

Lumbar spine segmentation in MR images: a dataset and a public benchmark

Jasper W. van der Graaf ^{*1,2}, Miranda L. van Hooff^{2,3}, Constantinus F. M. Buckens⁴, Matthieu Rutten^{1,5}, Job L. C. van Susante⁶, Robert Jan Kroeze⁷, Marinus de Kleuver², Bram van Ginneken¹, and Nikolas Lessmann¹

¹Diagnostic Image Analysis Group, Radboud University Medical Center Nijmegen, The Netherlands

²Department of Orthopedic surgery, Radboud University Medical Center Nijmegen, The Netherlands

³Department Research, Sint Maartenskliniek, Nijmegen, The Netherlands

⁴Department of Medical Imaging, Radboud University Medical Center Nijmegen, The Netherlands

⁵Department of Radiology, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands

⁶Department of Orthopedic Surgery, Rijnstate Hospital, Arnhem, the Netherlands

⁷Department of Orthopedic Surgery, Sint Maartenskliniek, Nijmegen, The Netherlands

June 23, 2023

Abstract

This paper presents a large publicly available multi-center lumbar spine magnetic resonance imaging (MRI) dataset with reference segmentations of vertebrae, intervertebral discs (IVDs), and spinal canal. The dataset includes 447 sagittal T1 and T2 MRI series from 218 patients with a history of low back pain. It was collected from four different hospitals and was divided into a training (179 patients) and validation (39 patients) set. An iterative data annotation approach was used by training a segmentation algorithm on a small part of the dataset, enabling semi-automatic segmentation of the remaining images. The algorithm provided an initial segmentation, which was subsequently reviewed, manually corrected, and added to the training data. We provide reference performance values for this baseline algorithm and nnU-Net, which performed comparably. We set up a continuous segmentation challenge to allow for a fair comparison of different segmentation algorithms. This study may encourage wider collaboration in the field of spine segmentation, and improve the diagnostic value of lumbar spine MRI.

1 Background & Summary

Low back pain (LBP) causes the largest burden of disease worldwide, with most years lived with disability of any disease.[1] As a consequence, lumbar spine magnetic resonance imaging (MRI) for LBP is one of the most used imaging procedures within musculoskeletal imaging.[2] In the United States, 93% of the lumbar MRI referrals were appropriate according to the American College of Radiology guidelines, even though only 13% of the scans contributed

in the clinical decision making.[3] Automatic image analysis might be the key to improve the diagnostic value of MRI by enabling more objective and quantitative image interpretation. A first step toward automatic assessment of lumbar spine MRI is segmentation of relevant anatomical structures, such as the vertebrae, intervertebral discs (IVDs) and the spinal canal.

With recent advances in machine learning and artificial intelligence (AI), state-of-the-art spine segmentation algorithms are generally learning-based algorithms that require well-curated training data. The development of vertebra segmentation algorithms for CT images has considerably benefitted from multiple large publicly available datasets with CT images and reference segmentations.[4, 5] Currently no comparable large high quality datasets are available for lumbar spine MRI. Existing available datasets are either small, only segment the vertebral body[6, 7], or are only annotated in the midsagittal slice (2D)[8, 9]. Moreover, most datasets are limited to only one of the many anatomical structures that are most relevant for assessing multifactorial disorders such as LBP, i.e., only the vertebrae[10, 11, 12, 13, 14] or the IVDs[15, 16, 17].

To advance the development of segmentation algorithms, and ultimately automatic image analysis, for lumbar spine MRI, this study has three primary goals:

1. To present a large multi-center lumbar spine MR dataset with reference segmentations of vertebrae, IVDs and spinal canal.
2. To introduce a continuous lumbar spine MRI segmentation challenge that allows algorithm developers to submit their models for evaluation.
3. To provide reference performance metrics for two algorithms that segment all three spinal structures auto-

*corresponding author: jasper.vandergraaf@radboudumc.nl

matically: a baseline AI algorithm, which was used in the data collection process, and the nnU-Net, a popular algorithm for 3D segmentation tasks for which training and inference code is publicly available.

