

REVIEW

Open Access



# Deep learning assisted segmentation of the lumbar intervertebral disc: a systematic review and meta-analysis

Aobo Wang<sup>1†</sup>, Congying Zou<sup>1†</sup>, Shuo Yuan<sup>1</sup>, Ning Fan<sup>1</sup>, Peng Du<sup>1</sup>, Tianyi Wang<sup>1</sup> and Lei Zang<sup>1\*</sup>

## Abstract

**Background** In recent years, deep learning (DL) technology has been increasingly used for the diagnosis and treatment of lumbar intervertebral disc (IVD) degeneration. This study aims to evaluate the performance of DL technology for IVD segmentation in magnetic resonance (MR) images and explore improvement strategies.

**Methods** We developed a PRISMA systematic review protocol and systematically reviewed studies that used DL algorithm frameworks to perform IVD segmentation based on MR images published up to April 10, 2024. The Quality Assessment of Diagnostic Accuracy Studies-2 tool was used to assess methodological quality, and the pooled dice similarity coefficient (DSC) score and Intersection over Union (IoU) were calculated to evaluate segmentation performance.

**Results** 45 studies were included in this systematic review, of which 16 provided complete segmentation performance data and were included in the quantitative meta-analysis. The results indicated that DL models showed satisfactory IVD segmentation performance, with a pooled DSC of 0.900 (95% confidence interval [CI]: 0.887–0.914) and IoU of 0.863 (95% CI: 0.730–0.995). However, the subgroup analysis did not show significant effects of factors on IVD segmentation performance, including network dimensionality, algorithm type, publication year, number of patients, scanning direction, data augmentation, and cross-validation.

**Conclusions** This study highlights the potential of DL technology in IVD segmentation and its further applications. However, due to the heterogeneity in algorithm frameworks and result reporting of the included studies, the conclusions should be interpreted with caution. Future research should focus on training generalized models on large-scale datasets to enhance their clinical application.

**Keywords** Deep learning, Intervertebral disc, Magnetic resonance imaging, Meta-analysis

<sup>†</sup>Aobo Wang and Congying Zou contributed equally to this work.

\*Correspondence:

Lei Zang  
zanglei@ccmu.edu.cn

<sup>1</sup>Department of Orthopedics, Beijing Chaoyang Hospital, Capital Medical University, 5 JingYuan Road, Shijingshan District, Beijing 100043, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Chronic low back pain is a leading cause of global disability, which increases medical burden [1, 2]. Lumbar intervertebral disc (IVD) degeneration is a common cause of pain, which is considered as the precursor to other lumbar degenerative diseases [3, 4]. Magnetic resonance imaging (MRI), with its unique advantages in soft tissue visualization, provides clear depictions of the IVDs' morphology and structure, establishing itself as the main diagnostic tool for IVD degeneration [5]. However, the interpretation of lumbar spine MRI is complex and time consuming, necessitating considerable surgical expertise [6]. The increasing number of patients in recent years has further amplified the demand for radiologists and spinal surgeons.

The development of artificial intelligence (AI) presents the potential for rapid, accurate, and stable imaging analysis [7–10]. AI, once trained on extensive datasets, can surpass human experts in medical image processing [8]. Despite the scarcity of AI systems for widespread clinical use in spinal surgery, there has been a significant increase in research concerning AI's role in IVD degeneration [8, 10]. In 2023, a systematic review and meta-analysis revealed that machine learning and deep learning (DL) algorithms can offer relatively accurate and repeatable diagnosis of lumbar disc herniation and degeneration grading [5]. These techniques are also applicable to diagnose other disc-related diseases, support clinical decision, and predict patient outcomes [11–13].

However, the further optimization and clinical application of these algorithms hinge on a fundamental requirement: image segmentation. In MRI assessment of IVDs, accurate segmentation delineates the regions of interest for diagnostic models, enhancing their precision and interpretability [14]. The IVD segmentation technique can be applied to quantitative imaging assessments, including the automatic measurement of disc height and protrusion distance. These assessments were previously performed manually by physicians, which was a tedious and time-consuming process with low consistency in measurement results [15]. Moreover, AI algorithms can use image segmentation data to construct three-dimensional models of IVDs for applications in CT/MRI image fusion, surgical planning, and navigation [16]. DL algorithm frameworks, such as U-net, have become the state-of-the-art and primary methods for image segmentation [6, 10, 17]. However, to our knowledge, there has been no systematic investigation or summary of the performance of DL technology in IVD segmentation and quantitative measurement within lumbar spine MRI.

This systematic review and meta-analysis aims to bridge this knowledge gap by evaluating the performance of DL models in segmenting and measuring IVDs in MRI scans, with a focus on segmentation accuracy. We believe

that this review will offer a comprehensive overview for further research and application in this critical area.

## Methods

### General guidelines

This systematic literature review strictly followed the guidelines outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, see supplementary file 1) [18, 19]. The protocol of this study has been registered in PROSPERO (<https://www.crd.york.ac.uk/prospero/>) under the registration number CRD42024534092. Given the nature of this systematic review and meta-analysis, ethical approval and informed consent from participants were not required.

### Search strategy and review process

A systematic literature search was conducted independently by two researchers (A.W. and C.Z.), with records collected from three major databases up to the search date of April 10, 2024. The databases included PubMed, Embase, and Web of Science (including Medline). The following key terms were used for literature search: "Deep Learning," "Artificial Intelligence," "Neural Networks," "Segmentation," "Feature extraction" "Intervertebral Disc," and "Lumbar Vertebrae." Additionally, references from included studies were reviewed to identify any relevant literature.

Titles and abstracts of the identified studies were screened for eligibility by the two researchers independently. A list of references from relevant studies and systematic reviews was also screened. Disagreements were resolved by a third researcher and co-author (L.Z.).

### Inclusion and exclusion criteria

Inclusion criteria for this review were: (1) studies involving adult participants; (2) utilization of MRI to assess IVDs; (3) application of DL methodologies with comprehensive data on segmentation performance; (4) acceptance of both retrospective and prospective study designs.

Exclusion criteria included: (1) reviews, letters, guidelines, editorials, or errata; (2) studies involving animals, cadavers, in vivo biomechanics, or patients with lumbar tumors or trauma; (3) studies with overlapping cohorts, which would be summarized but not included in meta-analyses; (4) use of other machine learning algorithms other than DL; (5) studies of low quality; and (6) non-English publications.

### Quality assessment

The quality of included studies was assessed using the second version of the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS-2) [20], which included four domains: patient selection, index test,

reference standard, and flow and timing. For patient selection, the focus was on the inclusion of a well-defined patient population with clear criteria for inclusion and exclusion. For the index test, the explicit description of the DL algorithm for segmenting and evaluating IVDs was scrutinized. The reference standard domain evaluated the reliability of the ground truth determination through manual segmentation and quantitative measurement. The flow and timing domain assessed the clarity of the research flow [21].

#### Data extraction

The following variables were extracted and recorded: (1) study attributes, including the primary author, publication year, study design, and duration; (2) medical data, including patient count and object of the study; (3) characteristics of MRI scanning; (4) DL specifics including the algorithm framework, dataset partition, and data augmentation strategies; (5) performance metrics for IVD segmentation, including dice similarity coefficient (DSC) score, precision, recall, and Intersection over Union (IoU).

For summarization and subgroup analysis, the algorithms applied in the included studies were categorized into U-Net variants, Deeplab variants, Generative Adversarial Networks (GAN) variants, and CNN variants. The U-Net variants included 2D/3D U-Net networks and those combined with frameworks like ResNet, as well as other U-Net-like algorithms. CNN variants referred to other CNN or FCN frameworks not covered by the aforementioned categories. Given the specialized features of most algorithms, these classifications may be very crude. For detailed characteristics of a particular algorithm, consultation of the original literature is strongly recommended.

