

UNIVERSITY OF HELSINKI

MRI ASSESSMENT OF LUMBAR DISC DEGENERATION WITH THE PFIRRMANN CLASSIFICATION

QUANTITATIVE, DEEP LEARNING, AND LONG-TERM PERSPECTIVES

NIKO MURTO

DISSERTATIONES UNIVERSITATIS HELSINGIENSIS | 118 2026

University of Helsinki
Dissertationes Universitatis Helsingiensis
118/2026

Doctoral Programme in Clinical Research, Faculty of Medicine
HUS Diagnostic Center, Radiology
University of Helsinki and Helsinki University Hospital
Helsinki, Finland

MRI ASSESSMENT OF LUMBAR DISC DEGENERATION WITH THE PFIRRMANN CLASSIFICATION

QUANTITATIVE, DEEP LEARNING, AND LONG-TERM PERSPECTIVES

Niko Murto

DOCTORAL DISSERTATION
To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, in Lecture Hall 3, Biomedicum, Haartmaninkatu 8, on 27 March 2026, at 12:30.

Helsinki 2026

Supervised by
Docent Liisa Kerttula, MD, PhD
Department of Radiology
University of Helsinki

Docent Teija Lund, MD, PhD
Department of Orthopaedics and Traumatology
University of Helsinki

Docent Katariina Luoma, MD, PhD
Department of Radiology
University of Helsinki

Reviewed by
Docent Pekka Niemi, MD, PhD
Department of Radiology
University of Turku

Docent Juhani Määttä, MD, PhD
Department of Research Unit of Health Sciences and Technology
University of Oulu

To be discussed with
Professor Jaakko Niinimäki, MD, PhD
Department of Radiology
University of Oulu

Publisher: University of Helsinki
Series: Dissertationes Universitatis Helsingiensis 118/2026

ISBN 978-952-84-2107-8 (paperback)
ISBN 978-952-84-2106-1 (PDF)
ISSN 2954-2898 (print)
ISSN 2954-2952 (PDF)

Unigrafia / PunaMusta
Joensuu 2026

ABSTRACT

Lumbar disc degeneration (DD) is a complex and slowly progressive condition that is associated with spinal ageing. The intervertebral disc (IVD) is an important structure for load distribution and mobility of the spine, and it is composed of the gelatinous nucleus pulposus (NP), fibrous annulus fibrosus (AF) and cartilaginous endplates. During DD, biochemical and structural changes in the disc occur, such as proteoglycan loss, reduced water content in the NP, and disruption of AF collagen architecture. This may eventually lead to disc collapse and loss of the disc's mechanical function.

Magnetic resonance imaging (MRI), especially T2-weighted imaging (T2WI), enables a detailed evaluation of disc hydration and structure. The Pfirrmann classification is the most widely used qualitative grading method for DD assessment, and it is based on disc MRI characteristics on T2WI: signal intensity (i.e., brightness) and homogeneity of the NP, distinction of the border between the NP and AF, and disc height (DH). Although visually assessed DD on MRI is more prevalent in individuals with low back pain (LBP), its clinical significance remains debated because it is also a common MRI finding in asymptomatic individuals. This discrepancy between imaging findings and symptoms has increased interest in more sensitive approaches, including quantitative MRI analysis of DD and artificial intelligence (AI)-based image analysis.

This doctoral thesis investigates three complementary aspects of MRI-based DD evaluation. The first study examined the reliability and diagnostic value of quantitative T2-weighted signal intensity measurements of the IVD using different region-of-interest (ROI) approaches. The second study evaluated the agreement between an AI-based model (SpineNet) and radiologists' assessments of the Pfirrmann classification in a 14-year follow-up. The third study characterised DD progression over 14 years, focusing on changes in Pfirrmann grade and individual disc component features, including signal intensity and inhomogeneity of the NP, AF border distinction, and DH.

This work demonstrates that the quantitative T2-weighted signal intensity measurement used in this study provides a reliable and objective method for DD assessment and may complement other DD grading methods. The AI model SpineNet demonstrated moderate agreement with radiologist assessment but tended to under- or overestimate the severity of DD. During the 14-year follow-up, overall DD progressed according to both Pfirrmann grading and quantitative T2-weighted signal intensity measurements. However, the rate of degeneration in individual disc components differed. This observed variation

in disc components' degeneration highlights the limitations of the Pfirrmann classification in detecting subtle structural changes and emphasises the need for more detailed grading of DD.

These findings support the use of quantitative MRI methods and AI as supplementary tools for evaluating DD and emphasise the importance of detailed, disc component–level analysis to capture the nuances of degenerative change over time.

TIIVISTELMÄ

Nikamavälilevyn rappeuma on monisyinen ja hitaasti etenevä, selkärangan ikääntymiseen liittyvä prosessi. Välilevy on tärkeä rakenne selkärangan kuormituksen jakautumisen ja liikkuvuuden kannalta. Välilevy koostuu hyytelömäisestä sisäosasta (nucleus pulposus, NP), sitä ympäröivästä fibroottisesta renkaasta (annulus fibrosus, AF) sekä rustoisista päätelevyistä. Välilevyn rappeutuessa siinä tapahtuu biokemiallisia ja rakenteellisia muutoksia, kuten proteoglykaanien ja vesipitoisuuden vähenemistä ja kollageenirakenteiden hajoamista. Tämä voi lopulta johtaa välilevyn kasaan painumiseen ja mekaanisen toiminnan heikentymiseen.

Magneettikuvauksella (MRI), erityisesti T2-painotteisilla magneettikuvilla, voidaan arvioida yksityiskohtaisesti välilevyn vesipitoisuutta ja rakennetta. Pfirrmannin luokitus on yleisimmin käytetty kvalitatiivinen menetelmä nikamavälilevyjen rappeuman arviointiin. Se perustuu välilevyn rakenteiden ominaisuuksien arviointiin T2-painotteisista magneettikuvista: välilevyn sisäosan signaalin voimakkuuteen eli kirkkauteen sekä homogeenisuuteen, sisäosan ja fibroottisen renkaan erotettavuuteen sekä välilevyn korkeuteen. Vaikka magneettikuvissa nähtävä välilevyrappeuma on yleinen löydös alaselkäkipuisilla, sen kliininen merkitys on kiistanalainen, koska vastaavia löydöksiä esiintyy myös oireettomilla henkilöillä. Tämän vuoksi on alettu kehittää tarkempia arviointimenetelmiä, kuten kvantitatiivisia menetelmiä ja tekoälypohjaisia kuvantulkintatyökaluja nikamavälilevyjen rappeuman arviointiin.

Tässä väitöskirjassa tarkastellaan kolmea magneettikuviin perustuvaa, nykyistä välilevyrappeumaluokittelua täydentävää menetelmää. Ensimmäisessä osatyössä arvioitiin välilevyn T2-signaalin kvantitatiivisen mittauksen toistettavuutta ja diagnostista arvoa erilaisilla region of interest (ROI) -menetelmillä. Toisessa osatyössä verrattiin tekoälymallin (SpineNet) ja radiologien tekemiä Pfirrmann-luokituksia 14 vuoden aikavälillä. Kolmannessa osatyössä arvioitiin rappeuman etenemistä 14 vuoden seurannan aikana arvioimalla Pfirrmannin luokan muutosta sekä välilevyn eri rakenteiden rappeuman etenemistä: sisäosan signaalin tummumista ja rakennetta, fibroottisen renkaan ja sisäosan välisen rajan erotettavuutta ja välilevyn korkeutta.

Tutkimus osoitti, että käytetty T2-painotteinen signaalin mittaus on luotettava ja objektiivinen menetelmä välilevyrappeuman arviointiin ja se voi täydentää perinteisiä välilevyrappeuman luokituksia. Tekoälymallin ja radiologien välillä todettiin kohtalaista yhtenevyyttä välilevyrappeuman luokittelussa, mutta tekoäly sekä ali- että yliarvioi rappeuman vaikeusastetta. Kolmannessa osatyössä todettiin, että välilevyrappeuma eteni 14 vuoden tarkastelujakson

aikana pääosin odotetusti, mutta välilevyn eri rakenteiden muutosten etenemisnopeus vaihteli. Tämä havaittu vaihtelu kuvastaa Pfirrmann-luokituksen rajoituksia yksityiskohtaisten rakenteellisten muutosten havaitsemisessa ja korostaa tarvetta kattavammille välilevyrappeuman arviointimenetelmille, joissa huomioidaan välilevyn eri rakenteet yksityiskohtaisesti. Saadut tulokset tukevat kvantitatiivisten menetelmien ja tekoälyn käyttöä täydentävinä työkaluina välilevyrappeuman arvioinnissa ja korostavat yksityiskohtaisen välilevyn rakenteiden arvioinnin merkitystä erityisesti pitkittäistutkimuksissa.

TABLE OF CONTENTS

ABSTRACT iii
TIIVISTELMÄ v
TABLE OF CONTENTS vii
ABBREVIATIONS ix
LIST OF ORIGINAL PUBLICATIONS x
1 INTRODUCTION 1
2 REVIEW OF THE LITERATURE 4
2.1 Anatomy of the lumbar spine and the intervertebral disc 4
2.2 Intervertebral disc degeneration 8
2.2.1 Normal disc composition 8
2.2.2 Structural and biochemical changes in disc degeneration 9
2.2.3 Risk factors for disc degeneration 10
2.2.4 Association of disc degeneration with pain 11
2.3 Imaging of lumbar disc degeneration 12
2.3.1 Radiographic imaging modalities 12
2.3.2 Magnetic resonance imaging 13
2.3.3 Pfirrmann classification 13
2.3.4 Other degenerative features on magnetic resonance imaging 15
2.3.5 Quantitative magnetic resonance imaging of disc degeneration 16
2.4 Artificial intelligence in spine magnetic resonance imaging analysis 21
2.4.1 Overview 21
2.4.2 SpineNet: architecture and initial performance 22
2.4.3 External validation of SpineNet and SpineNetV2 22
2.4.4 Other convolutional neural network approaches 23
2.4.5 Hybrid segmentation–classification pipelines 23
2.4.6 Automated quantitative pipelines 24
2.4.7 Incorporating axial images and multisequence magnetic resonance imaging 24
2.4.8 Radiomics and prognostic modelling 25
2.4.9 Strengths and limitations of artificial intelligence approaches 25

2.5 Longitudinal studies on disc degeneration 27
2.5.1 Population-based cohorts 27
2.5.2 Twin studies 28
2.5.3 Developmental and early-life determinants of disc degeneration 29
2.5.4 Asymptomatic cohorts 30
2.5.5 Symptomatic cohorts 30
2.5.6 Differences in degeneration patterns 31

3 AIMS OF THE STUDY 33

4 MATERIALS AND METHODS 34
4.1 Study population and design 34
4.2 Magnetic resonance imaging acquisition and image analysis 35
4.2.1 Magnetic resonance imaging scanning protocols 35
4.2.2 Image analysis 36
4.3 Statistical analysis 39

5 RESULTS 41
5.1 Reliability of quantitative magnetic resonance imaging measurements (Study I) 41
5.2 SpineNet versus radiologist grading (Study II) 44
5.3 Longitudinal changes in disc degeneration (Study III) 48

6 DISCUSSION 51
6.1 Validity of T2-weighted nucleus pulposus signal intensity measurements (Study I) 51
6.2 SpineNet versus radiologist evaluation (Study II) 53
6.3 Progression of disc degeneration, assessed with Pfirrmann components (Study III) 55
6.4 Limitations and methodological considerations 57
6.5 Synthesis across Studies I–III 59
6.6 Future prospects 59

7 CONCLUSIONS 61

ACKNOWLEDGEMENTS 62

REFERENCES 64

ABBREVIATIONS

ADC – apparent diffusion coefficient
AF – annulus fibrosus
AI – artificial intelligence
ANOVA – analysis of variance
AUC – area under the ROC curve
BMI – body mass index
CCC – concordance correlation coefficient
CI – confidence interval
CNN – convolutional neural network
CSF – cerebrospinal fluid
CT – computed tomography
DD – disc degeneration
DH – disc height
DICOM – Digital Imaging and Communications in Medicine
DL – deep learning
DWI – diffusion-weighted imaging
gagCEST – glycosaminoglycan chemical exchange saturation transfer
HIZ – high-intensity zone
ICC – intraclass correlation coefficient
INC – intranuclear cleft
INHIZ – intranuclear high-intensity zone
IVD – intervertebral disc
κ – kappa statistic
LBP – low back pain
MMP – matrix metalloproteinase
MRI – magnetic resonance imaging
NP – nucleus pulposus
OR – odds ratio
ROC – receiver operating characteristic
ROI – region of interest
SD – standard deviation
SE – standard error
STIR – short tau inversion recovery
T1ρ – T1-rho
T2* – T2-star
T2WI – T2-weighted imaging
TNF-α – tumour necrosis factor-α
TR/TE – repetition time/echo time
UTE – ultrashort echo time

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I Murto N, Luoma K, Lund T, Kerttula L. Reliability of T2-weighted signal intensity–based quantitative measurements and visual grading of lumbar disc degeneration on MRI. Acta Radiol. 2023 Jun;64(6):2145–51.

II Murto N, Lund T, Kautiainen H, Luoma K, Kerttula L. Comparison of lumbar disc degeneration grading between deep learning model SpineNet and radiologist: a longitudinal study with a 14-year follow-up. Eur Spine J. 2025 Sep;34(9):3767–73.

III Murto N, Lund TE, Kautiainen H, Luoma K, Kerttula L. Progression of lumbar disc degeneration: a 14-year follow-up study examining Pfirrmann grading and its individual disc components. Acta Radiol Open. 2025 Sep;14(9).

The publications are referred to in the text by their Roman numerals. Studies I and III are reprinted by Permission of SAGE Publications. Study II is reprinted under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

1 INTRODUCTION

Low back pain (LBP) is the leading cause of disability worldwide and its overall burden both in disability and financial cost is expected to increase in the future decades (1–3). Lumbar disc degeneration (DD) is considered a significant contributor to chronic LBP (4–7). DD is a complex process that is characterised by progressive dehydration and structural breakdown of the intervertebral disc (IVD) structures, nucleus pulposus (NP) and annulus fibrosus (AF). DD is often accompanied by other morphological changes, such as disc height (DH) loss and endplate changes (8,9).

Magnetic resonance imaging (MRI) is the preferred imaging modality for DD due to its superior soft-tissue contrast, capability to image in multiple planes and sensitivity to disc structural changes without using ionising radiation. On MRI, DD is classically indicated by a decrease of NP T2-weighted signal intensity (reflecting reduced proteoglycan and water content and increased fibrosis) and changes in disc morphology (10–12). Although numerous studies report associations between DD and LBP, the relationship is heterogeneous because degenerative features are also observed in asymptomatic individuals (4,13). Given the large field of spine research where MRI-based DD is frequently used as a predictor, accurate assessment of DD on MRI is of high research importance.

Qualitative grading systems have been developed to standardise the evaluation of DD on MRI (14). The most widely used is the Pfirrmann classification introduced in 2001 (11). Pfirrmann grades are defined based on NP and AF characteristics, as well as DH on sagittal T2-weighted MRI. The Pfirrmann classification is known for its high inter- and intra-observer reliability, ranging from substantial to almost perfect (11,15), as well as for being straightforward and easy to use. However, while the Pfirrmann classification offers a broad overview, it compresses a continuous degenerative process into five grades. Because it is a qualitative grading system, it is subjective, which means that there are often inconsistencies at grade boundaries. It has been suggested that the Pfirrmann classification may not detect early or advanced stages of degeneration or subtle structural changes in the discs (16–21).

There is a need for more objective and sensitive grading of DD. Quantitative MRI analysis offers one avenue for improvement. Rather than relying solely on visual interpretation, quantitative approaches measure MRI parameters within the disc. Prior studies have explored measuring the signal intensity of the NP on T2-weighted images and normalising it to cerebrospinal fluid (CSF) (22–26), adjacent vertebral bone marrow (27,28) or the spinal cord (29,30) to reduce

inter-scan variability. Other studies have used advanced MRI techniques, such as T2 mapping (quantifying actual T2 relaxation values) or T1ρ (proteoglycan content of the disc), to evaluate degeneration stages. These advanced techniques have shown promising results in detecting early biochemical changes not seen on conventional MRI (31–34). A quantitative approach could potentially provide continuous metrics that reflect the disc's structural composition and offer better distinction of DD than broad ordinal grades. However, there is a challenge to ensure that these measurements are reproducible and truly reflective of degeneration rather than technical variability (34,35). There is no universally accepted quantitative method for assessing DD on MRI, and studies have reported various methodologies without a consensus standard. Establishing a standardised and accessible quantitative measure could also facilitate artificial intelligence (AI) applications by providing additional features and ground-truth labels for AI algorithms.

The advent of AI and deep learning in medical imaging has also sparked interest in the automated grading of DD. AI models can be trained on large image datasets to identify features of DD on MRI (36,37). One notable development was by Jamaludin et al. (38), who created SpineNet, a multi-task convolutional neural network (CNN) that automatically labels several lumbar MRI features, including Pfirrmann grades. Their model has demonstrated performance nearly comparable to that of human radiologists, accurately grading DD in 55–79% of Pfirrmann grades compared to radiologists in external validations (39–41). Such AI tools could potentially offer a fast, reproducible analysis of DD without inter-observer variability, which is particularly valuable in large-scale epidemiological studies. However, applying these models across different MRI datasets requires caution. AI systems must be continuously refined and thoroughly validated in heterogeneous datasets before they are deployed for the routine assessment of DD in research settings. Furthermore, most existing models have been developed and tested using cross-sectional data; their performance in longitudinal assessments has remained less explored.

Another dimension of DD research is understanding the course of DD. Cross-sectional studies have clearly shown that degenerative changes increase with age (4,42), but individuals vary widely. Studies show that while many individuals exhibit distinct progression of DD during ageing, others of similar age show little or no change in degeneration (43–45). DD is known to be a multifaceted process in which the NP and AF may deteriorate through different mechanisms and at varying rates (9,46,47). Longitudinal cohort studies are relatively rare because they require decades of follow-up. Twin studies have demonstrated that DD is predominantly genetic; familial factors explain most of the variance in both baseline severity and progression, while environmental

exposures such as occupational loading or smoking also affect the course of DD (48–51).

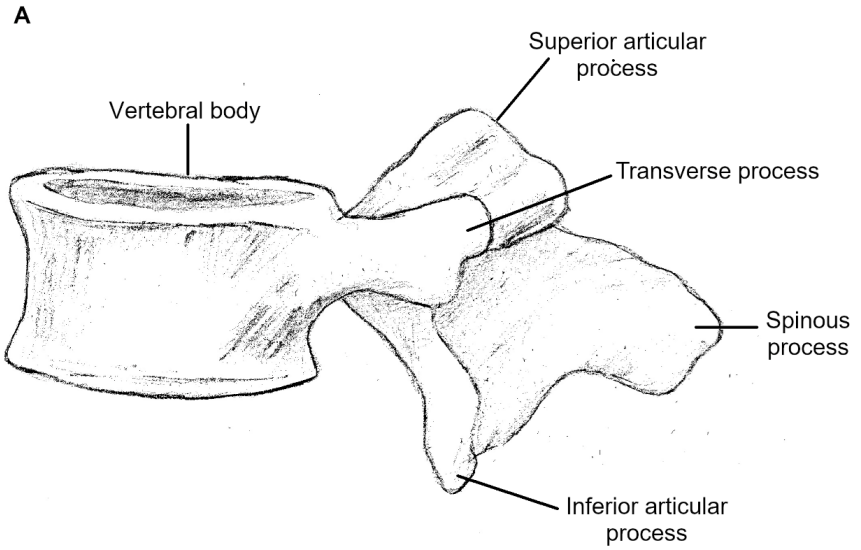
In summary, given the limitations of current DD grading and the lack of standardised quantitative metrics, MRI-based assessment of DD needs to be improved. The specific aims of the present thesis were as follows:

- to establish how consistent and informative NP T2 signal intensity measurements can be, compared to traditional visual grading (addressed in Study I)
- to validate AI-driven grading tools in a longitudinal dataset by comparing the deep learning model SpineNet with radiologists' grading (Study II)
- to characterise degeneration progression in detail: using a 14-year follow-up of 37- to 51-year-old individuals to map changes in Pfirrmann grades and disc components (NP and AF), thereby demonstrating what Pfirrmann grades do and do not capture about DD (Study III)

2 REVIEW OF THE LITERATURE

2.1 Anatomy of the lumbar spine and the intervertebral disc

The lumbar spine is composed of five lumbar vertebrae (L1–L5) at the lower part of the spine above the sacrum. Vertebrae articulate with each other with intervertebral discs (IVDs) and paired facet joints. Each vertebra is comprised of an anterior vertebral body, which carries most of the compressive axial loads, and posterior elements — including the pedicles, laminae, transverse processes, articular processes (superior and inferior) and spinous process (Figure 1). Together, the anterior and posterior vertebral elements form the vertebral arch. This arch forms the vertebral canal surrounding the spinal cord and provides broad surfaces for the attachment of deep back muscles and ligaments. The stability of the spine is maintained by a network of ligaments and muscles. The anterior and posterior longitudinal ligaments run along the front and back of the vertebral bodies, reinforcing the anterior part of the spine. Posteriorly, stability is provided by the ligamentum flavum, which connects adjacent laminae; the interspinous and supraspinous ligaments, which connect the spinous processes; and the intertransverse ligaments, which connect the transverse processes. Together with the paraspinal muscles, these ligaments work together to constrain excessive motion, support an upright posture, and protect the spinal cord and nerve roots (52).



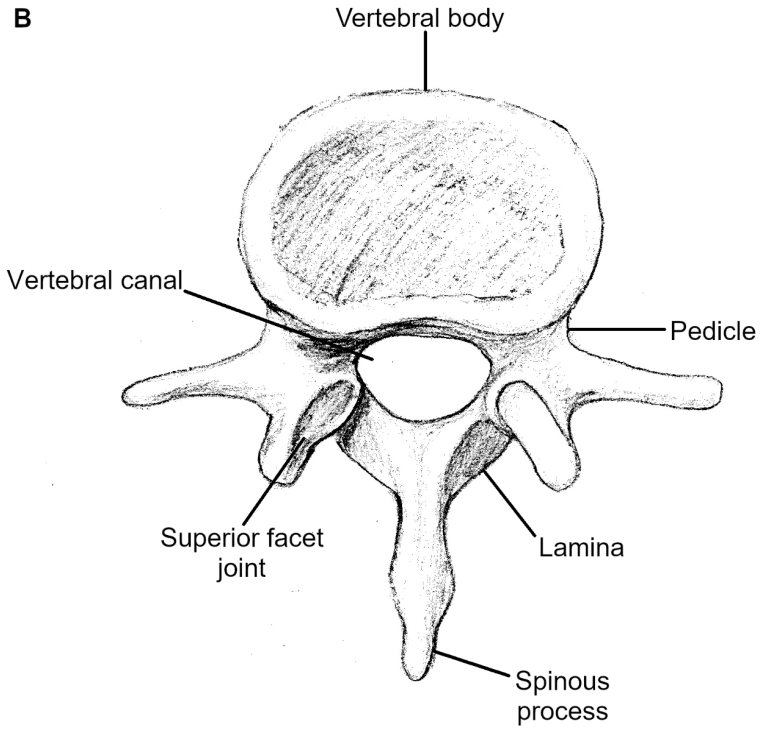


Figure 1 Anatomy of the lumbar vertebral column in the (A) lateral and (B) superior view. Illustration drawn by the author.

The intervertebral disc (IVD) is a crucial structural element of the spine, located between adjacent vertebral bodies (Figure 2). It consists of two main components: a gelatinous nucleus pulposus (NP) at the centre and a surrounding fibrocartilaginous annulus fibrosus (AF). The NP has a high water content and behaves like a hydrostatic gel under compressive loads, which allows it to absorb and redistribute forces, resist compression, and act as a shock absorber between vertebrae. The NP is surrounded by the AF, which is composed of concentric layers (lamellae) of collagen fibres. The collagen fibres in each lamella are oriented at alternating angles, which gives the AF high tensile strength in multiple directions. This lamellar structure enables the AF to resist tension and shear forces, maintaining the integrity of the disc during bending and rotation movements (53). Each IVD is attached to the adjacent vertebrae by thin cartilaginous endplates, which also serve as the pathway for nutrients to diffuse into the avascular NP (54).