2 Materials and methods

2.1 Data Collection

In total, 218 lumbar spine MRI studies from patients with a history of LBP were retrospectively collected, with each study consisting of up to three MRI series. The complete dataset comprises 447 series. The study was approved by the institutional review board at Radboud University Medical Center (IRB 2016-2275). Informed consent was exempted, given the retrospective scientific use of deidentified MRI scans and clinical data. Studies were collected from four different hospitals in the Netherlands, including one university medical center, two regional hospitals and one orthopedic hospital (data acquired between January 2019 and March 2022). Lumbar spine imaging at the university medical center includes a T2 SPACE sequence that produces images with almost isotropic spatial resolution. We included a random selection of lumbar spine studies with both a standard sagittal T1 and T2 sequence (voxel size: 3.30 x 0.59 x 0.59 mm) and a sagittal T2 SPACE sequence (voxel size: 0.90 x 0.47 x 0.47 mm). At the other three hospitals, we included random selections of lumbar spine studies with at least a sagittal T1 or a sagittal T2 sequence. The voxel size of these images ranged from 3.15 x 0.24 x 0.24 mm to 9.63 x 1.06 x 1.23 mm. Additional dataset characteristics are given in Table 1.

In all included MRI series, all visible vertebrae (excluding the sacrum), intervertebral discs, and the spinal canal were manually segmented. The segmentation was performed by a medical trainee who was trained and supervised by both a medical imaging expert and an experienced musculoskeletal radiologist. Three-dimensional MRI annotation is a complex and laborious task, especially for the vertebral arch of the lumbar vertebrae. Therefore, we worked with an iterative data annotation approach in which our automatic baseline segmentation method (baseline 1: iterative instance segmentation) was trained with a small part of the dataset, enabling semi-automatic segmentation of the remaining images. During semi-automatic segmentation, the automatic method was used to obtain an initial segmentation, which was subsequently reviewed and manually corrected. This process was repeated several times by retraining the automatic segmentation model until the entire dataset was annotated.

Initially, twenty randomly selected high resolution T2 (SPACE) series of the university medical center data were manually annotated using 3D Slicer version 5.0.3[18]. All structures were segmented in their entirety, which for the vertebrae also includes the vertebral arch. This was done since the vertebral arch is essential in the diagnosis of disorders such as foraminal stenosis, facet joint arthrosis, and spondylolysis. The initial manual annotations were per-

formed only on high resolution series because the near-isotropic resolution enables detailed viewing in sagittal, axial and coronal directions. Annotations of the corresponding standard sagittal T1 and T2 images were obtained by resampling the T2 SPACE segmentations to the resolution of the T1 and T2 images. The resampled segmentations were reviewed for misalignment due to patient movement between the acquisitions and corrected if needed. All other segmentations were created by first generating initial segmentations using the automatic segmentation method trained with already annotated data, followed by review and manual correction in 3D Slicer.

The vertebrae were not given a correct anatomical label, since accurately determining the anatomical type of vertebra requires information from multiple planes of MRI images, including axial and coronal views, in addition to sagittal views. These additional views are essential to ensure accurate identification of the ribs which is needed to determine the lowest thoracic vertebra and correctly label the lumbar levels. As only sagittal views were available for the majority of studies in this dataset, accurate anatomical labeling of vertebrae was considered infeasible. Therefore, the reference segmentations provided in this dataset are labeled from the bottom up with the most caudal vertebra (usually L5) labeled as 1.

The dataset was divided into a training set (179 out of 218 studies, 82%) and a validation set (39 out of 218 studies, 18%). This data-split was used during training of the iterative instance segmentation algorithm, however it is not mandatory to maintain the same training and validation split when using this dataset. Series belonging to the same patient were always placed in the same set.

2.2 Baseline 1: Iterative instance segmentation

By presenting this baseline algorithm, we establish a reference point for evaluating performance and provide users with an understanding of the algorithm employed in generating the dataset. This section summarizes the iterative instance segmentation (IIS) method. An automatic AI-based segmentation algorithm for vertebra segmentation[14] was extended to segment also the IVDs and the spinal canal. This algorithm uses a 3D patch-based iterative scheme to segment one pair of vertebra and the corresponding inferior IVD at a time, together with the segment of spinal canal covered by the image patch. A schematic image of the network architecture is shown in Figure 1.