#### Statistical analysis

Statistical analyses were performed using the Comprehensive Meta-Analysis software (version 3, Biostat, Englewood, NJ, USA). A random-effects model was applied for the meta-analysis, with  $p < 0.05$  indicating statistical significance. Forest plots were generated to visualize the estimated DSC and IoU scores and the overall performance. Subgroup analyses were performed to explore relationships between outcomes and potential influencing factors. Heterogeneity between studies was assessed using the Q-test and Higgins  $I^2$  statistics, categorized as follows: 0–25% (not important), 26–50% (low), 51–75% (moderate), and 76–100% (high). Publication bias was investigated using a funnel plot, with asymmetry evaluated by the Egger's test.

## Results

### Basic characteristics

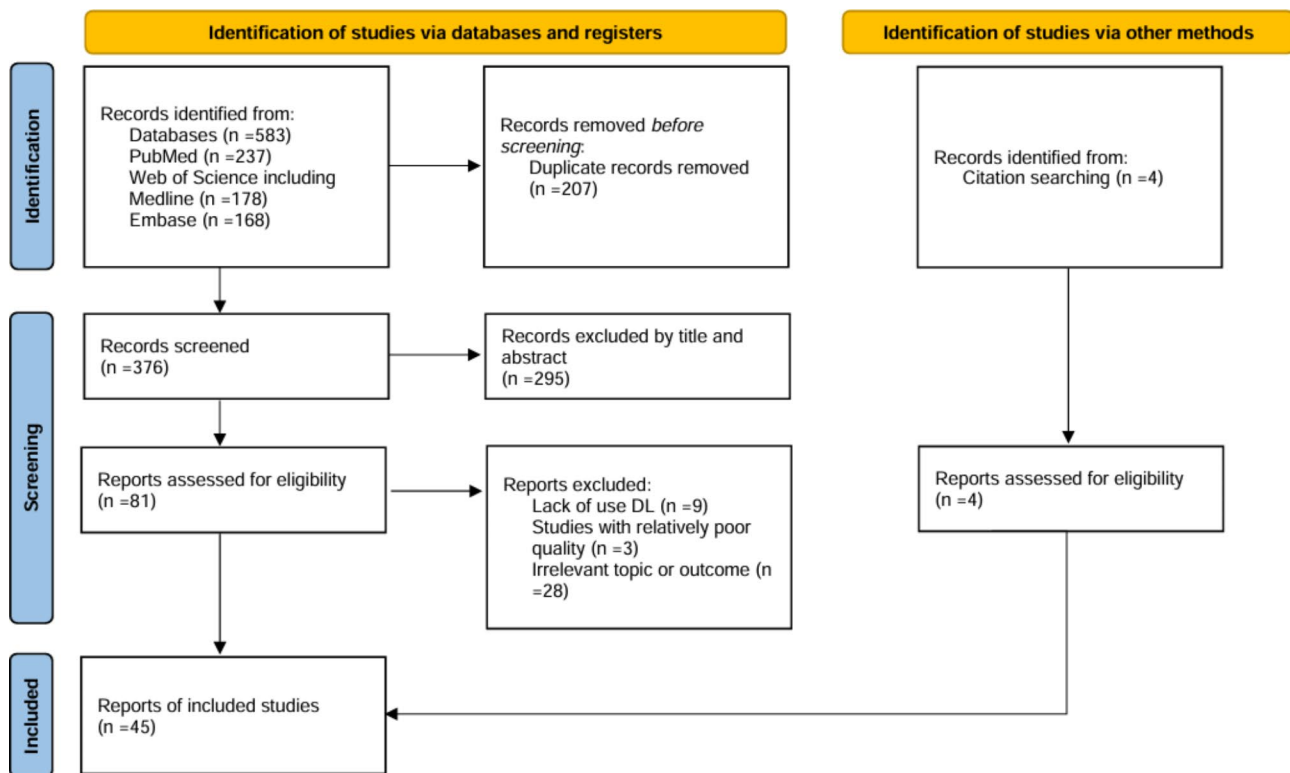
The PRISMA flowchart for the literature search is shown in Fig. 1. Initially, 583 publications were identified through database searching, and an additional 4 publications were retrieved through cross-referencing. After removing duplicates, 376 publications were screened, and 295 of them were excluded based on the titles and abstracts. Ultimately, 45 publications were included in the systematic review after full text screening [15, 22–65]. However, only 16 publications were eligible for the meta-analysis [22–37]. Since they provided sufficient quantitative data. It should be noted that 2 of the publications were based on the same cohort [29, 36], but differences in the MRI slices used for training and the algorithm frameworks led to their both inclusion in the meta-analysis. Attempts to contact corresponding authors of other publications did not obtain the necessary data.

Table 1 outlines the basic characteristics of the included studies and objects. Most studies were designed retrospectively and performed based on single-center datasets, including public datasets. However, since these studies only involve the processing of medical images, retrospective or prospective studies may not make a significant difference in data quality. The number of patients across the studies ranged considerably, from as few as 8 [57] to as many as 520 [55]. The study subjects primarily included healthy individuals and patients with various types of degenerative lumbar diseases. Many publications did not report the study durations.

Table 2 shows the information about the MR scans and DL strategies. 35 of the studies used sagittal slices for IVD segmentation, while axial and coronal slices were also used in other studies. The specific MRI slices selected for segmentation varied. 16 of the studies used mid- or para-sagittal slices, which can clearly show the IVD structures. Some studies used several or all sagittal or axial slices, or used 3D SPACE sequences. In terms of image capturing methods, 35 of the studies implemented T2 sequences, while other studies implemented both T1 and T2 sequences, or fused images produced by T2 images registered in T1 images. Although the scanner, slice thickness, Tesla, and image size may affect the IVD segmentation, many studies did not report detailed information about these items, especially those with multiple data sources. Therefore, this study did not further summarize these data.

### Data preprocessing and DL algorithms

The preprocessing of medical image data aims to enhance image quality and augment sample size to improve training effectiveness. Image cropping and resizing aims to standardize image dimensions for ease of training, or to pre-crop images to specific segments or regions of



**Fig. 1** PRISMA flowchart for the current meta-analysis

interest, and it was used in many studies [15, 22–25, 27, 29, 30, 33, 36, 39, 40, 42–44, 46, 48–51, 62, 64, 65]. Normalization, which standardizes the intensity values of images, was also employed in some studies [15, 24, 25, 28–31, 35, 36, 39, 44, 46, 48–51, 62, 64]. Data augmentation includes applying transformations such as rotation, flipping, and contrast enhancement, and some studies utilized this strategy to increase the amount of the training dataset [15, 22, 28–30, 32, 34, 36, 39, 43, 44, 46–48, 60, 62, 64]. Padding [24, 25, 36, 48, 62, 65] is also an optional preprocessing method. As shown in Table 2, all data were randomly or manually partitioned into training, testing, and validation sets. Some studies randomly grouped MR image slices, while others, especially those employing 3D algorithmic frameworks, grouped data by patient, meaning all data from a single patient's examination belonged exclusively to one dataset.

Most studies included in this review employed specifically designed or improved algorithms for IVD segmentation, with the U-net network and its variants being the most commonly applied models (28 studies), including classic 2D/3D U-net, V-net, ResUnet, etc. GANs and DeepLab segmentation networks were utilized in 4 and 2 studies, respectively, while the remaining studies employed other variants of CNNs or FCNs. The information about the algorithms is summarized in Table 2, and

we will further discuss the characteristics of the various algorithms in the [discussion](#) section.

DSC and IoU are the most commonly used performance metrics for automatic IVD segmentation. The reported DSC ranged from 0.810 [27] to 0.982 [40]. While the reported IoU ranged from 0.771 [42] to 0.972 [52]. Among the included studies, 3 studies performed IVD segmentation only at a specific segment (L4/5 or L5/S1). 5 studies also segmented other structures of the lumbar spine, such as the vertebral body and spinal canal, and reported only the overall segmentation results for all structures, with the reported DSC ranging from 0.803 [62] to 0.948 [58]. Other evaluation indexes of IVD segmentation, such as precision and recall, were also reported in several studies [24, 25, 29, 31, 41, 43, 46, 48, 65], with the reported precision ranging from 0.868 [41] to 0.986 [46], and the recall ranging from 0.904 [24] to 0.950 [46].

