Biomedical Signal Processing and Control 118 (2026) 109700

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Deep learning for lumbar spine segmentation in magnetic resonance imaging—A systematic review

Diogo Mendes [a], João Manuel R.S. Tavares [b,*]

[a] Faculdade de Engenharia, Universidade do Porto (FEUP), Rua Dr. Roberto Frias, s/n, Porto, 4200-465, Portugal
[b] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial (INEGI), Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto (FEUP), Rua Dr. Roberto Frias, s/n, Porto, 4200-465, Portugal

ARTICLE INFO

Keywords:
Machine learning
Medical imaging
Image analysis
CNN
MRI

ABSTRACT

Lumbar spine disorders are a major cause of disability worldwide, often requiring magnetic resonance imaging (MRI) for accurate diagnosis and treatment planning. Segmentation of spinal structures in MRI is a critical yet time-consuming task that is traditionally performed manually by experts. In recent years, deep learning (DL) has emerged as a powerful tool to automate this process, offering improvements in efficiency, consistency, and accuracy. This systematic review presents a comprehensive analysis of DL-based methods for lumbar spine segmentation in MRI, conducted by searching the Scopus and PubMed databases for peer-reviewed journal articles and reviews. Studies were included only if they employed MRI as the imaging modality, applied deep learning techniques for segmentation, and explicitly reported both the model architecture and segmentation results. A total of 56 studies met these criteria, comprising 49 original research articles and 7 review articles. The selected works were analyzed in two main dimensions: (1) dataset characteristics, including image orientation, modality, annotation methods, and data availability; and (2) deep learning frameworks, covering preprocessing, data augmentation, network architectures, and evaluation metrics. Convolutional neural networks (CNNs), particularly U-Net and its variants, are the most used architectures, often enhanced with residual blocks, attention mechanisms, and multi-scale feature extraction. Despite promising results, most studies relied on private datasets and lacked external validation, highlighting challenges to reproducibility.

Contents

1. Introduction 2
2. Search methodology 2
   2.1. Search strategy 2
   2.2. Quality and risk-of-bias assessment 3
3. Search results 3
   3.1. Risk-of-bias assessment 3
   3.2. Datasets characteristics 5
   3.3. Data preprocessing and augmentation 5
   3.4. Model architectures 5
   3.5. Loss function and hyperparameters 10
   3.6. Performance metrics 10
   3.7. Review articles 12
4. Discussion 13
5. Conclusion and limitations 17
CRediT authorship contribution statement 18
Declaration of competing interest 18
Appendix A. 19
Data availability 19

* Corresponding author.
E-mail address: tavares@fe.up.pt (J.M.R.S. Tavares).

https://doi.org/10.1016/j.bspc.2026.109700
Received 28 August 2025; Received in revised form 18 January 2026; Accepted 20 January 2026
Available online 30 January 2026
1746-8094/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

References 19

1. Introduction

The lumbar spine is the lower portion of the vertebral column, which consists of five vertebrae labeled L1 through L5. It plays a crucial role in supporting the upper body's weight, enabling mobility and flexibility, and protecting the spinal cord and nerve roots that branch into the lower limbs. Structurally, each lumbar vertebra is characterized by a large and robust vertebral body that bears mechanical loads and a vertebral arch that forms the spinal canal through which the cauda equina passes. Between adjacent vertebrae lie intervertebral discs, which act as shock absorbers and provide cushioning during motion. The posterior elements include facet joints that facilitate controlled movement and provide additional stability to the spine. The surrounding ligaments, muscles, and soft tissues further support the lumbar region, contributing to its complex biomechanics [1]. Given its anatomical and functional importance, the lumbar spine is particularly susceptible to degeneration, injury, and disease, often resulting in pain and disability.

Low back pain (LBP) is one of the leading causes of disability worldwide and is one of the main reasons for the use of healthcare in industrialized countries, with an estimated 1-year prevalence of 38% in the general population [2]. While the majority of LBP cases are self-limiting and nonspecific, a small but critical subset is attributable to structural abnormalities such as intervertebral disc herniation, spinal stenosis, neoplastic disease, infection, or inflammatory spondyloarthropathies [3]. Magnetic Resonance Imaging (MRI) is widely regarded as the imaging modality of choice for evaluating lumbar spine disorders due to its high-resolution soft tissue contrast, multiplanar capabilities, and absence of ionizing radiation [3]. It allows detailed visualization of key anatomical structures, including intervertebral discs, vertebral bodies, spinal cord, nerve roots, and paraspinal soft tissues, which are essential for identifying pathologies such as disc herniation, spinal stenosis, and degenerative disc disease [4]. Compared with other imaging techniques, such as X-ray and computed tomography (CT), MRI offers superior sensitivity for detecting early degenerative changes, inflammatory processes, and soft tissue abnormalities.

Typical MRI protocols include sagittal and axial planes, using T1- and T2-weighted sequences. Sagittal T1-weighted images provide excellent anatomical detail, especially of vertebral bone marrow and thecal sac, while T2-weighted images enhance the contrast of cerebrospinal fluid (CSF), making them ideal for visualizing disc hydration and spinal stenosis. Fig. 1 compares sagittal T1-weighted and T2-weighted MRI scans of the lumbar spine from the same subject, illustrating key anatomical structures and disc pathologies, including normal, herniated, and bulging discs, as well as the vertebrae and spinal canal. Axial T2-weighted images are critical for evaluating foraminal and central canal stenosis and identifying nerve root compression. Additional sequences, such as short tau inversion recovery (STIR) and fat-suppressed T2, are used to highlight edema and inflammation, which is particularly useful in conditions like infection and trauma. Gadolinium-enhanced sequences are reserved for suspected infections, tumors, or post-operative spine evaluations [5]. Although MRI is the preferred modality for evaluating lumbar spine disorders, its interpretation remains highly dependent on expert analysis, which is time-consuming and subject to considerable inter- and intra-observer variability [2]. Accurate identification and delineation of anatomical structures —such as vertebral bodies, intervertebral discs, and spinal canals —are fundamental not only for diagnosis but also for treatment planning, surgical navigation, and outcome monitoring.

Manual segmentation of structures in medical images is resource-intensive and impractical for large-scale applications, underscoring the need for automated, reproducible solutions. Traditional image processing techniques often rely on hand-engineered features and predefined rules, which may not generalize well across patient populations due to anatomical variability, degenerative changes, and imaging artifacts [7]. In recent years, deep learning has emerged as a transformative approach in medical image analysis, offering significant improvements in accuracy, efficiency, and automation over traditional techniques. According to IBM, Deep Learning is "a subset of machine learning that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain ". Its applications span a wide range of tasks, including classification, detection, segmentation, registration, and image synthesis across various imaging modalities such as MRI [8], CT [9], and X-ray [7, 10]. Among these, segmentation, i.e., the delineation of anatomical structures or pathological regions, is particularly critical, forming the basis for quantitative analysis, treatment planning, and disease monitoring [11]. Compared to classical approaches, deep learning methods offer significant improvements in terms of accuracy, automation, and scalability. These capabilities are especially relevant in spinal imaging, where precise segmentation is essential for morphological assessment, disease quantification, and biomechanical modeling. However, segmentation of the lumbar spine poses unique challenges due to its anatomical variability, partial-volume effects, and age-related or pathological changes [12]. Given the rapid pace of development in this area, a comprehensive review of current deep learning approaches for lumbar spine segmentation is essential to assess methodological trends, highlight performance benchmarks, and identify gaps in the existing literature.

2. Search methodology

2.1. Search strategy

A comprehensive search was conducted in the Scopus and PubMed databases for English-language journal research and review articles focused on deep learning methods for lumbar spine segmentation in magnetic resonance imaging. Searches were performed between January and February 2025. The query used in the Scopus search engine was *TITLE-ABS-KEY (deep AND learning AND lumbar AND spine AND segmentation AND (mr* OR (magnetic AND (imag* OR resonance)))) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (LANGUAGE, "English"))*, and in PubMed it was *(deep learning[Title/Abstract] AND lumbar spine[Title/Abstract] AND segmentation[Title/Abstract] AND ((MR [Title/Abstract] OR MRI[Title/Abstract]) OR (magnetic[Title/Abstract] AND (imag*[Title/Abstract] OR resonance [Title/Abstract])))) AND (english[Language]) AND (review[Publication Type] OR journal article[Publication Type])*.

Studies describing segmentation using deep learning (DL) approaches in MRI were eligible for inclusion. Segmentation was defined as the ability to label spinal structures at the pixel level. The placement of a bounding box to extract a region of interest (ROI) or the localization of key points was not considered a segmentation form.

Additionally, any article that did not present the segmentation model architecture or segmentation results was excluded. No time interval restrictions were applied. The screening process was initially conducted by a single reviewer, followed by validation by a second reviewer. The screening was performed by the authors without the use of automation tools. The review protocol, including the research question, inclusion and exclusion criteria, and databases to be searched, was developed a priori and not prospectively registered in a public database.

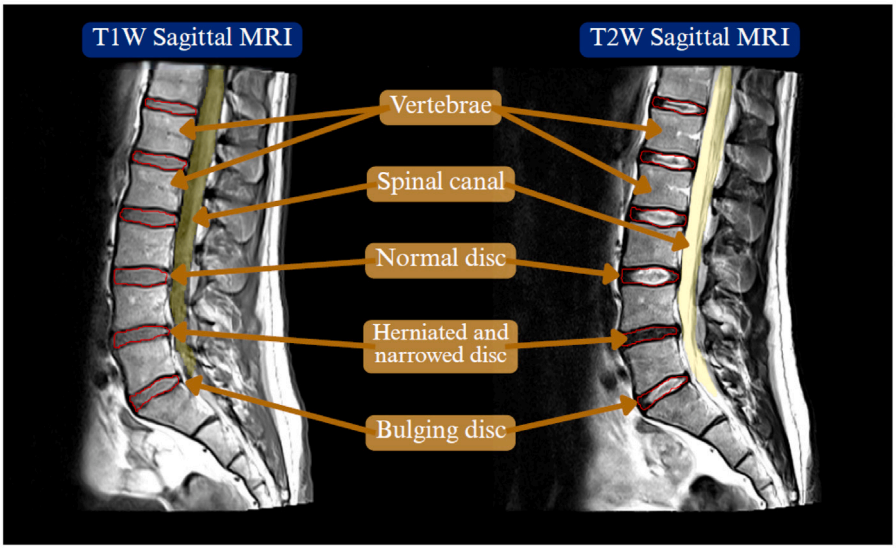

Fig. 1. Comparison of T1-weighted (T1 W) and T2-weighted (T2 W) sagittal MRI scans of the lumbar spine, illustrating key anatomical structures and disc conditions (both images were taken from the public SPIDER dataset [6]).

Table 1
Risk-of-bias assessment domains and criteria.

Domain	Low risk of bias	High risk of bias	Unclear
Dataset representativeness	≥50 subjects AND multi-center **or** public dataset	<50 subjects **or** single-center private dataset	Center or data source not reported
Ground-truth annotation	Expert annotation with clear description and/or validation	Non-expert or automated annotation without validation	Annotation process not clearly described
Validation strategy	Subject-level data split **AND** external validation on an independent dataset	Slice-level split **OR** no external validation	Validation strategy not clearly stated
Overall risk of bias (study-level)	**Low**: all three domains rated as low risk; **Moderate**: two domains rated as low risk; **High**: one or zero domains rated as low risk.		

2.2. Quality and risk-of-bias assessment

To evaluate the methodological quality of the included original studies, a formal risk-of-bias assessment was conducted. Given the heterogeneity of study designs, datasets, and evaluation protocols across deep learning-based lumbar spine segmentation studies, a simplified domain-based assessment was adopted rather than a task-specific, validated tool. The assessment focused on three key domains considered most relevant for this literature: dataset representativeness, ground-truth annotation quality, and validation strategy. These domains were selected based on common sources of bias in medical imaging artificial intelligence studies. The predefined criteria used to classify each domain as low, high, or unclear risk of bias are summarized in Table 1.

Each of the 49 original research articles was independently assessed across the three domains based on the information reported in the manuscript. Dataset representativeness is assessed based on sample size and data origin, favoring multi-center or publicly available datasets with at least 50 subjects. Ground-truth annotation quality was evaluated based on the level of expertise involved in the annotation and whether the annotation process was clearly described or validated. Validation strategy focused on whether external validation using independent data was performed, as this is critical for assessing generalizability. When insufficient methodological detail was provided to confidently judge a domain, the risk of bias was rated as unclear.

An overall risk-of-bias judgment was then derived for each study by aggregating the three domains. Studies were classified as low overall risk of bias when all three domains were rated as low risk, as moderate risk when two domains were rated as low risk, and as high risk otherwise. This approach provides a transparent, reproducible summary of study quality while avoiding overly subjective weighting of individual domains. The results of the study-level risk-of-bias assessment and the overall classifications are reported in the Results section and detailed in Appendix A.

3. Search results

A total of 85 records were retrieved from the Scopus database and 28 from PubMed, all of which were also included in the Scopus results. After the screening process, 29 records were excluded for the following reasons: 20 for not involving deep learning-based segmentation, 6 for not covering the lumbar spine region in MRI, and 3 for not presenting segmentation results. Consequently, 56 studies were included in this review, comprising 49 research articles and 7 review articles (Fig. 2).

The temporal distribution of the included studies is presented in Fig. 3. A clear upward trend in the number of publications is observed over time, with research activity beginning to increase notably in 2019. This growth reflects the expanding interest and adoption of deep learning methods for lumbar spine segmentation using MRI data. The number of articles peaked in 2022 and 2023, with 14 publications each year, including both research and review articles. Although the count for 2025 appears lower, it includes only publications indexed up to February 17 and is, therefore, not representative of the full year. Overall, this trend underscores growing scientific momentum toward automated, deep learning-based segmentation techniques in spinal imaging.

3.1. Risk-of-bias assessment

The results of the risk-of-bias assessment are summarized in Table 2. Dataset representativeness emerged as a major source of potential bias,

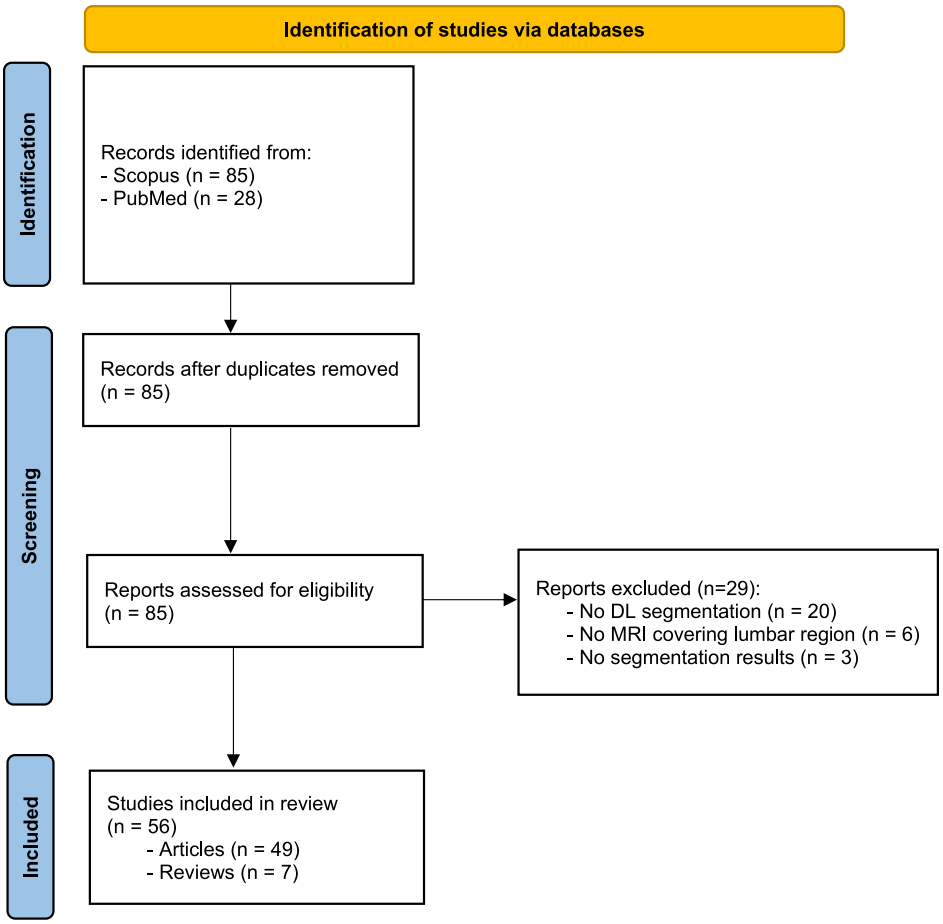

Fig. 2. PRISMA flow diagram of the conducted review.