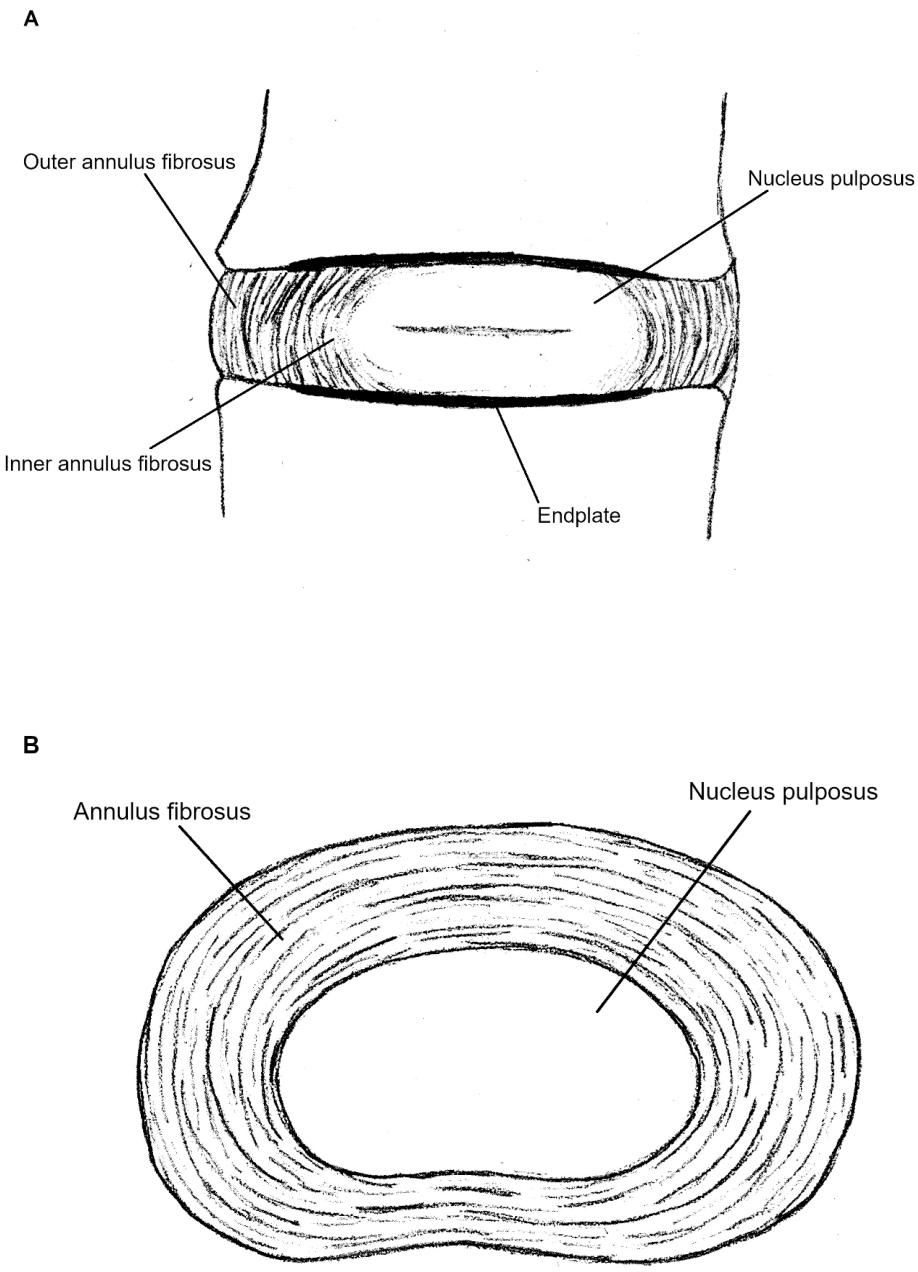


Figure 2 Anatomy of the intervertebral disc in the (A) lateral and (B) superior cross-sectional view. Illustration drawn by the author.

Posteriorly, each vertebra articulates with the vertebra above and below through a pair of facet (zygapophyseal) joints, which are synovial joints formed by the inferior articular processes of the upper vertebra and the superior articular

processes of the lower vertebra. These joints enable the controlled movement of the spine. In the lumbar region, the facet joint surfaces are mostly oriented in the sagittal plane, which permits significant flexion–extension movement but limits axial rotation and resists anterior–posterior shear between vertebrae that could damage the neural structures and IVDs (55). Together with the IVD, facet joints comprise a functional complex called a spinal motion segment (Figure 3), in which injury to one component increases stress on the others and may lead to the failure of the whole segment (56).

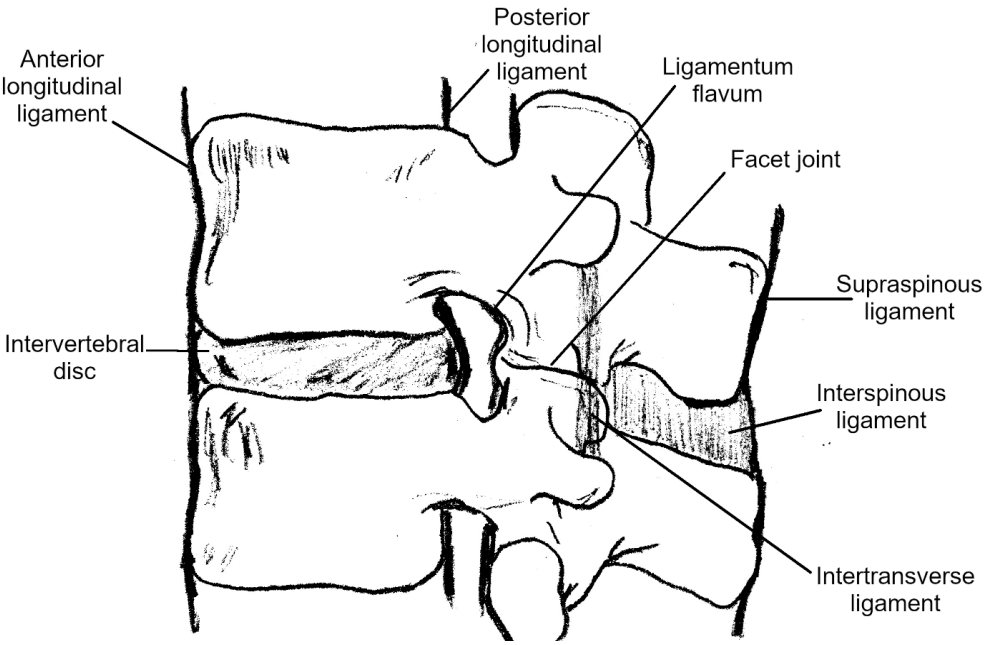


Figure 3 Anatomy of the lumbar spinal motion segment and ligaments. Illustration drawn by the author.

The biomechanics of the lumbar spine are characterised by its ability to accommodate flexion–extension, lateral bending, limited axial rotation, and load-bearing capacity, which are essential for upright posture. While permitting motion, the spine must also provide stability and protect the neural elements. In a neutral standing position, axial loads are mostly transmitted through the anterior column, vertebral bodies and IVDs. The distribution of load changes depending on the posture: posterior elements, including the facet joints, contribute to load sharing, particularly during extension and rotation (53,57).

2.2 Intervertebral disc degeneration

2.2.1 Normal disc composition

The nucleus pulposus (NP) contains a low number of cells within a water-rich, gelatinous extracellular matrix, which is rich in proteoglycans — primarily aggrecan — and type II collagen (58). Aggrecan contains highly sulphated glycosaminoglycan chains (chondroitin and keratan sulphate) that attract and hold water. This contributes to the high water content of the NP — up to 80–90% in healthy young individuals (59). This hydration generates hydrostatic pressure that allows the NP to act as a hydrostatic load-bearing core and the disc to resist compressive loads (60).

The annulus fibrosus (AF) is composed of concentric lamellae of collagen fibres that are arranged in alternating oblique orientations, forming a criss-crossed architecture that resists multidirectional forces (61). These collagen lamellae are separated by elastic fibres and an interlamellar matrix rich in proteoglycan. Elastin fibres contribute to shear resistance and help the annulus return to its original shape after bending (62). The AF surrounds the NP and restrains its outward expansion under axial loading, thereby converting compressive forces into circumferential tensile stress within the lamellae. Histologically and functionally, the outer and inner AF differ. The outer AF consists mostly of fibroblast-like cells and dense type I collagen, providing high tensile strength. In contrast, the inner AF is more fibrocartilaginous, with relatively higher proteoglycan and water content and number of chondrocyte-like cells, forming a progressive transition to the proteoglycan-rich NP and aiding in the transfer of the compressive load from the pressurised NP (8,62,63). In non-degenerated discs, only the outer AF is innervated, while the inner AF and NP have no nerve supply (64).

The cartilaginous endplate is a thin layer of hyaline cartilage that separates the disc from the vertebral body while allowing nutrient exchange. During childhood, the IVD, including the NP, is supplied via vascular channels that traverse the cartilaginous endplates (54). With skeletal maturation, these channels close, and the NP, together with the inner AF, become avascular, while only the outer AF retains sparse peripheral vascularity. After that, nutrients reach disc cells solely by diffusion from capillaries in the bony vertebral endplates (65). During daily activity, the disc undergoes diurnal fluid fluctuations: loading (e.g., upright posture) expels some water from the NP, whereas unloading (e.g., bed rest) allows water to be reabsorbed. This diurnal cycle of load-induced fluid transport may also affect nutrient transport (54).

2.2.2 Structural and biochemical changes in disc degeneration

Degeneration usually begins in the NP with a decrease in proteoglycan content, leading to the decline of hydration and compromising its ability to resist compressive loads (8,66). These changes are generally most evident at the lower lumbar levels, which tend to degenerate earlier than upper levels (45,67). The impaired nutrient supply contributes to the onset of DD. Because of the limited oxygen supply, disc cells rely predominantly on anaerobic glycolysis. At low oxygen tension, synthesis of matrix macromolecules such as sulphated glycosaminoglycans is markedly reduced, promoting matrix breakdown. Continuous glucose consumption by cells leads to lactic acid production. These lactic acids accumulate in the disc and lower the extracellular pH, making the environment in the disc acidic. Acidic conditions further suppress cellular energy production and matrix synthesis. Importantly, adequate glucose is a key determinant of cell survival, and even very low oxygen can be tolerated if the glucose supply is sufficient. When glucose becomes inadequate or the pH falls to very low levels, cells may die (54,65,68). Thus, any factor that hinders nutrient delivery can compromise disc cell function and potentially initiate degenerative changes in the disc matrix (54,65).

With age, the cell population of the NP also transforms. In early life, the NP consists mostly of large, vacuolated notochordal-like cells, which are gradually replaced by smaller, chondrocyte-like cells in adulthood. This cellular transition reduces water-holding matrix components, lowering NP hydration and pressure (8,69).

Degenerating discs also show a decrease in matrix build-up and an increase in breakdown, with increased activity of catabolic enzymes, including matrix metalloproteinases (MMP) and aggrecanases (12,70). These catabolic enzymes break down essential extracellular matrix components, including aggrecan and collagen, thereby accelerating matrix disintegration and contributing to further reduction of disc hydration. In parallel, NP cells, which are typically few in number and resemble chondrocyte cells, often proliferate and form clusters in degenerated discs (70,71). These cell clusters produce proinflammatory cytokines, including interleukins (e.g., interleukin-1α, interleukin-1β, interleukin-6, interleukin-8, interleukin-10) and tumour necrosis factor alpha (TNF-α), which promote further matrix breakdown and create a self-sustaining cycle of inflammation and degradation that accelerates degeneration (62,72,73). Collagen turnover also shows a characteristic shift. Aggrecan and type II collagen synthesis decline, while denatured type II collagen accumulates and type I collagen synthesis increases in the NP and inner annulus, which leads to gradual fibrotic transformation with advancing degeneration (66). As degeneration advances, clefts and fissures form in the NP and spread outwards, eventually leading to collapse and potentially cavity formation (74).

In the AF, degeneration is characterised by lamellar delamination, inward buckling of inner fibres, and fissuring, which permits ingrowth of nerve fibres and small blood vessels into regions normally poorly innervated and largely avascular (75–77). Annular tears may also lead to faster NP degeneration (46). In early life, the NP and inner AF are more clearly distinct structures. During degeneration, in addition to fissuring, myxoid degeneration of the AF occurs, which leads to a less distinctive border between NP and AF (12,77).

The structural consequences of these biochemical and cellular events include progressive fibrosis of the NP and disorganisation of the AF. In the NP, fibrosis is associated with fissure formation, reduced water content and a decrease in the size of NP and overall volume and height of the disc (78). The NP becomes more fibrous and less gel-like and the disc loses its ability to distribute stress (62,79). As a result, load sharing moves to the AF and facet joints, which causes abnormal joint loading and progressive zygapophyseal arthropathy, which in turn may accelerate spinal segment degeneration (80,81). As the disc loses its height and stress distributions in the disc change, the disc's ability to distribute loads evenly deteriorates, increasing its vulnerability to bulging and herniations (8).

During degeneration, endplates may develop cracks and fissures that disrupt nutrient supply and allow nuclear material to herniate into the vertebral body. In addition to impaired nutrient transport, these cracks provide pathways for vascular and nerve ingrowth, thereby contributing to DD progression (71,82). Advanced degeneration may also include endplate sclerosis, which further reduces nutrient diffusion through endplates, further worsening cell dysfunction and matrix degradation (8,12) (Figure 4).

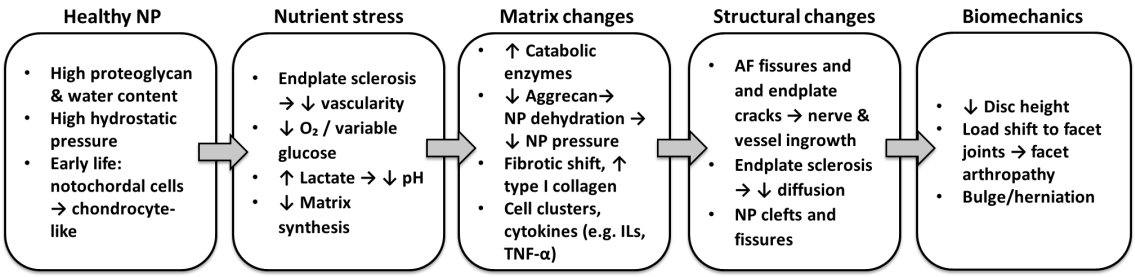


Figure 4 Simple pathway of disc degeneration. AF – annulus fibrosus, ILs – interleukins, NP – nucleus pulposus, O_2 – oxygen, TNF-α – tumour necrosis factor alpha.

2.2.3 Risk factors for disc degeneration

Risk factors for lumbar DD arise from both genetic predisposition and cumulative environmental or lifestyle factors. Genetic inheritance has a significant role

in lumbar DD development (83,84). At the same time, heavy physical loading of the spine — for example, years of manual labour or high-impact sports — can increase degenerative changes in the discs. The continuous compressive forces, torsional stress, and micro-traumas associated with heavy lifting or high-impact activities can damage disc structures over time (85,86). In addition to genetics and physical loading, various lifestyle habits also play a critical role in lumbar DD risk. Tobacco smoking has been consistently identified as a risk factor for accelerated DD. Smoking impairs blood flow and nutrient delivery to the IVDs (87). Obesity significantly increases the mechanical stress on the lumbar spine and is associated with a higher incidence of DD (83,88). Systemic conditions such as diabetes and metabolic syndrome may predispose individuals to multilevel DD, likely by inducing chronic low-grade inflammation (89). An initial injury to the AF that results in micro- or macrocracks is regarded as a primary trigger capable of initiating a degenerative cascade (90–92).

2.2.4 Association of disc degeneration with pain

The association between IVD degeneration and low back pain (LBP) has been demonstrated in several large-scale cohort studies across diverse populations (5,42,93). Nonetheless, the relationship is not straightforward. Not all degenerative findings or stages of degeneration are associated with pain (94,95). In general, more advanced degeneration shows stronger associations with LBP than early changes (94). Psychosocial factors, such as mental distress, can intensify the experience of LBP (6).

Discogenic pain arises from matrix breakdown, annular fissures, and inflammatory activation in the IVD (7,96). Loss of aggrecan and hydrostatic pressure, together with annular fissures and endplate cracks, permit nerve and vessel ingrowth into the inner annulus and the NP. This contributes to nociceptor sensitisation and discogenic pain. Inflammatory cytokines (e.g., TNF-α and interleukins) and neurotrophins (e.g., nerve growth factor, brain-derived neurotrophic factor) produced by disc cells and infiltrating macrophages increase the expression and sensitivity of nociceptors and their ion-channel mediators in these reinnervated tissues. Acidic, low-oxygen conditions within the disc further sensitise these nerves (97–99).

Endplate damage also provides a nociceptive source in DD. Defects in the cartilaginous and bony endplates allow leakage of NP material and inflammatory cytokines (e.g., TNF-α and interleukins) into the vascularised vertebral marrow, provoking reactive bone-marrow lesions with an ingrowth of sensory nerve fibres that are susceptible to chemical sensitisation and mechanical stress. These endplate marrow reactions correspond to Modic type 1 inflammatory change adjacent to degenerating discs and which has

been linked with pain (100). Decrease in disc height (DH) moves the axial load posteriorly onto the facet joints and produces osteoarthritic changes that can generate LBP provoked by extension and rotation. Disc herniations can compress the adjacent dorsal nerve roots and lead to radicular pain. More important than the mass effect, herniated NP irritates nerve roots through the release of inflammatory cytokines — especially interleukin-8 — which can sensitise nociceptors in the nerve root/dorsal root ganglion (101–103). Despite these associations, there are no clear morphological or biochemical criteria that can reliably tell the difference between normal age-related disc changes and pathological degeneration (62,70).

2.3 Imaging of lumbar disc degeneration

2.3.1 Radiographic imaging modalities

Plain radiography is used to visualise the bone structures. DD may be detected from plain radiography indirectly through secondary signs such as disc space narrowing and calcification, osteophyte formation, and endplate sclerosis. Its main advantage over other modalities is that it can be performed in the standing position, enabling evaluation of spinal alignment, mobility, and instability (104). Consequently, it remains the standard method for assessing the alignment of the spine in the standing position (105).

Computed tomography (CT) offers high spatial resolution of bone structures and allows for the assessment of bony endplate irregularities, vacuum phenomena, and intradiscal calcification. It is also useful for detecting an osseous spinal canal or foraminal stenosis and facet joint arthrosis. However, CT has poor soft tissue contrast and provides little information about disc composition (106).

Discography is a fluoroscopically guided procedure in which contrast medium is injected into the disc to evaluate morphology and provoke pain symptoms. It can delineate fissures in the NP and AF with high precision and has historically served as a reference standard for identifying painful discs. Nevertheless, discography is now rarely used due to false-positive pain responses even in asymptomatic individuals, the risk of post-procedural discitis and evidence that disc injection may accelerate DD (107–109).

All of these techniques rely on ionising radiation, which limits their use in population-based studies, particularly in younger individuals.

2.3.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is the gold standard for evaluating lumbar DD due to its good soft tissue contrast and ability to visualise structures without ionising radiation. The ability of MRI to acquire images in multiple planes allows for a comprehensive assessment of the lumbar spine. Standard lumbar spine MRI protocols include T2-weighted sequence in the sagittal and axial planes and T1-weighted sequence in the sagittal plane, typically with a slice thickness of 3 mm. In addition, T2-weighted fat-suppressed or short tau inversion recovery (STIR) sequences are used to detect bone marrow oedema and to identify inflammatory and infectious processes. Contrast-enhanced T1-weighted fat-suppressed imaging is mostly used for the evaluation of infectious or malignant diseases and postoperative complications.

On T2-weighted images (T2WIs), healthy NP tissue appears hyperintense (bright white) because of its high water and proteoglycan content, whereas a degenerated NP shows a reduced signal intensity because of its decreased water and proteoglycan content and increased fibrosis. This decrease in NP signal intensity has been shown to correlate with the biochemical decrease in proteoglycans and water in the disc (10,66,110). An adult NP may also show a horizontal intranuclear cleft (INC), a hypointense band seen on sagittal T2-weighted imaging (T2WI), which represents a normal, age-related fibrous structure (76). The AF, which is normally dark on T2WI, may also lose its sharp distinction from the NP as degeneration progresses. These characteristic MRI changes form the basis of most grading systems for lumbar DD (11). Most lumbar DD grading systems on MRI are based on the signal intensity of the NP on T2WI (11,12,18,111–120). Additionally, some systems incorporate combinations of NP signal intensity with disc height (DH), the distinction of the border between the NP and the AF (11,18,118), and disc shape features, such as bulge and herniations, and Schmorl nodes (12,111–115,117–120).

2.3.3 Pfirrmann classification

The Pfirrmann classification (11) (Figure 5) is widely used for research on DD. It was developed based on previously published literature (121–123). It categorises lumbar IVDs on mid-sagittal T2WIs into five categories:

Grade I: The structure of the disc is homogeneous, with a bright hyperintense white signal intensity and normal DH.

Grade II: The structure of the disc is inhomogeneous, with a hyperintense white signal. The distinction between the nucleus and the annulus is clear, and the DH is normal, with or without horizontal grey bands.

Grade III: The structure of the disc is inhomogeneous, with intermediate grey signal intensity. The distinction between the nucleus and the annulus is unclear, and the DH is normal or slightly decreased.

Grade IV: The structure of the disc is inhomogeneous, with hypointense dark grey signal intensity. The distinction between the nucleus and the annulus is lost, and the DH is either normal or moderately decreased.

Grade V: The structure of the disc is inhomogeneous, with hypointense black signal intensity. The distinction between the nucleus and the annulus is lost, and the disc space is collapsed.

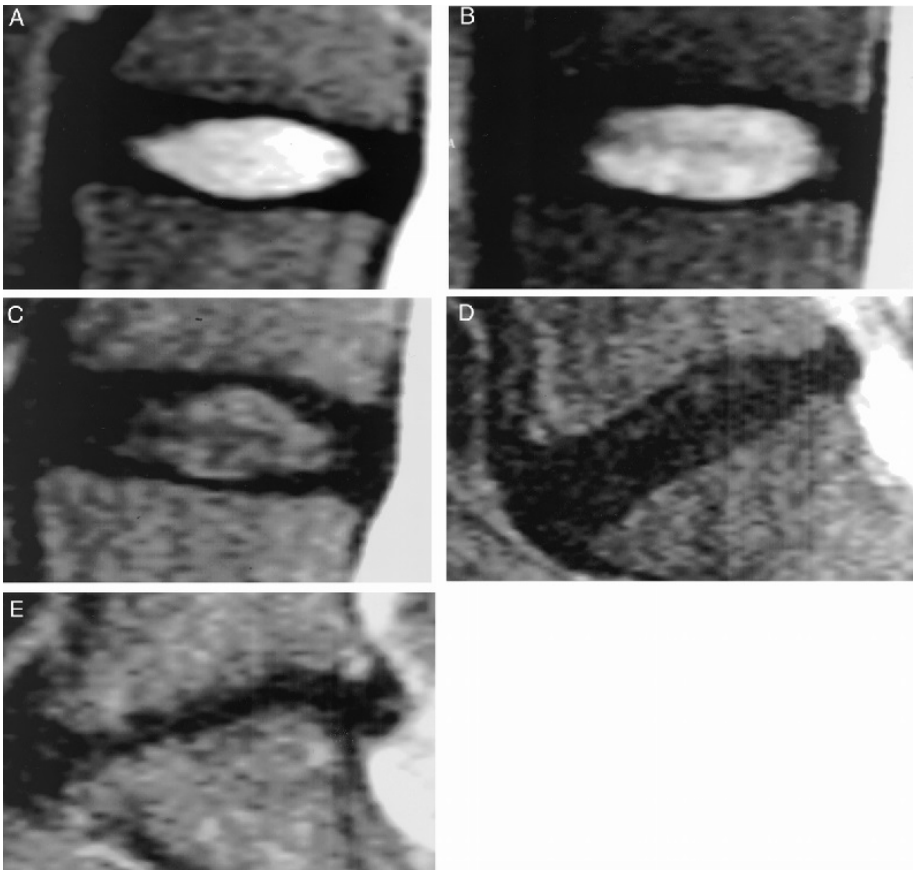


Figure 5 Pfirrmann grading (I–V) of lumbar disc degeneration on T2-weighted magnetic resonance imaging. Images correspond to grades I (A), II (B), III (C), IV (D), and V (E). Reprinted from: Pfirrmann CWA, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine. 2001;26(17):1873–1878. © 2001 Wolters Kluwer Health, Inc. Reprinted with permission from Wolters Kluwer Health, Inc. (License no. 6107050258124).

The Pfirrmann classification has gained popularity due to its simplicity and high inter- and intra-observer reliability (11,15). Several alternative DD gradings and modifications to the Pfirrmann classification have been proposed, but Pfirrmann's original five-grade classification remains the most commonly used in spine studies (14). Despite its utility, several concerns have been raised about the limitations of the Pfirrmann classification (124). It has limited sensitivity to early DD (12,20,21), a mildly degenerated, bright disc can be graded I or II. Conversely, in advanced age populations, many discs become grade IV or V, which compresses the range of differentiation and may reduce sensitivity to change in longitudinal studies (18). There can also be subjectivity at grade boundaries — for example, distinguishing a grade II from III, or III from IV, can be challenging (16,17,125). In addition to grading individual discs, the literature also uses sum scores, where the DD grades of all lumbar discs — e.g., Pfirrmann grades — are added together, instead of grading individual discs. This provides a single value that reflects the overall lumbar degeneration in that individual (126). However, such scores do not distinguish individuals with advanced degeneration limited to one or two levels from those with milder, multilevel changes.

2.3.4 Other degenerative features on magnetic resonance imaging

MRI also reveals degenerative features beyond the Pfirrmann criteria, many of which co-occur with DD and may be clinically relevant. Annular fissures can appear as focal T2-hyperintense high-intensity zones (HIZ) within the AF (127) and are more frequently reported in patients with LBP than in asymptomatic individuals (128). Disc bulges and herniations mean the displacement of disc material beyond the margins of the IVD space. Disc herniations are further classified as protrusion, extrusion, and sequestration (127). Disc herniations may cause spinal stenosis and may compress and irritate adjacent nerve roots and result in radiculopathy, but the clinical significance of disc herniations depends on the size and location (129,130). If disc material herniates through a defect in the vertebral endplate into the body of the vertebra, it is called an intravertebral herniation, commonly known as a Schmorl node. Vertebral endplate bone marrow signal changes (Modic changes) are also a common MRI finding associated with DD. These changes are seen in the vertebral bone marrow adjacent to degenerating discs and categorised as type 1 (bone marrow oedema), type 2 (fatty marrow), and type 3 (subchondral bone sclerosis) (131). Modic changes often occur together, most commonly as a mixed pattern of Modic types 1 and 2 (132). Notably, Modic type 1 endplate oedema has been strongly linked to inflammatory endplate changes and LBP (133).

Facet joint degeneration commonly coexists with DD. On MRI, degenerative facet changes may appear as facet hypertrophy, cartilage thinning, subchondral cysts or sclerosis, and joint effusion. Facet osteoarthritis can contribute to LBP and may further narrow the intervertebral foramina, worsening nerve compression (134). In advanced cases, degeneration of the facet joints and discs together can lead to degenerative spondylolisthesis, where one vertebra slips forward relative to the level below, possibly contributing to nerve root compression (135).