2.2.1 Instance memory

Because the MR volume is segmented by consecutively analyzing 3D patches, one vertebral level at a time, a method is needed to keep track of its progress. An instance memory volume is used to save the structures that have been segmented, and is used as an extra input channel to remind the network of the structures that can be ignored because they are already segmented. In contrast to the original vertebra-focused method, we introduced separate memory state vol-

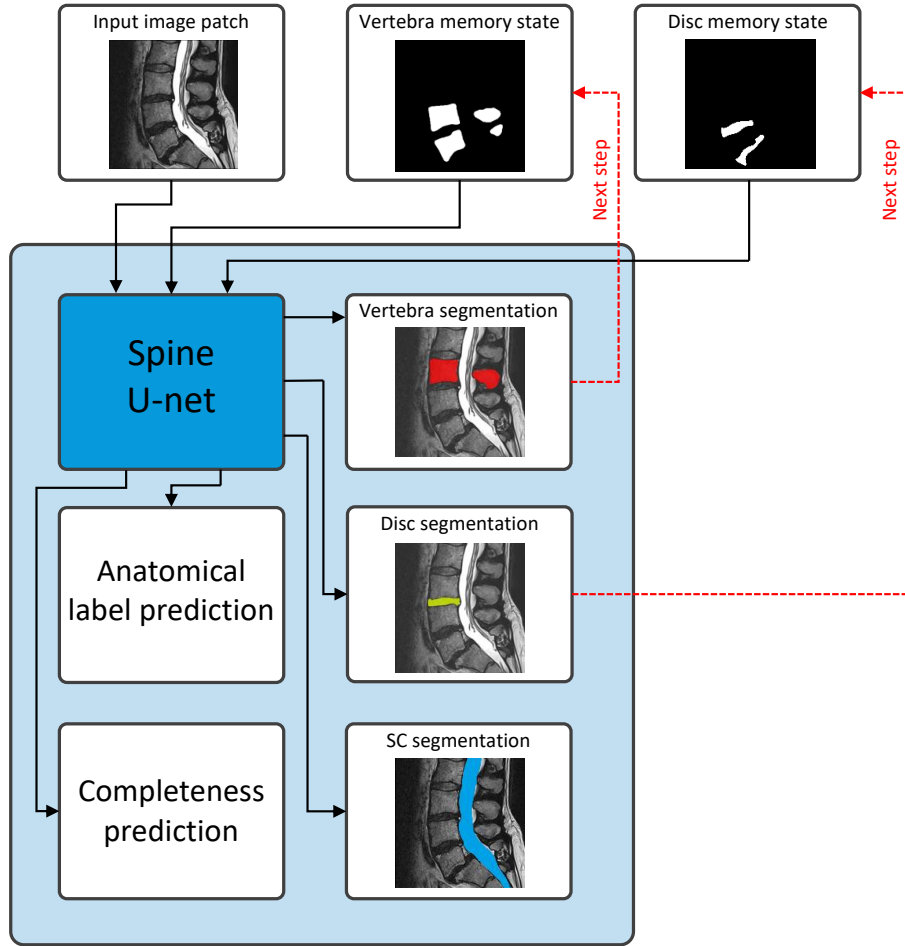


Figure 1: Schematic drawing of the network architecture. The input image and both memory states are fed into the Spine U-net, which produces predictions for the segmentation of the vertebrae (red), IVD (yellow), spinal canal (blue), as well as anatomical and completeness predictions. The vertebra and IVD segmentations are added to their respective memory states, which will be used as input to the network in the next iteration (see the section 2.2.3 for more detail)

umes for the vertebrae, IVDs, and the spinal canal. The spinal canal memory state is only used to save the segmentation progress, not as an extra input for the network, as the spinal canal is an elongated structure that cannot be covered by a single patch. Therefore, the network is trained to always segment any visible portion of the spinal canal, which is then stitched together for all patches that are fed through the network. In total, the network has three input channels, two memory states, and the corresponding image patch.

2.2.2 Network architecture

The segmentation approach is based on a single 3D U-net-like fully-convolutional neural network. Unlike the vertebra segmentation algorithm as described in the original paper[14], a patch size of $64 \times 192 \times 192$ voxels with a resolution of $2 \times 0.6 \times 0.6$ mm was used, as the created dataset contains sagittal MR images exclusively. These generally have a higher slice thickness compared to the data used by

Lessmann et al.[14] A higher in-plane resolution of the predicted segmentation is achieved while still ensuring the patch is large enough that a vertebra completely fits within one patch. The network has three output channels, one for each anatomical structure.