Several studies conducted automatic quantitative measurements of IVDs based on image segmentation [15, 49, 61], including measurements of disc height and area. These studies all reported good consistency between automatic segmentation and the gold standard (manual measurements). However, due to differences in measurement methods and evaluation metrics, this study did not summarize the performance of the quantitative measurements. The authors believe that quantitative

**Table 1** The patient and study characters

First author& publication year	Study design	Institution	Duration	Study subjects	Patients included (n)
Li X, 2018 [22]	Retrospective	Single-center	NA	Healthy subjects	12
Han Z, 2018 [64]	Retrospective	Multi-center	NA	Multiple pathologies included	253
Kim S, 2018 [52]	Retrospective	Single-center	NA	NA	20
Pang S, 2019 [50]	Retrospective	Single-center	NA	Multiple pathologies included	235
Rehman F, 2019 [34]	Retrospective	Single-center	NA	NA	20
Al-Kafri AS, 2019 [45]	Retrospective	Single-center	2015–2016	Low back pain patients	515
Huang J, 2020 [62]	Retrospective	Single-center	NA	Community people	100
Iriondo C, 2020 [31]	Retrospective	Single-center	NA	Healthy and low back pain patients	31
Han Z, 2021 [37]	Retrospective	Multi-center	NA	Multiple pathologies included	253
Cheng YK, 2021 [42]	Retrospective	Single-center	2016–2020	Multiple pathologies included	NA
Das P, 2021 [28]	Retrospective	Single-center	NA	NA	8
Gaonkar B, 2021 [30]	Retrospective	Single-center	NA	NA	103
Suri A, 2021 [47]	Retrospective	Single-center	2005–2014	Healthy and osteoporosis patients	NA
Mbarki W, 2021 [53]	Retrospective	Single-center	NA	NA	NA
Pang S, 2021 [36]	Retrospective	Single-center	NA	Multiple pathologies included	215
Liu Z, 2022 [25]	Prospective	Single-center	2020–2020	Lumbar disc herniation patients	100
Pang S, 2022 [29]	Retrospective	Single-center	NA	Multiple pathologies included	416
Zheng HD, 2022 [49]	Retrospective	Multi-center	2019–2020	NA	420
Sustersic T, 2022 [46]	Retrospective	Multi-center	NA	Healthy and disc herniation	243
Gong H, 2022 [33]	Retrospective	Single-center	NA	NA	62
Chen T, 2022 [24]	Retrospective	Single-center	2020–2020	Healthy, disc herniation, and spinal stenosis	100
Kuang X, 2022 [35]	Retrospective	Single-center	NA	Community people	40
Hou C, 2022 [61]	Retrospective	Single-center	NA	NA	NA
Altun S, 2022 [55]	Retrospective	Single-center	NA	Lumbar spinal stenosis patients	300
Wang H, 2023 [38]	Retrospective	Single-center	2019–2020	Lumbar disc herniation patients	70
Coppock JA, 2023 [40]	Retrospective	Single-center	NA	Healthy subjects	25
He S, 2023 [57]	Retrospective	Single-center	NA	Healthy and disc degeneration patients	214
Liu H, 2023 [58]	Retrospective	Single-center	NA	NA	8
Yilizati-Yilihamu EE, 2023 [63]	Retrospective	Single-center	NA	NA	172
Li H, 2023 [65]	Retrospective	Single-center	NA	NA	215
Hess M, 2023 [27]	Retrospective	Single-center	NA	Multiple pathologies included	27
Qinhong D, 2023 [32]	Retrospective	Multi-center	NA	NA	228
He S, 2023 [59]	Retrospective	Multi-center	NA	NA	NA
Bharadwaj UU, 2023 [26]	Retrospective	Single-center	2008–2019	Multiple pathologies included	200
Soydan Z, 2023 [44]	Retrospective	Single-center	2016–2020	Low back pain patients	363
Saenz-Gamboa JJ, 2023 [43]	Retrospective	Single-center	2015–2016	Multiple pathologies included	181
Altun I, 2023 [56]	Retrospective	Single-center	NA	Lumbar spinal stenosis patients	520
He S, 2023 [41]	Retrospective	Single-center	NA	Healthy and disc degeneration patients	263
Wang M, 2023 [48]	Prospective	Single-center	NA	Healthy and disc degeneration patients	50
Pang C, 2023 [39]	Retrospective	Single-center	2015–2016	Low back pain patients	515
He S, 2024 [60]	Retrospective	Single-center	NA	Healthy and disc degeneration patients	214
Cheng YK, 2024 [23]	Retrospective	Single-center	2016–2020	Multiple pathologies included	NA
Deng Y, 2024 [51]	Retrospective	Single-center	NA	NA	172
He S, 2024 [54]	Retrospective	Single-center	NA	Healthy and disc degeneration	276
Zhu Z, 2024 [66]	Prospective	Single-center	2020–2020	NA	50

measurements are also reflect of the accuracy of automatic segmentation.

#### Methodological quality

All integrated studies underwent quality assessment using the QUADAS-2 tool. Regarding bias risk within patient selection, 8 studies were classified as having a

high risk of bias since they did not report clear inclusion and exclusion criteria [23, 30, 34, 42, 52, 58, 59, 64]. The ambiguity of the subjects may limit the applicability of the results. 6 studies exhibited an indeterminate risk of bias [33, 38, 49, 50, 62, 63]. Concerning the reference standard, 5 studies were assessed with a high risk of bias due to the lack of description on ground truth establishment