2.3.5 Quantitative magnetic resonance imaging of disc degeneration

MRI not only enables qualitative grading (e.g., Pfirrmann grade) of DD but also offers quantitative techniques to objectively assess the disc's composition and structure. Traditional T2WIs provide a visual insight into disc hydration and morphology, but their interpretation is subjective and can vary between observers. In response, a range of quantitative MRI methods has been developed to characterise the structural and biochemical changes of lumbar DD more objectively. These include simple metrics like measuring DH (22,136) or normalised T2-weighted signal intensity of the NP (22,24,48,107,137), as well as advanced imaging sequences such as T2 relaxation time and T1ρ (T1-rho) mapping (31,32), diffusion-weighted imaging (DWI) (25,138,139), and glycosaminoglycan chemical exchange saturation transfer (gagCEST) (140–142). Collectively, these approaches provide complementary information about disc structure and composition with the goal of improving assessment consistency and sensitivity to detecting change in degeneration. Quantitative imaging techniques aim to supplement or replace subjective grading with objective data.

Broadly, quantitative MRI methods for DD evaluation can be divided into two approaches: (a) analysing conventional MRI images (usually T2-weighted) by measuring signal intensities or textures quantitatively, and (b) using advanced MRI sequences (T2 mapping, T1ρ, DWI, gagCEST) that produce direct quantitative maps of tissue properties.

2.3.5.1 T2-weighted signal intensity

A pragmatic quantitative approach is to measure the T2-weighted signal intensity of the disc, most often the NP, on mid-sagittal T2WI. Degenerated discs show a decreased NP T2 signal (a darker nucleus) due to loss of water and proteoglycans, and this change can be quantified (143,144). However, MRI signal intensity is an arbitrary unit representing the grey-level values

of the image, not a standardised value. Signal intensity is not only influenced by tissue properties but depends on the scanner and acquisition settings, so normalisation is essential for comparison of the signal intensity values. A common approach is to express the NP signal as a ratio to the cerebrospinal fluid (CSF) signal on the same image. CSF in the thecal sac provides a convenient internal reference that helps compensate for scanner and acquisition differences (23,25,145). However, factors like magnetic susceptibility, radiofrequency field inhomogeneity, pulsatile motion and flow artefacts can alter the apparent signal intensity of fluids such as CSF, introducing variability that may cause bias to the signal intensity measurements (146,147).

Other reference tissues have also been used — for example, adjacent lumbar vertebral body marrow (27–30). Some studies have found that the spinal cord provides a more consistent reference signal for disc intensity measurements, avoiding CSF flow issues (29,30). Beyond single-reference ratios, multivariable normalisation accounting for age, CSF signal, and disc level can improve the correlation of the quantitative disc signal measures with the radiologists' visual ratings and produce more normally distributed values (148).

Region of interest (ROI) selection for measuring signal intensity of the disc is also critical for the reliability of the measurements (149,150). Typically, disc signal intensity is measured on mid-sagittal T2WI by defining a ROI within the disc (26). The size and placement of this ROI can substantially influence the measurement. A common method is to use an ellipsoidal ROI covering the region of the NP on the mid-sagittal slice while avoiding the AF and endplates (25,149,151). Other approaches include measuring the whole disc area on the mid-sagittal slice (22,152) or drawing regional ROIs on axial images to sample the NP and AF separately (153,154). Including low-intensity structures like the collagenous outer AF, fibrotic structures such as the INC, and bony endplates into the ROI will lower the average signal of the disc and can obscure differences between healthy and degenerated discs.

In addition to simple ROI averages, histogram analysis of mid-sagittal T2WI is used to measure the signal intensity peaks of NP and AF to quantify their difference. A decreased difference between the peak intensities refers to increased degeneration. Histogram analysis may improve the objectivity of the DD analyses when only conventional images are available (155,156).

Overall, normalised NP T2 signal intensity metrics have demonstrated good correlation with qualitative degeneration grades (22,137). It should be noted, however, that T2-weighted signal intensity is not a true physical parameter but reflects image brightness. Thus, although an NP/CSF intensity ratio can serve as an indirect marker of disc water content and correlates with degeneration, it lacks the absolute consistency of disc composition such as the T2 relaxation time measurements. Small differences in the MRI sequence or

ROI placement could affect the signal intensity values. For this reason, although signal intensity–based measures are useful, especially for studies in which only conventional images are available, their interpretation requires proper normalisation and standardised ROI definition.

Some post-processing techniques have been explored to enhance the contrast between signal intensity measurements of the NP and AF on T2WI. For example, Eldaya et al. (157) introduced a custom post-processing algorithm that selectively enhances voxels within an intensity range corresponding to the inner AF on routine T2WI. By increasing the contrast-to-noise ratio between the inner AF and the NP, such methods may assist in defining the NP boundary and quantifying NP size or heterogeneity.

2.3.5.2 Advanced MRI techniques

T2 mapping is an MRI technique that measures the T2 relaxation time of tissues on a voxel-by-voxel basis, producing a quantitative map. In the context of IVD, the T2 relaxation time relates to water and collagen content, similarly to T2-weighted signal intensity. In a healthy young NP, T2 relaxation time is relatively high, reflecting high water and proteoglycan content. With degeneration, the NP loses proteoglycans and water, and its collagen network becomes more fibrous, and these changes shorten the T2 relaxation time (158). Quantitative T2 maps thus reflect the biochemical status of the disc. The regions of short T2 relaxation time indicate low hydration or high collagen density (e.g., fibrotic tissue), whereas long T2 relaxation time corresponds to well-hydrated tissue like a proteoglycan-rich matrix (33). Studies have shown strong associations between T2 relaxation times and conventional degeneration grading. For instance, the Pfirrmann grade correlates inversely with T2 values of the NP, discs with increasing grades of degeneration exhibit progressively shorter T2 relaxation times than the less degenerated discs (143,159,160). One important advantage of T2 mapping is that it can reveal signs of early degeneration (a slight shortening of the T2 relaxation time) before clear morphological changes appear on the T2WI (154).

Beyond conventional T2 mapping, **T2* (T2-star)** can be obtained using a multi-echo gradient-echo sequence. Unlike T2 mapping, which uses refocusing pulses to minimise the effects of magnetic-field inhomogeneities and therefore mainly reflects the true T2 relaxation, T2* also includes dephasing effects from local magnetic-susceptibility differences. As a result, T2* values are generally shorter than T2 and are influenced not only by hydration and collagen content but also by microstructural features such as collagen alignment, mineralisation, and other magnetic susceptibility variations (161). The gradient-echo acquisition allows shorter scan times than conventional

multi-echo spin-echo T2 mapping (33). In the IVD, T2* mapping has shown strong associations with glycosaminoglycan content (162).

T1ρ imaging measures the spin–lattice relaxation in the rotating frame. The T1ρ sequence applies a special spin-lock pulse that locks magnetisation and makes the signal relaxation sensitive to slow-motion interactions between water molecules and their local macromolecular environment. T1ρ mapping is particularly sensitive to proteoglycan content because the spin-lock pulse emphasises interactions between the free water and the large molecules like proteoglycan. In the young IVD, a proteoglycan-rich NP exhibits a prolonged T1ρ relaxation time, whereas with degeneration, the decrease of proteoglycan content leads to shortened T1ρ values (163). Numerous studies have underscored the value of T1ρ in disc assessment (32,33,164). T1ρ values have been shown to strongly correlate with Pfirrmann grades of degeneration (33). Importantly, T1ρ mapping can also reveal early biochemical changes in the disc matrix like proteoglycan depletion that precedes gross structural changes. T1ρ may drop in the NP as a result of the decreased proteoglycan content before the T2-weighted signal intensity decrease is evident on routine MRI (163).

Studies have presented conflicting findings regarding the relative sensitivity of MRI biomarkers in the early detection of DD. Some studies noted that T1ρ appeared more dynamic and had a wider range across degeneration than T2 maps, suggesting it might be more sensitive in early degeneration (165,166). In contrast, other work suggests T2 mapping often outperforms T1ρ in routine settings, with greater practicality, reproducibility, and strong performance for early collagen alterations (33). From a biological standpoint T1ρ's advantage is its specificity to proteoglycans — these molecules start decreasing early in DD, even before a significant decrease in water content or reorganisation of collagen content. T2, on the other hand, is influenced more by the content of unbound water and collagen, which change a bit later in the degenerative cascade.

Another application of T1ρ imaging is in identifying symptomatic 'painful' discs. Some studies have investigated whether quantitative MRI can distinguish discs that are sources of chronic LBP. In a prospective cohort that underwent provocative discography, painful levels demonstrated lower T1ρ than asymptomatic or non-painful levels (167). A case–control study also found that T1ρ alone separated painful from non-painful discs, and performance improved further when combined with a DH analysis (168). This may be explained by T1ρ shortening when the proteoglycan-rich, pressurised NP becomes depleted and dehydrated, and the matrix becomes more fibrotic, which often coexist with annular fissuring, nerve ingrowth, and inflammatory acidity. This altered macromolecular environment accelerates spin-lock relaxation — producing the low T1ρ signal (166–168).

Quantitative studies consistently show that T1ρ and T2 relaxation time decline with degeneration but exhibit little additional ability to differentiate between end stages of DD (Pfirrmann grades IV and V). T2* shows mixed evidence at this transition — some data report the ability to differentiate between grades IV–V, while others do not. Consequently, no single metric of T1ρ or T2/ T2* mapping seems uniformly superior in differentiation between end-stage DD. T1ρ and T2 might be more informative in differentiation between the earlier stages (lower grades) of DD, while T2* may be better between the later stages (higher grades) of DD, but not reliably across studies (33,169–171).

Diffusion-weighted imaging (DWI) assesses the mobility of unbound water molecules within tissues and quantifies it by the apparent diffusion coefficient (ADC). In the disc, water molecule movement is restricted by the dense extracellular matrix of proteoglycans and collagen. If the diffusion time allowed in the MRI sequence is sufficiently long, water molecules in the NP collide with collagen fibres and molecules that restrict their motion. As a result, the ADC measured in NP tissue is lower than that of free water. Studies have reported that degenerated lumbar discs have a lower ADC value than non-degenerated discs (25,138,172). The reduction in ADC observed in degenerated discs has been shown to correlate with decreased proteoglycan and water content in the NP. This makes DWI useful for detecting early degeneration — a mildly dehydrated NP can be identified by a reduced ADC even if conventional MRI looks normal (173).

Interestingly, in very advanced degeneration, the relationship can reverse in localised areas since the severely degenerated discs often develop fissures, cavities, or areas filled with fluid, and water mobility in those regions may increase. Consequently, some end-stage degenerated discs show paradoxically elevated ADC values in focal regions, likely to reflect clefts filled with fluid (25).

Ultrashort echo time (UTE) MRI is a technique with echo times shorter than 1 ms, enabling signal acquisition from tissues with very short T2 components that are invisible on conventional sequences. In the spine, UTE has been particularly useful for visualising cartilaginous endplates (174). UTE sequences allow direct assessment of cartilaginous endplate structure and thickness (175). UTE imaging has also been proposed to accelerate quantitative disc assessment, with applications such as UTE T2 and T1ρ mapping (176,177).

GagCEST utilises the exchange of hydroxyl protons on glycosaminoglycans with unbound water to generate a quantitative signal directly reflecting glycosaminoglycan content. In lumbar IVDs, gagCEST values are higher in the healthy, glycosaminoglycan-rich NP than in the AF and decline with degeneration, showing significant negative correlations with Pfirrmann grade (142). gagCEST provides a glycosaminoglycan-specific biomarker, with particular value for detecting early extracellular matrix loss before visual

morphological changes (178). GagCEST has shown the best correlation with Pfirrmann grades and the highest discriminant accuracy to Pfirrmann grades compared to other methods (T2 mapping, T1ρ and ADC) (179).

2.3.5.3 Applicability and limitations of quantitative techniques

Techniques like T2 mapping, T1ρ, DWI and gagCEST each offer distinct advantages in detecting the biochemical and early changes of DD. However, a common limitation of these methods is that they typically require specialised MRI sequences and are not used on standard imaging protocols. This restricts their use mainly to prospective research settings. Adding these sequences to the imaging protocol also prolongs the scanning time. They are not applicable to retrospective studies based on routine lumbar MRI scans. In particular, the feasibility of T2 and T1ρ mapping, as well as gagCEST is often further constrained by scanner-specific protocol requirements and the lack of cross-centre standardisation due to calibration differences, which often limits the sample size and generalisability in multicentre research.

In this context, quantitative assessment of T2-weighted signal intensity on conventional images remains useful. It can be applied retrospectively and in large cohorts since every routine lumbar MRI has T2WIs. Although normalised NP signal intensity ratios have shown correlation with qualitative DD grades, they remain indirect markers of tissue composition. T2-weighted signal intensity reflects image brightness, and it is not a true physical parameter like T2 relaxation time or T1ρ — thus, it is subject to scanner and acquisition variability.

2.4 Artificial intelligence in spine magnetic resonance imaging analysis

2.4.1 Overview

The advent of artificial intelligence (AI) has revolutionised radiological image interpretation, and its role in assessing DD has grown rapidly in recent years. In particular, deep learning (DL) — a subset of machine learning that trains multi-layered artificial 'neural networks' to learn patterns directly from large image datasets — has markedly improved the automatic detection and grading of DD. Most imaging DL systems use convolutional neural networks (CNNs). This network learns step by step through repeated layers, beginning with basic image characteristics such as edges and textures and advancing to complex anatomical patterns. This architecture enables the evaluation of complex

structures such as the IVD. As a result, DL models can deliver consistent and rapid assessments of spinal MRI by reducing observer-dependent variability and enabling large-scale evaluation of lumbar MRIs (36,180).

2.4.2 SpineNet: architecture and initial performance

A landmark application of applying DL to lumbar spine MRI was by Jamaludin et al. (38), who introduced a model called **SpineNet** in 2017. SpineNet is a multitask CNN designed to classify multiple degenerative features from sagittal T2-weighted lumbar MRIs, including the Pfirrmann grade, disc bulging, central canal stenosis, spondylolisthesis, and endplate and bone marrow changes. In their initial study, Jamaludin et al. (38) reported that SpineNet achieved near-human performance across these features, with agreement ranging from 76% to 95% compared with expert radiologists. For Pfirrmann grading, the model's predictions agreed with radiologist's assessments in approximately 70% of cases on their test set.

2.4.3 External validation of SpineNet and SpineNetV2

External validation studies have also shown that SpineNet performs reliably across various populations. In a cohort of 882 patients, Grob et al. (40) reported class-average accuracy of 55% (kappa (κ) = 0.72) for Pfirrmann grading with the updated SpineNet V1.1 and class-average accuracy of 56% (κ = 0.63) for spondylolisthesis and class-average accuracy of 43% (κ = 0.60) for central canal stenosis compared with an expert radiologist. Subsequently, a new version of SpineNet (SpineNetV2) was released and made publicly available (181). In the population-based Northern Finland Birth Cohort 1966 of 1331 individuals, McSweeney et al. (39) found with SpineNetV2 a balanced accuracy of 78% for Pfirrmann grading (κ = 0.68; Lin's concordance correlation coefficient (CCC) = 0.86) and a balanced accuracy of 86% (κ = 0.74) for Modic changes. Most recently, Nigru et al. (41) validated SpineNetV2 with 353 patients across all features that the algorithm is capable of evaluating: Pfirrmann grading, disc narrowing, central canal stenosis, spondylolisthesis, endplate defects and bone marrow changes, foraminal stenosis, and disc herniation. They reported a balanced accuracy of 79% (κ = 0.74; CCC = 0.95) for Pfirrmann grading and a balanced accuracy for other MRI findings ranging from 69% to 99% (κ = 0.46–0.80; CCC = 0.53–0.97). This high level of agreement, close to inter-rater reliability amongst radiologists, demonstrates that AI can provide a consistent interpretation of DD features on MRI.

2.4.3.1 Patterns of disagreement and labelling considerations

Similar discrepancies between SpineNet and radiologists were documented in all three external validation studies. Disagreement often arose in borderline cases between adjacent grades, particularly grade II versus III and grade IV versus V. Radiologists showed a central tendency, assigning a higher proportion of intermediate Pfirrmann grades, whereas SpineNet more often assigned end-scale grades. Additional sources of disagreement included collapsed discs and anatomical variants such as lumbosacral transitional vertebrae, in which SpineNet failed to label the disc level correctly (39–41).

2.4.4 Other convolutional neural network approaches

Beyond SpineNet, a variety of other AI models have been proposed for the automated assessment of lumbar DD (36,37). Most of these methods are built upon CNN architecture, but the implementation strategies vary. In 2021, Niemeyer et al. (182) demonstrated that a CNN trained on mid-sagittal lumbar MRI slices could achieve an excellent repeatability in Pfirrmann grading, in fact exceeding the inter-rater agreement between expert readers. Similarly, in 2021, Gao et al. (183) developed a CNN with a 'push–pull' regularisation strategy, which drives the features from the same Pfirrmann grade to cluster together (pull) and separates features from different grades further apart (push). This approach improved feature discrimination across the Pfirrmann grades, thereby reducing the common problem of class overlap in adjacent grades. Adding the push–pull regularisation increased overall accuracy by 8–10% across CNN backbones and delivered the largest benefit in the challenging differentiation between Pfirrmann grade II and III categories.

2.4.5 Hybrid segmentation–classification pipelines

Other research studies have explored hybrid frameworks that combine segmentation and classification. For example, in 2025, Soydan et al. (184) used a two-stage pipeline in which a U-Net first performs automatic disc segmentation on mid-sagittal T2WIs, after which a fine-tuned Inception-v3 classifier assigns Pfirrmann grades to the segmented discs. A model using consensus labels from four specialists (two orthopaedists, one neurosurgeon, one radiologist) to classify discs outperformed the individual physicians on the test set. When deployed as a computer-aided diagnosis tool for the human raters, it improved physicians' diagnostic accuracy by 8–22%.

2.4.6 Automated quantitative pipelines

Beyond categorical Pfirrmann grading, several groups have developed fully automated, quantitative pipelines for DD. Spine Explorer, by Huang et al. in 2020 (185), uses a U-Net–based segmentation to delineate vertebrae, discs, the sacrum, and CSF on sagittal T2WI, then computes morphometric and signal intensity–based metrics — including DH, disc and vertebral areas/diameters, disc bulge area, and CSF-adjusted signal intensity of the disc. The tool achieved excellent agreement with manual measurements done by human raters (intra-class correlation (ICC) = 0.81–1.00). Similarly, in 2022, Zheng et al. (186) introduced a high-precision pipeline built on BianqueNet (a DeepLabv3+ backbone enhanced with depth-feature extraction, Swin-Transformer skip connections, and multi-feature fusion) to segment IVDs and vertebral bodies and then compute quantitative parameters such as DH, DH index and disc height-to-diameter ratio. They also performed a normalised histogram analysis of the disc, quantifying NP–AF peak separation normalised with a CSF reference, which decreased with higher modified Pfirrmann grades and enabled automated, objective grade thresholds. Automated measurements showed an excellent agreement with the manual measurements (ICC = 0.91–0.96).

Extending beyond geometry and signal intensity ratios, in 2025, Li et al. (187) developed a three-stage pipeline that infers T1ρ-defined DD grades from routine sagittal T1-weighted images. Crucially, the ground-truth phases were created by measuring T1ρ of NP on actual T1ρ maps, while the network itself saw only T1-weighted images. The model learned to predict the T1ρ-derived categories from the T1-weighted images. In the test set, phase classification reached an accuracy of 84% with an average area under the receiver operating characteristic (ROC) curve (AUC = 0.87). This study illustrates how DL can approximate biochemical staging linked to proteoglycan loss without acquiring the actual T1ρ sequence.

2.4.7 Incorporating axial images and multisequence magnetic resonance imaging

Although most DL systems use sagittal T2WI, incorporating axial images is valuable for features that are difficult to judge on sagittal images, such as foraminal stenosis, herniation morphology, and posterior elements. Segmentation based on axial T2WI and classification tools for these features has been developed. In 2021, Gong et al. (188) introduced Axial-SpineGAN, which simultaneously segments discs, neural foramina, the thecal sac, and posterior arches on axial MRI and performs automated ‘normal versus abnormal’ diagnosis for each structure, reporting an overall diagnosis accuracy of 93.9%. In 2021, Pan et al. (189) used sagittal T2WI to index levels and

then axial images to identify disc bulging and herniations, achieving per-level accuracies of 84–93%.

2.4.8 Radiomics and prognostic modelling

More recently, object detection and multi-task models have been introduced to broaden the spectrum of captured degenerative features. Radiomics, a complementary AI-driven approach, involves extracting a high-dimensional set of quantitative features (including signal intensity–based textures and shape descriptors) from routine MRI scans. This technique has been applied to the evaluation of lumbar DD to capture subtle imaging patterns that are beyond human visual perception. For example, in 2025, McSweeney et al. (190) reported that a radiomics-based model could classify Pfirrmann grades with a balanced accuracy of 73% ($\kappa = 0.67$), outperforming a conventional model relying solely on DH and normalised T2-weighted signal intensity (balanced accuracy 66%, $\kappa = 0.55$). Notably, certain radiomic texture and shape features, such as 2D sphericity and intensity interquartile range, showed strong correlations with Pfirrmann grade, highlighting the ability of radiomics to see nuances in disc texture and disc shape features that the human readers cannot recognise. By revealing these formerly undetected imaging characteristics as objective markers, radiomic analysis can enhance DL models and potentially improve DD classification and the assessment of disc abnormalities with additional quantitative features.

A complementary use of hybrid DL radiomics might provide support in predicting prognosis and in decision-making. In 2024, Fan et al. (191) combined radiomics DL features from sagittal and axial MRI with clinical features into a nomogram that accurately predicted risk for surgery in patients with lumbar disc herniation. On their external testing, the clinical and DL–radiomics model nomogram showed high performance in predicting risk for surgery (AUC = 0.941). Similarly, in 2025, Zou et al. (192) combined CT and multi-sequence MRI radiomics and clinical data to predict adjacent segment degeneration after lumbar fusion surgery, outperforming the prediction of surgeons. Together, these studies show that radiomics can extend beyond DD grading to individualised risk assessment and outcome prediction.

2.4.9 Strengths and limitations of artificial intelligence approaches

These models offer clear advantages in speed, reproducibility, and capacity for large-scale evaluation. CNN-based classifiers can automatically learn hierarchical image features, while hybrid segmentation–classification

pipelines ensure anatomically localised, disc-specific predictions. Quantitative or radiomics-based tools can generate objective metrics that complement traditional categorical grading and thus reduce dependence on subjective interpretation.

Nevertheless, the limitations of these approaches must be acknowledged. A central limitation is generalisability. Most models are trained on homogeneous datasets, often collected from a single centre, scanner, or acquisition protocol, and therefore tend to perform optimally only within that context. When applied to external data, accuracy and agreement commonly decline due to variation in patient populations and MRI acquisition parameters (37). This problem is evident in the external validations of SpineNet, where agreement metrics varied widely across cohorts (39–41). Such performance variability underscores the need for multicentre, age-diverse training samples. Techniques such as transfer learning and data augmentation have shown promise to handle differences between datasets and improve generalisability across scanners (180). Both data augmentation and larger, heterogeneous datasets improve model stability (37).

Another important challenge for AI training is data imbalance. In both clinical and population-based samples, the non-degenerated (Pfirrmann grade I) and severely degenerated (Pfirrmann grade V) discs are relatively rare. When these grades are underrepresented, the model may not adequately learn the distinctive imaging features of truly normal discs and end-stage degeneration. Synthetic data generation or diffusion models have been proposed to expand underrepresented classes and improve model balance (180).

A further limitation concerns the ground-truth labels themselves. Nearly all published models rely on radiologist-assigned Pfirrmann grades, which exhibit only moderate inter- and intra-observer reliability. These inconsistencies in the training labels constrain the upper limit the performance model can achieve and may transfer human subjectivity into the models. Consensus labelling by multiple experts or integration of objective quantitative features may help to establish more reproducible reference standards for model training.

In addition, most of the CNN-based approaches operate exclusively on sagittal T2WI, disregarding other MRI sequences that supply complementary diagnostic information. T1-weighted and STIR sequences are particularly valuable for evaluating other degenerative changes, such as the Modic changes. Expanding model inputs to multisequence or multimodality data, and even combining imaging with clinical variables, has been proposed to enhance diagnostic relevance and support patient-specific prognostic assessment (180).

Finally, interpreting and validating model predictions remains an unresolved challenge. CNNs are often criticised as ‘black boxes’, providing limited insight into the reasoning behind predictions. Visual explanation tools can highlight image regions that most influence a model’s output, but they only

partly address the need for explainable and verifiable AI (180). Ethical and regulatory discussions increasingly emphasise that explainable processing and outputs will be required for clinical integration (36,180).

2.5 Longitudinal studies on disc degeneration

Understanding the course of DD is essential for interpreting MRI findings and identifying factors that influence the rate of DD progression. Cross-sectional studies provide only observations at certain time points, whereas longitudinal follow-up studies enable direct evaluation of structural changes within the same individuals. Repeated imaging offers evidence of longitudinal changes and the rate of DD progression and allows detection of early degenerative changes that may precede more advanced degeneration. Long-term MRI cohort studies have been used to clarify the natural history of lumbar DD and the determinants that influence its course.

2.5.1 Population-based cohorts

In a large general population cohort from Japan, the Wakayama Spine Study, Teraguchi et al. (89) followed 617 adults with MRI over 4 years (baseline age 21–97 years; mean 66). Progression of DD was observed in 52% of men and 60% of women, and new-onset of DD developed in approximately one-third of men (32%) and nearly half of women (45%). Ageing was found as the strongest determinant of DD at all lumbar levels, while females presented with faster progression in the upper lumbar region (odds ratio (OR) = 1.68). Diabetes was found as an independent risk factor for DD in the upper lumbar spine (OR = 6.8). Other examined factors — including obesity, smoking, and occupational lifting — were not significant in multivariable models. Extending the Wakayama cohort, Deguchi et al. (193) performed a 7-year longitudinal study with genetic analysis of 574 participants (baseline age 26–98 years, mean 62.3). They identified the THBS2 gene, in which the association between its risk alleles and MRI-based DD progression was age-dependent, with the strongest effects at the thoracolumbar junction (T12/L1) and lumbar level L3/L4. Individuals with two THBS2 risk alleles showed faster degeneration at younger ages, with the effect diminishing later in life, whereas other candidate genes showed only minor or segment-specific associations. Overall, DD progression displayed age- and level-specific genetic effects.