2.2.3 Iterative segmentation approach

The patch-based scheme is structured in such a way that only relevant parts of the MR volume are being processed. The patch systematically moves through the image until it finds a fragment of the first vertebra, in this case always the lowest vertebra. Subsequently, the patch moves to the center of mass of that fragment after which a new segmentation is made. This process continues until the vertebra’s volume stabilizes, which means that the detected vertebra is completely visible within the patch. Binary masks of that vertebra, its underlying IVD, and the spinal canal are then added to their respective memory states. The same patch is segmented again with the updated memory states as in-

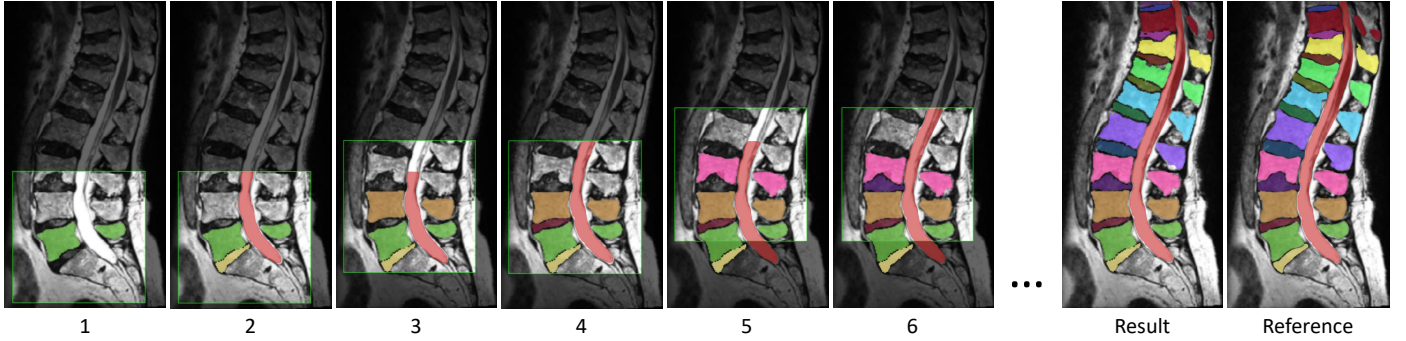


Figure 2: Illustration of the iterative segmentation approach’s functionality. The process involves traversing the 3D patch (depicted by the light area with a green border) along the spine with alternating steps for segmentation of the vertebrae (shown in images 1, 3, and 5) and IVD and spinal canal (displayed in images 2, 4, and 6). The right-hand side displays the final automatic segmentation result alongside the reference segmentation.

put, which causes a fragment of the next vertebra to be segmented. This iterative process, illustrated in Figure 2, continues until no more vertebra fragments are detected or when the top of the MR volume is reached.

2.2.4 Completeness and label prediction

The most cranial vertebra is often only partially visible within the field of view of the MR image. The segmentation method includes an additional compression path after the compression path of the U-net, which has a single binary value as output, predicting the completeness of a vertebra. The original vertebra segmentation method also contained a similar compression path for predicting the anatomical label. However, this output was not used in our experiments since no accurate anatomical labels regarding lumbosacral transitional vertebrae were present in our dataset.

2.2.5 Training of the algorithm

Preprocessing of the images consisted of resampling to a standard resolution of $2 \times 0.6 \times 0.6$ mm and orientation in axial slices. Standard data augmentation steps were implemented, such as random elastic deformation, the addition of random Gaussian noise, random Gaussian smoothing, and random cropping along the longitudinal axis. The loss function used during training consisted of three parts: (1) The segmentation error was defined by the weighted sum of false positives and false negatives combined with the binary cross-entropy loss. (2) The labeling error was defined by the absolute difference between the predicted label and the ground truth. (3) The completeness classification error was defined as the binary cross-entropy between the true label and the predicted probability.

2.3 Baseline 2: nnU-Net

In addition to adapting a segmentation method that was specifically developed for vertebra segmentation, reference results for nnU-Net are provided. nnU-Net is a self-configuring, deep learning-based framework for medical image segmentation.[19] It has been widely accepted in the

medical image analysis community as a state-of-the-art approach to 3D image segmentation tasks after winning the Medical Segmentation Decathlon[20] and performing well in several other segmentation challenges. A 3D full resolution nnU-Net was trained on the training and validation datasets with 5-fold cross validation, which is its recommended training strategy.[19] Data pre-processing, network architecture and other training details were automatically determined by the nnU-Net framework. The network was trained on both the T1- and T2-weighted MRI series after which the overall performance was compared to the IIS baseline algorithm.