**Table 2** Characteristics of used datasets and algorithms frameworks

First author & publication year	Train/validation/test dataset (n)	Scanning direction	Slice	Sequence	Algorithm	Classification of algorithm	Network dimensionality
Li X, 2018 [22]	18/0/6 <sup>a</sup>	Sagittal	Sagittal slices	T2	MsFCN	CNN variants	3D
Han Z, 2018 [64]	203/0/50 <sup>bc</sup>	Sagittal	Para-sagittal	T1/T2	Spine-GAN	GAN variants	2D
Kim S, 2018 [52]	20/0/5 <sup>bc</sup>	Sagittal	Mid-sagittal	T2	BSUNet	U-Net variants	2D
Pang S, 2019 [50]	188/0/47 <sup>bc</sup>	Sagittal	Mid-sagittal	T2	CARN	CNN variants	2D
Rehman F, 2019 [34]	20/0/5 <sup>b</sup>	Sagittal	Mid-sagittal	T2	U-Net	U-Net variants	2D
Al-Kafri AS, 2019 [45]	38,676/0/9669 <sup>b</sup>	Sagittal and axial	All sagittal and axial slices	T1/T2 registration	SegNet	CNN variants	2D
Huang J, 2020 [62]	50/0/50 <sup>a</sup>	Sagittal	Mid-sagittal and para-sagittal	T2	U-Net	U-Net variants	2D
Iriondo C, 2020 [31]	62/0/18 <sup>b</sup>	Sagittal	Sagittal	T1/T2 registration	CFCM	CNN variants	2D/3D
Han Z, 2021 [37]	203/0/50 <sup>bc</sup>	Sagittal	Para-sagittal	T1/T2	NSL	GAN variants	2D
Cheng YK, 2021 [42]	2674/0/308 <sup>b</sup>	Sagittal	Mid-sagittal	T2	MultiResUNet	U-Net variants	2D
Das P, 2021 [28]	16/0/8 <sup>a</sup>	Sagittal	Sagittal	T2	RIMNet	CNN variants	3D
Gaonkar B, 2021 [30]	15/0/67 <sup>a</sup>	Sagittal	Sagittal	T2	U-Net	U-Net variants	2D
Suri A, 2021 [47]	898/0/225 <sup>bc</sup>	Sagittal	Sagittal slices	T1/t2	SpineTK	CNN variants	2D
Mbarki W, 2021 [53]	1500/0/500 <sup>b</sup>	Axial	Axial slices	T2	CycleGAN	GAN variants	2D
Pang S, 2021 [36]	172/0/43 <sup>ac</sup>	Sagittal	Sagittal slices	T2	SpineParseNet	U-Net variants	3D
Liu Z, 2022 [25]	80/0/20 <sup>ac</sup>	Axial	3D SPACE	T2	3D U-Net	U-Net variants	3D
Pang S, 2022 [29]	373/0/43 <sup>bc</sup>	Sagittal	Mid-sagittal	T2	DGMSNet	Deeplab variants	2D
Zheng HD, 2022 [49]	316/0/104 <sup>c</sup>	Sagittal	Mid-sagittal	T2	BianqueNet	CNN variants	2D
Sustersic T, 2022 [46]	195/24/24 <sup>a</sup>	Sagittal and axial	Sagittal and axial	T2	DiscNet	U-Net variants	2D
Gong H, 2022 [33]	170/0/33 <sup>bc</sup>	Sagittal	Sagittal slices	T2	ResAttenGAN	GAN variants	2D
Chen T, 2022 [24]	70/10/20 <sup>ac</sup>	Axial	3D SPACE	T2	3D U-Net	U-Net variants	3D
Kuang X, 2022 [35]	20/10/10 <sup>a</sup>	Sagittal	7 or more sagittal slices	T2	Spine-GFlow	U-Net variants	3D
Hou C, 2022 [61]	1157/100/290 <sup>b</sup>	Axial	Axial slices	T1/T2 registration	FC-DenseNet	U-Net variants	2D
Altun S, 2022 [55]	270/0/30 <sup>a</sup>	Axial	All axial slices	T2	3D U-Net	U-Net variants	3D
Wang H, 2023 [38]	56/0/14 <sup>a</sup>	Axial	12 axial slices	T2	3D segmentation model	CNN variants	3D
Coppock JA, 2023 [40]	22/2/1 <sup>a</sup>	Sagittal	3D SPACE	T2	U-Net	U-Net variants	3D
He S, 2023 [57]	545/156/78 <sup>b</sup>	Sagittal	Mid-sagittal and para-sagittal	T2	DLS-Net	U-Net variants	2D
Liu H, 2023 [58]	13/0/3 <sup>a</sup>	Sagittal	Sagittal	T2	MLP-Res-Unet	U-Net variants	2D
Yilizati-Yilihamu EE, 2023 [63]	138/4/30 <sup>ac</sup>	Sagittal	NA	T2	SAFNet	CNN variants	3D
Li H, 2023 [65]	172/0/43 <sup>bc</sup>	Sagittal	Mid-sagittal	T2	SSCK-Net	CNN variants	2D
Hess M, 2023 [27]	160/80/16 <sup>b</sup>	Sagittal	8 sagittal slices	T1	2DV-Net	U-Net variants	2D
Qinhong D, 2023 [32]	182/0/46 <sup>a</sup>	Sagittal	Sagittal slices	T1/T2	MAS-Net	U-Net variants	3D
He S, 2023 [59]	621/0/155 <sup>bc</sup>	Sagittal	Sagittal slices	T2	U-Net variants	U-Net variants	2D
Bharadwaj UU, 2023 [26]	150/20/30 <sup>a</sup>	Axial	All axial slices	T2	2DV-Net	U-Net variants	2D
Soydan Z, 2023 [44]	327/0/36 <sup>b</sup>	Sagittal	Mid-sagittal	T2	U-Net	U-Net variants	2D
Saenz-Gamboa JJ, 2023 [43]	96/49/36 <sup>a</sup>	Sagittal	Sagittal slices	T1/T2 registration	U-Net variants	U-Net variants	2D
Altun I, 2023 [56]	1404/0/156 <sup>b</sup>	Axial	L3-5 axial slices	T2	ResUNet	U-Net variants	2D
He S, 2023 [41]	607/0/152 <sup>bc</sup>	Sagittal	Mid-sagittal and para-sagittal	T2	Lightweight ANN	U-Net variants	2D
Wang M, 2023 [48]	32/8/10 <sup>ac</sup>	Axial	3D SPACE	T2	3D Deeplab V3+	Deeplab variants	3D
Pang C, 2023 [39]	1082/154/309 <sup>b</sup>	Axial	Axial images	T1/T2 registration	SGRNet	U-Net variants	2D
He S, 2024 [60]	621/0/155 <sup>bc</sup>	Sagittal	Mid-sagittal and para-sagittal	T2	SALW-Net	CNN variants	2D

**Table 2** (continued)

First author & publication year	Train/validation/test dataset (n)	Scanning direction	Slice	Sequence	Algorithm	Classification of algorithm	Network dimensionality
Cheng YK, 2024 [23]	2674/0/308 <sup>b</sup>	Sagittal	Mid-sagittal	T2	MultiResUNet	U-Net variants	2D
Deng Y, 2024 [51]	1953/0/216 <sup>b</sup>	Sagittal	Sagittal	T2	U-Net and BiSeNet complementary network	U-Net variants	2D
He S, 2024 [54]	795/0/99 <sup>b</sup>	Sagittal	Mid-sagittal and para-sagittal	T2	DLS-Net	U-Net variants	2D
Zhu Z, 2024 [66]	40/0/10 <sup>ac</sup>	Coronal, sagittal, and axial	3D SPACE	T2	3D U-Net	U-Net variants	3D

<sup>a</sup> Grouped by number of patients, <sup>b</sup> Grouped by number of slices, <sup>c</sup> 5-fold cross validation

[22, 49, 50, 52, 62]. 2 studies exhibited an indeterminate risk of bias [38, 57]. All studies were considered to have low risk of bias in the index test and flow and timing, since an explicit algorithm model and a clarified research flow are necessary conditions for this type of research to be recognized. However, judging the performance of the proposed models solely based on the description in the articles may not be sufficiently accurate. The repeatability and applicability of applying algorithmic models for automatic IVD segmentation can only be confirmed through further external validation.

Regarding applicability, Subjects recruited from the community and patients with lower back or leg pain were considered to match the review question. 9 studies were assessed as having high concern in patient selection [22, 28, 32–34, 40, 52, 57, 64]. The datasets included images from specific treatment stages, reformatted images, images without clear patient information. 9 studies were assessed as having indeterminate concern [23, 30, 42, 49, 50, 58, 59, 62, 63]. The datasets were derived from hospital databases but without associated patient information. Concerning the reference standard, 5 studies were assessed with high concern [22, 49, 50, 52, 62] and one study exhibited an indeterminate concern [57] due to lack or insufficient description. The detailed information of quality assessment was shown in Figure S1 and Table S1.

#### Meta-analysis of the included studies

As is shown in Fig. 2, the pooled value of DSC from 14 studies was 0.900 (95% confidence interval [CI]: 0.887–0.914) [22, 24–32, 34–37]. The Higgins  $I^2$  statistic showed not important heterogeneity across the studies ( $I^2=20.501$ ). A sensitivity analysis confirmed the robustness of the results, as the overall effect sizes remained statistically significant even when any individual study was excluded from the analysis (Figure S2). In addition, 4 studies reported the IoU of IVD segmentation [23, 28,

33, 35], and the pooled value was 0.863 (95% CI: 0.730–0.995,  $I^2=0.000$ ,  $p=0.073$ , Fig. 3).