2.5.2 Twin studies

Twin studies have provided the most convincing evidence for the heritability of DD progression. In a 5-year MRI follow-up of adult monozygotic twins (baseline age 35–69 years, mean 49), Videman et al. (49) found progression of DD findings in only 7–13% of lumbar discs. Despite this modest overall change, familial factors accounted for 47–66% of the variance in progression, whereas occupational loading and resistance training together accounted for only 2–10%, and smoking and body measurements were not significant. In the same male monozygotic twin cohort using quantitative MRI measures including DH and disc bulging, familial aggregation was found to explain a smaller share of change (17% for DH change, 11% for anterior bulge, 0% for posterior bulge). Overall five-year changes were small though. Mean DH decreased by 2.2–3.6%, and bulges increased by about 7–10% in the L1/2–L3/4 discs and 4% in the L4/5–L5/S1 discs. Maximal occupational lifting and smoking during follow-up were associated with slightly greater DH loss, with additional explained variance of 4.9% and 3.5%, respectively (50). These results demonstrate that progression of DD in adults is slow and substantially influenced by genetic factors, with modest contributions from specific environmental exposures.

Extending work in the same Twin Spine cohort, quantitative MRI in a 15-year follow-up of Finnish men (baseline ages approximately 35–58 years; mean age 63 at follow-up) showed midsagittal DH decreases of 8.7% (1.0 mm) in the upper and 11.3% (1.3 mm) in the lower lumbar discs. Over the earlier 5-year interval, DH declined by about 3.3–3.4% (0.4 mm) in both regions, indicating a slow, progressive decrease in DH across mid- to late adulthood (194).

In a decade-long twin study by Williams et al. (51), 234 same-sex twin pairs were followed (baseline age 32–70 years; mean 53). They examined progression in four DD features based on MRI: DH, disc signal intensity, posterior disc bulge, and anterior osteophytes. Genetic influence on the progression of these DD features varied across age groups. Change in DH was not heritable at any age, whereas progression of posterior disc bulging showed modest heritability in all age strata (28–53%). Progression of decreased disc signal intensity and anterior osteophyte formation was only highly heritable in participants younger than 50 years at baseline (heritability 76% and 74%, respectively). Overall, genetic influences on progression were weaker and more selective than those typically seen and tended to diminish with age. They concluded that before midlife, progression in some features (e.g., decreased disc signal intensity and anterior osteophytes) are substantially shaped by genetic factors, whereas after midlife, the change over time appears to be driven largely by non-genetic, environmental influences, with DH loss in particular showing no detectable heritable component.

2.5.3 Developmental and early-life determinants of disc degeneration

Recent longitudinal evidence has highlighted the adolescent growth phase as a critical phase that determines the development of lumbar DD. Aavikko et al. (195) followed a randomly selected cohort of Finnish schoolchildren from childhood through adolescence into adulthood. Participants underwent MRI at ages 8, 11, 18, and 34 years. The Pfirrmann summary score — sum of grades across the three lowest lumbar discs — increased significantly with age, and participants who reported LBP at age 34 had already shown higher scores at age 18 compared with asymptomatic peers. Quantitative assessment of CSF-normalised NP signal intensity decreased uniformly with age but did not differ between pain and non-pain groups, suggesting that the overall extent of degeneration across discs, rather than changes in single discs, may distinguish individuals who later develop LBP.

The subsequent follow-up study (196) focused on MRI data at ages 18 and 34 years and confirmed that early degenerative burden after the pubertal growth spurt predicts future LBP. Each one-point increase in the Pfirrmann summary score at age 18 increased the odds of LBP at age 34 by 5.5-fold (adjusted OR 5.46). All but one participant with a Pfirrmann summary score > 6 or with at least one disc graded Pfirrmann ≥ III at baseline developed LBP by their mid-thirties. Over the 16-year follow-up, DD progressed on average among both participants who later reported LBP at age 34 and those who remained asymptomatic. By age 34, the initially asymptomatic group had reached a similar degree of degeneration as their symptomatic peers. This pattern suggests that shortly after the pubertal growth spurt is the critical period when early degenerative changes differentiate those who will later develop LBP from those who will not.

Makino et al. (45) performed a decade-long longitudinal study of lumbar discs in 84 young adult women to characterise early degeneration (baseline age 20–22 years; mean 20.9). At the initial MRI, about one-third of the cohort already had DD. Ten years later (around age 30), those with DD at baseline were significantly more likely to show further progression in DD (58% of that subgroup) compared to those with initially normal discs (31% developing new DD). Overall, the proportion with MRI-detectable DD increased from 33% at baseline to 52% at follow-up. Across all participants, the L5/S1 level was the most frequently affected, and overall degeneration severity scores increased modestly during the follow-up period.

2.5.4 Asymptomatic cohorts

Borenstein et al. (197) conducted a seven-year follow-up of initially asymptomatic adults (baseline age 26–68 years; mean 52). Over the follow-up period, LBP developed in 21 of the 50 respondents (42%). Among the 31 participants who underwent repeat MRI, ages at follow-up ranged from 26 to 68 years (mean 52). The documented baseline MRI findings — DD, disc herniations or bulging, and spinal canal stenosis — did not predict the development or duration of LBP over seven years. On repeat imaging, degenerative features frequently progressed, yet the largest subgroup that developed LBP had normal baseline scans. Pain correlated more with DD findings at follow-up —particularly disc herniation and moderate DD — than with baseline imaging, suggesting that these MRI findings had little predictive information about future LBP in asymptomatic individuals.

Elfering et al. (198) conducted a 5-year MRI follow-up in an initially asymptomatic adult cohort of 41 individuals (20–50 years old) to identify factors predicting DD. They reported that 17 of 41 previously pain-free individuals (41%) showed MRI-defined progression of DD over follow-up. Only a small subset (24%, n = 10) progressed by at least one degenerative grade. Notably, the occurrence of DD progression was only weakly associated with the development of LBP in this group. Instead, a baseline MRI finding of disc herniation, lack of regular sports activity, and night-shift work emerged as significant independent predictors of DD, even after adjusting for baseline disc status and age, sex, and body mass index.

2.5.5 Symptomatic cohorts

Findings from long-term clinical follow-ups further demonstrate how early degenerative changes predict later DD. Waris et al. (199) followed Finnish men with a history of LBP (baseline age 20 years). At baseline, 69% already had at least one degenerated disc. After 17 years, all participants showed DD, with the number of degenerated discs increasing on average from 1.1 to 3.0. The most distinct changes were found at lower lumbar levels L4/5 and L5/S1. Early DD was found to predict later herniation: 76% of degenerated discs in youth progressed to herniation by age 37, compared with 29% of initially healthy discs. However, early DD did not correlate with persistent pain or disability.

The same cohort was re-examined after 30 years by Sääksjärvi et al. (200). This study revealed that by age 50, almost all participants had multilevel DD. Discs that already showed a decreased NP signal intensity (>20% reduction relative to the participant's brightest disc) at age 20 were more likely to develop severely decreased NP signal intensity, DH loss, disc protrusion/bulging, HIZ/annular fissures, Modic changes, and severe endplate lesions

at follow-up. However, symptom burden remained modest, again illustrating that radiological DD does not always predict symptoms.

A five-year prospective MRI study by Schistad et al. (201) provided complementary evidence from symptomatic adults (baseline age 24–64 years; mean 46). Among 144 patients with LBP or radicular pain, 59% showed DD progression over 5 years. Higher age and lower baseline DD predicted greater change, whereas sex, body mass index (BMI), smoking, and candidate gene variants were not found to be related. Importantly, progression of DD or HIZ did not correlate with pain, while an increased number of herniated discs was associated with pain. This implies that in symptomatic adults, DD mainly appears to reflect age-related processes, and newly developed herniations appear to be more closely linked to pain than the progressive DD itself.

2.5.6 Differences in degeneration patterns

Longitudinal MRI studies indicate that the structural components of the IVD do not degenerate uniformly but follow different temporal trajectories. Progression of DD in adulthood is slow but almost universal, many individuals show some worsening over years to a decade, with considerable inter-individual variability (49,50,89). Crucially, not all MRI features change together or for the same reasons. Genetic influences are substantial for several degeneration changes, whereas others, like DH, show weaker heritability (49–51). Age remains the single strongest predictor of new degeneration, while metabolic and hormonal influences — such as diabetes and pubertal timing —modulate its onset (89,193,195,196). Regional variability is pronounced. The lower lumbar discs (L4/L5 and L5/S1) tend to degenerate earliest and most severely, whereas upper lumbar discs may progress faster in females or in association with metabolic risk factors (42,193).

In a retrospective longitudinal series of a mean 32-month interval (age 20–88 years; mean 53.6), Sharma et al. (2009) demonstrated that discs with baseline annular fissures had a significantly faster decrease in the NP signal intensity and an increase in Pfirrmann grade on follow-up than discs without fissures. Some annular fissures were present before detectable NP darkening, suggesting that AF failure may precede and accelerate subsequent NP dehydration (46). Also, in a 1-year prospective MRI study of patients with subchondral, degenerative Modic type 1 signal changes, progression of Modic 1 changes was associated with worsening adjacent bony endplate damage, decrease of DH and change in disc signal intensity. Progression of these features was uncommon at disc levels without Modic 1 changes, suggesting that the Modic 1 changes may be a marker of accelerated degeneration (202).

Taken together, these longitudinal observations suggest that NP dehydration, AF structural disruption, and vertebral findings such as Modic lesions may arise in sequence. Typically, the decrease of NP signal intensity and DH occur gradually with age, whereas annular fissures develop episodically, often in response to mechanical stress, and may further accelerate NP degeneration. These imaging patterns align with biomechanical–biochemical models, in which biochemical degradation of the NP reduces intradiscal pressure and redistributes stress to the annulus, leading to progressive fibre delamination and fissuring (9,47). Together, these findings underscore that although the NP and AF deteriorate through interrelated pathways, their degeneration may be temporally distinct. Therefore, progression of DD on MRI may be better interpreted as a multidimensional process rather than a single uniform trajectory of 'disc degeneration'.

3 AIMS OF THE STUDY

The overall aim of this thesis was to enhance the assessment of lumbar DD on MRI by combining methodological research (reliability of T2 signal intensity measurements, validation of AI technology) with longitudinal observation of degeneration. The specific aims were:

1. To evaluate the reliability and diagnostic value of quantitative T2WI measurements of the NP signal intensity in lumbar IVDs. This included comparing different ROI techniques and normalisation methods and determining how well these quantitative measures correlate with qualitative DD grading. *(Addressed in Study I)*

2. To assess the agreement between an AI-based automated grading system (SpineNet) and expert radiologists in grading lumbar DD, both in cross-sectional and longitudinal contexts. *(Addressed in Study II)*

3. To characterise the progression of lumbar DD over a 14-year period in mid-life, using MRI-based evaluation of disc components. This included determining how Pfirrmann grades of discs change over time, as well as examining changes in specific disc features (NP signal intensity, NP homogeneity, AF border distinction, and DH). *(Addressed in Study III)*

4 MATERIALS AND METHODS

4.1 Study population and design

All three studies utilised a common cohort. The original study population consisted of 75 male participants originally recruited at a mean age of 20 (standard deviation (SD) 1.0) years during the mandatory military service as part of a broader research project on the longitudinal evolution of DD. All the original participants had a history of LBP severe enough to interfere with their military service. Subsequently, each participant was invited to participate in lumbar MRI at the ages of 37 and again at 51, which served as the time points of this study. Of the original 75 participants, 6 could not be contacted (deceased, abroad, or no permanent address). Of the remaining 69, 32 attended MRI at age 37, 26 attended MRI at age 51, and 19 men completed MRI at both time points. These 19 form the longitudinal cohort for this study (Figure 6).

Study I is a cross-sectional reliability study that used MRI examinations of the 19 participants at the age of 51 to compare quantitative measurements of NP signal intensity on T2WI using different ROI methods with the visual grading.

Studies II–III are longitudinal and use MRI examinations acquired at ages 37 and 51 in the same 19 participants: in Study II, Pfirrmann grades assigned at both time points with an automated (SpineNet) method are compared with those assessed by a radiologist. In Study III, the 14-year changes in Pfirrmann grade are evaluated and compared to the grades of separately assessed degenerative disc features.

This study was performed in line with the principles of the Declaration of Helsinki. Approval was obtained from the ethics committee of Helsinki University Hospital (reference number HUS/1493/2017). Informed consent was obtained from all individual participants included in the study. Patients signed informed consent regarding the publishing of their data.

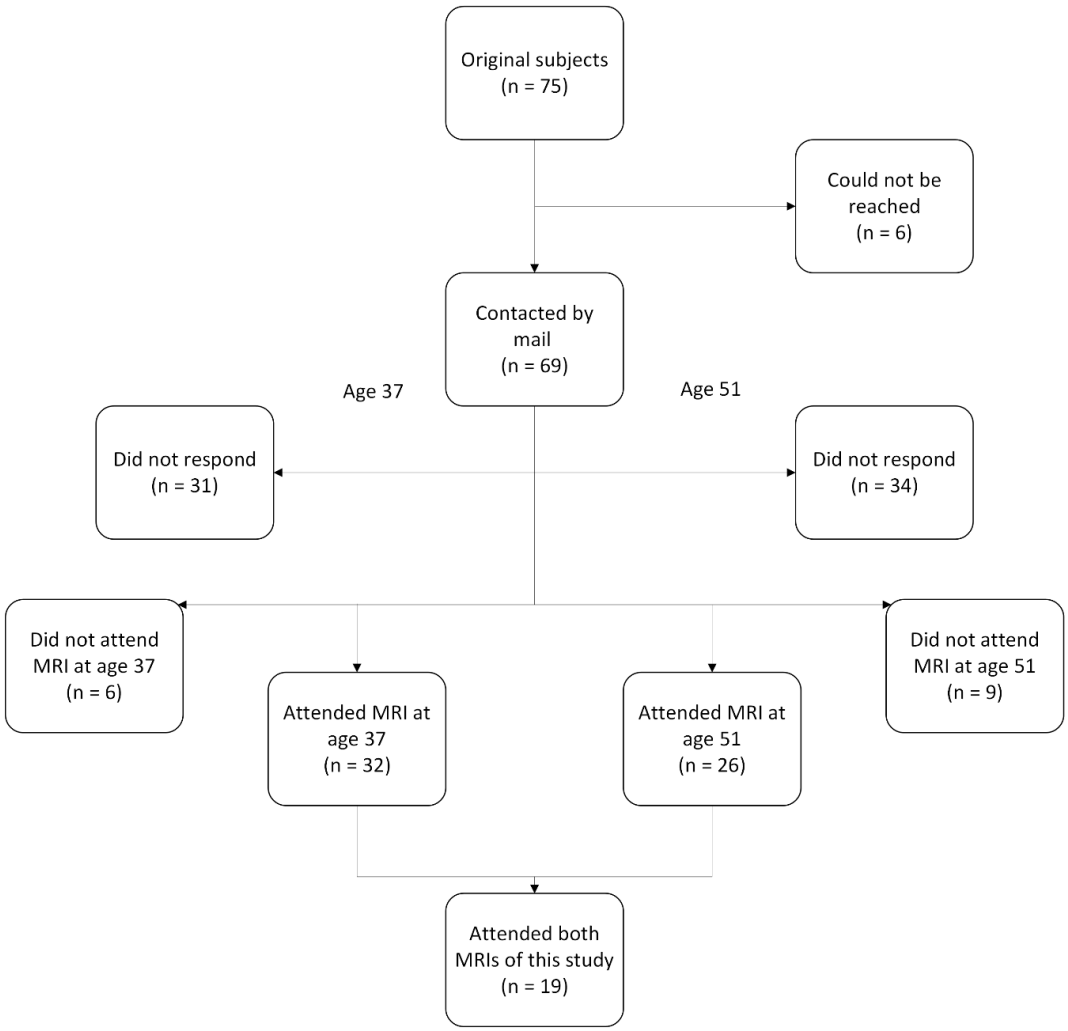


Figure 6 Flowchart of participant recruitment and follow-up. MRI – magnetic resonance imaging

4.2 Magnetic resonance imaging acquisition and image analysis

4.2.1 Magnetic resonance imaging scanning protocols

Lumbar spine MRI scans were performed at two time points for each participant using different MRI machines due to technological updates over the 14-year time period:

The MRI at age 37 was acquired using a 1.0 T GE Signa scanner. The primary sequence for this study was a sagittal T2-weighted fast spin-echo sequence (TR/TE 4000/96.9 ms, field of view 260 × 260 mm, slice thickness 4 mm, matrix 256 × 256). Additional sequences included sagittal T1-weighted and axial T2-weighted sequences at each disc level.

The MRI at age 51 was performed on a 1.5 T Siemens Magnetom Avanto Fit scanner. The sagittal T2-weighted sequence parameters were TR/TE 3520/117 ms, field of view 300 × 300 mm, slice thickness 4 mm, and matrix 326 × 384. Sagittal STIR, sagittal T1-weighted, and axial T2-weighted sequences were also obtained.

All images were stored in DICOM format. Visual grading of DD was performed on high-resolution medical displays using Siemens Syngo.Plaza software, and quantitative signal intensity measurements were performed using Siemens Syngo.via.

4.2.2 Image analysis

Study I

For this cross-sectional analysis, the MRI examinations acquired at age 51 years were used. Lumbar IVDs from L1/L2 to L5/S1 (total 95 discs in 19 participants) were evaluated on mid-sagittal T2WIs. DD was graded visually using the Pfirrmann classification (grades I–V) according to the criteria described in Section 2.3.3, and NP signal intensity was assessed using a five-level scale (0–4): 0 = bright signal equal to CSF; 1 = slightly decreased; 2 = clearly decreased; 3 = severely but inhomogeneously decreased with a small bright NP focus (intranuclear high-intensity zone, INHIZ); and 4 = severely decreased.

For quantitative analysis, the mean T2-weighted signal intensity of discs was measured using three region of interest (ROI) methods (Figure 7):

Targeted ROI (TARGET-ROI): A small freehand ROI targeted on the brightest, most homogeneous portion of the NP, avoiding the INC and dark regions of the NP.

Ellipsoid ROI (ELLIPS-ROI): An ellipsoid-shaped ROI placed centrally on the NP region, including as much NP as possible while remaining within its boundaries. In degenerated discs with poor NP/AF distinction, the ellipsoid was placed on the assumed NP region, avoiding the obvious AF margins.

Whole-disc ROI (WD-ROI): A freehand ROI drawn along the edge of the entire IVD, including NP and AF. Vertebral bone, endplates, and longitudinal ligaments were excluded.

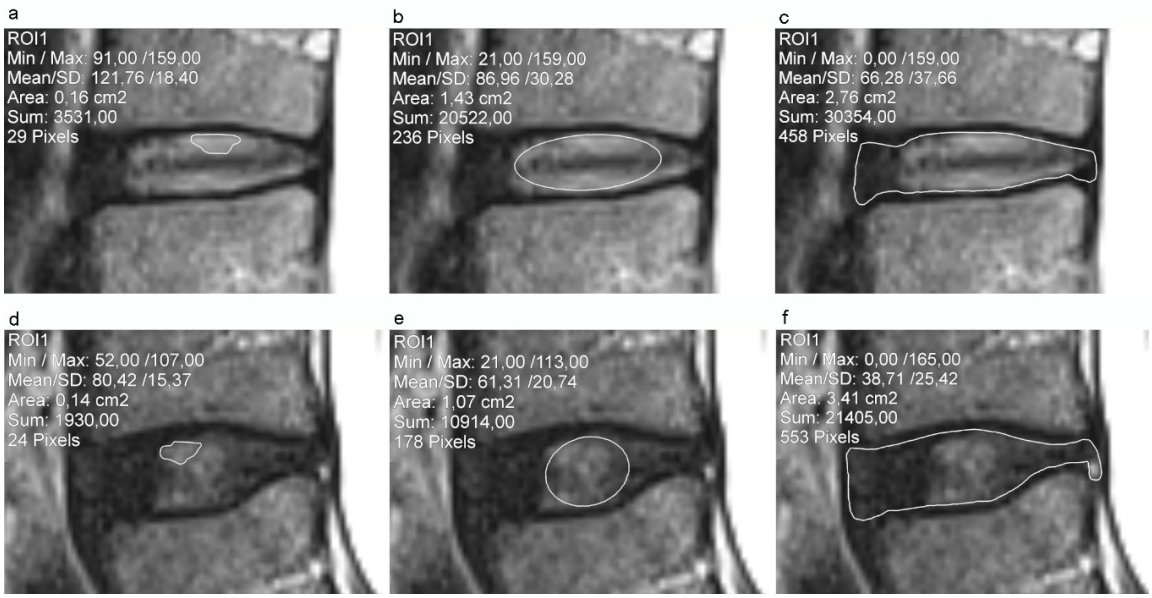


Figure 7 Examples of different signal intensity measurement methods: (a, d) freehand ROI on the most homogeneous and brightest area of the NP, excluding the INC (TARGET-ROI); (b, e) ellipsoid ROI on the NP, including the INC (ELLIPS-ROI); (c, f) ROI, including the whole IVD (WD-ROI). Reprinted from Murto N, Luoma K, Lund T, Kerttula L. Reliability of T2-weighted signal intensity-based quantitative measurements and visual grading of lumbar disc degeneration on MRI. *Acta Radiologica.* 2023;64(6):2145–2151. Copyright © 2023 SAGE Publications. Reprinted by permission of SAGE Publications.

The mean signal intensity of the CSF was measured with an ellipsoid ROI at the L3/L4 disc level, or at L2/L3 in cases of severe spinal canal narrowing or pulsation artefacts at L3/L4, to obtain a signal intensity reference. In addition, the mean signal intensity of the bone marrow of the L3 vertebral body was measured with an ellipsoid ROI, excluding areas with Modic changes or other focal abnormalities. Then, the mean signal intensity values of the disc were adjusted by dividing them by the mean signal intensity of the CSF as a reference and separately by the signal intensity of the bone marrow as a reference. The adjusted signal intensity values were then compared.

All signal intensity measurements were performed by a radiology resident (Observer 1), who repeated the measurements to assess the intra-observer agreement. A subset of 25 discs was additionally measured by an experienced musculoskeletal radiologist (Observer 2) to assess the inter-observer agreement. Pfirrmann grading and visual NP signal intensity grading was done twice by Observers 1 and 2 to assess intra- and inter-observer agreement. A

neuroradiologist (Observer 3) and an orthopaedic spine surgeon (Observer 4) graded 25 discs for additional inter-observer agreement analysis.

Study II

Two radiologists (one musculoskeletal radiology fellow and one experienced musculoskeletal radiologist) independently graded lumbar IVDs from L1/2 through L5/S1 on sagittal T2WI at ages 37 and 51 using the Pfirrmann classification (grades I–V). The radiologists were blinded to each other's assessments and graded each examination independently, with age-37 images not displayed during the grading of the age-51 images.

Automated Pfirrmann grading was performed using SpineNetV2, an open-source DL model available at https://github.com/rwindsor1/SpineNet, which detects lumbar discs and assigns Pfirrmann grades from sagittal T2WIs, among other DD findings. The SpineNet code was implemented in Python and executed within the HUS Acamedic secure virtual research environment. HUS Acamedic is a controlled analytics platform operated by Helsinki University Hospital, designed for projects involving sensitive health data. It meets the requirements of the Finnish Act on Secondary Use of Health and Social Data by ensuring that all processing occurs within a secure, access-controlled environment, with compliance oversight by the national data permit authority (Findata).

Study III

Two radiologists (one musculoskeletal fellow and one experienced musculoskeletal radiologist) independently graded all lumbar discs at ages 37 and 51 on mid-sagittal T2WIs using the Pfirrmann classification and by separately evaluating DD components according to the following criteria:

Heterogeneity of the nucleus pulposus (NP inhomogeneity):
- 0 = homogeneous, uniformly bright or grey
- 1 = inhomogeneous, speckled or patchy structure
- 2 = markedly inhomogeneous, uniformly dark or black

Annulus fibrosus border distinction (AF border):
- 0 = well-defined, homogeneous dark lamellar structure
- 1 = partially distinct with inhomogeneous, irregular structure
- 2 = indistinguishable, no clear border between AF and NP

Disc height (DH):

0 = not decreased;
1 = slightly decreased (1–33%)
2 = moderately decreased (34–66%)
3 = severely decreased or collapsed (67–100%)

Baseline and follow-up images were graded in separate sessions without side-by-side comparison. The radiologists were blinded to each other's assessments.

Quantitative NP signal intensity values of the targeted ROI method described in Section 4.2.2 (Study I image analysis) were presented using values normalised with the CSF signal intensity.

4.3 Statistical analysis

Study I: Agreement for visual grading scales (Pfirrmann grades and qualitative NP signal grades) was evaluated using weighted Cohen's kappa (κ). Interpretation thresholds were: κ = 0, less than chance; 0.01–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–0.99, almost perfect agreement (203). Reproducibility of continuous quantitative NP signal measurements was assessed using intraclass correlation coefficients (ICC), classified as < 0.50 poor, 0.50–0.75 moderate, 0.75–0.90 good, and > 0.90 excellent (204). For the analyses of the relationships between normalised NP signal intensity and visual gradings, both the quantitative measurements and the visual NP signal grades from Observer 1 were used. Relationships between normalised NP signal intensity and visual grades were tested with ANOVA, and pairwise differences between groups were examined using Dunnett's T3 post-hoc test. The effect size was determined with eta squared (η^2), interpreted as follows: η^2 = 0.01, small; 0.06, medium; 0.14, large (205,206).