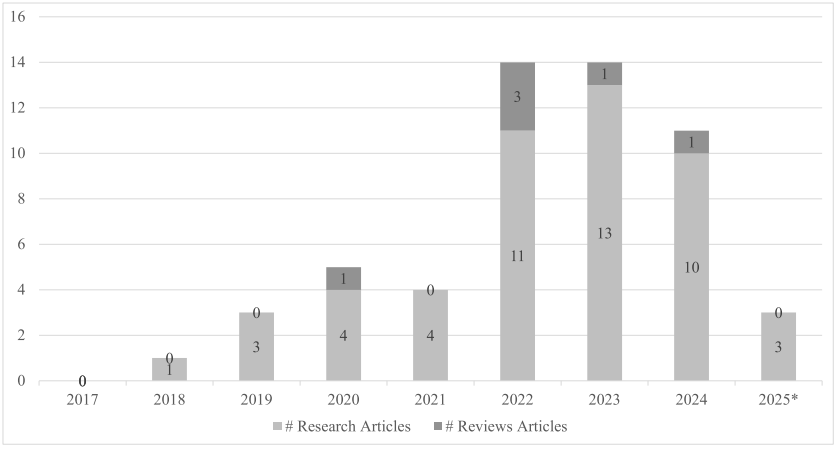

Fig. 3. Distribution of included studies by year of publication, separated by research (on the left) and review (on the right) article types (* 2025 only contains results up to 17 February).

with 61% of studies classified as high risk. This was primarily due to the widespread reliance on single-center and/or private datasets, which, despite often including more than 50 subjects or more than 250 image slices, may not adequately capture the heterogeneity of lumbar spine anatomy, imaging protocols, and pathological variability encountered in routine clinical practice. Only 35% of studies met the criteria for low risk in this domain, typically by using multi-center data or publicly available datasets with sufficiently large cohorts. These findings suggest that dataset-related constraints remain a key limitation affecting the generalizability of many proposed deep learning models.

In contrast, ground-truth annotation quality was generally strong across the reviewed literature, with 88% of studies rated as low risk. Most studies relied on expert annotations provided by radiologists, spine surgeons, or experienced clinicians, often supplemented by validation or review procedures, which supports the reliability of the reference segmentations used for training and evaluation.

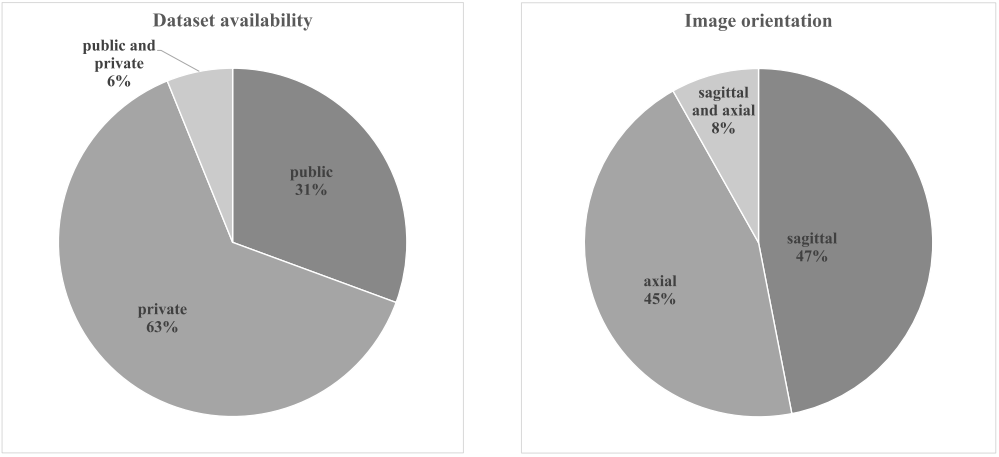

Fig. 4. Access type of the datasets (on the left) and image orientations (on the right) used in the selected studies.

Table 2
Summary of risk-of-bias assessment across the included studies.

Domain	Low n (%)	Unclear n (%)	High n (%)
Dataset representativeness	17 (35%)	2 (4%)	30 (61%)
Ground-truth annotation	43 (88%)	3 (6%)	3 (6%)
Validation strategy	9 (18%)	2 (4%)	38 (78%)
Overall risk of bias	**Low: 3 (6%)**	**Moderate: 17 (35%)**	**High: 29 (59%)**

The validation strategy was one of the most frequent sources of bias across the included studies. Although a majority of articles (69%) adopted subject-level data splitting to reduce information leakage between the training and test sets, external validation was rarely performed. In fact, 76% of studies did not evaluate their models on an independent external dataset, and only 18% implemented external validation. As a result, many studies that otherwise employed appropriate subject-level splits were still classified as high risk in this domain. The limited use of external validation restricts the assessment of model generalizability across different institutions, acquisition protocols, and patient populations.

3.2. Datasets characteristics

Table 3 summarizes the characteristics of the datasets used in the 49 original research articles included in this review. It outlines the imaging modalities, dataset availability, sample sizes, ground truth annotation methods, and the anatomical regions segmented. The majority of the datasets are private, often generated and annotated internally by medical experts or research teams. Public datasets, although fewer in number, are essential for reproducibility and benchmarking. This distribution is illustrated by the chart on the left of Fig. 4, where 63% of datasets are private, 31% are public, and only 6% combine both sources. On the other hand, the graph on the right of Fig. 4 shows the image orientations used in the reviewed works: sagittal (47%) and axial (45%) views were employed with nearly equal frequency, while a smaller proportion of studies used both orientations (8%).

Table 4 outlines the various preprocessing pipelines employed by the authors of the selected works, including data preprocessing steps, augmentation techniques, and dataset splitting strategies. Table 5 summarizes the key aspects of the reviewed models, including input types, architectural designs, loss functions, training hyperparameters, and quantitative performance metrics. Together, these tabulated overviews enable a comprehensive cross-comparison of methodological trends and innovations, highlighting both common practices and distinctive contributions across studies.

3.3. Data preprocessing and augmentation

Data preprocessing

Preprocessing refers to the set of operations applied to raw medical images to enhance consistency, quality, and suitability for neural network input. Among the reviewed studies, the most frequently used preprocessing steps were image resizing (e.g., to 256 × 256 or 512 × 512 pixels), normalization (e.g., Z-score or min–max scaling), cropping, and padding. These steps standardize data dimensions and intensity ranges, which is crucial for ensuring training stability and reproducibility. A subset of articles utilized more advanced techniques. Image registration, particularly between T1-weighted and T2-weighted modalities, was employed in studies such as [12,15,31]. Registration aligns anatomical structures across modalities to reduce spatial discrepancies, thereby improving the utility of multi-modal inputs for training. This is particularly important when combining information from images acquired at different times or sequences. One notable contribution comes from Bonnheim et al. [36], who implemented an automatic cropping algorithm specifically designed to localize the L5 vertebral body. This method extracts a fixed-size region of interest centered around L5, ensuring anatomical consistency across samples and reducing input redundancy.

Data augmentation

Data augmentation is a strategy for artificially increasing the size and diversity of training datasets by applying label-preserving transformations. Its primary goal is to enhance model generalization and reduce overfitting, especially when labeled data is limited. Among the 49 studies reviewed, 34 explicitly reported using data augmentation. The most common augmentation techniques included in-plane rotations (typically ±10° to ±20°), horizontal and vertical flipping, elastic deformations, brightness and contrast adjustments, and random cropping or zooming. These transformations simulate real-world variability in anatomy and acquisition conditions, making models more robust to spatial and intensity variations. 15 studies either did not mention augmentation or reported that no augmentation was used.

3.4. Model architectures

All 49 studies included in this review employed convolutional neural networks (CNNs) as the foundational architecture for segmentation tasks. CNNs are inherently suited to image analysis, as their learnable convolutional filters efficiently capture spatial hierarchies and patterns at multiple scales, both locally and globally.

Table 3
Summary of the characteristics of the datasets used in the selected research articles.

Author	Year	Dataset availability	Image type	Dataset size	Ground-Truth	Segmented anatomical ROIs
Lessmann et al.	2019	public[a]	Sagittal T2W MRI	23 subjects, with at least 7 vertebral bodies (T11–L5)	not mentioned	Vertebral bodies
Pang et al.	2021	public[b]	Sagittal T2W MRI	215 subjects	junior and senior expert via ITK-SNAP1	Vertebrae and Intervertebral Discs (IVDs)
Al-Kafri et al.	2019	public[c]	Axial T1W and T2W MRI	515 subjects, 48345 MRI slices	5 expert radiologists	IVDs, Posterior Element, Thecal Sac, AAP
Huang et al.	2020	private	Sagittal T2W MRI	486 subjects, 300 slices	two residents, adjusted by surgeon	Lumbar vertebrae, IVDs, Sacrum, CSF, Presacral fat area
Zheng et al.	2022	private	Sagittal T2W MRI	A: 223, B: 63, C: 1051 subjects	LabelMe software	Lumbar vertebrae, IVDs, Sacrum, CSF, Presacral fat area
Kim et al.	2018	public[d]	Sagittal T2W MRI	20 subjects - 25 images	by a spine MR researcher and reviewed by a radiologist with 10+ years experience	IVDs
Han et al.	2021	private	Sagittal T1W and T2W MRI	253 subjects	by board-certified radiologists	Vertebrae, IVDs and Neural Foramina
Mbarki et al.	2020	private	Axial T2W MRI	450 subjects	LabelMe software	IVDS and apophyse from lumbar spine
Li et al.	Li et al.	private	Axial T2W MRI	120 subjects, 1080 slices	by spine surgeons and imaging specialists, double-checked by experts with 10-30 years of experience	Vertebral bodies, lamina, and dural sac
Šušteršič et al.	2022	public[c] and private	Sagittal and Axial T1W and T2W MRI	243 subjects, 1169 sagittal slices and 557 axial slices	by neurologists	IVDs (L4/L5 and L5/S1)
Zhou et al.	2020	private	Sagittal T1W and T2W; Sagittal 3D spoiled GRE MRI to obtain IDEAL series (water, fat, R∗2, and fat fraction maps).	57 subjects	two trained operators under supervision by a musculoskeletal imaging fellowship-trained radiologist	Vertebral bodies (L1-L5)
Shen et al.	2021	private	Axial T2W MRI	120 slices	by a senior spine surgeon using LabelMe software	Paraspinal muscles: bilateral multifidus, erector spinae, and psoas major. L4/L5 disc and spinal canal
Rehman et al.	2019	public[d]	Sagittal T2W MRI	20 subjects, 25 slices	by radiological experts	IVDs
Masood et al.	2022	public[c]	Sagittal T1W and T2W MRI	514 subjects	by expert spinal surgeons	Vertebral bodies (L1-S1)
Bharadwaj et al.	2023	private	Axial T2W MRI	200 subjects, 987 slices	by a trained researcher, radiology trainee, and a board-certified neuroradiologist	IVDs and dural sac
Iriondo et al.	2020	private	Data set A: T1ρ-W images and T2W images (SE MRI), sagittal; Data set B: Multislice T1ρ/T2 maps (3D spoiled GE MRI), sagittal	Data set A: 16 subjects; Data set B: 15 subjects	Dataset A: fully manual; Dataset B: Semi-automated 3D region growing (with manual seeding and edits)	IVDs (L1/L2-L5/S1)
Hwang et al.	2022	private	Sagittal T1W MRI	44 healthy subjects (836 slices) and 56 diseased subjects (1064 slices)	by a board-certified musculoskeletal radiologist	Lumbar bone marrow
Liu et al.	2022	private	Axial T2W MRI	100 subjects, 100 slices	by a spine surgeon using Mimics Innovation Suite; reviewed by an expert radiologist and surgeon	Vertebrae L5 and S1; IVD L5/S1

(*continued on next page*)

U-Net and its variants

The U-Net architecture, introduced by [64], is the most widely adopted model in medical image segmentation due to its effectiveness in learning spatial hierarchies and preserving fine-grained anatomical details. Among the 49 studies reviewed, 33 employed U-Net or a variant, often modifying the architecture to address specific segmentation challenges and improve performance.

Among the several modifications proposed in the reviewed articles using U-Net-based architectures, the most common enhancements are the integration of residual blocks, attention mechanisms, and multiscale feature extractors.

Residual blocks, first proposed in 2016 by He et al. in the ResNet architecture [65], facilitate the training of deeper networks by enabling shortcut connections that bypass one or more convolutional layers. These identity mappings help mitigate vanishing gradients and improve gradient flow during backpropagation, enabling the model to learn more complex features. In U-Net variants, residual blocks are typically embedded within the encoder or decoder paths to enhance feature propagation and model convergence. [27,55] employed a hybrid model combining a ResNet encoder with a U-Net decoder. [18,53] replaced conventional convolutional blocks in both the encoder and decoder with residual blocks, facilitating deeper architectures and improved gradient flow. [43] proposed an ensemble of Residual U-Net models

Table 3 (*continued*).