2.4 Evaluation

The segmentation performance was evaluated using two metrics: (1) The Dice coefficient to measure the volume overlap, and (2) the average absolute surface distance (ASD) as an indication of the segmentation accuracy along the surface of all structures. Both metrics were calculated separately for all individual structures and were averaged per anatomical structure (vertebrae, IVDs, or spinal canal). Additionally, the average Dice coefficient and average ASD per MRI sequence (T1 vs. T2) were calculated for each anatomical structure. To ensure the Dice score and ASD are not influenced by labeling differences, the individual structures of the reference segmentation are matched to the structured in the predicted segmentation based on the largest found overlap. The completeness classification performance was determined by the percentage of accurate predictions, as well as the average number of false positives and false negatives. Evaluation was performed on a sequestered test set which is a subset of the presented dataset.

3 Data Records

To generate this dataset, a total of 218 lumbar MRI studies of patients with low back pain were included. Each study consisted of up to three sagittal MRI series which were either T1-weighted or T2-weighted (regular resolution, or high resolution generated using a SPACE sequence) with a total of 447 series. Of all included patients, 63% were female.

Table 1: Overview dataset

Abbreviations: UMC, University Medical Center; RH, Regional Hospital; OH, Orthopedic Hospital.

* This only applies to the regular T1 and T2 weighted images. The T2 SPACE sequence has a voxel size of 0.90 x 0.47 x 0.47 mm.

Hospital	Studies	T1	T2	T2 SPACE	Voxel size range (min – max)(mm)	Sex (% female)
UMC	41	39	39	41	(3.24 x 0.27 x 0.47) - (3.34 x 0.59 x 0.85)*	55%
RH1	43	43	37	0	(0.46 x 0.46 x 4.20) - (9.63 x 1.06 x 1.06)	58%
RH2	44	24	44	0	(0.46 x 0.46 x 4.20) - (5.17 x 1.00 x 1.23)	59%
OH	90	90	90	0	(3.15 x 0.24 x 0.24) - (3.39 x 0.83 x 1.02)	68%
Total	218	196	210	41	(3.15 x 0.24 x 0.24) - (9.63 x 1.06 x 1.23)*	63%

Table 2: Overview of the distribution of data between the training and validation set.

Abbreviation: IVDs, Intervertebral Discs.

	Training	Validation	Total
Studies	179	39	218
Series	360	87	447
Vertebrae	2512	613	3125
IVD's	2518	629	3147
Sex (% female)	63%	62%	63%

A total of 3125 vertebrae, 3147 IVDs, and 447 spinal canal segmentations were included over all series combined. An overview of the complete dataset divided by the different hospitals is shown in Table 1. An overview of the training and validation sets and all included structures is shown in Table 2.

All MR images and their corresponding segmentation masks used in this study are stored in MHA format in separate directories. Both files have the same name, which is a combination of the MRI study identifier and the specific sequence type (T1, T2, or T2 SPACE). It is important to note that all MRI series from the same MRI study have the same identifier.

4 Technical Validation

The performance of the IIS baseline algorithm, which was used to generate initial segmentation masks of unseen images from the dataset, was assessed on a hidden test set. These results are presented to assess the data-annotation strategy, as well as establish a reference performance for users of the dataset. The results for the different structures and the different sequences are shown in Table 3. The overall mean (SD) Dice score was 0.93 (± 0.05), 0.85 (± 0.10) and 0.92 (± 0.04) for the vertebrae, IVDs and spinal canal respectively. The overall mean (SD) ASD was 0.49 mm (± 0.95 mm), 0.53 mm (± 0.46 mm) and 0.39 (mm ± 0.45) mm for the vertebrae, IVDs and spinal canal respectively. The spinal canal was identified in all scans. One of the 656 vertebrae and nine (three in T1 images and six in T2 images) of the 688 IVDs were not found. The completeness prediction was correct in 650 of the 656 vertebrae (99.1%). A nnU-Net was trained on the same training data to enable comparison between the IIS baseline algorithm and nnU-Net baseline. The results of both networks is displayed in Table 4. Fig-

ure 3 shows a collection of segmentations obtained by both networks.

The IIS model demonstrates strong performance on our dataset, which is comparable to other MR segmentation methods in the literature.[7, 21, 22] These results are nearly identical to the results of the nnU-Net baseline model, which is considered the gold-standard in medical image segmentation. This indicates that the IIS baseline model is a reasonable benchmark for comparison and was an accurate tool in the iterative data annotation workflow.

The used iterative data annotation approach showed to be an effective strategy. One strength of this approach is its ability to improve the quality of the dataset over time by incorporating corrections from segmentation predictions into the training data. This helps to reduce errors and increase accuracy in subsequent iterations. Additionally, this approach was faster and more efficient compared to fully manual annotations. However, there are several limitations that should be addressed. Firstly, the iterative process of training the network on a small dataset, generating segmentation predictions on unseen images, and manually correcting the predictions before adding them to the dataset, can introduce bias in the final dataset. Moreover, the use of only high-resolution T2 series for the initial manual annotation may not be representative of the entire population, as it is limited to patients from one hospital who underwent this specific imaging modality.