Study II: All comparisons between SpineNet and human grading were performed against the experienced musculoskeletal radiologist, who served as the reference reader. Inter-rater agreement between the radiologist and the SpineNet model was analysed using weighted κ for individual disc grades and Lin's concordance correlation coefficient (CCC) for the Pfirrmann summary score (sum of Pfirrmann grades from L1/2 to L5/S1). The strength of agreement for the CCC was considered accordingly: < 0.90 = poor; 0.90–0.95 = moderate; 0.95–0.99 = substantial; > 0.99 = almost perfect (207,208). Differences in Pfirrmann summary score distributions were assessed with the exact McNemar's test. Bland–Altman plots were generated to visualise mean differences and limits of agreement, and the Bradley–Blackwood test

was applied to evaluate systematic and proportional bias (209). Given the small sample size, exact and resampling-based methods (Monte Carlo permutation and bootstrap) were used to compute significance levels and 95% confidence intervals (CI).

Study III: Within-participant changes from age 37 to 51 were analysed with a bootstrap-based paired t-test. Associations between changes in quantitative NP signal intensity and visual gradings were assessed using Spearman's rank correlation, with bias-corrected 95% CIs from 10,000 bootstrap replications. For interpretation, correlation magnitudes > 0.20, > 0.40 and > 0.60 were considered weak, moderate and strong, respectively (210). Inter-rater agreement for the visual gradings was summarised as overall accuracy (%) and weighted κ (ordinal weights), with bias-corrected bootstrap (10,000 iterations) used to derive 95% CIs. Variable distributions were evaluated graphically and with the Shapiro–Wilk W-test. All summary scores were calculated by summing the individual grades for each lumbar IVD (L1/2 to L5/S1) across the variables. Continuous data was summarised as means with standard deviations (SD) and categorical data as counts with percentages. Analyses were performed in Stata 18.0 (StataCorp LP; College Station, Texas, USA).

5 RESULTS

5.1 Reliability of quantitative magnetic resonance imaging measurements (Study I)

Repeatability of NP signal intensity measurements was excellent for all three ROI methods (ICC 0.94–0.99; $p < 0.001$) (Table 1). Inter-observer agreement for visual NP signal intensity grading ranged from moderate to almost perfect (weighted κ 0.50–0.88; $p < 0.001$) (Table 2). Pfirrmann grading showed substantial to almost perfect inter-observer agreement (weighted κ 0.74–0.90; $p < 0.001$).

No discs were graded visual NP signal intensity grade 0 (bright) or Pfirrmann grade I. The measured, CSF-normalised NP signal intensity differed across visual NP signal intensity grades and across Pfirrmann grades ($p < 0.001$). With TARGET-ROI, post-hoc testing showed significant differences between all adjacent visual grades (1 vs 2, 2 vs 3, 3 vs 4). For non-adjacent grades, all visual NP signal intensity grades differed significantly except for 1 vs 3 (Figure 8). Three degenerated discs graded as visual NP signal intensity grade 3 with severely decreased but inhomogeneous NP (with an INHIZ) showed higher TARGET-ROI signal intensity than discs visually graded 2 or 4. Using ELLIPS-ROI or WD-ROI, visual NP signal intensity grade 3 did not differ from grades 2 or 4 (Figure 8).

Compared with the Pfirrmann grades, CSF-normalised NP signal intensity differed significantly between grade III and IV for all ROI methods, whereas differences between grades II vs III and IV vs V were not significant (Figure 8).

All effect sizes were interpreted as large, with η^2 values for discriminating visual NP signal intensity grades of 0.67 (TARGET-ROI), 0.61 (ELLIPS-ROI), and 0.61 (WD-ROI), and for Pfirrmann grades of 0.53, 0.59, and 0.60, respectively.

Correlation analyses showed strong negative associations for the CSF-normalised NP signal intensity with the visual NP signal intensity for all ROI methods (ρ −0.82 to −0.84; 95% CI −0.74 to −0.89; $p < 0.001$) and slightly weaker — yet still strong — associations with the Pfirrmann (ρ −0.72 to −0.77; 95% CI −0.60 to −0.84; $p < 0.001$). All three CSF-normalised signal intensity measurement techniques showed comparable correlations with both the visual NP signal intensity grading and the Pfirrmann grading, with no significant differences between them. The vertebral marrow–normalised measures showed significantly weaker correlation with both references (visual NP signal intensity ρ −0.70 to −0.72; Pfirrmann ρ −0.59 to −0.67; $p < 0.001$ for all). Visual NP

signal intensity grades correlated strongly with Pfirrmann grading (ρ 0.78, p < 0.001).

ROC analysis on CSF-normalised measures for differentiating visual NP signal intensity grade 1 (slightly decreased) from grades 2 (clearly decreased) and 4 (severely decreased) showed high AUCs with no differences between the ROI methods (Figure 9).

Table 1 Intra- and inter-observer reliability of quantitative NP signal intensity measurements. CI – confidence interval, ELLIPS – ellipsoid, ICC – intraclass correlation coefficient, ROI – region of interest, TARGET – targeted, WD – whole disc.

Reprinted from Murto N, Luoma K, Lund T, Kerttula L. Reliability of T2-weighted signal intensity-based quantitative measurements and visual grading of lumbar disc degeneration on MRI. Acta Radiologica. 2023;64(6):2145–2151. Copyright © 2023 SAGE Publications. Reprinted by permission of SAGE Publications.

ROI	Intra-observer ICC	95% CI	Inter-observer ICC (1–2)	95% CI
TARGET-ROI	0.98	0.97–0.99	0.94	0.88–0.95
ELLIPS-ROI	0.99	0.98–0.99	0.95	0.90–0.98
WD-ROI	0.98	0.97–0.99	0.97	0.92–0.99

p < 0.001

Table 2 Intra- and inter-observer reliability of visual NP signal intensity. CI – confidence interval, κ – kappa. Reprinted from Murto N, Luoma K, Lund T, Kerttula L.

Reliability of T2-weighted signal intensity-based quantitative measurements and visual grading of lumbar disc degeneration on MRI. Acta Radiologica. 2023;64(6):2145–2151. Copyright © 2023 SAGE Publications. Reprinted by permission of SAGE Publications.

Observer(s)	Weighted κ	95% CI
Intra-observer		
1	0.56	0.27–0.84
2	0.71	0.48–0.94
Inter-observer		
1–2	0.72	0.59–0.85
1–3	0.75	0.52–0.97
1–4	0.50	0.24–0.75
2–3	0.88	0.73–1.04
2–4	0.69	0.45–0.94
3–4	0.69	0.45–0.94

p < 0.001

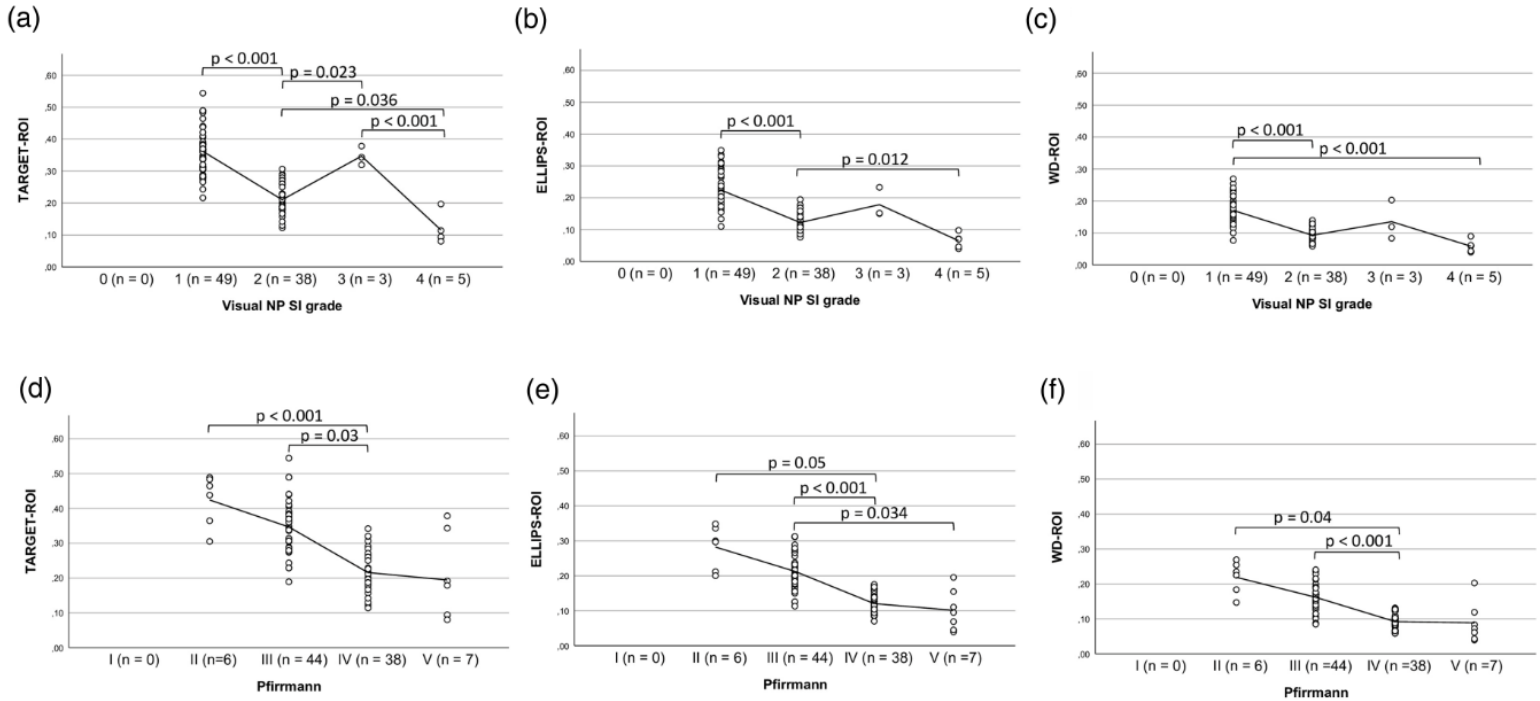


Figure 8 Scatterplots of CSF-normalised quantitative NP signal intensity measurements across visual grading categories, with lines connecting group means. (a) TARGET-ROI versus visual NP signal intensity grade; (b) ELLIPS-ROI versus visual NP signal intensity grade; (c) WD-ROI versus visual NP signal intensity grade; (d) TARGET-ROI versus Pfirrmann grade; (e) ELLIPS-ROI versus Pfirrmann grade; (f) WD-ROI versus Pfirrmann grade. Post-hoc pairwise comparisons between categories were performed with Dunnett's T3 test; statistically significant contrasts are indicated above the plots. CSF – cerebrospinal fluid, NP – nucleus pulposus, SI – signal intensity, TARGET – targeted, ROI – region of interest, ELLIPS – ellipsoid, WD – whole disc.

Reprinted from Murto N, Luoma K, Lund T, Kerttula L. Reliability of T2-weighted signal intensity-based quantitative measurements and visual grading of lumbar disc degeneration on MRI. *Acta Radiologica.* 2023;64(6):2145–2151. Copyright © 2023 SAGE Publications. Reprinted by permission of SAGE Publications.

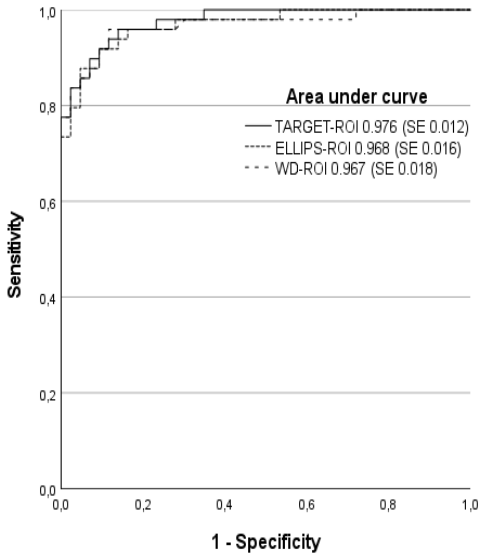


Figure 9 Receiver operating characteristic (ROC) curves for CSF-normalised quantitative NP signal intensity measurements using three ROI methods in differentiating visual NP signal intensity grade 1 from grades 2 and 4.

AUC – area under the curve, CSF – cerebrospinal fluid, ELLIPS – ellipsoid ROI, NP – nucleus pulposus, ROI – region of interest, SE – standard error, TARGET – targeted ROI, WD – whole disc. Reprinted from Murto N, Luoma K, Lund T, Kerttula L. Reliability of T2-weighted signal intensity-based quantitative measurements and visual grading of lumbar disc degeneration on MRI. *Acta Radiologica*. 2023;64(6):2145–2151. Copyright © 2023 SAGE Publications. Reprinted by permission of SAGE Publications.

5.2 SpineNet versus radiologist grading (Study II)

Inter-rater agreement between the two radiologists for Pfirrmann grading was substantial to almost perfect across disc levels (κ 0.73–0.96) (Table 3). The CCC for the Pfirrmann summary score between the two radiologists was 0.83 (95% CI 0.69–0.97) at baseline and 0.78 (95% CI 0.61–0.95) at follow-up.

Agreement between SpineNet and the experienced musculoskeletal radiologist for Pfirrmann grading ranged from fair to substantial across disc levels (κ 0.32–0.76 at both time points). The average κ values between SpineNet and the radiologist for all disc levels combined were κ 0.74 (95% CI 0.68–0.81) with 56% classification accuracy (range 43–64%) at baseline, and κ 0.68 (95% CI 0.58–0.77) with 56% accuracy (range 46–67%) at follow-up.

Systematic discrepancies were observed between SpineNet and radiologist assessments. SpineNet tended to underestimate grades at upper lumbar levels and slightly overestimate them at lower levels (Figure 10). Neither radiologist assigned any Pfirrmann grade I discs at either time point, whereas SpineNet assigned Pfirrmann grade I to several discs that radiologists graded as II or

III (including 5 discs graded III by both radiologists but I by SpineNet, and 33 discs graded I by SpineNet but grade II by both radiologists). SpineNet also labelled slightly more grade V discs at L4/L5 and L5/S1. In one case, a disc classified as Pfirrmann grade III by both radiologists was graded as V by SpineNet, and in five cases discs graded V by SpineNet were graded IV by radiologists. SpineNet was unable to grade one severely degenerated L5/S1 disc with spondylolisthesis and misidentified a small remnant S1/S2 disc as L5/S1. SpineNet assigned higher Pfirrmann grades at baseline than at follow-up for four discs, compared with two such cases among the radiologists.

Table 3 Agreement between SpineNet and the radiologist and inter-rater agreement between the radiologists across disc levels at baseline and follow-up.

CI – confidence interval.

Reprinted from Murto N, Lund T, Kautiainen H, Luoma K, Kerttula L. Comparison of lumbar disc degeneration grading between deep learning model SpineNet and radiologist: a longitudinal study with a 14-year follow-up. *European Spine Journal.* 2025;34(9):3767–3773. https://doi.org/10.1007/s00586-025-08900-2. © The Author(s) 2025. Distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

Disc level	Agreement between SpineNet and radiologist		Inter-rater agreement between radiologists	
	Kappa (95% CI)*	Accuracy %, (95% CI)**	Kappa (95% CI)*	Accuracy %, (95% CI)**
Baseline				
L1/2	0.57 (0.29 to 0.86)	42 (18 to 67)	0.82 (0.56 to 1.00)	96 (91 to 100)
L2/3	0.39 (0.13 to 0.79)	37 (13 to 61)	0.87 (0.58 to 1.00)	98 (95 to 100)
L3/4	0.71 (0.46 to 0.96)	74 (52 to 95)	0.74 (0.46 to 0.98)	96 (91 to 100)
L4/5	0.76 (0.64 to 0.88)	58 (33 to 82)	0.96 (0.87 to 1.00)	99 (98 to 100)
L5/S1	0.67 (0.48 to 0.86)	56 (29 to 82)	0.86 (0.69 to 0.94)	98 (96 to 100)
Follow-up				
L1/2	0.58 (0.26 to 0.89)	63 (39 to 87)	0.89 (0.71 to 1.00)	96 (91 to 100)
L2/3	0.45 (0.22 to 0.67)	37 (13 to 61)	0.76 (0.55 to 0.96)	91 (84 to 97)
L3/4	0.32 (0.10 to 0.54)	47 (23 to 72)	0.73 (0.46 to 0.99)	93 (63 to 99)
L4/5	0.74 (0.52 to 0.95)	68 (45 to 91)	0.80 (0.63 to 0.96)	96 (92 to 99)
L5/S1	0.71 (0.49 to 0.94)	67 (41 to 92)	0.91 (0.76 to 1.00)	98 (96 to 100)

* Kappa statistic (κ) with ordinal weights (confidence intervals obtained by bias-corrected bootstrapping)

** Percentage accuracy (confidence intervals obtained by bias-corrected bootstrapping)

The Pfirrmann summary score differed significantly between methods. At baseline, the mean was 13.5 (SD 1.8) for the radiologist and 12.2 (SD 2.9) for SpineNet (mean difference 1.3; 95% CI 0.3–2.3; p = 0.014). At follow-up, the

mean was 16.8 (SD 2.0) for the radiologist and 15.4 (SD 3.1) for SpineNet (mean difference 1.4; 95% CI 0.3–2.5; p = 0.016). Although both methods indicated progression over 14 years, agreement in the Pfirrmann summary score was poor (CCC 0.54, 95% CI 0.28–0.79 at baseline and 0.54, 95% CI 0.27–0.80 at follow-up) (Figure 11). The Bland–Altman plot of the longitudinal changes in Pfirrmann summary score further demonstrated poor agreement between SpineNet and the radiologist (CCC 0.41, 95% CI 0.01–0.80) (Figure 12).

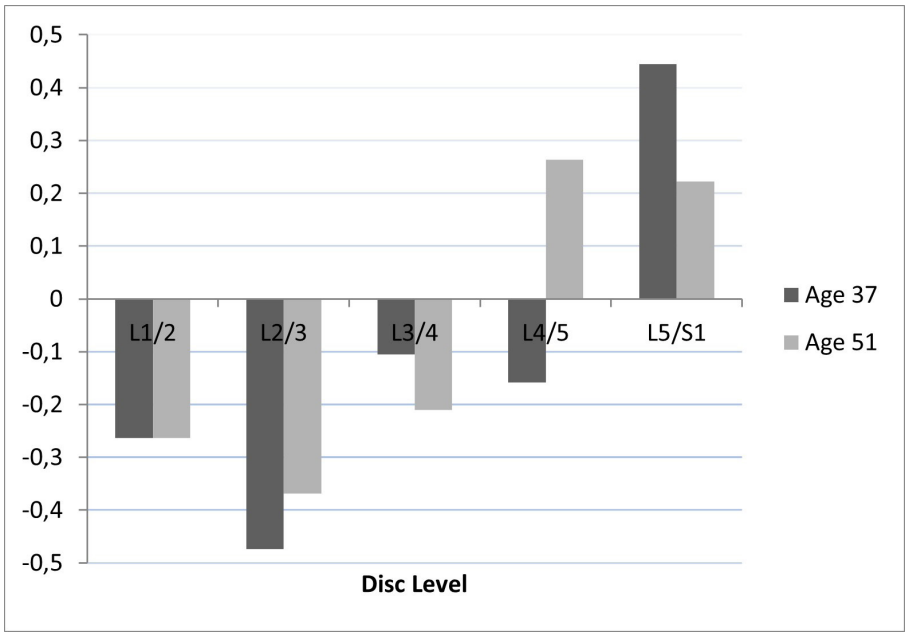


Figure 10 Level-wise mean difference in assessment between SpineNet and radiologist for Pfirrmann grading at two time points (age 37 and age 51). Positive values indicate higher Pfirrmann grades by SpineNet.

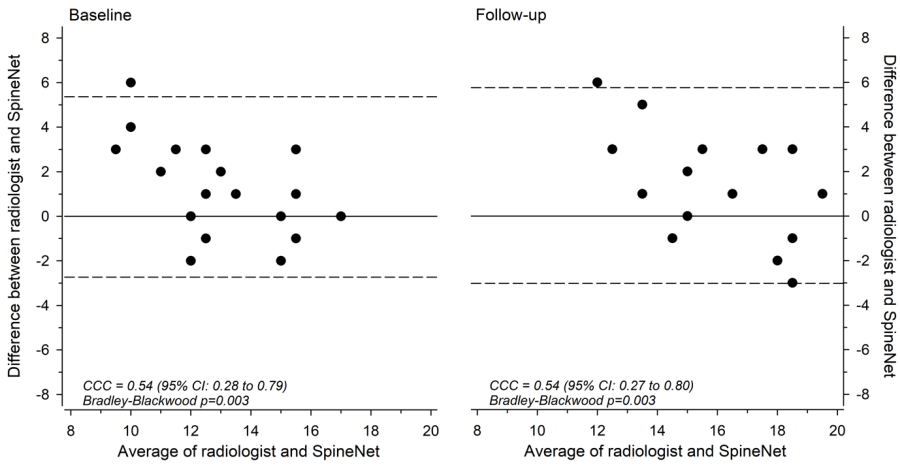


Figure 11 Bland–Altman plots and concordance CCC showing the agreement between SpineNet and the radiologist in the Pfirrmann summary score at baseline and follow-up. The dotted lines represent the 95% limits of agreement. CCC – concordance correlation coefficient, CI – confidence interval. Reprinted from Murto N, Lund T, Kautiainen H, Luoma K, Kerttula L. Comparison of lumbar disc degeneration grading between deep learning model SpineNet and radiologist: a longitudinal study with a 14-year follow-up. *European Spine Journal.* 2025;34(9):3767–3773. https://doi.org/10.1007/s00586-025-08900-2. © The Author(s) 2025. Distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

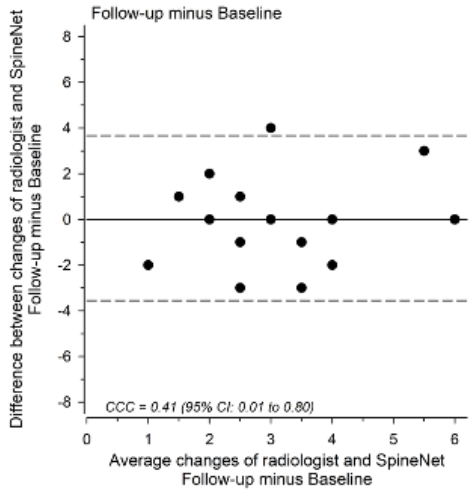


Figure 12 Bland–Altman plot and CCC showing the agreement between SpineNet and the radiologist in the change of the Pfirrmann summary score from baseline to follow-up. The dotted lines represent the 95% limits of agreement. CCC – concordance correlation coefficient, CI – confidence interval. Reprinted from Murto N, Lund T, Kautiainen H, Luoma K, Kerttula L. Comparison of lumbar disc degeneration grading between deep learning model SpineNet and radiologist: a longitudinal study with a 14-year follow-up. *European Spine Journal.* 2025;34(9):3767–3773. https://doi.org/10.1007/s00586-025-08900-2. © The Author(s) 2025. Distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

5.3 Longitudinal changes in disc degeneration (Study III)

Inter-rater agreement between the two radiologists was almost perfect for NP inhomogeneity (κ 0.87; 98% agreement) and AF border distinction (κ 0.90; 96% agreement), moderate for DH (κ 0.42; 87% agreement), and almost perfect for Pfirrmann grading (κ 0.89).

The distribution of Pfirrmann grades shifted over 14 years. At age 37, most discs were assessed as having grade II (62%), while fewer were at grades III (17%), IV (19%) or V (2%). By age 51, the percentage of discs with grade II had declined to 26%, while grades III and IV had increased to 34% and 33%, respectively, and grade V rose to 7%. No Pfirrmann grade I discs were seen at either time point. Across all discs, 51/95 (54%) increased by ≥ 1 Pfirrmann grade, 42/95 (44%) were unchanged, and 2/95 (2%) decreased. Among the 42 discs that remained in the same Pfirrmann grade, DH decreased in 9/42 (21%), and the NP signal intensity decreased by ≥ 10% in 30/42 (71%).

Across participants, the mean change in the Pfirrmann summary score was 3.0 points (95% CI 2.1–3.9; $p < 0.001$), from 13.1 (SD 1.5) to 16.1 (SD 2.1). At the participant level, the Pfirrmann summary score increased in 17/19 participants, and was unchanged in 2/19 participants.

Across participants, the mean change in the NP signal intensity summary score was -0.58 (95% CI −0.73 to −0.03; $p < 0.001$), from 2.08 (SD 0.24) to 1.50 (SD 0.24). Across individual discs, NP signal intensity changed by a mean of −23%, indicating that most discs darkened over 14 years (Figure 13). Discs with the brightest NP signal intensity at baseline showed the greatest decreases over follow-up.

AF border distinction worsened in 42/95 (44%) (38 by one grade; 4 by two grades), was unchanged in 52/95 (55%), and improved in 1/95 (1%). NP inhomogeneity grade increased in 13/95 (14%), decreased in 2/95 (2%), and was unchanged in 80/95 (84%). DH decreased in 34/95 (36%) (28 by one grade; 5 by two grades; 1 by three grades), and was unchanged in 61/95 (64%). The distribution of DH within Pfirrmann grades is shown in Table 4.

Correlation between the longitudinal change in Pfirrmann grade and the individual visually assessed disc components was strongest with AF border distinction ($r = 0.68$; 95% CI 0.55–0.77), moderate with NP signal intensity ($r = −0.40$; 95% CI −0.55 to −0.21) and NP inhomogeneity ($r = 0.38$; 95% CI 0.19–0.54), and weak with DH ($r = 0.22$; 95% CI 0.01–0.40). For summary scores, the correlations with the Pfirrmann summary score were $r = 0.67$ (AF border), $r = −0.57$ (NP signal intensity), $r = 0.21$ (NP inhomogeneity), and $r = 0.04$ (DH) (Table 5).

Although the mean NP signal intensity declined overall, a subset of discs (n = 15) showed increases in NP signal during follow-up. Most of these discs were at the lower lumbar levels L4/5 and L5/S1 (13/15) and exhibited a decrease in DH (11/15).