Author	Year	Dataset availability	Image type	Dataset size	Ground-Truth	Segmented anatomical ROIs
Gao et al.	2022	private	Sagittal T1W and T2W MRI	75 subjects	by research associates, a board-certified neuroradiologist and a junior musculoskeletal radiologist	Vertebral bodies and modic changes
Hess et al.	2023	private	Sagittal and Axial T1W MRI	206 subjects	by trained readers using MD.ai and a board-certified radiologist	Vertebral bodies, IVDs and paraspinal muscles
Baur et al.	2022	private	Axial T2W MRI	100 subjects, 3650 slices	by a spine surgery specialist using Materialize Mimics software	Lumbar spinal muscles: erector spinae and multifidus combined, intramuscular fatty tissue
Sáenz-Gamboa et al.	2023	public[e]	Sagittal T1W and T2W MRI	181 subjects, 41572 slices	by two radiologists with expertise in skeletal muscle pathologies	Vertebrae, Sacrum, IVD, Spinal Cavity, Subcutaneous Cellular Tissue, Epidural Fat, Intramuscular Fat, Retroperitoneal Fat, Nerve root, Blood vessels, Muscle
Niemeyer et al.	2022	private	Axial MRI	Training set: 60 axial images; External set: 22 subjects	by a single operator using ITK-SNAP software	Paraspinal muscles: psoas major, quadratus lumborum, erector spinae, and multifidus
Dourthe et al.	2022	private	Axial T1W MRI	6 healthy subjects, 126 slices; 8 subjects with Adult Spinal Deformity, 120 slices	by biomedical engineers, neuro-radiologists and spine surgery fellows	Vertebral bodies, multifidus/erector spinae, psoas major, subcutaneous fat
Bonnheim et al.	2023	private	Sagittal UTE MRI	83 subjects, 6972 slices	by two trained annotators	Superior and inferior spinal cartilage endplate (CEPs) at the L4-L5 and L5-S1 levels
Yilizati-Yilihamu et al.	2023	public[b]	Sagittal T2W MRI	172 slices	by experts	Vertebral bodies (T10-S1) and IVDs (T10/T11 - L5/S)
Hou et al.	2023	public[c]	Axial T2W MRI	1545 slices	by 5 expert radiologists	IVDs, dural sac, lamina and AAP
Morris et al.	2024	private	Axial T1W MRI	75 subjects	by an experienced observer	Bone marrow at four regions: spine (T8-L3), femoral head, total hip and femoral diaphysis
Saravi et al.	2023	public[c] and private	Axial T1W MRI	Main dataset: 515 subjects, external dataset: 50 subjects	by non-experts under supervision by expert radiologists	Dural sac
Chen et al.	2022	private	Axial T2W MRI	100 subjects, 100 slices	by an experienced surgeon and radiologist	Vertebrae L4 and L5 and IVD L4/L5
Altun et al.	2023	private	Axial T2W MRI	520 subjects, 1560 slices	by an expert	IVDs, spinal canal, posterior element and thecal sac
Laiwalla et al.	2023	private	Axial T2W MRI	379 slices	by students using ITK-Snap and verified by expert clinicians	Lumbar spinal canal
Mohanty et al.	2023	public[c]	Axial T1W MRI	5000 slices	by 5 expert radiologists	Spinal cord
Adibatti et al.	2023	public[c]	Axial T1W MRI	515 subjects, 48345 slices	by 5 expert radiologists	IVDs
Balzer et al.	2022	private	Sagittal T2W MRI	1000 subjects, 1000 slices	by a physician	IVDs
Zhou et al.	2024	public[c]	Axial T1W and T2W MRI	515 subjects, 48345 slices	by 5 expert radiologists	IVDs, Posterior Element, Thecal Sac, Area between Anterior and Posterior vertebrae elements (AAP)

(*continued on next page*)

and [14] used a Residual U-Net for segmentation refinement. [40] adopted the MultiResUNet architecture, introduced by Ibtehaz and Rahman [66], which augments the U-Net by integrating multi-resolution convolutional paths and residual connections within each block.

Attention mechanisms are another prevalent enhancement that aims to improve the model's ability to focus on clinically relevant regions while suppressing background noise. These modules dynamically assign weights to spatial or channel-wise features, refining the network's representational capacity. Attention blocks are particularly beneficial in medical imaging, where lesions or structures of interest may occupy small, heterogeneous regions. Variants such as Attention U-Net [40], CBAM-attention U-Net [39], and BUA-Net [47] demonstrate the integration of attention into the U-Net framework, enhancing segmentation accuracy through adaptive feature recalibration. [21] uses a dual-branch multi-scale attention module before every downsampling and upsampling stage. One branch corresponds to an attention block that sequentially applies channel and spatial attention mechanisms to refine feature representations, first emphasizing informative channels and then enhancing spatially relevant regions. [12] used attention gates to replace standard skip connections, introducing spatial attention mechanisms that selectively filter encoder features before fusion with decoder features. [39] incorporated the Convolutional Block Attention Module (CBAM), originally proposed by Woo et al. [67], into the encoding

Table 3 (*continued*).

Author	Year	Dataset availability	Image type	Dataset size	Ground-Truth	Segmented anatomical ROIs
Candito et al.	2024	private	Whole-Body DWI MRI	40 subjects, 14194 slices	by a medical physicist using ITK-SNAP software	Spinal cord
Jung et al.	2024	private	Axial T1W Whole-Body 3D VIBE MRI	UKBB dataset: 36,317 subjects; NAKO dataset: 23,725 subjects.	by a senior radiology resident and validated by a board-certified radiologist	subcutaneous adipose tissue (SAT), visceral adipose tissue (VAT), skeletal muscle (SM), skeletal muscle fat fraction (SMFF), and intramuscular adipose tissue (IMAT)
Wang et al.	2023	private	Axial MRI	70 subjects, 70 volumes (12 slices each)	by radiologists	IVDs, spinal canal, facet joint and herniated disc
Kok et al.	2023	private	Sagittal T1W MRI	8 subjects (6 healthy, 1 scoliosis, 1 spinal implants)	by experts	IVDs
Natalia et al.	2024	public[c]	Sagittal T2W MRI	515 subjects, 48345 slices	by 5 expert radiologists	Six non-overlapping regions — Anterior, PosteriorA, PosteriorB, Vertebrae, IVD, and Sacrum.
Liang et al.	2024	private	Sagittal and Axial T1W and T2W MRI	539 subjects	by 3 clinicians using ITK-SNAP software	Sagittal view: IVDs, vertebral bodies, dural sac. Axial view: IVDs, lamina, dural sac, yellow ligament
Xu et al.	2024	private	Axial T2W MRI	417 subjects, 1251 slices	Training/Validation: by three spinal surgeons; Internal Test: spinal expert and a surgeon; External Test: by the same surgeon	IVDs, vertebral arch (VA), psoas major (PM), quadratus lumborum (QL), longissimus (LO), iliocostalis (IL), multifidus (MF); and fat regions within paraspinal muscles
Giaccone et al.	2024	public[c] and private	Axial T2W MRI	private dataset: 91 subjects; public dataset: 515 subjects	by expert physicians using OsiriX software	Paraspinal muscles (erector spinae and multifidus) and fat infiltrations in the muscles at the L4-L5
Ding et al.	2024	private	Sagittal T2W MRI	94 subjects; 466 slices	by 2 experienced radiologists using 3D-Slicer software	IVDs
Basak et al.	2025	public[f]	Sagittal T1W	34 subjects, 2692 slices	by a medical professional	Vertebral bodies (L1-L5) and IVDs (L1/L2-L4/L5)
Yan et al.	2024	private	Sagittal T1W and T2W STIR MRI	370 MRI cases (307 compression fractures and 63 normal fractures)	by clinical experts	Normal and fractured vertebrae, and IVDs
Karthik et al.	2025	private	Sagittal and Axial T2W MRI	191 subjects; 231 scans	by several raters	Spinal cord and intramedullary lesions (encompassing edema and hemorrhage)
Lakshmi et al.	2025	public[c]	Sagittal T1W MRI	515 subjects, 1312 slices	by a radiologist	Vertebrae (L1-S1)

[a] [61].
[b] [14].
[c] [62].
[d] SpineWeb dataset 10.
[e] MIDAS dataset.
[f] [63].

stage, and [40] used the Attention U-Net architecture, which integrates attention gates into the decoder to selectively focus on relevant encoder features during skip connections. In [47], a boundary- and uncertainty-aware attention network (BUA-Net) is presented that applies a multiscale spatial and channel attention module to the encoder's lowest layer. It is also proposed an uncertainty-aware attention module at the end of the decoder that generates an attention mask, improving the segmentation of boundary and blurred regions.

Multi-scale CNNs enhance spine image segmentation by extracting features at multiple spatial scales, unlike fixed-scale CNNs that operate at a single resolution. According to [21], multi-scale approaches generally fall into three categories: (1) using input patches of varying sizes at the same resolution, (2) using inputs of the same region at different resolutions, and (3) applying convolutional kernels of different sizes.[43] employed the first approach to build an ensemble of Residual U-Nets with different fields of view, while [12,21,40], and [47] use the third approach.

In addition to the commonly adopted enhancements, several studies proposed other innovative modifications to the U-Net architecture, each designed to address specific segmentation challenges and imaging contexts. [13] proposed a 3D U-Net-based architecture with a shared encoder and task-specific decoders for vertebra instance segmentation, anatomical label regression, and completeness classification, incorporating an instance memory mask to guide segmentation of already-identified vertebrae. [18] proposed the BSU-pooling layer, which combines max-pooling with strided convolutions to retain local information during downsampling, reducing boundary detail loss in segmentation. [12,38] substituted the original convolution blocks with dense blocks, which connect each layer to all subsequent layers within the block, enhancing feature reuse, improving gradient flow, and enabling more compact and efficient representations. [57] used a cascaded Self-ONN-UNet to progressively refine segmentation outputs across multiple stages, and [60] incorporates a Meijing convolutional filter at the beginning of the U-Net architecture.

Table 4
Summary of the data preprocessing, augmentation, and split strategies used in the reviewed articles.

Author	Year	Data preprocessing	Data augmentation	Data split
Lessmann et al.	2019	normalization; resampling	elastic deformations, Gaussian noise/smoothing, cropping along z-axis	3-fold cross-validation
Pang et al.	2021	cropping, resizing, padding, normalization	rotation, elastic deformation, contrast adjustment	5-fold cross-validation
Al-Kafri et al.	2019	registration, bias field correction, resizing	not mentioned	multiple ratios tested (20:80 to 80:20)
Huang et al.	2020	not mentioned	not mentioned	50:50 split
Zheng et al.	2022	not mentioned	not mentioned	training: 223 subjects from set A and 93 from set B. test: 80 subjects from set A and 24 from set B
Kim et al.	2018	resizing, normalization	not mentioned	5-fold cross-validation
Han et al.	2021	not applied	not applied	80:20 + 5-fold CV on train set
Mbarki et al.	2020	not applied	image flipping	training: 1320, test: 320
Li et al.	2021	brightness/contrast adjustment; normalization	not mentioned	70:20:10 split
Šušteršič et al.	2022	brightness adjustment, CLAHE	horizontal flip	80:10:10 subjects-based
Zhou et al.	2020	not applied	not applied	training: 1230 subjects; test: 205 subjects
Shen et al.	2021	not applied	not mentioned	75:25 split
Rehman et al.	2019	padding and resizing	rotations, shifting, intra-slice extraction	5-fold cross-validation
Masood et al.	2022	not mentioned	not mentioned	70:30 split
Bharadwaj et al.	2023	not mentioned	not mentioned	75:10:15 subjects-level split
Iriondo et al.	2020	AHE; standardization	rotations, elastic deformation, Gaussian brightening	5-fold CV subjects-level split
Hwang et al.	2022	resampling, normalization, mask alignment	flipping, rotation, zoom, shear	50:50 split
Liu et al.	2022	cropping, standardization, padding	not mentioned	5-fold cross-validation
Gao et al.	2022	image alignment, standardization, padding	not mentioned	50:42.5:7.5 train-val-test split
Hess et al.	2023	not mentioned	not mentioned	75:15:10 train-val-test split
Baur et al.	2022	not mentioned	shift, scale, rotation, elastic, optical distortion	10-fold cross-validation
Sáenz-Gamboa et al.	2023	registration, standardization, patch extraction	rotation, zoom, shift, flip	80:20 + 3-fold CV on train
Niemeyer et al.	2022	not mentioned	not mentioned	80:20 train-validation + external
Dourthe et al.	2022	not mentioned	rotations, elastic deformation, cropping	5-fold cross-validation
Bonnheim et al.	2023	cropping around L5	rotation, shift, zoom, shear	80:20 subjects-based
Yilizati-Yilihamu et al.	2023	cropping, resampling, padding, standardization	rotation, contrast, elastic deformation	train:138, val:4, test:30 subjects
Hou et al.	2023	downsampling (bicubic)	flipping, clipping	train:1157, val:100, test:290 subjects
Morris et al.	2024	not mentioned	not mentioned	80:20 + external validation
Saravi et al.	2023	normalization, label binarization	not mentioned	5-fold cross-validation

(*continued on next page*)

The nnU-Net framework proposed by [68], which automatically configures model architecture and training parameters based on dataset-specific characteristics, was also employed by [51,59].

These architectural adaptations demonstrate the flexibility and extensibility of the U-Net framework, enabling tailored solutions for diverse segmentation tasks. By integrating residual learning, attention mechanisms, multi-scale feature extraction, and specialized modules, the reviewed studies demonstrate how U-Net variants continue to evolve in response to the specific demands of medical image analysis.

Other architectures

While U-Net and its variants were the predominant choice, several studies explored alternative CNN-based architectures. V-Net, originally proposed by [69], was developed for 3D volumetric segmentation and retains a similar encoder–decoder structure to U-Net, but differs in several key aspects. It replaces max pooling with strided convolutions for downsampling, adopts PReLU instead of ReLU as the activation function, and integrates residual connections throughout. These design choices make V-Net more expressive but also increase its parameter count and computational complexity. Although originally intended for 3D inputs, 2D adaptations were employed by [27,31,32], and [35], while [50] implemented an extended 3D version incorporating squeeze-and-excitation modules and multi-scale feature extraction.

SegNet was another architecture identified in this review, used by [15,26], and [45]. Originally proposed by [70], SegNet follows an encoder–decoder structure but differs from U-Net by omitting skip connections. Instead, it uses max-pooling indices from the encoder to guide non-learnable upsampling in the decoder, thereby reducing memory consumption while maintaining spatial structure.

Mask R-CNN was used by [24,44,45,54]. It was first introduced by He et al. [71], and extends Faster R-CNN by adding a parallel branch for pixel-wise segmentation masks, enabling simultaneous object detection and instance segmentation through a two-stage architecture that incorporates Region Proposal Networks (RPNs), bounding box regression, classification, and a fully convolutional mask prediction head.

DeepLabv3, introduced by [72], is another architecture adopted in this review, known for its use of atrous convolutions and the Atrous Spatial Pyramid Pooling (ASPP) module, which enables multi-scale feature extraction without reducing spatial resolution. Pang et al. [14]

Table 4 (*continued*).

Author	Year	Data preprocessing	Data augmentation	Data split
Chen et al.	2022	cropping, standardization, zero-padding	not mentioned	5-fold CV: 70:10:20
Altun et al.	2023	resizing	not mentioned	90:10 split
Laiwalla et al.	2023	normalization, resampling	not mentioned	train: 100, test: 279 subjects
Mohanty et al.	2023	not mentioned	applied (not specified)	60:10:30 split
Adibatti et al.	2023	resizing, AHE	not mentioned	70:30 split
Balzer et al.	2022	resizing, normalization	not mentioned	60:20:20 split
Zhou et al.	2024	resizing	flip, Gaussian noise, scale/shift intensity	train: 1230, test: 205 subjects
Candito et al.	2024	misalignment correction, normalization	not mentioned	70:30 split + test on 8 subjects
Jung et al.	2024	Dixon swap correction	not mentioned	not mentioned
Wang et al.	2023	resizing	not mentioned	80:20 split
Kok et al.	2023	not mentioned	not mentioned	6 train, 2 test (scoliotic + implant)
Natalia et al.	2024	resampling (cubic interpolation)	not mentioned	60:20:20 split
Liang et al.	2024	contrast/intensity adjustments	brightness, contrast, rotation, translation	70:10:20 subjects-based
Xu et al.	2024	standardization, ROI cropping	not mentioned	70:30 split
Giaccone et al.	2024	inhomogeneity correction, normalization	elastic/grid distortions	10-fold cross-validation
Ding et al.	2024	resizing	not mentioned	train: 300 ; val: 72; test: 94 discs
Basak et al.	2025	Gaussian blur, CLAHE, heuristic slice trimming	translation, rotation, shear	10-fold cross-validation (30 subjects)
Yan et al.	2024	resizing, standardization	flipping, rotation	5-fold cross-validation
Karthik et al.	2025	orientation correction, resampling, normalization	extensive transformations	80:20 subjects-based split
Lakshmi et al.	2025	grayscale conversion, threshold ops	applied (not specified)	60:20:20 split

Abbreviations: AHE Adaptive Histogram Equalization; **CLAHE** Contrast Limited Adaptive Histogram Equalization; **CV** Cross-Validation; **ROI** Region of Interest; **subjects** Subjects; **val.** Validation set.

employed a modified DeepLabv3+ encoder with ASPP and atrous convolutions within a 3D Graph Convolutional Network (GCN) framework, designed to model long-range spatial dependencies for coarse segmentation. Zheng et al. [17] used DeepLabv3+ as the backbone for BianqueNet, integrating transformer-based skip connections and specialized fusion modules. [52] evaluated DeepLabv3 with various encoders – ResNet-18, ResNet-50, MobileNetV2, Xception, and Inception-ResNetV2 – to balance segmentation accuracy and computational efficiency.