#### Subgroup analysis

Subgroup analyses were conducted to determine if factors such as network dimensionality, type of algorithm, publication year, number of patients included, scanning direction, data augmentation, and cross validation might influence the effect of IVD segmentation (DSC). The detailed results are shown in Table 3. Although the Higgins  $I^2$  statistics indicated moderate heterogeneity between subgroups for network dimensionality ( $I^2 = 73.174$ ) and publication year ( $I^2 = 65.760$ ), the Q-test suggested no significant difference between subgroups when stratified by these factors ( $p>0.05$ ). The forest plots of the subgroup analyses are presented in Figures S3–S8.

#### Publication bias

Publication bias analysis was conducted for the DSC using a funnel plot (Figure S9), as the number of studies available for other outcome evaluation metrics was limited. The  $p$ -value of the Egger's test was 0.458, suggesting no significant publication bias.

#### Discussion

The global prevalence of low back pain is 18%, with IVD pathology identified as a significant contributor [66]. The interpretation of imaging for IVD diseases is often time-consuming and challenging. In light of recent advancements in AI, particularly DL, the application of these technologies to medical imaging has the potential to improve current medical practices. This systematic review and meta-analysis showed that the pooled DSC for lumbar IVD segmentation in MRI using various DL techniques was 0.900, with a 95% CI of 0.887–0.914, indicating a satisfactory level of accuracy. This technology can be further applied in diagnosis, measurement and evaluation, and surgical planning. To the best of our knowledge, this is the first systematic review and meta-analysis

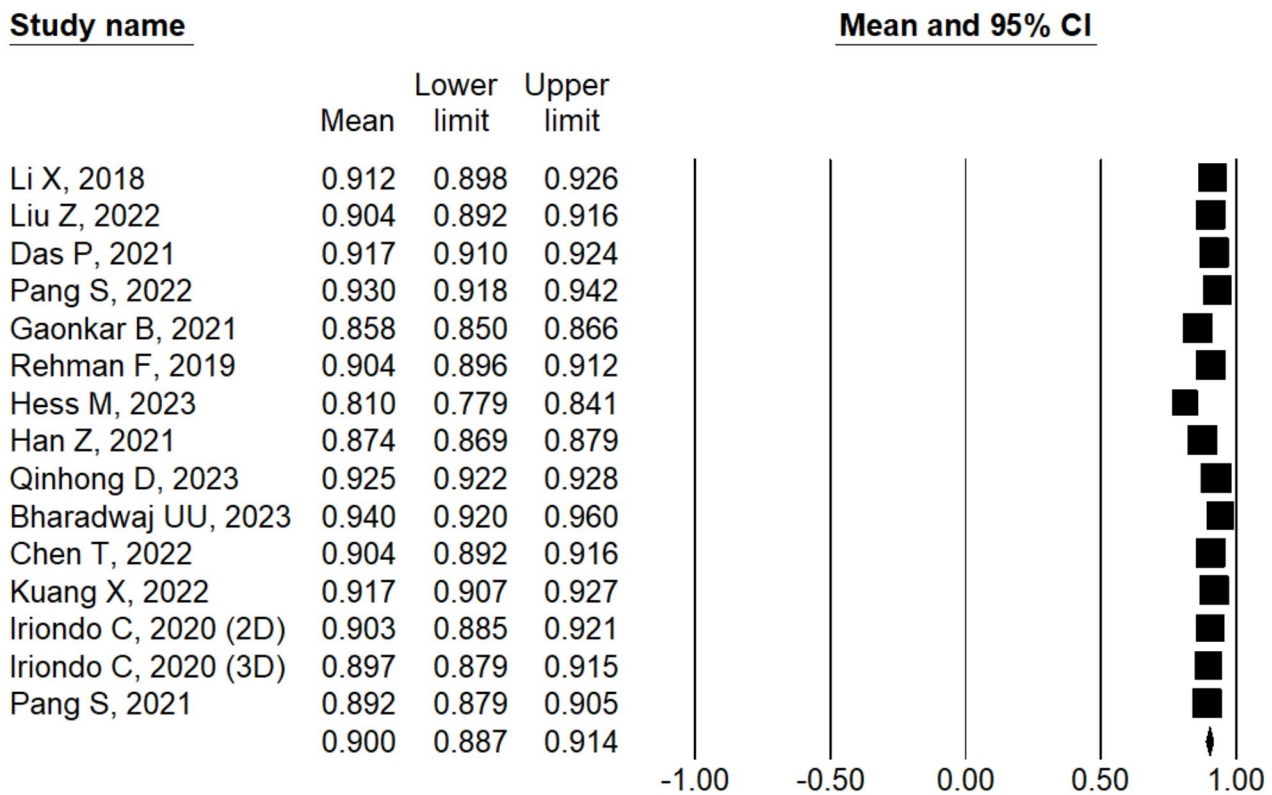


Fig. 2 Forest plot of deep learning algorithms' dice similarity coefficient

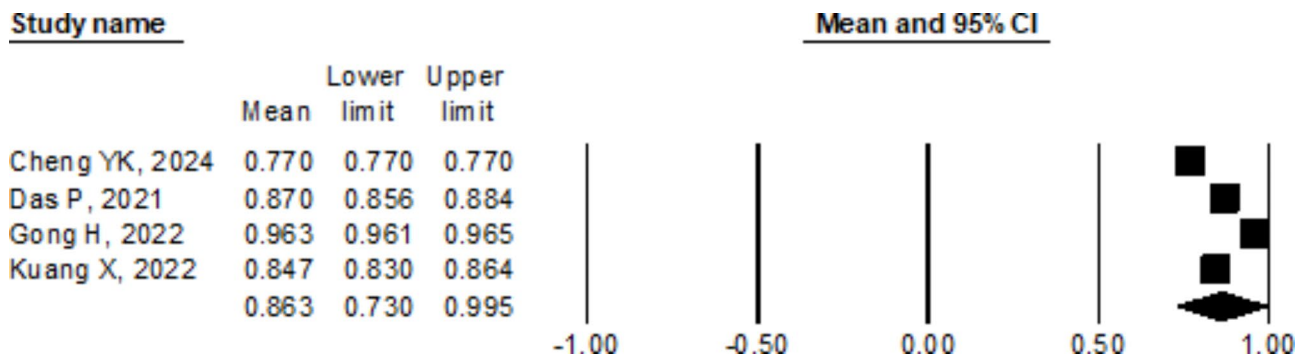


Fig. 3 Forest plot of deep learning algorithms' Intersection over Union score

to address this topic. However, due to inconsistencies in reporting metrics and algorithm frameworks among the included studies, the interpretation of the results should be approached with caution.

In this systematic review and meta-analysis, the reported DSC for IVD segmentation ranged from 0.810 to 0.982. The studies exhibited no significant heterogeneity. This may be because the structural contours of IVDs in lumbar MR images are usually clear, making automatic IVD segmentation easier and more stable than structures that are harder to distinguish, such as tumors [67]. The included studies were divided into different subgroups based on various criteria, but no significant statistical

differences were found between any of the subgroups. We think that most studies used high-quality lumbar MRI datasets and followed similar research designs and processes. Additionally, all studies were based on limited datasets, and the algorithm design is the main factor affecting IVD segmentation performance. Although the applied algorithms can be broadly classified, almost no two studies used identical algorithm frameworks. For instance, the U-Net network, widely used for medical image segmentation due to its symmetric encoder-decoder structure, captures both global context and local details [68]. With technological advancements, there are now several common variants like U-Net++ and V-Net.