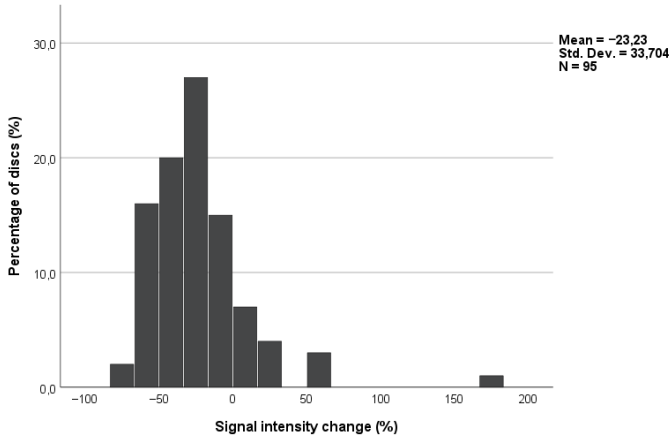


Figure 13 Histogram of the percentage change in NP signal intensity between the first and second MRI. The distribution is centred below zero, since most discs showed a decrease in NP signal intensity over 14 years. However, there were some discs that exhibited paradoxical signal increase.

Table 4 Disc height distribution within Pfirrmann grades at age 37 (A) and 51 (B). Each panel shows the count (percentage within row) of discs with normal height, slight height decrease (1–33%), moderate height decrease (34–66%), and severe height loss (67–100%) for Pfirrmann grades II–V (n = 95 discs per timepoint).

A

Pfirrmann grade	Disc height				
	Normal	Slight (1–33% decreased)	Moderate (34–66% decreased)	Severe (67–100% decreased)	Total
II	44 (75%)	12 (20%)	3 (5%)	0	59 (100%)
III	1 (6%)	10 (63%)	5 (31%)	0	16 (100%)
IV	1 (6%)	10 (55%)	7 (39%)	0	18 (100%)
V	0	0	0	2 (100%)	2 (100%)

B

Pfirrmann grade	Disc height				
	Normal	Slight (1–33% decreased)	Moderate (34–66% decreased)	Severe (67–100% decreased)	Total
II	21 (84%)	4 (16%)	0	0	25 (100%)
III	8 (25%)	16 (50%)	8 (25%)	0	32 (100%)
IV	2 (7%)	10 (32%)	15 (48%)	4 (13%)	31 (100%)
V	0	0	0	7 (100%)	7 (100%)

Table 5 The Spearman correlation coefficients between changes in the Pfirrmann grade and nucleus pulposus (NP) signal intensity, NP inhomogeneity, annulus fibrosus (AF) border distinction and disc height during follow-up A) across all 95 discs (representing all lumbar discs from the participants) and B) the summary scores (calculated as the total scores of the five lumbar discs for each participant). Reprinted from: Murto N, Lund T, Kautiainen H, Luoma K, Kerttula L, Progression of lumbar disc degeneration: A 14-year follow-up study examining Pfirrmann grading and its individual disc components. *Acta Radiologica Open.* 2025;14(9). Copyright © 2025 by the Author(s). Reprinted by permission of SAGE Publications.

A

	Correlation with Pfirrmann change (n = 95) (r)	**95% Confidence interval (CI)**
NP signal intensity change (n = 95)	-0.40	-0.55 to -0.21
NP inhomogeneity change (n = 95)	0.38	0.19 to 0.54
AF border distinction change (n = 95)	0.68	0.55 to 0.77
Disc height change (n = 95)	0.22	0.01 to 0.40

B

	Correlation with Pfirrmann summary score (r)	**95% Confidence interval (CI)**
NP signal intensity summary score	-0.57	-0.81 to -0.15
NP inhomogeneity summary score	0.21	-0.27 to 0.60
AF border distinction summary score	0.67	0.31 to 0.86
Disc height summary score	0.04	-0.42 to 0.48

6 DISCUSSION

6.1 Validity of T2-weighted nucleus pulposus signal intensity measurements (Study I)

Our first study demonstrated that the quantitative method for measuring the T2-weighted signal intensity of the NP is a repeatable and valid method for assessing DD. The intra- and inter-observer agreement of the measured signal intensity of NP was excellent. The values also showed a strong negative correlation with the qualitative grading scales — the Pfirrmann grade and the visually assessed NP signal intensity grade. This finding aligns with prior studies in which the decreased signal intensity of NP on T2WI — a known hallmark of degeneration caused by proteoglycan loss and dehydration and increased fibrosis in the NP — can be captured objectively (22,24,26).

Crucially, the validity of the quantitative approach depended on how the ROI was defined. We found that a targeted ROI placed on the brightest, most homogeneous NP region, excluding the AF and any dark INC, provided a good differentiation between degeneration grades. The targeted ROI improved the discrimination of visual NP signal intensity grades. This result indicates that including low-intensity structures — such as the collagenous outer AF, fibrous INC, or vertebral endplates — in the ROI lowers the average NP signal and can obscure differences between healthy and degenerated discs.

The targeted measurement of the NP region with the highest T2 signal more specifically reflects the water content of the NP, thereby improving sensitivity to detect degenerative changes. This agrees with earlier studies like Nagashima et al. (137), who reported that focusing on the high-intensity centre of the disc provides the most accurate discrimination of degeneration severity. Another application of the highest signal intensity measurement is the histogram analysis, which quantifies the separation of the NP and AF intensity peaks; a reduced peak separation refers to increasing degeneration (155,156). Both of these methods, CSF normalised signal intensity and histogram analysis of the disc, have been successfully integrated into AI applications as well (185,186).

Likewise, our study highlighted the value of choosing an appropriate reference for normalising NP signal intensity. Because the raw MRI signal varies with scanner settings, we normalised the NP signal intensity with an internal reference. We found that normalising the signal intensity values of the NP by using the CSF in the thecal sac as the reference provided more consistent results than normalising using the adjacent vertebral marrow. The NP/CSF

signal intensity ratio had a stronger correlation with visual grades than the NP/bone marrow signal intensity ratio. This is likely because CSF T2 signal intensity provides a relatively uniform signal, whereas the adjacent vertebral marrow provides a reference that reflects both image characteristics and status of the bone marrow (145,211). Our findings therefore support the widely adopted practice of CSF normalisation for quantitative disc signal assessments (22,24,48,107,137). We note, however, that even with proper normalisation, T2-weighted signal intensity remains a relative metric rather than an absolute physical measure. Differences in acquisition parameters between scans can affect the intensity values.

The quantitative NP signal intensity metric proved to be an objective complement to conventional grading. It essentially translates the observer's impression of disc darkening into a continuous scale. One advantage is its ability to detect subtle changes within the NP signal intensity — i.e., the continuum of degeneration that underlies the coarse categorical grades. That said, our findings also illustrate that the quantitative NP signal intensity is not a sufficient DD marker on its own. For instance, a targeted ROI measures the brightest NP region, but if a degenerated disc contains focal areas of high signal intensity, the measurement may be paradoxically high despite the overall degeneration. In our data, such cases could be identified by visual inspection as a bright focal spot that we termed the intranuclear high-intensity zone (INHIZ) within a generally dark disc on T2WI. Thus, the quantitative assessment should be interpreted alongside morphologic evaluation to avoid misinterpretation of focal abnormalities.

Discrimination between the adjacent Pfirrmann categories by the signal intensity measurements was also limited. Differences between grades II versus III and IV vs V were non-significant for all ROI approaches, whereas differences between grades III vs IV were significant across every ROI method. This likely reflects that the Pfirrmann classification incorporates NP/AF border distinction and DH, in addition to NP signal intensity, so grade boundaries do not depend on signal intensity assessments only. In our cohort, some Pfirrmann grade V discs showed a relatively broad range of NP signal intensity values. These findings suggest that quantitative NP signal intensity captures one sign of degeneration, NP hydration, extremely well, but cannot alone reflect all structural changes.

Therefore, a combined approach may be more informative. Quantitative NP signal may serve as a reliable marker of NP matrix status, while other structural changes — such as annular integrity and DH — are assessed separately. Overall, Study I establishes that a carefully standardised NP signal intensity measurement is a valid and sensitive indicator of DD, supporting its use in research alongside other disc gradings to reduce subjectivity.

6.2 SpineNet versus radiologist evaluation (Study II)

Study II compared an open-source DL model (SpineNetV2) with an expert radiologist for Pfirrmann grading at 2 time points separated by 14 years. Agreement between SpineNet and the radiologist ranged from fair to substantial across disc levels (κ = 0.32–0.76 at both assessments). When averaged across levels, κ was 0.74 at baseline and 0.68 at follow-up, which is in the same range as prior external validations reporting moderate-to-substantial agreement for Pfirrmann grading (39–41).

Notably, discrepancies were observed. The model applied Pfirrmann grades I and V more liberally than the radiologists. It labelled several discs as Pfirrmann grade I (healthy) that the radiologists considered grade II or III, and it overlabelled grade V in some cases where the human readers graded a III or IV. A comparable tendency to over-assign grades I and V was reported in external validation of SpineNet in an older cohort (mean age 72 ± 8.8 years) by Grob et al. (40). In the validation study by McSweeney et al. (39), Pfirrmann grades I and II were combined — so the frequency of grade I in this study cannot be determined — but the model still assigned more grades I+II and slightly more grade V than radiologists. By contrast, in the validation by Nigru et al. (41) of middle-aged participants (mean age 54 ± 15.4 years), radiologists assigned more grade I than SpineNet, whereas SpineNet assigned more grade V than the radiologists.

It is likely that some discrepancies reflect the model inheriting the subjective bias of its training labels. SpineNet was originally trained on the Genodisc dataset (12,018 individual discs, mean participant age 50.9 ± 14.6 years), graded by a single radiologist. The prevalence of Pfirrmann grade I labels in that training set (n = 3862) is noteworthy, given that grade I discs are expected to be uncommon in middle-aged adults (212) and the INC, which is one of the features of Pfirrmann grade II, is commonly seen in individuals aged 30 years and older (76). This highlights a general concern: an AI trained on one reader's assessments will inherently reproduce that reader's interpretation rather than providing an objective assessment.

Disc-level patterns were also evident in our data. The model tended to underestimate degeneration at upper lumbar levels and slightly overestimate at lower lumbar levels. The same disagreement pattern persisted at both time points, suggesting differences in the model calibration compared to our radiologist grading rather than random error. This consistent disc-level discrepancy also likely reflects calibration inherited from the training data, with SpineNet tending to assign milder grades at upper levels and higher grades at lower levels.

However, SpineNet captured the expected progression of degeneration on average. Both radiologists and the model detected an overall increase in summary Pfirrmann scores over time, indicating that the model can reliably capture gross changes. Importantly, despite the scanner change (1.0 T to 1.5 T), agreement metrics were comparable at both time points, indicating stable performance across different kinds of image data. However, disc-level discrepancies between SpineNet and the radiologist were amplified when summed across five levels: agreement for the Pfirrmann summary score was poor (CCC = 0.54), and agreement for change in the summary score over 14 years was also poor (CCC = 0.41), indicating systematic disagreements.

These results underscore both the potential and current limitations of AI-based grading. Automated grading is most useful as a standardised labelling tool for large MRI cohorts, reducing inter-observer variability. Looking forward, training on age-diverse, multi-centre datasets with multi-reader consensus is needed before such models match radiologist performance (180). We also note that the 'black-box' nature of CNNs obscures the reasoning behind the model's predictions. Labelling errors occurred with anatomical variants such as a small remnant S1/S2 disc or lumbosacral transitional vertebrae, as documented in the previous validation studies as well (39,40). Additionally, the model is limited to the degenerative features it was trained to recognise and cannot identify unexpected pathologies such as infection, neoplasm, or inflammatory changes. In practice, therefore, an AI like SpineNet should be used to assist or complement radiologist assessment rather than replace it.

In summary, Study II illustrates that while automated grading of lumbar DD is feasible and can reproduce broad patterns of degeneration, it has limitations and is not yet interchangeable with an expert radiologist. Algorithm refinement and broader validation of the AI model are required to improve accuracy and generalisability. Nevertheless, the AI's consistency and capacity to process large datasets efficiently make it valuable for large-scale epidemiologic studies of DD, where it can markedly accelerate image analysis while maintaining standardised criteria across thousands of images.

6.3 Progression of disc degeneration, assessed with Pfirrmann components (Study III)

Study III characterised how lumbar DD evolved from the late 30s to early 50s in our cohort by comparing the change in the degenerative grade of the discs between the two time points using both the Pfirrmann grade and the grades of individual disc features (NP signal intensity, NP inhomogeneity, AF border distinction, and DH). Over the 14-year follow-up, the overall degeneration progressed as expected. The mean Pfirrmann summary score increased by three points. At age 37, the majority of discs were Pfirrmann grade II (mild degeneration), with fewer in grades III or IV and almost none in grade V. By age 51, most discs had progressed into grades III–IV, with only a small minority reaching grade V. In practical terms, many discs that were initially mildly degenerated became moderately degenerated over 14 years, consistent with age-related DD (212). Overall, the CSF-normalised T2 signal intensity of the NP decreased substantially across lumbar levels (mean decrease of about 23%). Notably, a subset of discs — mostly at the lower lumbar levels — showed paradoxical increases in NP T2 signal intensity on follow-up.

By disaggregating the Pfirrmann classification into its component features, Study III revealed heterogeneity in how different structures of the discs change over time. The change in the AF border distinction was the single strongest predictor of change in the Pfirrmann grade. In our data, blurring of the normally sharp NP–AF boundary had the highest correlation with an increase in Pfirrmann grade (Spearman r = 0.68). Most discs that progressed to a higher Pfirrmann grade also exhibited a blurring or loss of the clear NP–AF demarcation over time. This is expected since the AF-border distinction is a key discriminator between mid-range Pfirrmann grades II and III: grade II discs already have a lightly inhomogeneous NP but still have a clear NP–AF boundary, whereas in grade III the AF border is unclear and in grade IV indistinct (11).

By contrast, the changes in the other Pfirrmann components showed only moderate to weak associations with Pfirrmann grade progression. Changes in NP inhomogeneity of the discs were only modestly correlated with Pfirrmann grade change (r = 0.38). This moderate correlation can be understood considering the baseline condition of many discs. There were no Pfirrmann grade I discs, most of the grade II discs already had an inhomogeneous NP at age 37, and in only a small proportion of discs the NP inhomogeneity grade increased.

Similarly, change in DH only had a weak correlation with Pfirrmann grade change (r = 0.22). This is expected because loss of DH is generally a late-stage phenomenon in degeneration. In our cohort, DH varied across Pfirrmann

grades II–IV and often remained normal or only mildly reduced. Even many of the grade IV discs did not show a significantly decreased DH. Moreover, in 64% of the discs the DH grade did not change. This pattern is consistent with a prior quantitative study that DH does not substantially vary between Pfirrmann grades I–IV and is markedly decreased only at grade V (22). This raises an argument that DH should be assessed independently of the Pfirrmann grade. Given that inter-rater agreement for visual DH was weak ($\kappa \approx 0.42$), and that even some Pfirrmann grade II discs received DH grades of slight or moderate decrease in our readings, a quantitative DH assessment (e.g., measured as absolute height or a height index) may be the preferred approach for evaluating DH, especially in longitudinal studies (213).

Finally, our incorporation of quantitative NP signal intensity measurements provided an additional perspective on the progression of DD. On average, the NP T2-weighted signal intensity decreased significantly over the 14 years. This reflects the expected loss of proteoglycans and water content and increased fibrosis of the NP during ageing of the disc (10,66,110). However, we observed some paradoxical increases in NP signal intensity in a small subset of discs. Most of these were detected in lower lumbar discs (L4/5 or L5/S1). These findings of signal intensity increase are not a sign of regeneration or improvement in the disc health. Instead, they likely represent intradiscal fluid accumulation in vacuum clefts, a phenomenon associated with advanced DD and supine positioning (214,215). While the T2-weighted signal is sensitive in detecting a decrease in water content and the associated progression of DD, the increase in signal intensity as a result of fluid accumulation in degenerated discs can reduce its specificity. Therefore, structural markers — like DH — are still needed when interpreting changes in DD findings in longitudinal studies. Proteoglycan- or glycosaminoglycan-sensitive techniques, such as T1ρ mapping and gagCEST, might be less affected by intradiscal fluid accumulation and therefore represent promising biomarkers for longitudinal follow-up of DD (142,163).

However, degeneration did not progress in every disc according to Pfirrmann grading — roughly half of the discs remained stable in Pfirrmann grade over the decade-long follow-up, in accordance with observations of a longitudinal MRI study in which a subset of discs shows little or no progression over a decade (45). In our cohort, even among discs that remained stable in Pfirrmann grade, DH decreased in 21% and the NP signal intensity decreased by ≥ 10% in 71%.

These component-specific observations suggest that the Pfirrmann grading may have limited sensitivity for detecting subtle longitudinal changes. An important insight from this study is that the Pfirrmann grading system compresses the multi-dimensional degenerative changes into a single ordinal scale, which makes it relatively insensitive to component-level changes. A

disc's Pfirrmann grade is a coarse summary of the NP signal, NP homogeneity, AF border distinction, and DH, but possibly it may not capture when one component changes while others do not. This underscores that relying solely on Pfirrmann grades may underestimate the true progression and variability of degeneration in longitudinal studies.

In summary, by analysing each Pfirrmann component separately, this study provided a more detailed characterisation of DD progression. The AF border distinction was found as the most sensitive indicator of a disc advancing to a higher Pfirrmann grade in our study population, distinguishing moderate from mild degeneration. In contrast, changes in NP inhomogeneity were often minimal if the NP was already inhomogeneous at baseline, and DH typically changed more slowly and correlated more weakly with the change of Pfirrmann grade. These findings highlight the value of component-level assessment. Component-based grading can reveal changes that an overall Pfirrmann grade might overlook. Evaluating individual features — whether with refined visual grading or quantitative measures — may therefore provide a more sensitive method to study the progression of DD. For instance, following the NP signal-intensity ratio over time provides a continuous metric of biochemical degeneration, and measuring actual DH can detect subtle height reductions objectively. Such a level of detail is particularly important as AI is integrated into spinal imaging: models trained only on coarse Pfirrmann grades may overlook the multifaceted nature of DD.

Given these insights, a multi-label, component-based grading approach may be more recommended for future research and AI applications. Instead of predicting a single Pfirrmann grade, an algorithm could estimate each disc component separately — e.g., NP signal intensity, NP homogeneity, AF border, and DH. This would allow capturing the full spectrum of degenerative changes with greater sensitivity. A multi-parameter assessment reflects that DD is not a binary 'grade up or not' process, but a continuous, multifaceted progression.

6.4 Limitations and methodological considerations

Across Studies I–III, all analyses were based on a small population. The longitudinal series consisted of 19 participants from the original cohort of 75 male who completed both MRI examinations at the ages of 37 and 51, indicating considerable participant dropout over the 14-year follow-up. Such losses to follow-up reduce statistical power and introduce selection bias. As a result, variability and the apparent rate of degeneration may be underestimated, and generalisability to broader populations is limited. Generalisability is further

limited by the homogeneity of the population, as all the participants were male and approximately the same age.

Baseline and follow-up MRI were acquired on different scanners, 1.0 T versus 1.5 T. Although we used the same slice orientation, applied identical grading criteria at both time points and signal intensities were normalised with an internal reference, differences in resolution, signal-to-noise ratio, and contrast may still have affected both the visual gradings and the quantitative signal measurements. Because MR signal intensities were expressed in arbitrary, sequence-dependent units, the obtained absolute signal intensity values in this study should not be transferred to different protocols.

In Study I, validation of quantitative NP signal intensity relied on qualitative visual grades (Pfirrmann and NP brightness) as the reference. Because these visual assessments are subjective, they inevitably introduce some misclassification and thereby limit confidence that the quantitative measure accurately reflects the degree of DD. The sample did not contain any Pfirrmann grade I discs and relatively few grade V discs, so the full range of degeneration was incompletely represented. This limits the ability to evaluate the method's performance in non-degenerated discs and at the most advanced stage of degeneration. Manual ROI placement — despite excellent repeatability — is time-consuming and operator dependent. Automated and standardised ROI selection could make the method more feasible for larger datasets and reduce inter-rater variance.

In Study II, evaluation of the DL model was based on 19 participants (95 discs), which is modest for external validation. Although inter-rater agreement between two radiologists was substantial to almost perfect, we did not establish a multi-reader consensus ground truth. Moreover, only Pfirrmann grades were assessed, so the model's performance for other degenerative features (e.g., endplate changes) was not examined.

In Study III, generalisability is constrained by an all-male, pre-selected symptomatic cohort (history of LBP in early adulthood). In our study, the reading of baseline and follow-up MRIs was performed in separate sessions. For the longitudinal study design, a paired, side-by-side reading of baseline and follow-up MRIs may be preferable to increase the sensitivity to detect subtle structural changes between time points. The analysis emphasised disc markers and did not systematically grade other DD findings (e.g., Modic/endplate changes, bulging discs or protrusions, annular fissures, facet arthrosis, central canal or foraminal stenosis) that may relate to disc changes and symptoms. Finally, longitudinal imaging relied on conventional T2-weighted sequences. Advanced quantitative techniques (e.g., T2 or T1ρ mapping) were not included.

6.5 Synthesis across Studies I–III

The overall aim of this thesis was to characterise lumbar DD on MRI with three approaches: quantitative assessment of NP T2-weighted signal intensity, automated AI-based grading, and longitudinal evaluation of changes in Pfirrmann grades and individual disc components.

Study I introduced a reliable, standardised NP signal intensity measurement that can provide an objective and continuous method for DD assessment and complement qualitative grading methods such as the Pfirrmann classification. A limitation of the NP signal intensity measurement is that it reflects mainly reduced proteoglycan and water content and increased fibrosis and therefore cannot capture other structural changes in the IVD. This underscores the need to assess other disc components alongside quantitative NP signal intensity.

Study II further demonstrated that the categorical nature of the Pfirrmann classification can constrain the performance of AI models used for DD grading. The DL model showed persistent level-specific disagreements with the radiologist and a tendency to assign end-of-scale grades. This suggests that AI systems trained only on Pfirrmann labels are likely to inherit the subjectivity bias of categorical grading from their training data. When considered alongside Study I, the findings suggest that supplementing categorical DD grades with continuous metrics such as quantitative NP signal intensity could reduce this bias originating from the model's training labels. In the context of Study III, incorporating quantitative measures may be necessary to enable AI models to detect subtle longitudinal changes that Pfirrmann grading alone does not capture.

Study III deepened the understanding of DD progression by showing that DD is a multidimensional process in which AF border loss, NP inhomogeneity, and DH reduction may evolve at different rates. Importantly, many discs that did not change Pfirrmann grade nevertheless exhibited a decrease in NP signal intensity or DH, highlighting that the Pfirrmann system lacks sensitivity to disc-component-specific progression. These findings support the use of component-based DD assessment in both radiological evaluation and AI development.

6.6 Future prospects

This thesis shows that lumbar DD is a multidimensional process, and future research should focus on methods that capture this complexity more accurately. Future studies should focus on developing quantitative, component-based, and AI-assisted approaches that could serve as more standardised reference methods for DD assessment in spine research. A practical approach is to

expand the use of quantitative MRI. Study I showed that quantitative NP signal intensity can be measured reliably on routine sagittal T2WIs, making it suitable for retrospective and longitudinal datasets. Future cohorts should incorporate this metric alongside advanced quantitative sequences, such as T2 or T1ρ mapping or gagCEST, to detect biochemical changes in the disc more sensitively. Study III indicates that assessing the degenerative changes in NP, AF, and DH separately can provide a more detailed view of DD progression than relying on a single composite grade. Such an approach could better characterise longitudinal changes and improve sensitivity to detecting DD progression. Study II demonstrated that AI models may reproduce the limitations of their training labels, underscoring the need to improve AI training (180). Future AI development should include age-diverse training samples and rely on multi-reader consensus grading to reduce subjective bias. Models should aim to incorporate quantitative MRI measures and predict changes of individual disc components to improve sensitivity to longitudinal DD changes.

7 CONCLUSIONS

Study I demonstrated that quantitative NP signal intensity measured on routine T2-weighted MRI is a reliable, normalisation-dependent marker of DD. When the NP signal was measured with an ROI targeted to the brightest central NP and normalised with CSF, the metric showed excellent repeatability and strong correspondence with Pfirrmann grade and visual grading of NP signal intensity. As a continuous metric, it complements visual grading by improving objectivity and sensitivity to change, although we acknowledge its vulnerability to the heterogeneity of the signal intensity in the degenerated disc.

Study II showed that DL-based automated DD grading is feasible but not yet interchangeable with radiologist assessment. The model achieved fair-to-substantial agreement with the radiologist but assigned lower Pfirrmann grades at the upper lumbar levels and higher grades at the lower levels, with a tendency to assign end-of-scale grades. These findings highlight that, at present, AI systems should be viewed as adjuncts for DD evaluation, capable of rapid and standardised analysis of large datasets, rather than replacements for radiologists.

Study III revealed that over 14 years, overall Pfirrmann grades increased, but changes in NP, AF borders, and DH evolved at different rates. Using component-level grading and quantitative NP signal-intensity measurement, we found that the Pfirrmann grade alone was insufficient to detect the subtle progression of DD.

Taken together, the findings from Studies I–III suggest that MRI-based DD assessment may benefit from combining quantitative, component-based, and automated methods rather than relying solely on a single composite grading such as the Pfirrmann classification. Quantitative MRI offers objective and continuous metrics of disc changes, while component-level grading captures structural variation across NP, AF, and DH that Pfirrmann grading may overlook. AI methods can provide scalable and standardised analysis. By demonstrating the strengths and limitations of these complementary approaches to DD evaluation, this thesis provides a framework for developing more precise, reproducible, and comprehensive methods for assessing lumbar DD in future spine research.

ACKNOWLEDGEMENTS

This doctoral thesis was carried out at the Helsinki University Hospital. I wish to express my deepest gratitude to my supervisors, Docents Liisa Kerttula, Teija Lund and Katariina Luoma, for their invaluable guidance, expertise, and encouragement throughout this project. Their insights into research were instrumental in shaping this thesis. I am also thankful to Hannu Kautiainen, for statistical consultation, particularly in navigating the complex longitudinal data analysis.

I wish to express my sincere gratitude to the official reviewers of this dissertation, Docents Pekka Niemi and Juhani Määttä for their thorough and insightful evaluation of this work. Their constructive comments and valuable suggestions improved the clarity and scientific quality of the thesis, and I am grateful for the time and expertise they devoted to this thesis.