[37,58] proposed specialized fusion networks – SAFNet and MHFN – that incorporate multi-scale feature extraction, attention modules, and hierarchical fusion schemes to improve the integration of spatial and modality-specific information for more accurate medical image segmentation.

Some researchers integrated non-CNN components to enhance model capabilities. [19] proposed an Adversarial Graph Network that combines a generative adversarial framework with symbolic graph reasoning embedded at the encoder stage to impose anatomical consistency via structured semantic priors. [28] developed a coarse-to-fine context memory (CFCM) network that incorporates convolutional LSTM units to model spatial–temporal dependencies and iteratively refine segmentation across image slices.

3.5. Loss function and hyperparameters

In deep learning, the loss function is a mathematical function that measures the distance between the model's predicted segmentation maps and the ground truth annotations. During training, the model tries to minimize this loss, thereby improving its predictions. An optimization method in machine learning is an algorithm or technique used to adjust a model's parameters during training to minimize the loss function. These algorithms use gradients of the loss function with respect to the model parameters to determine how the parameters should be updated at each training step.

Among the reviewed studies, cross-entropy-based loss functions were the most common, used in 19 articles [12–15,19,21,25,26,31,36–38,40,45–47,50,54,60], followed by Dice-based losses in 14 articles [13,17,21,28,31–33,35,39,43,46,47,55,58].

In addition to cross-entropy and Dice-based losses, several less common loss functions were identified. These include Jaccard distance loss [23], Intersection-over-Union loss [34], focal loss [33], focal Tversky loss [48], and Gaussian divergence loss [38]. A few studies employed hybrid or task-specific combinations, such as attention supervision loss [47], and custom multi-part losses that integrate components such as bounding box regression and segmentation mask accuracy [24,54].

Regarding the optimization algorithm, the Adam optimizer was the most commonly employed, reported in 31 of the selected articles [12–14,17,18,20,21,28–33,35–41,43,45–48,50,53–55,58,60]. The SGD optimizer was used in 6 studies [15,21,22,25,41,59], while Adadelta was reported in two studies [26,53], RMSProp and AMSGrad, a variant of Adam, were used in one study each [19,46], respectively.

3.6. Performance metrics

Model performance was primarily evaluated using overlap-based metrics. The Dice Similarity Coefficient (DSC) was the most frequently reported, appearing in 38 studies [13,14,17–19,21–23,25–33,35–37,39,41–43,46,48,49,51,53–60]. Intersection over Union (IoU), also referred to as the Jaccard Index, was found in 23 studies [12,15–17,20,22–24,26,29,34,38,42,50,52,53,56,57,60]. Other evaluation metrics included accuracy [15,19,22,26,40,45,52,57,60], precision and recall (or sensitivity) [14,19,22,26,31,33,36,40,41,45,47,48,54,57], specificity [19,36], and F1-score [40]. Surface-based metrics such as the Hausdorff Distance (HD) and Average Symmetric Surface Distance (ASSD) were used in a subset of studies [13,18,21,29,43,47]. These were especially relevant for evaluating boundary precision.

Across studies, segmentation performance shows consistent structure-dependent trends rather than uniform gains across all targets. Large, well-defined anatomical regions such as vertebral bodies and intervertebral discs typically achieve the highest accuracy, with Dice Similarity Coefficient values frequently reported above 0.90 and often

Table 5
Summary of the model architectures used in the reviewed articles.

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Lessmann et al.	2019	3D 128 × 128 × 128 patch	3D U-Net-based FCN with shared encoder and separate decoder branches for vertebrae segmentation, label regression, and completeness classification. Uses instance memory concatenated with the input patch.	segmentation error (Dice loss variant) + labeling error (L1 norm) + completeness classification error (binary cross-entropy)	Adam optimizer, LR=1e−3, momentum=0.99, 100000 iterations	DSC: 0.944 ±0.033; ASSD: 0.4 ±0.3 mm
Pang et al.	2021	3D volumes and 2D slices	3D GCSN (based on DeepLabv3+) for coarse segmentation + 2D Res-U-Net for refinement	Multi-class cross-entropy (used with deep supervision)	Adam optimizer; LR=1e-3 (3D), LR=1e-4 (2D), LR drops at epochs 16, 33, 66	DSC: 0.873-0.878
Al-Kafri et al.	2019	3-channel 360 × 480 input (T1, T2, and Manhattan diff)	SegNet-TL (transfer learning from VGG16) and SegNet-FS (from scratch)	Cross-entropy loss	SGD optimizer, LR=1e−3, 100 epochs	IOUc: IVD=0.92, PE=0.78, TS=0.85; BF-Score (d_i=1/2/3): IVD=0.21/0.60/0.81
Huang et al.	2020	2D slices	Adapted U-Net	Not mentioned	Not mentioned	IoU: VB=0.947, IVDs=0.926 (elderly: 0.941 and 0.925)
Zheng et al.	2022	Model A: 512 × 512; Model B: 320 × 320	BianqueNet (DeepLabv3+ backbone) with Swin Transformer, Skip-Connection (ST-SC), Multi-scale Fusion (MFF), and Depth Feature Extraction (DFE) modules	Weighted Dice loss	Not mentioned	DSC: VB=0.970, IVD=0.948
Kim et al.	2018	2D 256 × 256 images	BSU-net: U-Net with res blocks and BSU pooling	not mentioned	Adam optimizer with LR set to 1e-3; 200 epochs	BSU-Net: DSC (area): 0.894 ±0.214, DSC (boundary 1px): 0.546 ±0.459, DSC (boundary 2px): 0.749 ±0.320, MHD: 0.81 ±0.10 mm
Han et al.	2021	2D slices	Adversarial Graph Network: Generative Adversarial Network + Symbolic Graph Reasoning embedded on the bottom of the generator (at the end of the encoder)	Hybrid loss function: generator loss function (weighted multi-class cross-entropy) + discriminator loss function (binary cross-entropy)	Generator: RMSProp (decay=0.9, momentum=0.9, e=1e-10), LR=1e-2 Discriminator: Adam (decay = 0.9, momentum=0.999, e=1e-8), LR=1e-3; Xavier weight initialization	Pixel accuracy: 0.965 ±0.004 DSC:0.879 ±0.003 Specificity: 0.903 ±0.012 Sensitivity: 0.872 ±0.023
Mbarki et al.	2020	2D slices	U-Net	not mentioned	Adam optimizer (LR=1e-3 for convolution layers, LR=1e-4 for deconvolution layers)	IoU: 0.933
Li et al.	2021	2D slices 512 × 512	Multi-scale Attention U-Net (MANet): dual-branch multi-scale feature extraction module and attention-based selection module	Hybrid loss: Generalized Dice loss + Cross-entropy loss	SGD with momentum 0.9 and LR=1e-3 with decay=0.1, mini-batch size=4; Xavier weight initialization	Improvement over U-Net: DSC ↑ 0.901 → 0.925, ASD ↓ 6.4 mm → 2.71 mm
Šušteršič et al.	2022	2D slices 256 × 256	Adapted U-Net	not mentioned	SGD optimizer with different learning rates	Test Performance: Axial view - Acc: 0.996, DSC: 0.961, IoU: 0.925, Prec: 0.986, Rec: 0.954 Sagittal view - Acc: 0.997, DSC: 0.897, IoU: 0.813, Prec: 0.976, Rec: 0.860

(*continued on next page*)

exceeding 0.95 in controlled settings [16,17,26,60]. In contrast, elongated or anatomically complex structures, including the spinal canal, nerve roots, laminae, and posterior elements, show more variable and generally lower performance, with DSC values often ranging between 0.75 and 0.90 [12,24,38,47]. Very small or low-contrast targets, such as nerve roots, epidural fat, or fine muscle boundaries, remain particularly challenging, as reflected by lower IoU values and higher surface distance errors even in advanced multi-class models [12,47]. These trends suggest that segmentation accuracy is strongly influenced by object size, contrast, and anatomical variability, rather than solely by network architecture.

3D models generally demonstrate improved spatial consistency and robustness for volumetric structures, particularly vertebrae and discs, but at the cost of higher computational demands and increased sensitivity to limited training data [13,30,41]. Attention mechanisms, including Attention U-Net, CBAM, and uncertainty-aware attention, are repeatedly associated with improved boundary delineation and reduced false positives, especially in overlapping or low-contrast regions [21,39,47]. Similarly, multi-scale feature extraction, whether through parallel convolutions, pyramid structures, or ensembles with varying fields of view, appears to enhance robustness to anatomical variability and disc degeneration [12,40,43]. Data augmentation, although inconsistently reported, is frequently credited with stabilizing training and improving

Table 5 (*continued*).

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Zhou et al.	2020	2D slices 256 x 256 pixels, with 4 channels: water, fat, fat fraction, and R2°maps	Standard U-Net	Jaccard distance loss	Adam optimizer with LR= 1e-4; Batch Size: 4	DSC/IoU (Mean ± SD): Set 1A: 0.956 ± 0.076/0.928 ± 0.105 Set 1B: 0.886 ± 0.040/0.798 ± 0.059 Set 2A: 0.838 ± 0.198/0.757 ± 0.216 Set 2B&3B: 0.849 ± 0.091/0.747 ± 0.118
Shen et al.	2021	2D 400 × 400 slices	Based on Mask R-CNN (region proposal network + ROI Align + mask head) with two postprocessing rules to handle small regions and reassign mislabeled muscle candidates.	Standard Mask R-CNN multi-part loss: bounding box regression loss + classification loss + segmentation mask loss	initial LR of 1e−4, momentum = 0.9; optimizer not specified	IoU: 0.823–0.922 (Best: Disc, Worst: Spinal canal); DSC: 0.902–0.959 (Best: Disc, Worst: Spinal canal)
Rehman et al.	2019	2D 128 × 128 slices	Adapted U-Net. A parametric level set was integrated with the deep network to refine segmentation.	Cross-entropy loss	decaying learning rate (initial value not specified), momentum = 0.9 and dropout=0.2, optimizer not specified	DSC: 0.904 ± 0.009
Masood et al.	2022	2D slices	Multiple encoder–decoder models (e.g., UNet, ResNet-UNet, SegNet, PSPNet, MobileNets).	Cross-entropy	Adadelta optimizer, batch size = 2, 50 epochs	For ResNet-UNet: mpA = 0.991, mpP = 0.983, IoU = 0.86, DSC = 0.97;
Bharadwaj et al.	2023	2D slices	2D V-Net; separate models were trained for the dural sac and IVD segmentation.	not mentioned	not mentioned	DSC: Dural sac 0.93, IVD 0.94
Iriondo et al.	2020	2D 256 × 256 slices	A coarse-to-fine context memory (CFCM) network based on convolutional LSTM.	Dice loss	Adam optimizer, LR of 1e-6 and ϵ of 1e-8; batch size = 20; Xavier weight initialization; trained for 8000 epochs	DSC: >0.85 overall; set A: 0.87-0.93, set B: 0.88-0.91
Hwang et al.	2022	2D 528 × 528 slices	2D U-Net: Three models were trained on different datasets: UNET_HC (healthy only), UNET_HD (diseased only), and UNET_HCHD (combined healthy + diseased)	not mentioned	Adam optimizer, LR of 1e-2; batch size of 32; trained for 100 epochs	DSC: Highest for UNET_HD/HCHD (0.950), lowest for UNET_HC (0.830) IoU: Best with UNET_HD (0.913), worst with UNET_HC (0.774) HSD: Lowest for UNET_HD (1.589 mm), highest for UNET_HC (2.124 mm)
Liu et al.	2022	3D 128 × 180 × 180 volumes	3D U-Net	not mentioned	Adam optimizer; batch size = 2; trained for 100 epochs; initial LR = 5e-4 (reduced by a factor of 5 at epochs 33 and 66).	DSC (mean ± SD): L5 vert.: 0.935 ± 0.0293; IVDs: 0.904 ± 0.062; S1 vert.: 0.933 ± 0.015; All L5/S1 structures: 0.924 ± 0.028.
Gao et al.	2022	Vert. segmentation: 2D T1W slices Modic segmentation: 2D 100 × 100 slices with 2 channels (stacked T1 and aligned T2)	2D V-Net	Vert. segmentation: Dice loss; Modic segmentation: Weighted Cross-entropy combined with Dice loss (loss weights of 20:1 for foreground:background)	Adam optimizer; for Vert. segmentation: Batch size=3; LR=1e-4; Dropout = 0.8; Modic segmentation: Batch size=128; LR = 1e-4; Dropout = 0.2	Vert. segmentation: DSC = 0.882 ± 0.018 on the test set (inter-rater DSC = 0.927 ± 0.011) Modic segmentation: Sensitivity = 0.71 ± 0.072, Specificity = 0.95 ± 0.022, Cohen's kappa = 0.63

(*continued on next page*)

generalization, particularly in small private datasets [12,28,33]. Due to substantial variation in datasets, segmentation targets, evaluation strategies, and model architectures, direct comparison of metric values across studies was not conducted. This heterogeneity limits the interpretability of pooled performance and highlights the need for standardized benchmarks in future research.

3.7. Review articles

Table 6 summarizes seven review articles included in this review. Azimi et al. [73] and Ren et al. [74] offer wide-ranging overviews, highlighting applications beyond imaging, such as biomechanical modeling and clinical decision support. Cui et al. [75] provide a narrative integration of AI methods within spinal imaging workflows, without formal inclusion criteria. More specialized reviews, such as those by Vrtovec et al. [76] and Chen et al. [77], target spinopelvic sagittal balance measurements and bibliometric trends, respectively. Vrettos et al. [78] explore the role of generative adversarial networks (GANs) in spinal image synthesis and enhancement, while Verheijen et al. [79] compare conventional ML and DL techniques specifically for lumbar spinal stenosis.

The number of included studies across reviews varies widely, ranging from as few as 27 to nearly 300 articles, and spans a publication period from the early 1990s to 2023. Most reviews utilized databases

Table 5 (*continued*).

