *Deep learning in histopathology: the path to the clinic.* Nature Medicine, 2021. **27**(5): p. 775-784.
152. Scapicchio, C., et al., *A deep look into radiomics.* Radiol Med, 2021. **126**(10): p. 1296-1311.
153. Zwanenburg, A., et al., *The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping.* Radiology, 2020. **295**(2): p. 328-338.

154. Zhou, B., et al. *Learning deep features for discriminative localization*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
155. Gunning, D. and D. Aha, *DARPA's explainable artificial intelligence (XAI) program.* AI magazine, 2019. **40**(2): p. 44-58.
156. Ribeiro, M.T., S. Singh, and C. Guestrin. *" Why should i trust you?" Explaining the predictions of any classifier*. in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.
157. Lundberg, S.M. and S.-I. Lee, *A unified approach to interpreting model predictions.* Advances in neural information processing systems, 2017. **30**.
158. Das, A. and P. Rad, *Opportunities and challenges in explainable artificial intelligence (xai): A survey.* arXiv preprint arXiv:2006.11371, 2020.
159. Kanaan, S.F., et al., *Evaluating the effectiveness of a comprehensive education on low back pain treatment outcomes: A controlled clinical study.* Clinical Rehabilitation, 2023. **37**(1): p. 98-108.
160. Hoy, D., et al., *The Epidemiology of low back pain.* Best Practice & Research Clinical Rheumatology, 2010. **24**(6): p. 769-781.
161. Hestbaek, L., C. Leboeuf-Yde, and C. Manniche, *Low back pain: what is the long-term course? A review of studies of general patient populations.* European Spine Journal, 2003. **12**(2): p. 149-165.