**Table 3** Results of subgroup analyses

Group by	Subgroup	Number of studies	DSC (95% CI)	I <sup>2</sup> (%)	p
Network dimensionality	2D	7	0.891 (0.875–0.901)	73.174	0.098
	3D	8	0.909 (0.894–0.923)	0	
Type of algorithm	U-Net variants	9	0.897 (0.879–0.914)	46.056	0.439
	CNN variants	4	0.908 (0.881–0.934)	0	
	Deeplab variants	1	0.930 (0.878–0.982)	0	
	GAN variants	1	0.874 (0.823–0.925)	0	
Publication year	Before 2022	7	0.908 (0.891–0.925)	65.760	0.230
	After 2022	8	0.894 (0.879–0.910)	0	
Number of patients	< 100	7	0.897 (0.875–0.918)	31.430	0.688
	≥ 100	8	0.903 (0.883–0.923)	0	
Scanning direction	Sagittal	12	0.896 (0.881–0.912)	25.037	0.294
	Axial	3	0.915 (0.884–0.947)	0	
Data augmentation	Yes	7	0.905 (0.886–0.925)	0	0.465
	No	8	0.895 (0.877–0.914)	44.260	
Cross validation	Yes	6	0.901 (0.880–0.922)	0	0.913
	No	9	0.900 (0.882–0.917)	48.859	

Researchers can optimize the performance of specific algorithms by adjusting the number of convolutional layers, replacing convolution kernels and pooling layers, and introducing other structures. It may be difficult to quantify the specific impact of these strategies on segmentation performance.

Based on the above discussion, it is worth noting the improvement strategies proposed by studies to achieve a precise and practical algorithm model for IVD segmentation. Researchers have made several representative improvements in various aspects:

(1) To mitigate poor training quality or overfitting caused by limited datasets, in addition to common strategies such as cross-validation, Pang [15] proposed a method called adaptive local shape-constrained manifold regularization, this method forces the output of the cascade amplifier regression network to lie on the target output manifold using local linear representation, which reduces the overfitting of the model. Das [28] and Li [22] conducted IVD segmentation using the MACCAI Challenge datasets, which contained multimodal MR images of a few patients. They used region-to-image matching and dropout strategies to improve feature learning and generalization, maximizing the utilization of multiple MRI sequences. Another solution is to make the data more useful. Gaonkar [30] designed the Eigenrank by Committee (EBC) algorithm. EBC can choose images that are harder to classify for training, which improves the effectiveness of manual annotation and gives better results compared to randomly partitioned training sets. Although data augmentation is a common strategy to increase the sample size. Some scholars believe that methods such as rotation and contrast adjustment may change key information in the MRI and therefore may not be suitable for such rigorous medical images

[5]. In the subgroup analysis of this study, data enhancement showed no significant effect on segmentation performance.

(2) To improve the performance and usability of algorithms, a common strategy is to use multi-scale feature fusion [32, 33, 43, 49, 50, 62]. By utilizing multi-branch structures, such as BiSeNet and PSPNet, it combines high-resolution details from low-level features and semantic information from high-level features. This approach fully leverages the small inter-class differences and large intra-class variations in spinal anatomy features, enhancing detection capability. Multi-scale feature fusion also helps the model understand the broader anatomical context and relationships within the spinal structure, which allows it to segment various spinal structures at the same time [50]. Another popular approach is semi-supervised learning (SSL), which combines labeled data with weakly labeled or unlabeled data. Common methods include self-training, pseudo-labeling, and generative models. SSL has many advantages, such as increasing the amount of data, reducing the workload of expert labeling, enhancing the model's generalization ability, and allowing AI systems to identify imaging features that may be undetectable by human doctors [29, 64]. Other strategies include the level set approach [34] and residual refinement attention [33] (for tracking and refining image boundaries), mixed loss functions [63] (to enhance model robustness), and ensemble learning [43] (to combine outputs from multiple models), etc. Additionally, He [38, 41, 53, 56, 58, 59], Wang [38], and Liu [57] proposed reducing convolutional parameters and calculation complexity. These approaches minimize the algorithm's size and memory usage while maintaining segmentation performance, which is crucial for applying and deploying such algorithms in further clinical practice.

(3) Several studies have further explored the use of IVD segmentation. The segmented IVDs can be used to diagnose degenerative diseases such as IVD herniation [26, 44]. Recent advancements include the application of meta-interpretive learning, dimensionality reduction and integration of MR images, and multi-input, multi-class algorithms, aiming for more precise diagnosis and report generation [37, 41, 46]. To improve the appearance of segmentation, Hou [60] introduced Gaussian divergence loss and contour loss to address issues such as irregular edges and isolated segments. Meanwhile, He [58] developed a filling algorithm for sparse segmented images, which leverages contextual information from adjacent slices to generate interpolated slices and smooth the 3D reconstructed image. 3D reconstruction of the IVD and surrounding structures has been explored for clinical applications, particularly for morphological evaluation and surgical planning of the lumbar spine [24, 25, 65]. However, these applications still rely on manual planning by physicians, and there is a lack of systematic automated surgical planning algorithms.

Through our review of the studies in this field, we have identified areas that require improvement to drive technological advancements. First, the training data rather than the algorithm framework is the essence of determining algorithm performance. However, due to gaps in specialized knowledge and the confidentiality requirement of medical data, we have not yet encountered research that can truly be considered “big data”. Such research should encompass diverse ages, races, and other variables to minimize bias and include a broader range of pathologies, such as internal fixation, infections, and deformities, to ensure applicability. Second, most related studies are conducted by engineers rather than clinicians, resulting in a primary focus on algorithm design. Many studies lack detailed reports on patient inclusion and ground truth establishment, which may limit the models’ applicability. Therefore, closer collaboration between engineers and clinicians is essential for further research. Third, although no significant differences were observed in the subgroup analysis, we still recommend that future studies employ more 3D MRI data, as it provides richer detail and aids subsequent applications. A considerable number of studies train segmentation on specific slices, such as midsagittal slices, which, while providing critical information, are not sufficiently suitable for direct clinical application. Finally, as many scholars have recently highlighted, the practical issues of software implementation of models, ethical approval, and cost-effectiveness must be addressed in future research [7, 10]. Despite these challenges, current advancements demonstrate a promising outlook for the application of DL technology. Therefore, continuing to explore ways for DL technology

to provide tangible benefits to clinicians and patients remains worthwhile.

Additionally, this systematic review and meta-analysis have some limitations. First, there are certain discrepancies in the reporting metrics of the reviewed studies, and many do not provide complete data on segmentation performance, such as standard deviations or confidence intervals, resulting in a limited number of studies that can be included in quantitative summaries. Second, image segmentation requires less complete patient baseline data compared to diagnostic and prognostic studies, however, the heterogeneity of datasets may still limit the significance of this study’s results. Third, there is currently very limited peer review and external validation of the models.

## Conclusion

In conclusion, the DL algorithm enables automatic segmentation of IVDs in MRI imaging with relatively satisfactory performance. This technology has potential applications in diagnosis, measurement and evaluation, and surgical planning. However, the current results should be interpreted with caution due to limitations, such as small sample sizes, differences in reporting metrics, and lack of external validation of the algorithms.

In future studies, it is recommended to use larger and more diverse datasets for training, and to promote external validation and applied research of the algorithms. Clinicians and DL experts can work together to guide this technology and bring tangible benefits to patients and clinical practice.

## Abbreviations

IVD	Lumbar intervertebral disc
AI	Artificial intelligence
DL	Deep learning
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2	Quality Assessment Tool for Diagnostic Accuracy Studies
DSC	Dice similarity coefficient
IoU	Intersection over Union
GAN	Generative adversarial networks
CNN	Convolutional neural network
EBC	Eigenrank by Committee
SSL	Semi-supervised learning

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13018-024-05002-5>.

Supplementary Material 1

Supplementary Material 2

## Acknowledgements

Not applicable.

### Author contributions

L.Z. conceived and designed the study. A.W. and C.Z. performed analysis and drafted the manuscript. S.Y., N.F., and P.D. performed the literature research and interpretation of data and revised the manuscript. T.W. contributed to the data extraction and critical revision of the paper. All authors read and approved the final manuscript.

### Funding

This work was supported by the Beijing Natural Science Foundation (7242059).

### Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 11 July 2024 / Accepted: 16 August 2024

Published online: 21 August 2024

### References

1. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. *Lancet*. 2021;398(10294):78–92.
2. Wenig CM, Schmidt CO, Kohlmann T, Schweikert B. Costs of back pain in Germany. *Eur J Pain*. 2009;13(3):280–6.
3. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)*. 2001;26(17):1873–8.
4. Leone A, Guglielmi G, Cassar-Pullicino VN, Bonomo L. Lumbar intervertebral instability: a review. *Radiology*. 2007;245(1):62–77.
5. Compte R, Granville Smith I, Isaac A, Danckert N, McSweeney T, Liantis P, Williams FMK. Are current machine learning applications comparable to radiologist classification of degenerate and herniated discs and Modic change? A systematic review and meta-analysis. *Eur Spine J*. 2023;32(11):3764–87.
6. Hirschmann A, Cyriac J, Stieltjes B, Kober T, Richiardi J, Omoumi P. Artificial Intelligence in Musculoskeletal Imaging: review of current literature, challenges, and Trends. *Semin Musculoskelet Radiol*. 2019;23(3):304–11.
7. Bousson V, Benoist N, Guetat P, Attane G, Salvat C, Perronne L. Application of artificial intelligence to imaging interpretations in the musculoskeletal area: where are we? Where are we going? *Joint Bone Spine*. 2023;90(1):105493.
8. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine*. 2019;2(1):e1044.
9. Ren G, Yu K, Xie Z, Wang P, Zhang W, Huang Y, Wang Y, Wu X. Current applications of machine learning in spine: from clinical view. *Global Spine J*. 2022;12(8):1827–40.
10. Hornung AL, Hornung CM, Mallow GM, Barajas JN, Espinoza Orias AA, Galbusera F, Wilke HJ, Colman M, Phillips FM, An HS. Artificial intelligence and spine imaging: limitations, regulatory issues and future direction. *Eur Spine J*. 2022;31(8):2007–21.
11. Jamaludin A, Kadir T, Zisserman A. SpineNet: automated classification and evidence visualization in spinal MRIs. *Med Image Anal*. 2017;41:63–73.
12. Wilson B, Gaonkar B, Yoo B, Salehi B, Attiah M, Villaroman D, Ahn C, Edwards M, Laiwalla A, Ratnaparkhi A, et al. Predicting spinal surgery candidacy from Imaging Data using machine learning. *Neurosurgery*. 2021;89(1):116–21.
13. Kim JK, Wang MX, Chang MC. Deep learning algorithm trained on lumbar magnetic resonance imaging to Predict outcomes of Transforaminal Epidural Steroid Injection for Chronic Lumbosacral Radicular Pain. *Pain Physician*. 2022;25(8):587–92.
14. Sustersic T, Rankovic V, Milovanovic V, Kovacevic V, Rasulic L, Filipovic N. A deep learning model for automatic detection and classification of disc herniation in magnetic resonance images. *IEEE J Biomed Health Inf*. 2022;26(12):6036–46.
15. Pang S, Su Z, Leung S, Nachum IB, Chen B, Feng Q, Li S. Direct automated quantitative measurement of spine by cascade amplifier regression network with manifold regularization. *Med Image Anal*. 2019;55:103–15.
16. Liu J, Cui Z, Desrosiers C, Lu S, Zhou Y. Grayscale self-adjusting network with weak feature enhancement for 3D lumbar anatomy segmentation. *Med Image Anal*. 2022;81:102567.
17. Martin-Noguerol T, Onate Miranda M, Amrhein TJ, Paulano-Godino F, Xiberta P, Vilanova JC, Luna A. The role of Artificial intelligence in the assessment of the spine and spinal cord. *Eur J Radiol*. 2023;161:110726.
18. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
19. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the, Clifford P-DTAG, Cohen T, Deeks JF, Gatsonis JJ et al. C : Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* 2018;319(4):388–396.
20. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. Group Q-: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–36.
21. Wang JC, Shu YC, Lin CY, Wu WT, Chen LR, Lo YC, Chiu HC, Ozcakar L, Chang KV. Application of deep learning algorithms in automatic sonographic localization and segmentation of the median nerve: a systematic review and meta-analysis. *Artif Intell Med*. 2023;137:102496.
22. Li X, Dou Q, Chen H, Fu CW, Qi X, Belavý DL, Armbrecht G, Felsenberg D, Zheng G, Heng PA. 3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR Images. *Med Image Anal*. 2018;45:41–54.
23. Cheng YK, Lin CL, Huang YC, Lin GS, Lian ZY, Chuang CH. Accurate Intervertebral Disc Segmentation Approach based on deep learning. *Diagnostics (Basel Switzerland)* 2024, 14(2).
24. Chen T, Su Z-h, Liu Z, Wang M, Cui Z-f, Zhao L, Yang L-j, Zhang W-c, Liu X, Liu J, et al. Automated magnetic resonance image segmentation of spinal structures at the L4-5 level with deep learning: 3D Reconstruction of lumbar intervertebral foramen. *Orthop Surg*. 2022;14(9):2256–64.
25. Liu Z, Su Z, Wang M, Chen T, Cui Z, Chen X, Li S, Feng Q, Pang S, Lu H. Computerized characterization of spinal structures on MRI and clinical significance of 3D Reconstruction of Lumbosacral Intervertebral Foramen. *Pain Physician*. 2022;25(1):E27–.
26. Bharadwaj UU, Christine M, Li S, Chou D, Pedoia V, Link TM, Chin CT, Majumdar S. Deep learning for automated, interpretable classification of lumbar spinal stenosis and facet arthropathy from axial MRI. *Eur Radiol*. 2023;33(5):3435–43.
27. Hess M, Allaire B, Gao KT, Tibrewala R, Inamdar G, Bharadwaj U, Chin C, Pedoia V, Boussein M, Anderson D, et al. Deep learning for Multi-tissue Segmentation and fully automatic personalized biomechanical models from BACPAC clinical lumbar spine MRI. *Pain Med*. 2023;24(5):139–48.
28. Das P, Pal C, Acharyya A, Chakrabarti A, Basu S. Deep neural network for automated simultaneous intervertebral disc (IVDs) identification and segmentation of multi-modal MR images. *Comput Methods Programs Biomed*. 2021;205:106074.
29. Pang S, Pang C, Su Z, Lin L, Zhao L, Chen Y, Zhou Y, Lu H, Feng Q. DGMSNet: spine segmentation for MR image by a detection-guided mixed-supervised segmentation network. *Med Image Anal*. 2022;75:102261.
30. Gaonkar B, Beckett J, Attiah M, Ahn C, Edwards M, Wilson B, Laiwalla A, Salehi B, Yoo B, Bui AAT et al. Eigenrank by committee: Von Neumann entropy based data subset selection and failure prediction for deep learning based medical image segmentation. *Med Image Anal* 2021, 67.
31. Iriondo C, Pedoia V, Majumdar S. Lumbar intervertebral disc characterization through quantitative MRI analysis: an automatic voxel-based relaxometry approach. *Magn Reson Med*. 2020;84(3):1376–90.
32. QinHong D, Yue H, Wendong B, Yukun D, Huan Y, Yongming X. MAS-Net: multi-modal Assistant Segmentation Network for lumbar intervertebral disc. *Phys Med Biol* 2023, 68(17).
33. Gong H, Liu J, Chen B, Li S. ResAttenGAN: simultaneous segmentation of multiple spinal structures on axial lumbar MRI image using residual attention and adversarial learning. *Artif Intell Med* 2022, 124.
34. Rehman F, Shah SIA, Riaz N, Gilani SO. A Robust Scheme of Vertebrae Segmentation for Medical diagnosis. *Ieee Access*. 2019;7:120387–98.