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Hess et al.	2023	2D slices; sagittal 512 × 512 images and axial 256 × 256 images.	3 separate 2D V-Net for IVD, vert. body and paraspinal muscles segmentation	Weighted Dice loss	Adam optimizer, LR: 1e-4. Batch Size: 8 slices for disc and vert. body segmentation, 32 slices for muscle segmentation in the standard split; 8 slices in the shared split. Dropout: 0.2 for disc and muscle networks; 0.05 for vert. body network.	DSC (Standard/Shared): Disc: 0.88/0.86, VB: 0.88/0.76, Muscle: 0.81–0.95/0.78–0.88
Baur et al.	2022	2D 320 × 320 slices	U-Net	Combination of dice loss and focal loss	not mentioned	Overall: DSC 0.91, CSA error 0.084, Precision 0.92, Recall 0.90; Muscle tissue: DSC 0.83, CSA error 0.14, Precision 0.82, Recall 0.86; Fatty tissue: DSC 0.51, CSA error 0.41, Precision 0.54, Recall 0.53
Sáenz-Gamboa et al.	2023	2D 256 × 256 × 2 slices (2 channels: T1W and T2W)	Variants derived from U-Net combining different blocks: U-Net convolution blocks, VGG16 convolution blocks, and Dense blocks; Multi-kernel input block; Spatial attention gates and Deep supervision blocks. Ensembles of several network variants were also evaluated.	Categorical cross-entropy loss	optimizer: Primarily Adam (RMSprop or Adadelta were also used). Most configurations use a LR of 3.33e−4	Best topology: UMD (U-Net + multi-kernel + dense block after decoder) with TH labeling. Best ensemble: E12 (12 U-Net variants, TH labeling). Best IoU (without Bg): 0.753 (E12, TH); with Bg: 0.767. Strongest classes: SCT, Muscle, IVDs (IoUs >0.89). Challenging classes: Nerve Root (IoU ≈ 0.53), Epi-Fat (0.60)
Niemeyer et al.	2022	2D slices	Modified 2D U-Net	Intersection-over-Union loss	Adam opt with a LR of 1e-3 and a batch size of 16.	Median IoU: 0.862-0.935
Dourthe et al.	2022	2D 224 × 160 slices	Two-stage framework: Stage 1 - 2D multiscale pyramid CNN Stage 2 - Based on a 2D V-Net	Multiclass Dice loss - sum of Dice coefficients from two stages	Adam optimizer with a LR of 1e-4; trained for 500 epochs	Healthy groups: DSC >0.87 for all classes ASD group: DSC 0.76–0.92
Bonnheim et al.	2023	2D slices 128 × 128	Original U-Net. Four independent models were trained (one per sublevel: L4-L5 superior/inferior, L5-S1 superior/inferior)	Binary cross-entropy	Adam optimizer with LR of 1e-4 and batch size = 4	Sensitivity: 0.80–0.91, Specificity: 0.99, DSC: 0.77–0.85, ROC-AUC: 0.99–1.00, PR-AUC: 0.56–0.77
Yilizati-Yilihamu et al.	2023	3D MR 18 × 256 × 128 volumes	Scene-Aware Fusion Network (SAFNet). Key Blocks: Feature extraction network, Atrous Spatial Pyramid Pooling (ASPP), self-attention modules (position and channel attention), multi-scale fusion, and dimension splicing.	Pixel-wise cross-entropy loss	Adam optimizer with initial LR of 1e−3, batch size = 8, weight decay = 1e−4, trained for 50 epochs. A ReduceLROnPlateau scheduler (patience = 10, factor = 0.5) was used to dynamically adjust LR.	DSC: 0.803 ± 0.05
Hou et al.	2023	2D 224 × 224 slices	A variant of U-Net, FC-DenseNet, composed of dense blocks, transition down and up blocks, and incorporating a two-phase training strategy.	Combined loss function: Cross-entropy loss, Contour loss and Gaussian divergence loss	Adam optimizer; growth rate=12; dropout=0.8 Stage 1: LR = 1e−3, batch size = 8, L2 regularization weight = 1e−4, 30 epochs, using both random flipping and clipping Stage 2: Learning rate = 1e−4, only random flipping used, 40 epochs	Pixel Accuracy = 0.9935, IoU = 0.8493

(*continued on next page*)

such as PubMed, Scopus, Web of Science, and MEDLINE, often combining MeSH terms and free-text queries to achieve comprehensive retrieval.

4. Discussion

This review synthesizes current research on deep learning methods for lumbar spine segmentation in MRI, analyzing 49 original studies across dataset characteristics, model architectures, training pipelines, and evaluation strategies. The analysis reveals a field that is rapidly evolving, with growing methodological sophistication and increasing adoption of advanced neural architectures. Nonetheless, several areas warrant further refinement to enhance robustness, reproducibility, and clinical utility.

A notable strength of the reviewed literature is the widespread adoption of convolutional neural networks, particularly the U-Net

Table 5 (*continued*).

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Morris et al.	2024	3D volumes	Lightweight Attention-based 3D U-Net (termed CBAM-ROI-attention U-Net).	Combined loss function: Weighted Dice loss for full image segmentation, Dice loss for ROI segmentation, and ROI center localization loss	Adam optimizer	Dice Scores: Spine: 91.2%, Femoral head: 94.5%, Total hip: 91.2% Femoral diaphysis: Improved from 69.0% (U-Net) → 86.6% (ROI-attention U-Net)
Saravi et al.	2023	2D 320 × 320 slices	3 U-Net based architectures: standard U-Net, Attention U-Net, and MultiResUNet.	Weighted binary cross-entropy (weights: positive = 20, negative = 1)	Adam optimizer, batch size = 32, 20 epochs	MultiResUNet: Acc. 0.9996, Recall 0.9996, Precision 0.8966, F1 0.9453, MAE 17.95 mm^2 Attention U-Net: Acc. 0.9992, Recall 0.9903, Precision 0.7966, F1 0.8829, MAE 39.41 mm^2 U-Net: Acc 0.9990, Recall 0.9943, Precision 0.7673, F1 0.8662, MAE 46.60 mm^2
Chen et al.	2022	3D 128 × 180 × 180 volumes	3D U-Net	not mentioned	Adam optimizer with initial LR of 5e-4 (reduced by a factor of 5 at epochs 33 and 66) and batch size = 2	Mean DSC: 0.918 overall (Healthy: 0.922, Unhealthy: 0.916) Precision: 0.922 Recall: 0.917
Altun et al.	2023	2D 128 × 128 slices	Modified U-Net with different encoders.	not mentioned	not mentioned	Best Model (ResUNet): DSC = 0.93, IoU = 0.892
Laiwalla et al.	2023	2D slices; models trained on input fields of view of 128 × 128, 160 × 160, and 192 × 192 pixels	Deep U-Net ensemble comprising three Res U-Net models with different FOV to capture multi-scale features.	Dice loss	Adam optimizer with fixed LR of 1e-5; Dropout rate of 0.25; trained for 1000 epochs.	DSC: Machine vs Expert 1: 0.88 ± 0.04, Expert 2: 0.89 ± 0.04 Hausdorff Distance (mm): 11.7 ± 13.8 (Exp 1), 13.1 ± 16.3 (Exp 2) Avg. Surface Distance (mm): 0.18 ± 0.13 (Exp 1), 0.18 ± 0.16 (Exp 2) Inter-rater DSC: 0.94 ± 0.02
Mohanty et al.	2023	2D slices	Multiple Mask Regional Convolutional Neural Networks (MRCNNs).	not mentioned	not mentioned	Average Segmentation PSNR: 47.62 dB
Adibatti et al.	2023	2D 320 × 320 slices	Two-stage framework: Stage 1: Mask R-CNN for IVD localization Stage 2: SegNet for segmentation	Binary Cross-entropy	Adam optimizer with LR of 1e-1; Batch size =128; Epochs: 100; Momentum: 0.8; Weight tuning achieved via AROA	Accuracy = 0.985, Precision = 0.979, Recall = 0.978
Balzer et al.	2022	2D 512 × 512 slices	Modified U-Net: inception modules on the decoder branch	Combination of binary weighted cross-entropy loss and Dice loss	Adam with AMSGrad optimizer with LR=1e−4, β_1 = 0.9, β_2 = 0.999, ϵ = 1e-7; Batch size = 6; Epochs = 30; Weight Initialization: He normal approach	DSC = 0.918 ± 0.004
Zhou et al.	2024	2D 320 × 320 slices	BUA-Net – a boundary and uncertainty-aware attention network that uses a single encoder with two decoders. It incorporates a multi-scale spatial and channel attention (MSCA) module, an uncertainty-aware attention module, and a fine segmentation (FS) module.	Attention supervision loss: cross-entropy loss for the boundary decoder; cross-entropy and dice losses for the uncertainty-aware attention decoder	Adam optimizer (β_1 = 0.9, β_2 = 0.999, ϵ = 1e-8) with LR = 1e-4 with decay by a factor of 10 every 50 epochs; Batch Size: 6; Weight Decay: 1e−5	Precision/Recall – IVD: 0.966/0.966, TS: 0.896/0.752, PE: 0.917/0.975, AAP: 0.949/0.815, BG: 0.998/0.998 95HD (mm): IVD: 10.71, TS: 18.81, PE: 5.36, AAP: 4.12, BG: 1.43 ASD (mm): IVD: 0.62, TS: 0.62, PE: 0.59, AAP: 0.85, BG: 0.21

(*continued on next page*)

and its derivatives. These models have proven effective in preserving anatomical detail and have been adapted with innovations such as residual blocks, attention mechanisms, and multi-scale processing to better address the challenges posed by spinal anatomy and MRI variability. While architectural convergence around U-Net may reflect its proven utility, it also suggests that future work could benefit from exploring more diverse approaches, such as transformer-based architectures or hybrid CNN—transformer models. Transformers can explicitly model long-range spatial dependencies and global anatomical context through self-attention, which may be particularly advantageous for capturing inter-vertebral relationships and multi-level spinal structure beyond the receptive field of conventional convolutions. In addition, emerging diffusion-based models may support uncertainty-aware segmentation and advanced data augmentation strategies, offering new avenues to improve robustness and generalizability in data-limited clinical settings.

Table 5 (*continued*).

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Candito et al.	2024	2D 256 × 256 slices; 2-channel input (ADC map and S0)	Adapted U-Net	Focal Tversky loss	Adam optimizer, Batch size=8, trained for 150 epochs	DSC: 0.88, Precision: 0.89, Recall: 0.90
Jung et al.	2024	3D MR volumes	Convolutional Neural Networks	not mentioned	not mentioned	DSC ≥ 0.88 in NAKO dataset and ≥ 0.86 in the UKBB dataset
Wang et al.	2023	3D 256 × 256 × 12 volumes	Modified V-Net with 3D squeeze-and-excitation (SE) blocks and multi-scale feature extraction (MFE) blocks (using parallel atrous convolutions).	Weighted cross-entropy	Adam optimizer, LR= 1e−4, batch size = 5, trained for 500 epochs.	Mean IoU: 0.777
Kok et al.	2023	Multi-channel 3D MR volumes (in-phase, out-of-phase, fat, and water channels)	Self-configured nn-Unet	not mentioned	not mentioned	DSC: 0.93 (scoliotic subject) and 0.95 (implant subject)
Natalia et al.	2024	2D 384 × 384 slices	DeepLabv3 with different encoders: ResNet-18, ResNet-50, MobileNetv2, Xception, and InceptionResNetv2.	not mentioned	For the best model: Adam optimizer with initial LR of 1e−3, LR drop factor = 0.3, LR drop period = 10 epochs; L2 regularization factor = 5e-3; Gradient decay factor = 0.9; batch size = 16; Number of epochs = 120	Best Model (ResNet-50), (Acc/IoU/BFScore): IVD: 0.977/0.882/0.991, VB.: 0.972/0.924/0.995
Liang et al.	2024	2D 320 × 320 slices	Residual U-Net	not mentioned	Adadelta optimizer, LR of 1e-3; batch size of 10; trained for 2000 epochs	IoU/DSC: Sagittal: Disc 0.91/0.91, VB 0.93/0.93, Dural sac 0.93/0.87 Axial: Disc 0.82/0.82, Lamina 0.85/0.84, Dural sac 0.82/0.82, Yellow lig. 0.83/0.83
Xu et al.	2024	2D 512 × 512 slices	Segmentation Model: Cascade Mask R-CNN with Feature Pyramid Network (FPN); backbone based on ResNet101 with pre-trained weights from Open-MMLab FI Model: U-Net based architecture	Segmentation Model: Combination of smooth L1 loss (for bounding box regression) and cross-entropy loss (for target classification) FI Model: Custom Focal Loss with $\gamma = 5$	Segmentation Model: SGD optimizer, batch size = 16, initial LR = 2e-2. FI Model: Adam optimizer, batch size = 16, initial LR = 1e−3	segmentation Model: Int. set: DSC 0.92 ± 0.10, Recall 0.926, mAP 0.98 Ext. set: DSC 0.91 ± 0.09, Recall 0.908, mAP 0.98 FI Classification Model: Int. set: Precision 0.743, Kappa 0.76, AUC 0.97 Ext. set: Precision 0.577, Kappa 0.61, AUC 0.93
Giaccone et al.	2024	2D 256 × 256 slices	U-Net architecture with a ResNet34 backbone. Separate models were used for left and right paraspinal muscle segmentation	Smoothed Dice loss (with a regularization factor set to 2000, determined via a grid search).	Adam optimizer trained over 100 epochs, batch size=50.	DSC Internal set: Right 0.952 ± 0.02, Left 0.951 ± 0.02 DSC External set: Right 0.945 ± 0.02, Left 0.943 ± 0.02

(*continued on next page*)

The interpretability of deep learning models remains a critical concern for their adoption in clinical practice, and only a limited subset of the reviewed studies explicitly addressed this issue. Gao et al. proposed an interpretable pipeline for voxel-wise mapping of Modic changes, combining deep learning-based segmentation with rule-based signal-intensity classification to generate spatially explicit Modic maps that can be directly inspected and validated by radiologists, and demonstrated improved inter-rater agreement in an AI-assisted reading experiment. Bonnheim et al. focused on clinical transparency by linking U-Net-based segmentations of cartilage endplates to quantitative, biophysically meaningful T2* biomarkers, allowing clinicians to trace model outputs back to established imaging biomarkers rather than abstract feature representations. Han et al. introduced a hybrid approach that integrates neural networks with symbolic graph reasoning, explicitly encoding anatomical and semantic relationships between spinal structures, thereby constraining the model's predictions to anatomically plausible configurations and improving interpretability at a structural level. Finally, two-stage frameworks such as SpineParseNet, developed by Pang et al., decomposed the segmentation task into anatomically meaningful sub-steps with intermediate semantic representations, making the intermediate outputs interpretable and easier for clinicians to audit.

A major limitation identified across the reviewed literature is the reliance on private, single-center datasets. In this review, 63% of studies used exclusively private data, and only 9 of the 49 original studies (18%) evaluated their models on an independent external test set. While many reported high segmentation performance, such evaluations are likely to provide optimistic estimates of model accuracy due to limited anatomical, pathological, and acquisition variability. Incorporating publicly available datasets and multi-center validation would likely lead to an initial decline in commonly reported performance metrics, such as Dice Similarity Coefficient or Intersection-over-Union, reflecting the effects of domain shift due to differences in scanner vendors, imaging protocols, patient demographics, and disease prevalence. However, this decline should be interpreted as a necessary step toward realistic clinical generalizability rather than a weakness of the methods themselves. Models demonstrating stable performance across heterogeneous datasets are more likely to translate into routine clinical practice. Dataset size, image quality, and annotation fidelity play a critical role

Table 5 (*continued*).