I am honoured to have Professor Jaakko Niinimäki serve as the opponent of this dissertation. I sincerely thank him for accepting this role and for his interest in my work.

I would also like to thank all my colleagues at the HUS Diagnostic Center, Department of Radiology, for providing a supportive and collegial working environment. I am grateful for the time and support you have given me during the different phases of this project. Special thanks to all my coworkers at Jorvi Hospital — thank you for the support, humour, and teamwork you bring to everyday work. You have made even the most challenging times easier and more rewarding.

I extend my appreciation to the University of Helsinki and the Doctoral Programme in Clinical Research and the support from the Musculoskeletal Radiologists of Finland Society, the Radiological Society of Finland and Pehr Oscar Klingendahl Fund and the funding from HUS Diagnostic Center that made this work possible — I gratefully acknowledge these institutions.

I sincerely thank the participants of the longitudinal study. Their commitment over 14 years to undergo MRI examinations made this thesis possible.

I am also grateful to my friends, who have shared both the challenges and joys of life with me and helped me take my mind off the demands of work and research. Special thanks to some of my course mates from the University of Oulu Faculty of Medicine — you are truly remarkable friends.

Finally, most importantly, I am forever grateful to my family. To my mom and late father, thank you for your love and encouragement to pursue education — without your guidance and support, I would not have come this far. To my

wife Veera, for your love, understanding, and encouragement that provided me the strength to balance life's challenges. Lastly, to our Väinö and newborn O, who are a constant source of joy — you remind me of the important things.

Thank you all for being there for me. This thesis is as much a result of your support as it is of my own efforts.

Helsinki, January 2026

Niko Murto

REFERENCES

1. Buchbinder R, van Tulder M, Öberg B, Costa LM, Woolf A, Schoene M, et al. Low back pain: a call for action. Lancet. 2018 Jun 9;391(10137):2384–8.
2. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol. 2023 May 22;5(6):e316–e329.
3. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet. 2018 Jun 9;391(10137):2356–2367.
4. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, et al. MRI findings of disc degeneration are more prevalent in adults with low back pain than in asymptomatic controls: A systematic review and meta-analysis. AJNR Am J Neuroradiol. 2015 Dec 1;36(12):2394–9.
5. Takatalo J, Karppinen J, Niinimäki J, Taimela S, Näyhä S, Mutanen P, et al. Does lumbar disc degeneration on magnetic resonance imaging associate with low back symptom severity in young Finnish adults? Spine (Phila Pa 1976). 2011 Dec 1;36(25):2180–9.
6. Mertimo T, Karppinen J, Niinimäki J, Blanco R, Määttä J, Kankaanpää M, et al. Association of lumbar disc degeneration with low back pain in middle age in the Northern Finland Birth Cohort 1966. BMC Musculoskelet Disord. 2022 Apr 15;23(1):359.
7. Mohd Isa IL, Teoh SL, Mohd Nor NH, Mokhtar SA. Discogenic Low Back Pain: Anatomy, Pathophysiology and Treatments of Intervertebral Disc Degeneration. Int J Mol Sci. 2022 Dec 22;24(1):208.
8. Adams MA, Roughley PJ. What is intervertebral disc degeneration, and what causes it? Spine (Phila Pa 1976). 2006 Aug 15;31(18):2151–61.
9. Kirnaz S, Capadona C, Lintz M, Kim B, Yerden R, Goldberg JL, et al. Pathomechanism and Biomechanics of Degenerative Disc Disease: Features of Healthy and Degenerated Discs. Int J Spine Surg. 2021 Apr;15(s1):10–25.
10. Tertti M, Paajanen H, Laato M, Aho H, Komu M, Kormano M. Disc degeneration in magnetic resonance imaging. A comparative biochemical, histologic, and radiologic study in cadaver spines. Spine (Phila Pa 1976). 1991 Jun;16(6):629–34.
11. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. Spine (Phila Pa 1976). 2001 Sep 1;26(17):1873–8.
12. Benneker LM, Heini PF, Anderson SE, Alini M, Ito K. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. Eur Spine J. 2005 Feb;14(1):27–35.

13. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, et al. Systematic literature review of imaging features of spinal degeneration in asymptomatic populations. AJNR Am J Neuroradiol. 2015 Apr 1;36(4):811–6.

14. Esposito D, Brown B, Hancock MJ, King SSG, Searant IGT, Jenkins H. MRI-Based Grading Systems for Assessing Lumbar Disc Degeneration: A Scoping Review. JOR Spine. 2025 Sep 15;8(3):e70113.

15. Urrutia J, Besa P, Campos M, Cikutovic P, Cabezon M, Molina M, et al. The Pfirrmann classification of lumbar intervertebral disc degeneration: an independent inter- and intra-observer agreement assessment. Eur Spine J. 2016 Sep 15;25(9):2728–33.

16. Nagy SA, Juhasz I, Komaromy H, Pozsar K, Zsigmond I, Perlaki G, et al. A statistical model for intervertebral disc degeneration: determination of the optimal T2 cut-off values. Clin Neuroradiol. 2014 Dec;24(4):355–63.

17. Soydan Z, Bayramoglu E, Urut DU, Iplikcioglu AC, Sen C. Tracing the disc: The novel qualitative morphometric MRI based disc degeneration classification system. JOR Spine. 2024 Mar;7(1):e1321.

18. Griffith JF, Wang YXJ, Antonio GE, Choi KC, Yu A, Ahuja AT, et al. Modified Pfirrmann Grading System for Lumbar Intervertebral Disc Degeneration. Spine (Phila Pa 1976). 2007 Nov;32(24):E708–12.

19. Vadapalli R, Mulukutla R, Vadapalli AS, Vedula RR. Quantitative Predictive Imaging Biomarkers of Lumbar Intervertebral Disc Degeneration. Asian Spine J. 2019 Aug;13(4):527–34.

20. Auerbach JD, Johannessen W, Borthakur A, Wheaton AJ, Dolinskas CA, Balderston RA, et al. In vivo quantification of human lumbar disc degeneration using T(1rho)-weighted magnetic resonance imaging. Eur Spine J. 2006 Aug;15 Suppl 3:S338-44.

21. Samartzis D, Borthakur A, Belfer I, Bow C, Lotz JC, Wang HQ, et al. Novel diagnostic and prognostic methods for disc degeneration and low back pain. The Spine Journal. 2015 Sep;15(9):1919–32.

22. Salamat S, Hutchings J, Kwong C, Magnussen J, Hancock MJ. The relationship between quantitative measures of disc height and disc signal intensity with Pfirrmann score of disc degeneration. Springerplus 2016 Dec 1;5(1):829.

23. Luoma EK, Raininko R, Nummi PJ, Luukkonen R, Manninen HI, Riihimäki HA. Suitability of cerebrospinal fluid as a signal-intensity reference on MRI: evaluation of signal-intensity variations in the lumbosacral dural sac. Neuroradiology. 1997 Oct 2;39(10):728–32.

24. Fan SW, Zhou ZJ, Hu ZJ, Fang XQ, Zhao FD, Zhang J. Quantitative MRI analysis of the surface area, signal intensity and MRI index of the central bright area for the evaluation of early adjacent disc degeneration after lumbar fusion. Eur Spine J. 2012 Sep;21(9):1709–15.

25. Niinimäki J, Korkiakoski A, Ojala O, Karppinen J, Ruohonen J, Haapea M, et al. Association between visual degeneration of intervertebral discs and the apparent diffusion coefficient. Magn Reson Imaging. 2009 Jun;27(5):641–7.

26. Hu X, Chen M, Pan J, Liang L, Wang Y. Is it appropriate to measure age-related lumbar disc degeneration on the mid-sagittal MR image? A quantitative image study. Eur Spine J. 2018 May 1;27(5):1073–81.

27. Kotilainen E, Alanen A, Erkintalo M, Valtonen S, Kormano M. Association between decreased disc signal intensity in preoperative T2-weighted MRI and a 5-year outcome after lumbar minimally invasive discectomy. Minim Invasive Neurosurg. 2001 Mar;44(1):31–6.

28. Gervais J, Périé D, Parent S, Labelle H, Aubin CE. MRI signal distribution within the intervertebral disc as a biomarker of adolescent idiopathic scoliosis and spondylolisthesis. BMC Musculoskelet Disord. 2012 Dec 3;13:239.

29. Kamei N, Nakamae T, Nakanishi K, Tamura T, Tsuchikawa Y, Morisako T, et al. Evaluation of intervertebral disc degeneration using T2 signal ratio on magnetic resonance imaging. Eur J Radiol. 2022 Jul;152:110358.

30. Niemeläinen R, Videman T, Dhillon SS, Battié MC. Quantitative measurement of intervertebral disc signal using MRI. Clin Radiol. 2008 Mar;63(3):252–5.

31. Hwang D, Kim S, Abeydeera NA, Statum S, Masuda K, Chung CB, et al. Quantitative magnetic resonance imaging of the lumbar intervertebral discs. Quant Imaging Med Surg. 2016 Dec;6(6):744–55.

32. Yoon MA, Hong SJ, Kang CH, Ahn KS, Kim BH. T1rho and T2 mapping of lumbar intervertebral disc: Correlation with degeneration and morphologic changes in different disc regions. Magn Reson Imaging. 2016 Sep;34(7):932–9.

33. Yang L, Sun C, Gong T, Li Q, Chen X, Zhang X. T1ρ, T2 and T2* mapping of lumbar intervertebral disc degeneration: a comparison study. BMC Musculoskelet Disord. 2022 Dec 27;23(1):1135.

34. Tamagawa S, Sakai D, Nojiri H, Sato M, Ishijima M, Watanabe M. Imaging Evaluation of Intervertebral Disc Degeneration and Painful Discs-Advances and Challenges in Quantitative MRI. Diagnostics (Basel). 2022 Mar 14;12(3):707.

35. Menezes-Reis R, Salmon CEG, Carvalho CS, Bonugli GP, Chung CB, Nogueira-Barbosa MH. T1ρ and T2 mapping of the intervertebral disk: comparison of different methods of segmentation. AJNR Am J Neuroradiol. 2015 Mar;36(3):606–11.

36. Liawrungrueang W, Park JB, Cholamjiak W, Sarasombath P, Riew KD. Artificial Intelligence-Assisted MRI Diagnosis in Lumbar Degenerative Disc Disease: A Systematic Review. Global Spine J. 2025 Mar;15(2):1405–18.

37. Compte R, Granville Smith I, Isaac A, Danckert N, McSweeney T, Liantis P, et al. Are current machine learning applications comparable to radiologist classification of degenerate and herniated discs and Modic change? A systematic review and meta-analysis. Eur Spine J. 2023 Nov 8;32(11):3764–87.

38. Jamaludin A, Lootus M, Kadir T, Zisserman A, Urban J, Battié MC, 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 May 1;26(5):1374–83.

39. McSweeney TP, Tiulpin A, Saarakkala S, Niinimäki J, Windsor R, Jamaludin A, et al. External Validation of SpineNet, an Open-Source Deep Learning Model for Grading Lumbar Disk Degeneration MRI Features, Using the Northern Finland Birth Cohort 1966. Spine (Phila Pa 1976). 2023 Apr 1;48(7):484–91.

40. Grob A, Loibl M, Jamaludin A, Winklhofer S, Fairbank JCT, Fekete T, et al. External validation of the deep learning system "SpineNet" for grading radiological features of degeneration on MRIs of the lumbar spine. Eur Spine J. 2022 Aug 14;31(8):2137–48.

41. Nigru AS, Benini S, Bonetti M, Bragaglio G, Frigerio M, Maffezzoni F, et al. External validation of SpineNetV2 on a comprehensive set of radiological features for grading lumbosacral disc pathologies. North American Spine Society journal. 2024;20:100564.

42. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Minamide A, et al. Prevalence and distribution of intervertebral disc degeneration over the entire spine in a population-based cohort: the Wakayama Spine Study. Osteoarthritis Cartilage. 2014 Jan;22(1):104–10.

43. Battié MC, Videman T, Parent E. Lumbar disc degeneration: epidemiology and genetic influences. Spine (Phila Pa 1976). 2004 Dec 1;29(23):2679–90.

44. Cheung KMC, Karppinen J, Chan D, Ho DWH, Song YQ, Sham P, et al. Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. Spine (Phila Pa 1976). 2009 Apr 20;34(9):934–40.

45. Makino H, Kawaguchi Y, Seki S, Nakano M, Yasuda T, Suzuki K, et al. Lumbar disc degeneration progression in young women in their 20's: A prospective ten-year follow up. J Orthop Sci. 2017 Jul;22(4):635–40.

46. Sharma A, Pilgram T, Wippold FJ. Association between Annular Tears and Disk Degeneration: A Longitudinal Study. AJNR Am J Neuroradiol. 2009 Mar;30(3):500–6.

47. Vergroesen PPA, Kingma I, Emanuel KS, Hoogendoorn RJW, Welting TJ, van Royen BJ, et al. Mechanics and biology in intervertebral disc degeneration: a vicious circle. Osteoarthritis Cartilage. 2015 Jul;23(7):1057–70.

48. Battié MC, Videman T, Gibbons LE, Fisher LD, Manninen H, Gill K. 1995 Volvo Award in clinical sciences. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. Spine (Phila Pa 1976). 1995 Dec 15;20(24):2601–12.

49. Videman T, Battié MC, Ripatti S, Gill K, Manninen H, Kaprio J. Determinants of the progression in lumbar degeneration: a 5-year follow-up study of adult male monozygotic twins. Spine (Phila Pa 1976). 2006 Mar 15;31(6):671–8.

50. Videman T, Battié MC, Parent E, Gibbons LE, Vainio P, Kaprio J. Progression and determinants of quantitative magnetic resonance imaging measures of lumbar disc degeneration: a five-year follow-up of adult male monozygotic twins. Spine (Phila Pa 1976). 2008 Jun 1;33(13):1484–90.

51. Williams FMK, Popham M, Sambrook PN, Jones AF, Spector TD, MacGregor AJ. Progression of lumbar disc degeneration over a decade: a heritability study. Ann Rheum Dis. 2011 Jul;70(7):1203–7.

52. Panjabi MM, White AA. Basic biomechanics of the spine. Neurosurgery. 1980 Jul;7(1):76–93.

53. Galbusera F, van Rijsbergen M, Ito K, Huyghe JM, Brayda-Bruno M, Wilke HJ. Ageing and degenerative changes of the intervertebral disc and their impact on spinal flexibility. Eur Spine J. 2014 Jun;23 Suppl 3:S324-32.

54. Urban JPG, Smith S, Fairbank JCT. Nutrition of the intervertebral disc. Spine (Phila Pa 1976). 2004 Dec 1;29(23):2700–9.

55. Jaumard NV, Welch WC, Winkelstein BA. Spinal facet joint biomechanics and mechanotransduction in normal, injury and degenerative conditions. J Biomech Eng. 2011 Jul;133(7):071010.

56. Gellhorn AC, Katz JN, Suri P. Osteoarthritis of the spine: the facet joints. Nat Rev Rheumatol. 2013 Apr;9(4):216–24.

57. Adams MA, Hutton WC. The effect of posture on the lumbar spine. J Bone Joint Surg Br. 1985 Aug;67(4):625–9.

58. Roughley PJ. Biology of Intervertebral Disc Aging and Degeneration. Spine (Phila Pa 1976). 2004 Dec;29(23):2691–9.

59. Iatridis JC, MacLean JJ, O'Brien M, Stokes IAF. Measurements of Proteoglycan and Water Content Distribution in Human Lumbar Intervertebral Discs. Spine (Phila Pa 1976). 2007 Jun;32(14):1493–7.

60. Roughley PJ, Alini M, Antoniou J. The role of proteoglycans in aging, degeneration and repair of the intervertebral disc. Biochem Soc Trans. 2002 Nov;30(Pt 6):869–74.

61. Marchand F, Ahmed AM. Investigation of the laminate structure of lumbar disc anulus fibrosus. Spine (Phila Pa 1976). 1990 May;15(5):402–10.

62. Urban JPG, Roberts S. Degeneration of the intervertebral disc. Arthritis Res Ther. 2003;5(3):120–30.

63. Eyre DR, Muir H. Quantitative analysis of types I and II collagens in human intervertebral discs at various ages. Biochim Biophys Acta. 1977 May 27;492(1):29–42.

64. Groh AMR, Fournier DE, Battié MC, Séguin CA. Innervation of the Human Intervertebral Disc: A Scoping Review. Pain Medicine. 2021 Jun 4;22(6):1281–304.

65. De Geer CM. Intervertebral Disk Nutrients and Transport Mechanisms in Relation to Disk Degeneration: A Narrative Literature Review. J Chiropr Med. 2018 Jun;17(2):97–105.

66. Antoniou J, Steffen T, Nelson F, Winterbottom N, Hollander AP, Poole RA, et al. The human lumbar intervertebral disc: evidence for changes in the biosynthesis and denaturation of the extracellular matrix with growth, maturation, ageing, and degeneration. Journal of Clinical Investigation. 1996 Aug 15;98(4):996–1003.

67. Evans W, Jobe W, Seibert C. A cross-sectional prevalence study of lumbar disc degeneration in a working population. Spine (Phila Pa 1976). 1989 Jan;14(1):60–4.

68. Horner HA, Urban JP. 2001 Volvo Award Winner in Basic Science Studies: Effect of nutrient supply on the viability of cells from the nucleus pulposus of the intervertebral disc. Spine (Phila Pa 1976). 2001 Dec 1;26(23):2543–9.

69. Hunter CJ, Matyas JR, Duncan NA. The notochordal cell in the nucleus pulposus: a review in the context of tissue engineering. Tissue Eng. 2003 Aug;9(4):667–77.

70. Boos N, Weissbach S, Rohrbach H, Weiler C, Spratt KF, Nerlich AG. Classification of age-related changes in lumbar intervertebral discs: 2002 Volvo Award in basic science. Spine (Phila Pa 1976). 2002 Dec 1;27(23):2631–44.

71. Vo NV, Hartman RA, Patil PR, Risbud M V, Kletsas D, Iatridis JC, et al. Molecular mechanisms of biological aging in intervertebral discs. J Orthop Res. 2016 Aug;34(8):1289–306.

72. Weiler C, Nerlich AG, Bachmeier BE, Boos N. Expression and Distribution of Tumor Necrosis Factor Alpha in Human Lumbar Intervertebral Discs: A Study in Surgical Specimen and Autopsy Controls. Spine (Phila Pa 1976). 2005 Jan;30(1):44–53.

73. Baumgartner L, Sadowska A, Tío L, González Ballester MA, Wuertz-Kozak K, Noailly J. Evidence-Based Network Modelling to Simulate Nucleus Pulposus Multicellular Activity in Different Nutritional and Pro-Inflammatory Environments. Front Bioeng Biotechnol. 2021 Nov 10:9:734258.

74. Quint U, Wilke HJ. Grading of degenerative disk disease and functional impairment: imaging versus patho-anatomical findings. Eur Spine J. 2008 Dec;17(12):1705–13.

75. Perry J, Haughton V, Anderson PA, Wu Y, Fine J, Mistretta C. The value of T2 relaxation times to characterize lumbar intervertebral disks: preliminary results. AJNR Am J Neuroradiol. 2006 Feb;27(2):337–42.

76. Aguila LA, Piraino DW, Modic MT, Dudley AW, Duchesneau PM, Weinstein MA. The intranuclear cleft of the intervertebral disk: magnetic resonance imaging. Radiology. 1985 Apr;155(1):155–8.

77. Schiebler ML, Camerino VJ, Fallon MD, Zlatkin MB, Grenier N, Kressel HY. In vivo and ex vivo magnetic resonance imaging evaluation of early disc degeneration with histopathologic correlation. Spine (Phila Pa 1976). 1991 Jun;16(6):635–40.

78. Prescher A. Anatomy and pathology of the aging spine. Eur J Radiol. 1998 Jul;27(3):181–95.

79. Buckwalter JA. Aging and degeneration of the human intervertebral disc. Spine (Phila Pa 1976). 1995 Jun 1;20(11):1307–14.

80. Ghezelbash F, Schmidt H, Shirazi-Adl A, El-Rich M. Internal load-sharing in the human passive lumbar spine: Review of in vitro and finite element model studies. J Biomech. 2020 Mar 26;102:109441.

81. Ito K, Creemers L. Mechanisms of intervertebral disk degeneration/injury and pain: a review. Global Spine J. 2013 Jun;3(3):145–52.

82. Kyere KA, Than KD, Wang AC, Rahman SU, Valdivia-Valdivia JM, La Marca F, et al. Schmorl's nodes. Eur Spine J. 2012 Nov;21(11):2115–21.

83. Livshits G, Popham M, Malkin I, Sambrook PN, MacGregor AJ, Spector T, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. Ann Rheum Dis. 2011 Oct;70(10):1740–5.

84. Munir S, Rade M, Määttä JH, Freidin MB, Williams FMK. Intervertebral Disc Biology: Genetic Basis of Disc Degeneration. Curr Mol Biol Rep. 2018;4(4):143–50.

85. Luoma K, Riihimäki H, Raininko R, Luukkonen R, Lamminen A, Viikari-Juntura E. Lumbar disc degeneration in relation to occupation. Scand J Work Environ Health. 1998 Oct;24(5):358–66.

86. Hangai M, Kaneoka K, Hinotsu S, Shimizu K, Okubo Y, Miyakawa S, et al. Lumbar intervertebral disk degeneration in athletes. Am J Sports Med. 2009 Jan;37(1):149–55.

87. Nasto LA, Ngo K, Leme AS, Robinson AR, Dong Q, Roughley P, et al. Investigating the role of DNA damage in tobacco smoking-induced spine degeneration. The Spine Journal. 2014 Mar;14(3):416–23.

88. Samartzis D, Karppinen J, Chan D, Luk KDK, Cheung KMC. The association of lumbar intervertebral disc degeneration on magnetic resonance imaging with body mass index in overweight and obese adults: A population□based study. Arthritis Rheum. 2012 May 26;64(5):1488–96.

89. Teraguchi M, Yoshimura N, Hashizume H, Yamada H, Oka H, Minamide A, et al. Progression, incidence, and risk factors for intervertebral disc degeneration in a longitudinal population-based cohort: the Wakayama Spine Study. Osteoarthritis Cartilage. 2017 Jul;25(7):1122–31.

90. Haschtmann D, Stoyanov JV., Gédet P, Ferguson SJ. Vertebral endplate trauma induces disc cell apoptosis and promotes organ degeneration in vitro. Eur Spine J. 2008 Feb 10;17(2):289–99.

91. Holm S, Holm AK, Ekström L, Karladani A, Hansson T. Experimental disc degeneration due to endplate injury. J Spinal Disord Tech. 2004 Feb;17(1):64–71.

92. Karchevskaya AE, Poluektov YM, Korolishin VA. Understanding Intervertebral Disc Degeneration: Background Factors and the Role of Initial Injury. Biomedicines. 2023 Oct 6;11(10):2714.

93. Samartzis D, Karppinen J, Mok F, Fong DYT, Luk KDK, Cheung KMC. A population-based study of juvenile disc degeneration and its association with overweight and obesity, low back pain, and diminished functional status. J Bone Joint Surg Am. 2011 Apr 6;93(7):662–70.

94. Dragsbæk L, Kjaer P, Hancock M, Jensen TS. An exploratory study of different definitions and thresholds for lumbar disc degeneration assessed by MRI and their associations with low back pain using data from a cohort study of a general population. BMC Musculoskelet Disord. 2020 Dec 17;21(1):253.

95. Tonosu J, Oka H, Higashikawa A, Okazaki H, Tanaka S, Matsudaira K. The associations between magnetic resonance imaging findings and low back pain: A 10-year longitudinal analysis. Espinoza Orías AA, editor. PLoS One. 2017 Nov 15;12(11):e0188057.

96. Fujii K, Yamazaki M, Kang JD, Risbud MV, Cho SK, Qureshi SA, et al. Discogenic Back Pain: Literature Review of Definition, Diagnosis, and Treatment. JBMR Plus. 2019 Mar 4;3(5):e10180.

97. Brisby H. Pathology and possible mechanisms of nervous system response to disc degeneration. J Bone Joint Surg Am. 2006 Apr;88 Suppl 2:68–71.

98. Diwan AD, Melrose J. Intervertebral disc degeneration and how it leads to low back pain. JOR Spine. 2023 Mar;6(1):e1231.

99. Wuertz K, Haglund L. Inflammatory mediators in intervertebral disk degeneration and discogenic pain. Global Spine J. 2013 Jun;3(3):175–84.

100. Lotz JC, Fields AJ, Liebenberg EC. The role of the vertebral end plate in low back pain. Global Spine J. 2013 Jun;3(3):153–64.

101. Ahn SH, Cho YW, Ahn MW, Jang SH, Sohn YK, Kim HS. mRNA expression of cytokines and chemokines in herniated lumbar intervertebral discs. Spine (Phila Pa 1976). 2002 May 1;27(9):911–7.

102. Brisby H, Olmarker K, Larsson K, Nutu M, Rydevik B. Proinflammatory cytokines in cerebrospinal fluid and serum in patients with disc herniation and sciatica. Eur Spine J. 2002 Feb;11(1):62–6.

103. Lee S, Moon CS, Sul D, Lee J, Bae M, Hong Y, et al. Comparison of growth factor and cytokine expression in patients with degenerated disc disease and herniated nucleus pulposus. Clin Biochem. 2009 Oct;42(15):1504–11.

104. Kuoppala A, Määttä J, Hanhivaara J, Niinimäki J, Nevalainen M. Lumbosacral transitional vertebra alters the mobility of the lumbar spine on flexion-extension radiographs. Int Orthop. 2025 Oct;49(10):2559–65.

105. Ruiz Santiago F, Láinez Ramos-Bossini AJ, Wáng YXJ, López Zúñiga D. The role of radiography in the study of spinal disorders. Quant Imaging Med Surg. 2020 Dec;10(12):2322–55.

106. Kushchayev S V., Glushko T, Jarraya M, Schuleri KH, Preul MC, Brooks ML, et al. ABCs of the degenerative spine. Insights Imaging. 2018 Apr 22;9(2):253–74.