35. Kuang X, Cheung JPY, Wong KK, Lam WY, Lam CH, Choy RW, Cheng CP, Wu H, Yang C, Wang K, et al. Spine-GFlow: a hybrid learning framework for robust multi-tissue segmentation in lumbar MRI without manual annotation. *Comput Med Imaging Graph*. 2022;99:102091.
36. Pang S, Pang C, Zhao L, Chen Y, Su Z, Zhou Y, Huang M, Yang W, Lu H, Feng Q. SpineParseNet: spine parsing for volumetric MR image by a two-stage Segmentation Framework with semantic image representation. *IEEE Trans Med Imaging*. 2021;40(1):262–73.
37. Han Z, Wei B, Xi X, Chen B, Yin Y, Li S. Unifying neural learning and symbolic reasoning for spinal medical report generation. *Med Image Anal* 2021, 67.
38. Wang H, Chen Y, Jiang T, Bian H, Shen X. 3D multi-scale feature extraction and recalibration network for spinal structure and lesion segmentation. *Acta Radiol (Stockholm Sweden)*: 1987. 2023;64(12):3015–23.
39. Pang C, Su Z, Lin L, Lin G, He J, Lu H, Feng Q, Pang S. Automated measurement of spine indices on axial MR images for lumbar spinal stenosis diagnosis using segmentation-guided regression network. *Med Phys*. 2023;50(1):104–16.
40. Coppock JA, Zimmer NE, Spritzer CE, Goode AP, DeFrate LE. Automated segmentation and prediction of intervertebral disc morphology and uniaxial deformations from MRI. *Osteoarthr Cartil open*. 2023;5(3):100378.
41. He S, Li Q, Li X, Zhang M. Automatic aid diagnosis report generation for lumbar disc MR image based on lightweight artificial neural networks. *Biomed Signal Process Control* 2023, 86.
42. Cheng YK, Lin CL, Huang YC, Chen JC, Lan TP, Lian ZY, Chuang CH. Automatic Segmentation of Specific Intervertebral Discs through a two-stage MultiResUNet Model. *J Clin Med* 2021, 10(20).
43. Saenz-Gamboa JJ, Domenech J, Alonso-Manjarres A, Gomez JA, Iglesia-Vaya, Mdl. Automatic semantic segmentation of the lumbar spine: clinical applicability in a multi-parametric and multi-center study on magnetic resonance images. *Artif Intell Med* 2023, 140.
44. Soydan Z, Bayramoglu E, Karasu R, Sayin I, Salturk S, Uvet H. An Automated Deep Segmentation and classification model for lumbar disk degeneration and clarification of its impact on clinical decisions. *Global Spine J* 2023;21925682231200783.
45. Al-Kafri AS, Sudirman S, Hussain A, Al-Jumeily D, Natalia F, Meidia H, Afriliana N, Al-Rashdan W, Bashtawi M, Al-Jumaily M. Boundary delineation of MRI images for lumbar spinal stenosis detection through semantic segmentation using deep neural networks. *IEEE Access*. 2019;7:43487–501.
46. Sustersic T, Rankovic V, Milovanovic V, Kovacevic V, Rasulic L, Filipovic N. A deep learning model for automatic detection and classification of disc herniation in magnetic resonance images. *IEEE J Biomedical Health Inf*. 2022;26(12):6036–46.
47. Suri A, Jones BC, Ng G, Anabaraonye N, Beyrer P, Domi A, Choi G, Tang S, Terry A, Lechner T et al. A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. *Bone* 2021, 149.
48. Wang M, Su Z, Liu Z, Chen T, Cui Z, Li S, Pang S, Lu H. Deep learning-based automated magnetic resonance image segmentation of the lumbar structure and its adjacent structures at the L4/5 level. *Bioeng (Basel Switzerland)* 2023, 10(8).
49. Zheng H-D, Sun Y-L, Kong D-W, Yin M-C, Chen J, Lin Y-P, Ma X-F, Wang H-S, Yuan G-J, Yao M et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. *Nat Commun* 2022, 13(1).
50. Deng Y, Gu F, Zeng D, Lu J, Liu H, Hou Y, Zhang Q. An effective U-Net and BiSeNet complementary network for spine segmentation. *Biomed Signal Process Control* 2024, 89.
51. Kim S, Bae WC, Masuda K, Chung CB, Hwang D. Fine-grain segmentation of the intervertebral discs from MR Spine images using deep convolutional neural networks: BSU-Net. *Appl Sci (Basel Switzerland)* 2018, 8(9).
52. Mbarki W, Bouchouicha M, Tshienda FT, Moreau E, Sayadi M. Herniated lumbar disc generation and classification using cycle generative adversarial networks on Axial View MRI. *Electronics* 2021, 10(8).
53. He S, Li Q, Li X, Zhang M. A lightweight convolutional neural network based on dynamic level-set loss function for spine MR Image Segmentation. *J Magn Reson Imaging*. 2024;59(4):1438–53.
54. Altun S, Alkan A. LSS-net: 3-dimensional segmentation of the spinal canal for the diagnosis of lumbar spinal stenosis. *Int J Imaging Syst Technol*. 2022;33(1):378–88.
55. Altun I, Altun S, Alkan A. LSS-UNET: lumbar spinal stenosis semantic segmentation using deep learning. *Multimedia Tools Appl*. 2023;82(26):41287–305.
56. He S, Li Q, Li X, Zhang M. LSW-Net: lightweight deep neural network based on small-world properties for spine MR Image Segmentation. *J Magn Reson Imaging: JMRI*. 2023;58(6):1762–76.
57. Liu H, Lu S, Zhao F. MLP-Res-Unet: MLPs and residual blocks-based U-shaped network intervertebral disc segmentation of multi-modal MR spine images. *Curr Med Imaging* 2023.
58. He S, Li Q, Li X, Zhang M. An optimized segmentation convolutional neural network with dynamic energy loss function for 3D reconstruction of lumbar spine MR images. *Comput Biol Med*. 2023;160:106839.
59. He S, Li Q, Li X, Zhang M. SALW-Net: a lightweight convolutional neural network based on self-adjusting loss function for spine MR image segmentation. *Med Biol Eng Comput*. 2024;62(4):1247–64.
60. Hou C, Zhang W, Wang H, Liu F, Liu D, Chang J. A semantic segmentation model for lumbar MRI images using divergence loss. *Appl Intell*. 2022;53(10):12063–76.
61. Huang J, Shen H, Wu J, Hu X, Zhu Z, Lv X, Liu Y, Wang Y. 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*. 2020;20(4):590–9.
62. Yilizati-Yilhamu EE, Yang J, Yang Z, Rong F, Feng S. A spine segmentation method based on scene aware fusion network. *BMC Neurosci*. 2023;24(1):49.
63. Han Z, Wei B, Mercado A, Leung S, Li S. Spine-GAN: semantic segmentation of multiple spinal structures. *Med Image Anal*. 2018;50:23–35.
64. Li H, Wang Z, Shen W, Li H, Li H, Yu P. SSSCK-Net: spine segmentation in MRI based on cross attention and key-points recognition-assisted learner. *Biomed Signal Process Control* 2023, 86.
65. Zhu Z, Liu E, Su Z, Chen W, Liu Z, Chen T, Lu H, Zhou J, Li Q, Pang S. Three-Dimensional Lumbosacral Reconstruction by an Artificial Intelligence-based Automated MR Image Segmentation for selecting the Approach of Percutaneous endoscopic lumbar discectomy. *Pain Physician*. 2024;27(2):E245–54.
66. Vlaeyen JWS, Maher CG, Wiech K, Van Zundert J, Meloto CB, Diatchenko L, Battie MC, Goossens M, Koes B, Linton SJ. Low back pain. *Nat Rev Dis Primers*. 2018;4(1):52.
67. Wang TW, Hsu MS, Lee WK, Pan HC, Yang HC, Lee CC, Wu YT. Brain metastasis tumor segmentation and detection using deep learning algorithms: a systematic review and meta-analysis. *Radiother Oncol*. 2024;190:110007.
68. Zhang J, Liu Y, Wu Q, Wang Y, Liu Y, Xu X, Song B. SWTRU: Star-shaped window Transformer Reinforced U-Net for medical image segmentation. *Comput Biol Med*. 2022;150:105954.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.