Author	Year	Model input	Model architecture	Loss function	Hyperparameters	Results
Ding et al.	2024	2D 512 × 512 slices	Standard U-Net	not mentioned	SGD optimizer	Dice = 0.920; IoU = 0.853
Basak et al.	2025	2D slices	A cascaded framework with two stages: Stage 1: ROI extraction using the YOLOv8 medium model Stage 2: segmentation using a Self-ONN-based encoder–decoder network. The encoder is a DenseNet121, and the decoder is a Self-ONN-U-Net	not mentioned	Best model: LR of 1e-3; optimizer not specified	segmentation: DSC 0.910, IoU 0.837, Sensitivity 0.914, Acc. 0.998 ROI Extraction (YOLOv8m): mAP(0.5:0.95) = 0.916, mAP(0.5) = 0.993
Yan et al.	2024	2D 256 × 256 slices	Multi-modality Hierarchical Fusion Network (MHFN) comprising: two modality-specific encoders, one each for T1W and T2W-STIR, using a multi-scale feature encoder with a split strategy and dilated convolutions; an Adaptive Group Fusion Module (AGFM) that splits and fuses cross-modal features at a granular level; a hierarchical fusion structure that connects features across layers; a decoder with skip connections to generate the final segmentation map	Dice loss with class weights set to 1/C	Adam optimizer, LR=1e-3; batch size = 4	With AGFM (g = 4): Normal Vert. DSC: 0.865 IVDS DSC: 0.897 Fractured Vert. DSC: 0.851
Karthik et al.	2025	2D and 3D inputs were separately tested	The model is based on a region-based training strategy adapted from nnU-Net.	not mentioned	SGD optimizer with a polynomial LR scheduler; trained for 1000 epochs	3D Model Performance (outperformed 2D model): Spinal Cord segmentation: DSC 0.92 ± 0.07, RVE 0.35 ± 13.45 Lesion Segmentation: DSC 0.61 ± 0.27
Lakshmi et al.	2025	2D 320 × 320 slices	Modified U-Net - MU-Net that incorporates a Meijing convolutional layer.	Categorical cross-entropy loss	Adam optimizer, LR =1e-3 and dropout=0.1, Trained for 60 epochs	Pixel Accuracy: 0.988, DSC: 0.987, IoU: 0.974, Mean IoU: 0.926

Abbreviations: **AAP**: Area between Anterior and Posterior vertebral elements, **Acc.** Accuracy, **ASD**: Average Surface Distance, **ASSD**: Average Symmetric Surface Distance, **BF-Score**: Boundary F1 Score, **BG**: Background, **CSA**: Cross-sectional Area, **DSC**: Dice Similarity Coefficient, **FOV**: Field of View, **FI**: Fat Infiltration, **HSD**: Hausdorff Surface Distance, **IoU/IOU/IOUc**: Intersection over Union/Class-wise IoU, **IVD**: Intervertebral Disc, **LR**: Learning Rate, **MAE**: Mean Absolute Error, **MHD**: Mean Hausdorff Distance, **mpA/mpP**: Mean Pixel Accuracy/Mean Pixel Precision, **mAP**: Mean Average Precision, **mDice**: Mean Dice Score, **PSNR**: Peak Signal-to-Noise Ratio, **Precision**: Precision, **PR-AUC**: Precision–Recall Area Under the Curve, **ROI**: Region of Interest, **segmentation**: Segmentation, **SGD**: Stochastic Gradient Descent, **T1W/T2W**: T1-weighted/T2-weighted MRI sequences, **TH**: Thresholded (in labeling strategy), **TS**: Thecal Sac or Total Spine (context-dependent), **VB**: Vertebral Body **Vert.**: Vertebrae/Vertebral

in determining segmentation performance and model robustness. In this review, 11 of the 49 original studies relied on relatively small datasets (fewer than 50 subjects or fewer than 250 image slices), which may limit variability in anatomy, pathology, and acquisition conditions encountered during training and evaluation, thereby contributing to optimistic performance estimates. Reproducibility is further limited by the scarce availability of open-source implementations. Among the 49 original studies reviewed, only two [39,40] explicitly reported making their code publicly available via GitHub. The lack of accessible code and trained models hinders independent verification, benchmarking, and downstream methodological comparisons, particularly given the substantial heterogeneity in datasets, preprocessing pipelines, and evaluation protocols. Open-source implementations, when combined with publicly available or multi-center datasets, would substantially improve transparency, facilitate external validation, and accelerate progress toward clinically robust and reproducible lumbar spine segmentation frameworks.

Although this review focuses primarily on MRI-based segmentation, a small number of studies have explored alternative or complementary imaging modalities. For instance, Liu et al. incorporated datasets from CT, MRI, and X-ray imaging to address spinal analysis tasks, demonstrating that different modalities provide complementary structural information depending on tissue type and clinical objective. Similarly, Lessmann et al. included both CT and MRI data in their study; however, these modalities were treated as separate datasets with independently trained models rather than within a unified multi-modal framework. These examples suggest that true multi-modality fusion – particularly combining MRI with CT for bony anatomy or leveraging cross-modal priors – remains largely unexplored in lumbar spine segmentation. Future studies may benefit from systematically investigating multimodal learning strategies that jointly leverage complementary information across imaging modalities to improve segmentation accuracy, robustness, and clinical applicability.

Training strategies and loss function choices demonstrate a degree of convergence around standard practices, such as the use of the Adam optimizer and losses based on the Dice or cross-entropy metric. A growing number of studies employ composite or task-specific loss functions, a promising direction particularly for addressing class imbalance and

Table 6
Description of the reviews included in this review.

Author	Year	Description	Search	Document type	Databases	No. of articles	Years covered
Azimi et al.	2020	The review highlights that ANNs – including deep learning models – have been applied successfully not only for automated segmentation, e.g., in lumbar spine MRI and CT data, but also for diagnosis, outcome prediction, and biomechanical assessments.	The search combined keywords such as "artificial neural networks," "spine," "back pain," "prognosis," "grading," "classification," "prediction," "segmentation," "biomechanics," "deep learning," and "imaging."	Full-text publications in English biomedical journals	Medline (via PubMed) and Scopus (Elsevier)	79 from 3653	1993–2020
Ren et al.	2022	The review summarizes ML applications across several areas, including image processing, diagnosis, decision support, operative assistance, rehabilitation, outcome prediction, complications, and cost management.	The search used a combination of terms: "spine," "spinal," "lumbar," "cervical," "thoracic," and "machine learning."	Peer-reviewed articles	PubMed	292 from 1738	2006–2020
Cui et al.	2022	This narrative review discusses the integration of AI techniques – including machine learning and deep learning – into spinal imaging workflows.	N/A	N/A	N/A	N/A	N/A
Vrtovec and Ibragimov	2022	Current state of deep learning-based methods for automated spinopelvic sagittal parameter measurement from medical imaging.	Keywords included: (artificial intelligence OR deep learning) AND (sagittal balance OR spinopelvic balance OR pelvic incidence OR lumbar lordosis OR thoracic kyphosis OR cervical lordosis OR Cobb angle).	Published records in English	PubMed, Web of Science, Scopus, plus additional citation and web searches	34 from 529	2017–2022
Chen et al.	2023	Comprehensive bibliometric and visualization overview of deep learning applications in spine research.	TS = (Spine OR Spinal Column OR Vertebral Column OR Vertebra) AND TS = (Deep Learning)	English-language research articles	Web of Science Core Collection	273 of 586	2016–2022
Vrettos et al.	2024	The review details multiple GAN applications in spine imaging, including advanced segmentation, e.g., ResAttenGAN, Spine-GAN, Axial-SpineGAN, image-to-image translation, and synthesis of fat-suppressed or modality-converted images.	Keywords such as "GANs," "spine," "intervertebral disc," and "generative adversarial networks."	Not specified	MEDLINE, SCOPUS, ARXIV	33 from MEDLINE, 4 from ARXIV, 40 from SCOPUS (duplicates possible)	2017–2023
Verheijen et al.	2025	Provides a comparison of current conventional machine learning and deep learning methods used in segmentation and classification of lumbar spinal stenosis.	Search combining MeSH and free-text terms for AI-related techniques, e.g., algorithms, machine learning, deep learning, and lumbar spine pathologies, e.g., spinal stenosis, disc degeneration, and herniation.	Original research articles	Cochrane Library, Embase, Emcare, PubMed, Web of Science	27 from 1561	2010–2023

improving boundary accuracy. However, few studies provide systematic comparisons or ablation analyses to evaluate the specific contributions of these modifications. Future work could benefit from more rigorous methodological reporting, particularly regarding hyperparameter tuning, cross-validation design, and data split strategies.

Evaluation practices are predominantly centered on overlap-based metrics, particularly the Dice Similarity Coefficient and the Intersection-over-Union ratio. These remain useful for general performance assessment, but are limited in their ability to detect boundary inaccuracies or topological errors. Surface-based metrics, such as the Hausdorff Distance and Average Symmetric Surface Distance, although less frequently reported, provide valuable complementary insights, especially for surgical planning or biomechanical modeling, where precision at structural interfaces is critical. Establishing standardized evaluation protocols and incorporating clinically relevant endpoints would enhance the interpretability and translational relevance of future studies.

5. Conclusion and limitations

This systematic review analyzed 56 journal articles (49 original studies and 7 review articles) on deep learning for lumbar spine MRI segmentation. Overall, the literature shows rapid progress in segmentation accuracy and methodological sophistication, with U-Net and its variants remaining the dominant paradigm, frequently enhanced by residual, attention, and multi-scale components. While high performance is consistently reported for larger, well-defined structures (e.g., vertebral bodies and intervertebral discs), segmentation of smaller or low-contrast targets (e.g., nerve roots, posterior elements, and fine muscular boundaries) remains more variable and challenging.

Importantly, the evidence base remains constrained by limited generalizability: most studies relied on private, single-center datasets, and external validation on independent test sets was uncommon. In addition, many of the highest-performing approaches employ complex 3D or multi-stage architectures, which, although beneficial for spatial consistency, increase computational demands and sensitivity to limited training data. As highlighted by the risk-of-bias analysis, addressing these limitations will require a stronger emphasis on dataset diversity, external evaluation, and reproducible experimental design rather than further architectural complexity alone. Future research should therefore prioritize multicenter validation, transparent reporting of preprocessing and data-splitting strategies, and broader adoption of publicly available benchmarks to enable meaningful cross-study comparisons and support clinical translation.

Author	Dataset representativeness	Ground-truth	Validation strategy	Overall
Lessmann N. et al.	High	Unclear	High	High
Pang S. et al.	Low	Low	High	Moderate
Al-Kafri A.S. et al.	Low	Low	High	Moderate
Huang J. et al.	Low	Low	High	Moderate
Zheng H.-D. et al.	High	Unclear	Low	High
Kim S. et al.	High	Low	High	High
Han Z. et al.	Low	Low	Unclear	Moderate
Mbarki W. et al.	High	Unclear	High	High
Li H. et al.	High	Low	High	High
Šušteršič T. et al.	Low	Low	High	Moderate
Zhou J. et al.	High	Low	High	High
Shen H. et al.	Unclear	Low	High	High
Rehman F. et al.	High	Low	Unclear	High
Masood R.F. et al.	Low	Low	High	Moderate
Bharadwaj U.U. et al.	High	Low	High	High
Iriondo C. et al.	High	High	High	High
Hwang E.-J. et al.	High	Low	High	High
Liu Z. et al.	High	Low	High	High
Gao K.T. et al.	High	Low	High	High
Hess M. et al.	High	Low	High	High
Baur D. et al.	High	Low	High	High
Sáenz-Gamboa J.J. et al.	Low	Low	High	Moderate
Niemeyer F. et al.	High	High	Low	High
Dourthe B. et al.	High	Low	High	High
Bonnheim N.B. et al.	High	Low	High	High
Yilizati-Yilihamu E.E. et al.	High	Low	High	High
Hou C. et al.	Low	Low	High	Moderate
Morris D.M. et al.	Unclear	Low	High	High
Saravi B. et al.	Low	Low	Low	Low
Chen T. et al.	High	Low	High	High
Altun İ. et al.	High	Low	High	High
Laiwalla A.N. et al.	High	Low	Low	Moderate
Mohanty R. et al.	Low	Low	High	Moderate
Adibatti S. et al.	Low	Low	High	Moderate
Balzer I. et al.	High	Low	High	High
Zhou L. et al.	Low	Low	High	Moderate
Candito A. et al.	High	Low	Low	Moderate
Jung M. et al.	Low	Low	Low	Low
Wang H. et al.	High	Low	High	High
Kok J. et al.	High	Low	High	High
Natalia F. et al.	Low	Low	High	Moderate
Liang Y.-W. et al.	High	Low	High	High
Xu Y. et al.	High	Low	Low	Moderate
Giaccone P. et al.	Low	Low	Low	Low
Ding Z. et al.	High	Low	High	High
Basak P. et al.	High	Low	High	High
Yan H. et al.	High	Low	High	High
Karthik E.N. et al.	Low	High	Low	Moderate
Lakshmi S.V.V. et al.	Low	Low	High	Moderate

Fig. 5. Study-level risk-of-bias assessment for included original studies.

This review has several limitations. The search was restricted to Scopus and PubMed and to English-language journal articles, which may have excluded relevant conference proceedings, preprints, and non-indexed studies. Additionally, substantial heterogeneity in segmentation targets, datasets, evaluation metrics, and validation strategies precluded a quantitative meta-analysis. Screening was conducted by a single reviewer with subsequent validation by a second reviewer, which may increase the risk of selection bifas compared to fully independent duplicate screening. Finally, although the review protocol was developed a priori, it was not prospectively registered (e.g., in PROSPERO), which should be acknowledged as a limitation.

CRediT authorship contribution statement

Diogo Mendes: Writing – original draft, Visualization, Validation, Investigation. **João Manuel R.S. Tavares:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Fig. 5 depicts the study-level risk-of-bias assessment for the included original studies.

Data availability

No data was used for the research described in the article.

References

[1] N.A. Ebraheim, A. Hassan, M. Lee, R. Xu, Functional anatomy of the lumbar spine, Semin. Pain Med. 2 (2004) 131–137, http://dx.doi.org/10.1016/j.spmd.2004.08.004.

[2] D. Hoy, C. Bain, G. Williams, L. March, P. Brooks, F. Blyth, A. Woolf, R. Buchbinder, A systematic review of the global prevalence of low back pain, Best Pr. Res. Clin. Rheumatol. 24 (2010) 769–781, http://dx.doi.org/10.1016/j.berh.2010.10.002.

[3] R. Chou, A. Qaseem, V. Snow, D. Casey, J.T. Cross Jr., P. Shekelle, D.K. Owens, Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American college of physicians and the American pain society, Ann. Intern. Med. 147 (2007) 478–491, http://dx.doi.org/10.7326/0003-4819-147-7-200710020-00006.

[4] J. Feger, A. Er, J. Yap, et al., Lumbar spine protocol (mri), 2025, http://dx.doi.org/10.53347/rID-147093, Reference article, Radiopaedia.org (Accessed March 2025).

[5] J.G. Jarvik, R.A. Deyo, Lumbar spine mri for low back pain: Indications and yield, Am. J. Roentgenol. 195 (2010) 550–559, http://dx.doi.org/10.2214/AJR.10.4367.

[6] J. van der Graaf, M. van Hooff, C. Buckens, et al., Lumbar spine segmentation in mr images: a dataset and a public benchmark, Sci. Data 11 (2024) 264, http://dx.doi.org/10.1038/s41597-024-03090-w, URL: DOI: 10.1038/s41597-024-03090-w.

[7] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B.van. Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, http://dx.doi.org/10.1016/j.media.2017.07.005.

[8] M.U. Saeed, W. Bin, J. Sheng, H.M. Albarakati, A. Dastgir, Msff: An automated multi-scale feature fusion deep learning model for spine fracture segmentation using mri, Biomed. Signal Process. Control. 91 (2024a) 105943, http://dx.doi.org/10.1016/j.bspc.2024.105943.

[9] M.U. Saeed, W. Bin, J. Sheng, et al., An automated multi-scale feature fusion network for spine fracture segmentation using computed tomography images, J. Digit. Imaging 37 (2024b) 2216–2226, http://dx.doi.org/10.1007/s10278-024-01091-0.

[10] A. Dastgir, W. Bin, M.U. Saeed, J. Sheng, S. Saleem, Mafmv3: An automated multi-scale attention-based feature fusion mobilenetv3 for spine lesion classification, Image Vis. Comput. 155 (2025) 105440, http://dx.doi.org/10.1016/j.imavis.2025.105440.

[11] Q. Xia, H. Zheng, H. Zou, D. Luo, H. Tang, L. Li, B. Jiang, A comprehensive review of deep learning for medical image segmentation, Neurocomputing 613 (2025) 128740, http://dx.doi.org/10.1016/j.neucom.2024.128740.

[12] J.J. Sáenz-Gamboa, J. Domenech, A. Alonso-Manjarrés, J.A. Gómez, M. de la Iglesia-Vayá, Automatic semantic segmentation of the lumbar spine: Clinical applicability in a multi-parametric and multi-center study on magnetic resonance images, Artif. Intell. Med. 140 (2023) http://dx.doi.org/10.1016/j.artmed.2023.102559, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85154596558&doi=10.1016%2fj.artmed.2023.102559&partnerID=40&md5=d5314f94d89428fa198608d20c30cec1, cited by: 15; All Open Access, Green Open Access, Hybrid Gold Open Access.