107. Carragee EJ, Don AS, Hurwitz EL, Cuellar JM, Carrino JA, Herzog R. 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 Oct 1;34(21):2338–45.

108. Wolfer LR, Derby R, Lee JE, Lee SH. Systematic review of lumbar provocation discography in asymptomatic subjects with a meta-analysis of false-positive rates. Pain Physician. 2008;11(4):513–38.

109. Sharma SK, Jones JO, Zeballos PP, Irwin SA, Martin TW. The prevention of discitis during discography. Spine J. 2009 Nov;9(11):936–43.

110. Wagner M, Sether LA, Yu S, Ho PSP, Haughton VM. Age changes in the lumbar intervertebral disc studied with magnetic resonance and cryomicrotomy. Clin Anat. 1988;1(2):93–103.

111. Desigan S, Hall-Craggs MA, Ho CP, Eliahoo J, Porter JB. Degenerative disc disease as a cause of back pain in the thalassaemic population: A case-control study using MRI and plain radiographs. Skeletal Radiol. 2006 Feb;35(2):95–102.

112. Sorensen SJ, Kjaer P, Jensen ST, Andersen P. Low-field magnetic resonance imaging of the lumbar spine: Reliability of qualitative evaluation of disc and muscle parameters. Acta Radiol. 2006 Jan 1;47(9):947–53.

113. Askar Z, Wardlaw D, Muthukumar T, Smith F, Kader D, Gibson S. Correlation between inter-vertebral disc morphology and the results in patients undergoing Graf ligament stabilisation. Eur Spine J. 2004 Dec;13(8):714–8.

114. Southern EP, Fye MA, Panjabi MM, Patel TC, Cholewicki J. Disc degeneration: a human cadaveric study correlating magnetic resonance imaging and quantitative discomanometry. Spine (Phila Pa 1976). 2000 Sep 1;25(17):2171–5.

115. Frobin W, Brinckmann P, Kramer M, Hartwig E. Height of lumbar discs measured from radiographs compared with degeneration and height classified from MR images. Eur Radiol. 2001 Jan 16;11(2):263–9.

116. Schneiderman G, Flannigan B, Kingston S, Thomas J, Dillin WH, Watkins RG. Magnetic resonance imaging in the diagnosis of disc degeneration: correlation with discography. Spine (Phila Pa 1976). 1987 Apr;12(3):276–81.

117. Riesenburger RI, Safain MG, Ogbuji R, Hayes J, Hwang SW. A novel classification system of lumbar disc degeneration. Journal of Clinical Neuroscience. 2015 Feb 1;22(2):346–51.

118. Butler D, Trafimow JH, Andersson GB, McNeill TW, Huckman MS. Discs degenerate before facets. Spine (Phila Pa 1976). 1990 Feb;15(2):111–3.

119. Videman T, Battié MC, Gill K, Manninen H, Gibbons LE, Fisher LD. Magnetic Resonance Imaging Findings and Their Relationships in the Thoracic and Lumbar Spine. Spine (Phila Pa 1976). 1995 Apr;20(8):928–35.

120. Raininko R, Manninen H, Battié MC, Gibbons LE, Gill K, Fisher LD. Observer Variability in the Assessment of Disc Degeneration on Magnetic Resonance Images of the Lumbar and Thoracic Spine. Spine (Phila Pa 1976). 1995 May;20(9):1029–35.

121. Boos N, Wallin A, Schmucker T, Aebi M, Boesch C. Quantitative MR imaging of lumbar intervertebral disc and vertebral bodies: methodology, reproducibility, and preliminary results. Magn Reson Imaging. 1994;12(4):577–87.

122. Masaryk TJ, Ross JS, Modic MT, Boumphrey F, Bohlman H, Wilber G. High-resolution MR imaging of sequestered lumbar intervertebral disks. AJR Am J Roentgenol. 1988 May;150(5):1155–62.

123. Thompson JP, Pearce RH, Schechter MT, Adams ME, Tsang IK, Bishop PB. Preliminary evaluation of a scheme for grading the gross morphology of the human intervertebral disc. Spine (Phila Pa 1976). 1990 May;15(5):411–5.

124. Wang YXJ. Several concerns on grading lumbar disc degeneration on MR image with Pfirrmann criteria. J Orthop Translat. 2022 Jan 1;32:101–2.

125. Rim DC. Quantitative Pfirrmann Disc Degeneration Grading System to Overcome the Limitation of Pfirrmann Disc Degeneration Grade. Korean J Spine. 2016;13(1):1.

126. Dragsbæk L, Jensen TS, Kjær P, Hancock M, Clausen S, Jensen RK. Associations Between Sum Scores or Combinations of MRI Findings in the Lumbar Spine and Low Back Pain-Related Outcomes: A Systematic Review. Eur J Pain. 2025 Aug;29(7):e70076.

127. Fardon DF, Williams AL, Dohring EJ, Murtagh FR, Rothman SLG, Sze GK. Lumbar disc nomenclature: Version 2.0: Recommendations of the combined task forces of the North American spine society, the american society of spine radiology, and the american society of neuroradiology. Spine (Phila Pa 1976). 2014 Nov 15;39(24):E1448–65.

128. Teraguchi M, Yim R, Cheung JPY, Samartzis D. The association of high-intensity zones on MRI and low back pain: a systematic review. Scoliosis Spinal Disord. 2018;13:22.

129. Lee GY, Lee JW, Choi HS, Oh KJ, Kang HS. A new grading system of lumbar central canal stenosis on MRI: an easy and reliable method. Skeletal Radiol. 2011 Aug;40(8):1033–9.

130. Miskin N, Isaac Z, Lu Y, Makhni MC, Sarno DL, Smith TR, et al. Simplified Universal Grading of Lumbar Spine MRI Degenerative Findings: Inter-Reader Agreement of Non-Radiologist Spine Experts. Pain Medicine. 2021 Jul 25;22(7):1485–95.

131. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology. 1988 Jan;166(1):193–9.

132. Wang Y, Videman T, Battié MC. Modic changes: prevalence, distribution patterns, and association with age in white men. Spine J. 2012 May;12(5):411–6.

133. Saukkonen J, Määttä J, Oura P, Kyllönen E, Tervonen O, Niinimäki J, et al. Association Between Modic Changes and Low Back Pain in Middle Age. Spine (Phila Pa 1976). 2020 Oct 1;45(19):1360–7.

134. Varlotta GP, Lefkowitz TR, Schweitzer M, Errico TJ, Spivak J, Bendo JA, et al. The lumbar facet joint: a review of current knowledge: part 1: anatomy, biomechanics, and grading. Skeletal Radiol. 2011 Jan;40(1):13–23.

135. Koslosky E, Gendelberg D. Classification in Brief: The Meyerding Classification System of Spondylolisthesis. Clin Orthop Relat Res. 2020 May;478(5):1125–30.

136. Luoma K, Vehmas T, Riihimäki H, Raininko R. Disc height and signal intensity of the nucleus pulposus on magnetic resonance imaging as indicators of lumbar disc degeneration. Spine (Phila Pa 1976). 2001 Mar 15;26(6):680–6.

137. Nagashima M, Abe H, Amaya K, Matsumoto H, Yanaihara H, Nishiwaki Y, et al. A method for quantifying intervertebral disc signal intensity on T2-weighted imaging. Acta Radiol. 2012 Nov;53(9):1059–65.

138. Kerttula L, Kurunlahti M, Jauhiainen J, Koivula A, Oikarinen J, Tervonen O. Apparent diffusion coefficients and T2 relaxation time measurements to evaluate disc degeneration. A quantitative MR study of young patients with previous vertebral fracture. Acta Radiol. 2001 Nov;42(6):585–91.

139. Zhang W, Ma X, Wang Y, Zhao J, Zhang X, Gao Y, et al. Assessment of apparent diffusion coefficient in lumbar intervertebral disc degeneration. Eur Spine J. 2014 Sep 1;23(9):1830–6.

140. Borthakur A, Mellon E, Niyogi S, Witschey W, Kneeland JB, Reddy R. Sodium and T1rho MRI for molecular and diagnostic imaging of articular cartilage. NMR Biomed. 2006 Nov;19(7):781–821.

141. Majumdar S. Magnetic resonance imaging and spectroscopy of the intervertebral disc. NMR Biomed. 2006 Nov;19(7):894–903.

142. Schleich C, Müller-Lutz A, Eichner M, Schmitt B, Matuschke F, Bittersohl B, et al. Glycosaminoglycan Chemical Exchange Saturation Transfer of Lumbar Intervertebral Discs in Healthy Volunteers. Spine (Phila Pa 1976). 2016 Jan;41(2):146–52.

143. Antoniou J, Pike GB, Steffen T, Baramki H, Poole AR, Aebi M, et al. Quantitative magnetic resonance imaging in the assessment of degenerative disc disease. Magn Reson Med. 1998 Dec;40(6):900–7.

144. Videman T, Nummi P, Battié MC, Gill K. Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. Spine (Phila Pa 1976). 1994 Jan 15;19(2):192–8.

145. Luoma K, Raininko R, Nummi P, Luukkonen R. Is the signal intensity of cerebrospinal fluid constant? Intensity measurements with high and low field magnetic resonance imagers. Magn Reson Imaging. 1993;11(4):549–55.

146. Kim UH, Kim HJ, Seo J, Chai JW, Oh J, Choi YH, et al. Cerebrospinal fluid flow artifact reduction with deep learning to optimize the evaluation of spinal canal stenosis on spine MRI. Skeletal Radiol. 2024 May;53(5):957–65.

147. Lisanti C, Carlin C, Banks KP, Wang D. Normal MRI appearance and motion-related phenomena of CSF. AJR Am J Roentgenol. 2007 Mar;188(3):716–25.

148. King S, Magnussen J, Elliott J, Hancock MJ. Development of normalized quantitative measures of lumbar disc degeneration. JOR Spine. 2023 Sep;6(3):e1278.

149. Abdollah V, Parent EC, Battié MC. Is the location of the signal intensity weighted centroid a reliable measurement of fluid displacement within the disc? Biomed Tech (Berl). 2019;64(1):111–7.

150. Mok GSP, Zhang D, Chen SZ, Yuan J, Griffith JF, Wang YXJ. Comparison of three approaches for defining nucleus pulposus and annulus fibrosus on sagittal magnetic resonance images of the lumbar spine. J Orthop Translat. 2016 Jul;6:34–41.

151. Jiang Y, Yu L, Luo X, Lin Y, He B, Wu B, et al. Quantitative synthetic MRI for evaluation of the lumbar intervertebral disk degeneration in patients with chronic low back pain. Eur J Radiol. 2020 Mar:124:108858.

152. Bechara BP, Agarwal V, Boardman J, Perera S, Weiner DK, Vo N, et al. Correlation of pain with objective quantification of magnetic resonance images in older adults with chronic low back pain. Spine (Phila Pa 1976). 2014 Mar 15;39(6):469–75.

153. Videman T, Battié MC, Gibbons LE, Gill K. A new quantitative measure of disc degeneration. Spine Journal. 2017 May 1;17(5):746–53.

154. Watanabe A, Benneker LM, Boesch C, Watanabe T, Obata T, Anderson SE. Classification of intervertebral disk degeneration with axial T2 mapping. AJR Am J Roentgenol. 2007 Oct;189(4):936–42.

155. Waldenberg C, Hebelka H, Brisby H, Lagerstrand KM. MRI histogram analysis enables objective and continuous classification of intervertebral disc degeneration. Eur Spine J. 2018 May;27(5):1042–8.

156. Li Z, Wu Y, Liu M, Miao D, Zhang D, Ding W. A novel quantitative method to evaluate lumbar disc degeneration: MRI histogram analysis. Eur Spine J. 2024 Jun 5;33(6):2420–9.

157. Eldaya RW, Parsons MS, Rhea L, Sharma A. Delineation of Inner Annulus Fibrosus and Nucleus Pulposus on Routine T2-weighted MR Images. Spine (Phila Pa 1976). 2022 Jul 1;47(13):954–60.

158. Wang YXJ, Griffith JF, Leung JCS, Yuan J. Age related reduction of T1rho and T2 magnetic resonance relaxation times of lumbar intervertebral disc. Quant Imaging Med Surg. 2014 Aug;4(4):259–64.

159. Takashima H, Takebayashi T, Yoshimoto M, Terashima Y, Tsuda H, Ida K, et al. Correlation between T2 relaxation time and intervertebral disk degeneration. Skeletal Radiol. 2012 Feb;41(2):163–7.

160. Marinelli NL, Haughton VM, Muñoz A, Anderson PA. T2 relaxation times of intervertebral disc tissue correlated with water content and proteoglycan content. Spine (Phila Pa 1976). 2009 Mar 1;34(5):520–4.

161. Welsch GH, Trattnig S, Paternostro-Sluga T, Bohndorf K, Goed S, Stelzeneder D, et al. Parametric T2 and T2* mapping techniques to visualize intervertebral disc degeneration in patients with low back pain: initial results on the clinical use of 3.0 Tesla MRI. Skeletal Radiol. 2011 May 28;40(5):543–51.

162. Ellingson AM, Nagel TM, Polly DW, Ellermann J, Nuckley DJ. Quantitative T2* (T2 star) relaxation times predict site specific proteoglycan content and residual mechanics of the intervertebral disc throughout degeneration. Journal of Orthopaedic Research. 2014 Aug;32(8):1083–9.

163. Johannessen W, Auerbach JD, Wheaton AJ, Kurji A, Borthakur A, Reddy R, et al. Assessment of human disc degeneration and proteoglycan content using T1rho-weighted magnetic resonance imaging. Spine (Phila Pa 1976). 2006 May 15;31(11):1253–7.

164. Zobel BB, Vadalà G, Del Vescovo R, Battisti S, Martina FM, Stellato L, et al. T1ρ magnetic resonance imaging quantification of early lumbar intervertebral disc degeneration in healthy young adults. Spine (Phila Pa 1976). 2012 Jun 15;37(14):1224–30.

165. Paul CPL, Smit TH, de Graaf M, Holewijn RM, Bisschop A, van de Ven PM, et al. Quantitative MRI in early intervertebral disc degeneration: T1rho correlates better than T2 and ADC with biomechanics, histology and matrix content. PLoS One. 2018;13(1):e0191442.

166. Blumenkrantz G, Zuo J, Li X, Kornak J, Link TM, Majumdar S. In vivo 3.0-tesla magnetic resonance T1rho and T2 relaxation mapping in subjects with intervertebral disc degeneration and clinical symptoms. Magn Reson Med. 2010 May;63(5):1193–200.

167. Borthakur A, Maurer PM, Fenty M, Wang C, Berger R, Yoder J, et al. T1ρ magnetic resonance imaging and discography pressure as novel biomarkers for disc degeneration and low back pain. Spine (Phila Pa 1976). 2011 Dec 1;36(25):2190–6.

168. Fenty M, Crescenzi R, Fry B, Squillante D, Turk D, Maurer PM, et al. Novel imaging of the intervertebral disk and pain. Global Spine J. 2013 Jun;3(3):127–32.

169. Takashima H, Yoshimoto M, Ogon I, Takebayashi T, Imamura R, Akatsuka Y, et al. T1rho, T2, and T2* relaxation time based on grading of intervertebral disc degeneration. Acta Radiol. 2023 Mar 25;64(3):1116–21.

170. Griffith JF. Advanced Quantitative Spine Imaging. Semin Musculoskelet Radiol. 2020 Aug;24(4):413–27.

171. Zhang X, Yang L, Gao F, Yuan Z, Lin X, Yao B, et al. Comparison of T1ρ and T2* Relaxation Mapping in Patients with Different Grades of Disc Degeneration at 3T MR. Med Sci Monit. 2015 Jul 4;21:1934–41.

172. Sood A, Mishra GV, Suryadevara M, Parihar P, Khandelwal S, Manuja N, et al. Role of Apparent Diffusion Coefficient in Evaluating Degeneration of the Intervertebral Disc: A Narrative Review. Cureus. 2023 Aug;15(8):e43340.

173. Antoniou J, Demers CN, Beaudoin G, Goswami T, Mwale F, Aebi M, et al. Apparent diffusion coefficient of intervertebral discs related to matrix composition and integrity. Magn Reson Imaging. 2004 Sep;22(7):963–72.

174. Chen KC, Tran B, Biswas R, Statum S, Masuda K, Chung CB, et al. Evaluation of the disco-vertebral junction using ultrashort time-to-echo magnetic resonance imaging: inter-reader agreement and association with vertebral endplate lesions. Skeletal Radiol. 2016 Sep;45(9):1249–56.

175. Berg-Johansen B, Han M, Fields AJ, Liebenberg EC, Lim BJ, Larson PE, et al. Cartilage Endplate Thickness Variation Measured by Ultrashort Echo-Time MRI

Is Associated with Adjacent Disc Degeneration. Spine (Phila Pa 1976). 2018 May 15;43(10):E592–600.

176. Sun D, Wu G, Zhang W, Gharaibeh NM, Li X. Visualizing Preosteoarthritis: Updates on UTE-Based Compositional MRI and Deep Learning Algorithms. J Magn Reson Imaging. 2025 Jul;62(1):40–57.

177. Han J, Meng XH, Ji Z, Niu F, Zhu N, Zhao T, et al. Quantitative assessment of lumbar intervertebral disc degeneration and its correlation with clinical symptoms: a study utilizing ultrashort time-of-echo and T2 mapping as biomarkers. BMC Musculoskelet Disord. 2025 Sep 1;26(1):829.

178. Togao O, Hiwatashi A, Wada T, Yamashita K, Kikuchi K, Tokunaga C, et al. A Qualitative and Quantitative Correlation Study of Lumbar Intervertebral Disc Degeneration Using Glycosaminoglycan Chemical Exchange Saturation Transfer, Pfirrmann Grade, and T1-ρ. AJNR Am J Neuroradiol. 2018 Jul;39(7):1369–75.

179. Xiong X, Zhou Z, Figini M, Shangguan J, Zhang Z, Chen W. Multi-parameter evaluation of lumbar intervertebral disc degeneration using quantitative magnetic resonance imaging techniques. Am J Transl Res. 2018;10(2):444–54.

180. Zhao P, Zhu S. Advances and challenges in AI-assisted MRI for lumbar disc degeneration detection and classification. Eur Spine J. 2025 Jul 25.

181. Windsor R, Jamaludin A, Kadir T, Zisserman A. SpineNetV2: Automated Detection, Labelling and Radiological Grading of Clinical MR Scans. 2022 May 3.

182. Niemeyer F, Galbusera F, Tao Y, Kienle A, Beer M, Wilke HJ. A Deep Learning Model for the Accurate and Reliable Classification of Disc Degeneration Based on MRI Data. Invest Radiol. 2021 Feb 1;56(2):78–85.

183. Gao F, Liu S, Zhang X, Wang X, Zhang J. Automated Grading of Lumbar Disc Degeneration Using a Push-Pull Regularization Network Based on MRI. Journal of Magnetic Resonance Imaging. 2021 Mar 1;53(3):799–806.

184. Soydan Z, Bayramoglu E, Karasu R, Sayin I, Salturk S, Uvet H. An Automatized Deep Segmentation and Classification Model for Lumbar Disk Degeneration and Clarification of Its Impact on Clinical Decisions. Global Spine J. 2025 Mar;15(2):554–63.

185. Huang J, Shen H, Wu J, Hu X, Zhu Z, Lv X, et al. Spine Explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. Spine Journal. 2020 Apr 1;20(4):590–9.

186. Zheng HD, Sun YL, Kong DW, Yin MC, Chen J, Lin YP, et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. Nat Commun. 2022 Feb 11;13(1):841.

187. Li Y, Hu M, Chen J, Ling Z, Zou X, Cao W, et al. Deep Learning Assisted Classification of T1ρ-MR Based Intervertebral Disc Degeneration Phases. J Magn Reson Imaging. 2025 Mar;61(3):1492–500.

188. Gong H, Liu J, Li S, Chen B. Axial-SpineGAN: simultaneous segmentation and diagnosis of multiple spinal structures on axial magnetic resonance imaging images. Phys Med Biol. 2021 Jun 7;66(11):115014.

189. Pan Q, Zhang K, He L, Dong Z, Zhang L, Wu X, et al. Automatically Diagnosing Disk Bulge and Disk Herniation with Lumbar Magnetic Resonance Images by Using Deep Convolutional Neural Networks: Method Development Study. JMIR Med Inform. 2021 May 21;9(5):e14755.

190. McSweeney T, Tiulpin A, Kowlagi N, Määttä J, Karppinen J, Saarakkala S. Robust Radiomic Signatures of Intervertebral Disc Degeneration from MRI. Spine (Phila Pa 1976). 2025 Dec 15;50(24):1737-1746.

191. Fan Z, Wu T, Wang Y, Jin Z, Wang T, Liu D. Deep-Learning-Based Radiomics to Predict Surgical Risk Factors for Lumbar Disc Herniation in Young Patients: A Multicenter Study. J Multidiscip Healthc. 2024;17:5831–51.

192. Zou C, Wang T, Wang B, Fei Q, Song H, Zang L. Developing a Deep Learning Radiomics Model Combining Lumbar CT, Multi-Sequence MRI, and Clinical Data to Predict High-Risk Adjacent Segment Degeneration Following Lumbar Fusion: A Retrospective Multicenter Study. Global Spine J. 2025 Jun 9;21925682251342532.

193. Deguchi T, Hashizume H, Terao C, Nakajima M, Teraguchi M, Yamada H, et al. A longitudinal population-based study identifies THBS2 as a susceptibility gene for intervertebral disc degeneration. Eur Spine J. 2024 Sep;33(9):3334–42.

194. Videman T, Battié MC, Gibbons LE, Gill K. Aging changes in lumbar discs and vertebrae and their interaction: a 15-year follow-up study. Spine J. 2014 Mar 1;14(3):469–78.

195. Aavikko A, Lohman M, Ristolainen L, Kautiainen H, Österman K, Schlenzka D, et al. ISSLS prize in clinical science 2022: accelerated disc degeneration after pubertal growth spurt differentiates adults with low back pain from their asymptomatic peers. Eur Spine J. 2022 May;31(5):1080–7.

196. Aavikko A, Ristolainen L, Kautiainen H, Lohman M, Schlenzka D, Lund T. Relationship of disc degeneration after pubertal growth spurt to future low back pain: a longitudinal cohort study. Eur Spine J. 2024 Jun 20.

197. Borenstein DG, O'Mara JW, Boden SD, Lauerman WC, Jacobson A, Platenberg C, et al. The value of magnetic resonance imaging of the lumbar spine to predict low-back pain in asymptomatic subjects: a seven-year follow-up study. J Bone Joint Surg Am. 2001 Sep;83(9):1306–11.

198. Elfering A, Semmer N, Birkhofer D, Zanetti M, Hodler J, Boos N. Risk factors for lumbar disc degeneration: a 5-year prospective MRI study in asymptomatic individuals. Spine (Phila Pa 1976). 2002 Jan 15;27(2):125–34.

199. Waris E, Eskelin M, Hermunen H, Kiviluoto O, Paajanen H. Disc degeneration in low back pain: a 17-year follow-up study using magnetic resonance imaging. Spine (Phila Pa 1976). 2007 Mar 15;32(6):681–4.

200. Sääksjärvi S, Kerttula L, Luoma K, Paajanen H, Waris E. Disc Degeneration of Young Low Back Pain Patients: A Prospective 30-year Follow-up MRI Study. Spine (Phila Pa 1976). 2020 Oct 1;45(19):1341–7.

201. Schistad EI, Bjorland S, Røe C, Gjerstad J, Vetti N, Myhre K, et al. Five-year development of lumbar disc degeneration-a prospective study. Skeletal Radiol. 2019 Jun;48(6):871–9.

202. Kerttula L, Luoma K, Vehmas T, Grönblad M, Kääpä E. Modic type I change may predict rapid progressive, deforming disc degeneration: A prospective 1-year follow-up study. Eur Spine J. 2012;21(6):1135–42.

203. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977 Mar;33(1):159.

204. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun 1;15(2):155–63.

205. Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. CBE Life Sci Educ. 2013;12(3):345–51.

206. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013 Nov 26;4:863.

207. Lin LIK. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 1989 Mar;45(1):255.

208. Lin LI, McBride G, Bland JM, Altman DG. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. NIWA Client Report. 2005;45(1).

209. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307–10.

210. Evans JD. Straightforward statistics for the behavioral sciences. Belmont, CA, US: Thomson Brooks/Cole Publishing Co;1996. 600 p.

211. Vilanova C, Martín-Noguerol T, García-Figueiras R, Baleato-González S, Vilanova JC. Bone marrow magnetic resonance imaging (MRI): morphological and functional features from reconversion to infiltration. Quant Imaging Med Surg. 2024 Nov 1;14(11):7969–82.

212. Oh CH, Yoon SH. Whole Spine Disc Degeneration Survey according to the Ages and Sex Using Pfirrmann Disc Degeneration Grades. Korean J Spine. 2017 Dec 31;14(4):148–54.

213. Abdollah V, Parent EC, Battié MC. Reliability and validity of lumbar disc height quantification methods using magnetic resonance images. Biomed Tech (Berl). 2019 Feb 25;64(1):111–7.

214. D'Anastasi M, Birkenmaier C, Schmidt GP, Wegener B, Reiser MF, Baur-Melnyk A. Correlation Between Vacuum Phenomenon on CT and Fluid on MRI in Degenerative Disks. AJR Am J Roentgenol. 2011 Nov;197(5):1182–9.

215. Cawley DT, McDonnell A, Simpkin A, Doyle T, Habash M, McNamee C, et al. Intra-discal vacuum phenomenon with advanced lumbar spine disc degeneration: complementary findings from both MRI and CT. BMC Med Imaging. 2025 Mar 20;25(1):94.

ISBN 978-952-84-2107-8 (Paperback)
ISBN 978-952-84-2106-1 (PDF)
ISSN 2954-2898 (Paperback)
ISSN 2954-2952 (PDF)
Unigrafia, Helsinki
Printed in PunaMusta, Joensuu 2026