[13] N. Lessmann, B.van. Ginneken, P.A. de Jong, I. Išgum, Iterative fully convolutional neural networks for automatic vertebra segmentation and identification, Med. Image Anal. 53 (2019) 142–155, http://dx.doi.org/10.1016/j.media.2019.02.005, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85061381552&doi=10.1016%2fj.media.2019.02.005&partnerID=40&md5=649d1475c470448c64192309e5759d15, cited by: 197; All Open Access, Green Open Access.

[14] S. Pang, C. Pang, L. Zhao, Y. Chen, Z. Su, Y. Zhou, M. Huang, W. Yang, H. Lu, Q. Feng, Spineparsenet: Spine parsing for volumetric mr image by a two-stage segmentation framework with semantic image representation, IEEE Trans. Med. Imaging 40 (2021) 262–273, http://dx.doi.org/10.1109/TMI.2020.3025087, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85098870371&doi=10.1109%2fTMI.2020.3025087&partnerID=40&md5=6163e3b17cb54eb433d31b2472ed991d, cited by: 126.

[15] Ala S. Al-Kafri, Sud Sudirman, Abir Hussain, Dhiya Al-Jumeily, Friska Natalia, Hira Meidia, Nunik Afriliana, Wasfi Al-Rashdan, Mohammad Bashtawi, Mohammed Al-Jumaily, Boundary delineation of mri images for lumbar spinal stenosis detection through semantic segmentation using deep neural networks, IEEE Access 7 (2019) 43487–43501, http://dx.doi.org/10.1109/ACCESS.2019.2908002, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064873387&doi=10.1109%2fACCESS.2019.2908002&partnerID=40&md5=dd389f816fa0fc003f780b04f76c7944, cited by: 77; All Open Access, Gold Open Access.

[16] J. Huang, H. Shen, J. Wu, X. Hu, Z. Zhu, X. Lv, Y. Liu, Y. Wang, Spine explorer: a deep learning based fully automated program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine mr images, Spine J. 20 (2020) 590–599, http://dx.doi.org/10.1016/j.spinee.2019.11.010, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076562362&doi=10.1016%2fj.spinee.2019.11.010&partnerID=40&md5=9c1c7191bd8764da0c23abb3430c2f21, cited by: 72.

[17] H.D. Zheng, Y.L. Sun, D.W. Kong, M.C. Yin, J. Chen, Y.P. Lin, X.F. Ma, H.S. Wang, G.J. Yuan, M. Yao, X.J. Cui, Y.Z. Tian, Y.J. Wang, Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from mri, Nat. Commun. 13 (2022) http://dx.doi.org/10.1038/s41467-022-28387-5, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124577372&doi=10.1038%2fs41467-022-28387-5&partnerID=40&md5=132c7d98a3637fe69f1d5b9e8cc1885e, cited by: 73; All Open Access, Gold Open Access, Green Open Access.

[18] S. Kim, W.C. Bae, K. Masuda, C.B. Chung, D. Hwang, Fine-grain segmentation of the intervertebral discs from mr spine images using deep convolutional neural networks: Bsu-net, Appl. Sci. (Switzerland) 8 (2018) http://dx.doi.org/10.3390/app8091656, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053425285&doi=10.3390%2fapp8091656&partnerID=40&md5=708d61be4d45e2bb96d377f31384a2d7, cited by: 48; All Open Access, Gold Open Access, Green Open Access.

[19] Z. Han, B. Wei, X. Xi, B. Chen, Y. Yin, S. Li, Unifying neural learning and symbolic reasoning for spinal medical report generation, Med. Image Anal. 67 (2021) http://dx.doi.org/10.1016/j.media.2020.101872, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85094323461&doi=10.1016%2fj.media.2020.101872&partnerID=40&md5=6fb4c90947d3fd6fce24756a93bfd9b9, cited by: 44; All Open Access, Green Open Access.

[20] W. Mbarki, M. Bouchouicha, S. Frizzi, F. Tshibasu, L.B. Farhat, M. Sayadi, Lumbar spine discs classification based on deep convolutional neural networks using axial view mri, Interdiscip. Neurosurg.: Adv. Tech. Case Manag. 22 (2020) http://dx.doi.org/10.1016/j.inat.2020.100837, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85088994037&doi=10.1016%2fj.inat.2020.100837&partnerID=40&md5=8809dfbe1737f71f265270f3d41d05de, cited by: 41; All Open Access, Gold Open Access, Green Open Access.

[21] H. Li, H. Luo, W. Huan, Z. Shi, C. Yan, L. Wang, Y. Mu, Y. Liu, Automatic lumbar spinal mri image segmentation with a multi-scale attention network, Neural Comput. Appl. 33 (2021) 11589–11602, http://dx.doi.org/10.1007/s00521-021-05856-4, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102443276&doi=10.1007%2fs00521-021-05856-4&partnerID=40&md5=36523af8dbc44c3e9672277a86615ef6, cited by: 38; All Open Access, Bronze Open Access, Green Open Access.

[22] T. Šušteršič, V. Ranković, V. Milovanović, V. Kovačević, L. Rasulić, N. Filipović, A deep learning model for automatic detection and classification of disc herniation in magnetic resonance images, IEEE J. Biomed. Health Informatics 26 (2022) 6036–6046, http://dx.doi.org/10.1109/JBHI.2022.3209585, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85139498992&doi=10.1109%2fJBHI.2022.3209585&partnerID=40&md5=660f1ede55776a5943ec1cd14caa0775, cited by: 32.

[23] J. Zhou, P.F. Damasceno, R. Chachad, J.R. Cheung, A. Ballatori, J.C. Lotz, A.A. Lazar, T.M. Link, A.J. Fields, R. Krug, Automatic vertebral body segmentation based on deep learning of dixon images for bone marrow fat fraction quantification, Front. Endocrinol. 11 (2020) http://dx.doi.org/10.3389/fendo.2020.00612, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090997745&doi=10.3389%2ffendo.2020.00612&partnerID=40&md5=639e4a2df30851ffa2af856eb58f214f, cited by: 29; All Open Access, Gold Open Access, Green Open Access.

[24] H. Shen, J. Huang, Q. Zheng, Z. Zhu, X. Lv, Y. Liu, Y. Wang, A deep-learning-based, fully automated program to segment and quantify major spinal components on axial lumbar spine magnetic resonance images, Phys. Ther. 101 (2021) http://dx.doi.org/10.1093/ptj/pzab041, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85107711294&doi=10.1093%2fptj%2fpzab041&partnerID=40&md5=c5f3fb687a6ebf49ba3571053dfb2354, cited by: 23; All Open Access, Bronze Open Access.

[25] F. Rehman, S.I. Ali Shah, N. Riaz, S.O. Gilani, A robust scheme of vertebrae segmentation for medical diagnosis, IEEE Access 7 (2019) 120387–120398, http://dx.doi.org/10.1109/ACCESS.2019.2936492, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087795945&doi=10.1109%2fACCESS.2019.2936492&partnerID=40&md5=230801e3dcf6ea1e86120cae4c780cac, cited by: 23; All Open Access, Gold Open Access.

[26] R.F. Masood, I.A. Taj, M.B. Khan, M.A. Qureshi, T. Hassan, Deep learning based vertebral body segmentation with extraction of spinal measurements and disorder disease classification, Biomed. Signal Process. Control. 71 (2022) http://dx.doi.org/10.1016/j.bspc.2021.103230, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85121012221&doi=10.1016%2fj.bspc.2021.103230&partnerID=40&md5=1720b6151abea11de18f65a8e854e3e7, cited by: 23.

[27] U.U. Bharadwaj, M. Christine, S. Li, D. Chou, V. Pedoia, T.M. Link, C.T. Chin, S. Majumdar, Deep learning for automated, interpretable classification of lumbar spinal stenosis and facet arthropathy from axial mri, Eur. Radiol. 33 (2023) 3435–3443, http://dx.doi.org/10.1007/s00330-023-09483-6, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85150170484&doi=10.1007%2fs00330-023-09483-6&partnerID=40&md5=de7731ad1a43812b817aa88645945860, cited by: 21; All Open Access, Green Open Access.

[28] C. Iriondo, V. Pedoia, S. Majumdar, Lumbar intervertebral disc characterization through quantitative mri analysis: An automatic voxel-based relaxometry approach, Magn. Reson. Med. 84 (2020) 1376–1390, http://dx.doi.org/10.1002/mrm.28210, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079458578&doi=10.1002%2fmrm.28210&partnerID=40&md5=5d347f4365e03167b9b3d0bd636184fd, cited by: 18; All Open Access, Green Open Access, Hybrid Gold Open Access.

[29] E.J. Hwang, S. Kim, J.Y. Jung, Fully automated segmentation of lumbar bone marrow in sagittal, high-resolution t1-weighted magnetic resonance images using 2d u-net, Comput. Biol. Med. 140 (2022) http://dx.doi.org/10.1016/j.compbiomed.2021.105105, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85120403967&doi=10.1016%2fj.compbiomed.2021.105105&partnerID=40&md5=476c256d8c276ae5507a2e6dc9c3f3e3, cited by: 15; All Open Access, Hybrid Gold Open Access.

[30] Z. Liu, Z. Su, M. Wang, T. Chen, Z. Cui, X. Chen, S. Li, Q. Feng, S. Pang, H. Lu, Computerized characterization of spinal structures on mri and clinical significance of 3d reconstruction of lumbosacral intervertebral foramen, Pain Physician 25 (E27 – E35) (2022) URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123746131&partnerID=40&md5=88c86516a284376bb762396d9380b66f. cited by: 15.

[31] K.T. Gao, R. Tibrewala, M. Hess, U.U. Bharadwaj, G. Inamdar, T.M. Link, C.T. Chin, V. Pedoia, S. Majumdar, Automatic detection and voxel-wise mapping of lumbar spine modic changes with deep learning, JOR Spine 5 (2022) http://dx.doi.org/10.1002/jsp2.1204, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85131602709&doi=10.1002%2fjsp2.1204&partnerID=40&md5=1bc457994fc6864d8fc09d6b28fa1fc3, cited by: 14; All Open Access, Green Open Access.

[32] M. Hess, B. Allaire, K.T. Gao, R. Tibrewala, G. Inamdar, U. Bharadwaj, C. Chin, V. Pedoia, M. Bouxsein, D. Anderson, S. Majumdar, Deep learning for multi-tissue segmentation and fully automatic personalized biomechanical models from bacpac clinical lumbar spine mri, Pain Med. (United States) 24 (S139 – S148) (2023) http://dx.doi.org/10.1093/pm/pnac142, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146125451&doi=10.1093%2fpm%2fpnac142&partnerID=40&md5=79d276badd8dfd9960e0d77b2d14efcf, cited by: 13; All Open Access, Green Open Access, Hybrid Gold Open Access.

[33] D. Baur, R. Bieck, J. Berger, J. Neumann, J. Henkelmann, T. Neumuth, C.E. Heyde, A. Voelker, Analysis of the paraspinal muscle morphology of the lumbar spine using a convolutional neural network (cnn), Eur. Spine J. 31 (2022) 774–782, http://dx.doi.org/10.1007/s00586-021-07073-y, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85120874056&doi=10.1007%2fs00586-021-07073-y&partnerID=40&md5=531fb5802cab5a4ed4b97f97525c6410, cited by: 11; All Open Access, Hybrid Gold Open Access.

[34] F. Niemeyer, A. Zanker, R. Jonas, Y. Tao, F. Galbusera, H.J. Wilke, An externally validated deep learning model for the accurate segmentation of the lumbar paravertebral muscles, Eur. Spine J. 31 (2022) 2156–2164, http://dx.doi.org/10.1007/s00586-022-07320-w, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134475570&doi=10.1007%2fs00586-022-07320-w&partnerID=40&md5=d26eb46e2ec8b7f24cd475014a770f5b, cited by: 10; All Open Access, Green Open Access.

[35] B. Dourthe, N. Shaikh, S.A. Pai, S. Fels, S.H. Brown, D.R. Wilson, J. Street, T.R. Oxland, Automated segmentation of spinal muscles from upright open mri using a multiscale pyramid 2d convolutional neural network, Spine 47 (2022) 1179–1186, http://dx.doi.org/10.1097/BRS.0000000000004308, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135598813&doi=10.1097%2fBRS.0000000000004308&partnerID=40&md5=9c6bee822a155d17ef0b607a998bfd06, cited by: 10.

[36] N.B. Bonnheim, L. Wang, A.A. Lazar, R. Chachad, J. Zhou, X. Guo, C. O'Neill, J. Castellanos, J. Du, H. Jang, R. Krug, A.J. Fields, Deep-learning-based biomarker of spinal cartilage endplate health using ultra-short echo time magnetic resonance imaging, Quant. Imaging Med. Surg. 13 (2023) 2807–2821, http://dx.doi.org/10.21037/qims-22-729, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85168995822&doi=10.21037%2fqims-22-729&partnerID=40&md5=63851717529bff99dbd5b0c39b8ccd7e, cited by: 11; All Open Access, Gold Open Access, Green Open Access.

[37] E.E. Yilizati-Yilihamu, J. Yang, Z. Yang, F. Rong, S. Feng, A spine segmentation method based on scene aware fusion network, BMC Neurosci. 24 (2023) http://dx.doi.org/10.1186/s12868-023-00818-z, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85171279235&doi=10.1186%2fs12868-023-00818-z&partnerID=40&md5=8dfcc4af66759bc6a3486f6033f53be2, cited by: 7; All Open Access, Gold Open Access, Green Open Access.

[38] C. Hou, W. Zhang, H. Wang, F. Liu, D. Liu, J. Chang, A semantic segmentation model for lumbar mri images using divergence loss, Appl. Intell. 53 (2023) 12063–12076, http://dx.doi.org/10.1007/s10489-022-04118-7, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85138321159&doi=10.1007%2fs10489-022-04118-7&partnerID=40&md5=206a5694bc25d1f63bf879a93d2d5135, cited by: 5.

[39] D.M. Morris, C. Wang, G. Papanastasiou, C.D. Gray, W. Xu, S. Sjöström, S. Badr, J. Paccou, S.I. Semple, T. MacGillivray, W.P. Cawthorn, A novel deep learning method for large-scale analysis of bone marrow adiposity using uk biobank dixon mri data, Comput. Struct. Biotechnol. J. 24 (2024) 89–104, http://dx.doi.org/10.1016/j.csbj.2023.12.029, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85183353088&doi=10.1016%2fj.csbj.2023.12.029&partnerID=40&md5=0c543f2c27fa7c0d239b20277c1a7081, cited by: 4; All Open Access, Gold Open Access, Green Open Access.

[40] B. Saravi, A. Zink, S. Ülkümen, S. Couillard-Despres, J. Wollborn, G. Lang, F. Hassel, Automated detection and measurement of dural sack cross-sectional area in lumbar spine mri using deep learning, Bioengineering 10 (2023) http://dx.doi.org/10.3390/bioengineering10091072, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85172256760&doi=10.3390%2fbioengineering10091072&partnerID=40&md5=a5f7234edec06cb49f6ea0fa86f7e5d6, cited by: 5; All Open Access, Gold Open Access, Green Open Access.

[41] T. Chen, Z.h. Su, Z. Liu, M. Wang, Z.f. Cui, L. Zhao, L.j. Yang, W.c. Zhang, X. Liu, J. Liu, S.y. Tan, S.l. Li, Q.j. Feng, S.m. Pang, H. Lu, Automated magnetic resonance image segmentation of spinal structures at the l4-5 level with deep learning: 3d reconstruction of lumbar intervertebral foramen, Orthop. Surg. 14 (2022) 2256–2264, http://dx.doi.org/10.1111/os.13431, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135914168&doi=10.1111%2fos.13431&partnerID=40&md5=c93982995f2826d680b13757c5b32f27, cited by: 6; All Open Access, Gold Open Access, Green Open Access.

[42] S. Altun, S. Altun, A. Alkan, Lss-unet: Lumbar spinal stenosis semantic segmentation using deep learning, Multimedia Tools Appl. 82 (2023) 41287–41305, http://dx.doi.org/10.1007/s11042-023-15205-3, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85152360465&doi=10.1007%2fs11042-023-15205-3&partnerID=40&md5=daea13ddb9fc9ae209996a8cd20c0ad8, cited by: 5.

[43] A.N. Laiwalla, A. Ratnaparkhi, D. Zarrin, K. Cook, I. Li, B. Wilson, T. Florence, B. Yoo, B. Salehi, B. Gaonkar, J. Beckett, L. Macyszyn, Lumbar spinal canal segmentation in cases with lumbar stenosis using deep-u-net ensembles, World Neurosurg. 178 (2023) e135 – e140, http://dx.doi.org/10.1016/j.wneu.2023.07.009, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85169846424&doi=10.1016%2fj.wneu.2023.07.009&partnerID=40&md5=c16412fd6335d9c317f683cdc5db7958, cited by: 5; All Open Access, Bronze Open Access.

[44] R. Mohanty, S. Allabun, S.S. Solanki, S.K. Pani, M.S. Alqahtani, M. Abbas, B.O. Soufiene, Namstcd: A novel augmented model for spinal cord segmentation and tumor classification using deep nets, Diagnostics 13 (2023) http://dx.doi.org/10.3390/diagnostics13081417, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85153941197&doi=10.3390%2fdiagnostics13081417&partnerID=40&md5=92c2d5306d2038053bcf745227658a7b, cited by: 4; All Open Access, Gold Open Access, Green Open Access.

[45] S. Adibatti, K. Sudhindra, J.Manisha. Shivaram, Segmentation and classification of intervertebral disc using capsule stacked autoencoder, Biomed. Signal Process. Control. 86 (2023) http://dx.doi.org/10.1016/j.bspc.2023.105311, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85167786634&doi=10.1016%2fj.bspc.2023.105311&partnerID=40&md5=1b3160f609a8338afa8d80316afd4373, cited by: 5.

[46] I. Balzer, M. Mühlemann, M. Jokeit, I.S. Rawal, J.G. Snedeker, M. Farshad, J. Widmer, A deep learning pipeline for automatized assessment of spinal mri, Comput. Methods Programs Biomed. Updat. 2 (2022) http://dx.doi.org/10.1016/j.cmpbup.2022.100081, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85168737116&doi=10.1016%2fj.cmpbup.2022.100081&partnerID=40&md5=70b9538d8ba5328a95dc2e8a78364f3b, cited by: 3; All Open Access, Gold Open Access, Green Open Access.

[47] L. Zhou, Y. Liu, Y. Zhang, Z. Lin, Bua-net: Boundary and uncertainty-aware attention network for lumbar multi-region magnetic resonance imaging segmentation, Biomed. Signal Process. Control. 94 (2024) http://dx.doi.org/10.1016/j.bspc.2024.106267, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85188806735&doi=10.1016%2fj.bspc.2024.106267&partnerID=40&md5=5421c21ca3b78083f2cb19e9c306d718, cited by: 2.

[48] A. Candito, R. Holbrey, A. Ribeiro, A. Dragan, C. Messiou, N. Tunariu, M.D. Blackledge, D.M. Koh, Deep learning assisted atlas-based delineation of the skeleton from whole-body diffusion weighted mri in patients

with malignant bone disease, Biomed. Signal Process. Control. 92 (2024) http://dx.doi.org/10.1016/j.bspc.2024.106099, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185402513&doi=10.1016%2fj.bspc.2024.106099&partnerID=40&md5=71fe9872ebb404aefa3fc749b4a2d6d7, cited by: 2.

[49] M. Jung, V.K. Raghu, M. Reisert, H. Rieder, S. Rospleszcz, T. Pischon, T. Niendorf, H.U. Kauczor, H. Völzke, R. Bülow, M.F. Russe, C.L. Schlett, M.T. Lu, F. Bamberg, J. Weiss, Deep learning-based body composition analysis from whole-body magnetic resonance imaging to predict all-cause mortality in a large western population, EBioMedicine 110 (2024) http://dx.doi.org/10.1016/j.ebiom.2024.105467, xURL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85210544590&doi=10.1016%2fj.ebiom.2024.105467&partnerID=40&md5=5ae95cfd7505e8074e0d44014e1b35e6, cited by: 2; All Open Access, Gold Open Access.

[50] H. Wang, Y. Chen, T. Jiang, H. Bian, X. Shen, 3D multi-scale feature extraction and recalibration network for spinal structure and lesion segmentation, Acta Radiol. 64 (2023) 3015–3023, http://dx.doi.org/10.1177/02841851231204214, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85173777141&doi=10.1177%2f02841851231204214&partnerID=40&md5=f1431119a22ab4c6ec2e988b0831e464, cited by: 1.

[51] J. Kok, Y.M. Shcherbakova, T.P.C. Schlösser, P.R. Seevinck, T.A. van der Velden, R.M. Castelein, K. Ito, B. van Rietbergen, Automatic generation of subject-specific finite element models of the spine from magnetic resonance images, Front. Bioeng. Biotechnol. 11 (2023) http://dx.doi.org/10.3389/fbioe.2023.1244291, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85171529259&doi=10.3389%2ffbioe.2023.1244291&partnerID=40&md5=a32842ba72a1f83c6a3104742b93a848, cited by: 1; All Open Access, Gold Open Access, Green Open Access.

[52] F. Natalia, S. Sudirman, D. Ruslim, A. Al-Kafri, Lumbar spine mri annotation with intervertebral disc height and pfirrmann grade predictions, in: 19, PLoS ONE (2024) http://dx.doi.org/10.1371/journal.pone.0302067, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192912485&doi=10.1371%2fjournal.pone.0302067&partnerID=40&md5=847dfc4bb6db0e47ea3ba091aad16cec, cited by: 3; All Open Access, Gold Open Access, Green Open Access.

[53] Y.W. Liang, Y.T. Fang, T.C. Lin, C.R. Yang, C.C. Chang, H.K. Chang, C.C. Ko, T.H. Tu, L.Y. Fay, J.C. Wu, W.C. Huang, H.W. Hu, Y.Y. Chen, C.H. Kuo, The quantitative evaluation of automatic segmentation in lumbar magnetic resonance images, Neurospine 21 (2024) 665–675, http://dx.doi.org/10.14245/ns.2448060.030, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85197951792&doi=10.14245%2fns.2448060.030&partnerID=40&md5=f7692aae441fe84b6550b6d18f6b4a9e, cited by: 1; All Open Access, Gold Open Access.

[54] Y. Xu, S. Zheng, Q. Tian, Z. Kou, W. Li, X. Xie, X. Wu, Deep learning-based structure segmentation and intramuscular fat annotation on lumbar magnetic resonance imaging, JOR Spine 7 (2024) http://dx.doi.org/10.1002/jsp2.70003, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85204599638&doi=10.1002%2fjsp2.70003&partnerID=40&md5=a1fe7eb0552019fa2badf8009b03ada3, cited by: 0; All Open Access, Gold Open Access.

[55] P. Giaccone, F. D'Antoni, F. Russo, M. Volpecina, C.A. Mallio, G.F. Papalia, G. Vadalà, V. Denaro, L. Vollero, M. Merone, Fully automated evaluation of paraspinal muscle morphology and composition in patients with low back pain, Intelligence-Based Med. 9 (2024) http://dx.doi.org/10.1016/j.ibmed.2023.100130, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179001004&doi=10.1016%2fj.ibmed.2023.100130&partnerID=40&md5=e706ef5f78445d6e00cf5be4afca00e3, cited by: 0; All Open Access, Gold Open Access.

[56] Z. Ding, H. Li, L. Chen, B. Chen, H. Sun, W. Hou, C. Xia, Fully automatic classification of lumbar disc degeneration based on deep learning, Chin. J. Clin. Res. 37 (2024) 709–713, http://dx.doi.org/10.13429/j.cnki.cjcr.2024.05.039, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85196005146&doi=10.13429%2fj.cnki.cjcr.2024.05.039&partnerID=40&md5=8ff6ca2e1e7ba013499cd062509b652c, cited by: 0..

[57] P. Basak, R. Sarmun, S. Kabir, I. Al-Hashimi, E.H. Bhuiyan, A. Hasan, M.S. Khan, M.E. Chowdhury, Machine-agnostic automated lumbar mri segmentation using a cascaded model based on generative neurons, Expert Syst. Appl. 264 (2025) http://dx.doi.org/10.1016/j.eswa.2024.125862, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209894702&doi=10.1016%2fj.eswa.2024.125862&partnerID=40&md5=e69f78a8ae6f14987931e02fa88aa2ac, cited by: 0; All Open Access, Green Open Access.

[58] H. Yan, G. Zhang, W. Cui, Z. Yu, Multi-modality hierarchical fusion network for lumbar spine segmentation with magnetic resonance images, Control. Theory Technol. 22 (2024) 612–622, http://dx.doi.org/10.1007/s11768-024-00231-9, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85204004655&doi=10.1007%2fs11768-024-00231-9&partnerID=40&md5=c2868f21344f592a88fe0878ca6c6ae1, cited by: 0.

[59] E.N. Karthik, J. Valošek, A.C. Smith, D. Pfyffer, S. Schading-Sassenhausen, L. Farner, K.A. Weber, P. Freund, J. Cohen-Adad, Sciseg: Automatic segmentation of intramedullary lesions in spinal cord injury on t2-weighted mri scans, Radiol.: Artif. Intell. 7 (2025) http://dx.doi.org/10.1148/ryai.240005, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85216468766&doi=10.1148%2fryai.240005&partnerID=40&md5=f089d8fd236316afbec232e067018483, cited by: 1.

[60] S. Lakshmi, S.E. Darmanayagam, S.R.R. Cyril, Semantic segmentation of lumbar vertebrae using meijering u-net (mu-net) on spine magnetic resonance images, CMES - Comput. Model. Eng. Sci. 142 (2025) 733–757, http://dx.doi.org/10.32604/cmes.2024.056424, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85212584145&doi=10.32604%2fcmes.2024.056424&partnerID=40&md5=8f27d17bfe17c07e9a3eae18616c0d9a, cited by: 0; All Open Access, Gold Open Access.

[61] C. Chu, D. Belavy, W. Yu, G. Armbrecht, M. Bansmann, D. Felsenberg, G. Zheng, Fully automatic localization and segmentation of 3d vertebral bodies from ct/mr images via a learning-based method, PLoS ONE 10 (2015) e0143327, http://dx.doi.org/10.1371/journal.pone.0143327, URL: DOI: 10.1371/journal.pone.0143327.

[62] S. Sudirman, A.Al. Kafri, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi, M. Al-Jumaily, Label image ground truth data for lumbar spine mri dataset, 2019, http://dx.doi.org/10.17632/zbf6b4pttk.2, DOI: 10.17632/zbf6b4pttk.2.

[63] Y.A. Khalil, Lumbar vertebral body and intervertebral disc segmentation in multi-scanner and multi-modal mri – a ground truth database, 2025, http://dx.doi.org/10.17605/OSF.IO/QX5RT, URL: .

[64] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention, MICCAI, Springer, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[65] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[66] N. Ibtehaz, M.S. Rahman, Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation, Neural Netw. 121 (2020) 74–87, http://dx.doi.org/10.1016/j.neunet.2019.08.002.

[67] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer., 2018, pp. 3–19, http://dx.doi.org/10.1007/978-3-030-01234-2_1.

[68] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2021) 203–211, http://dx.doi.org/10.1038/s41592-020-01008-z.

[69] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: Proceedings of the Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571, http://dx.doi.org/10.1109/3DV.2016.79.

[70] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder–decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495, http://dx.doi.org/10.1109/TPAMI.2016.2644615.

[71] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2961–2969, http://dx.doi.org/10.1109/ICCV.2017.322.

[72] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587 URL: https://arxiv.org/abs/1706.05587.

[73] P. Azimi, T. Yazdanian, E.C. Benzel, H.N. Aghaei, S. Azhari, S. Sadeghi, A. Montazeri, A review on the use of artificial intelligence in spinal diseases, Asian Spine J. 14 (2020) 543–571, http://dx.doi.org/10.31616/asj.2020.0147, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090896595&doi=10.31616%2fasj.2020.0147&partnerID=40&md5=943d8f37e3b97ca78c1c971b7d1d3fc2, cited by: 44; All Open Access, Gold Open Access, Green Open Access.

[74] G. Ren, K. Yu, Z. Xie, P. Wang, W. Zhang, Y. Huang, Y. Wang, X. Wu, Current applications of machine learning in spine: From clinical view, Glob. Spine J. 12 (2022) 1827–1840, http://dx.doi.org/10.1177/21925682211035363, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85116938225&doi=10.1177%2f21925682211035363&partnerID=40&md5=3319b2bacfb93b4e959b39c2f5e1c59d, cited by: 28; All Open Access, Gold Open Access, Green Open Access.

[75] Y. Cui, J. Zhu, Z. Duan, Z. Liao, S. Wang, W. Liu, Artificial intelligence in spinal imaging: Current status and future directions, Int. J. Environ. Res. Public Health 19 (2022) http://dx.doi.org/10.3390/ijerph191811708, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85138341621&doi=10.3390%2fijerph191811708&partnerID=40&md5=25f3b52625d750a477fd51e75d202a3d, cited by: 20; All Open Access, Gold Open Access, Green Open Access.

[76] T. Vrtovec, B. Ibragimov, Spinopelvic measurements of sagittal balance with deep learning: systematic review and critical evaluation, Eur. Spine J. 31 (2022) 2031–2045, http://dx.doi.org/10.1007/s00586-022-07155-5, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126363964&doi=10.1007%2fs00586-022-07155-5&partnerID=40&md5=58d65fba6bafbe0f433b50e96ef281bb, cited by: 13.

[77] K. Chen, X. Zhai, S. Wang, X. Li, Z. Lu, D. Xia, M. Li, Emerging trends and research foci of deep learning in spine: bibliometric and visualization study, Neurosurg. Rev. 46 (2023) http://dx.doi.org/10.1007/s10143-023-01987-5, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85151459348&doi=10.1007%2fs10143-023-01987-5&partnerID=40&md5=fec52596baefcd72777f9f46a8ae406b, cited by: 13; All Open Access, Green Open Access.

[78] K. Vrettos, E. Koltsakis, A.H. Zibis, A.H. Karantanas, M.E. Klontzas, Generative adversarial networks for spine imaging: A critical review of current applications, Eur. J. Radiol. 171 (2024) http://dx.doi.org/10.1016/j.ejrad.2024.111313, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85183563695&doi=10.1016%2fj.ejrad.2024.111313&partnerID=40&md5=27033725c9da058005e36bd76dfd9f10, cited by: 4.

[79] E. Verheijen, T. Kapogiannis, D. Munteh, J. Chabros, M. Staring, T. Smith, C. Vleggeert-Lankamp, Artificial intelligence for segmentation and classification in lumbar spinal stenosis: an overview of current methods, Eur. Spine J. 34 (2025) 1146–1155, http://dx.doi.org/10.1007/s00586-025-08672-9, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85217156537&doi=10.1007%2fs00586-025-08672-9&partnerID=40&md5=6d33b4795662f7d7e30d2e25b215bb43, cited by: 0; All Open Access, Hybrid Gold Open Access.