In the era of machine learning and AI algorithms, lumbar spine segmentation can serve as the basis for automated, accurate lumbar spine MR analysis, assisting clinical radiologists and imaging-minded spinal surgeons in their daily practice. It will be able to generate robust, quantitative MR results that can serve as inputs into larger models of lumbar spine disease in clinical practice and research settings. The availability of public datasets and benchmarks plays a crucial role in advancing the field. While datasets

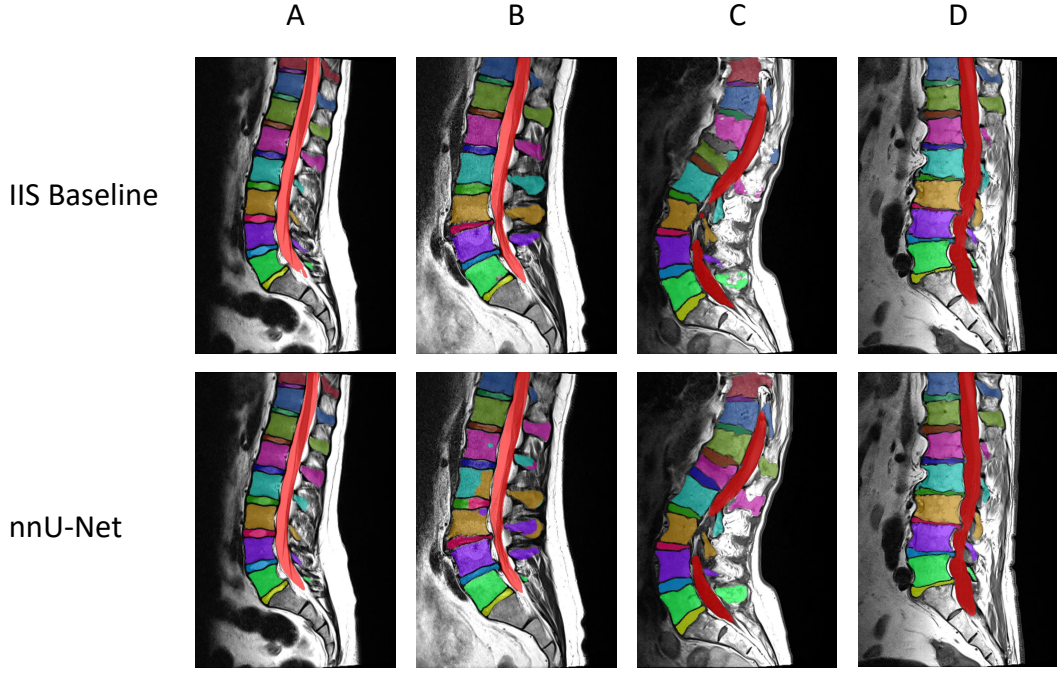


Figure 3: Examples of cases segmented by both baseline algorithms. Each column represents one case. Column A shows a case without any major pathologies and without significant segmentation errors. Columns B and C show cases where nnU-Net (B) and the IIS baseline algorithm (C) made mistakes, indicated by a white arrow. Column D shows a case with severe degenerative features present and no substantial segmentation errors.

Table 3: Overview of all results of the IIS baseline algorithm.

Abbreviations: IVDs, Intervertebral Discs; ASD, Average absolute Surface Distance; SD, Standard Deviation.

		T1		T2		Total	
		Mean	SD	Mean	SD	Mean	SD
Vertebrae	Dice	0.93	0.04	0.92	0.06	0.93	0.05
	ASD (mm)	0.45	0.84	0.51	1.01	0.48	0.95
	Detection%	100.0%		99.7%		99.8%	
	Completeness accuracy%	98.9%		99.2%		99.1%	
IVDs	Dice	0.85	0.10	0.84	0.10	0.84	0.10
	ASD (mm)	0.48	0.47	0.58	0.43	0.54	0.45
	Detection%	99.2%		98.7%		98.9%	
Spinal canal	Dice	0.93	0.03	0.92	0.04	0.92	0.04
	ASD (mm)	0.29	0.17	0.46	0.56	0.39	0.45
	Detection%	100.0%		100.0%		100.0%	

Table 4: Comparison between the iterative segmentation algorithm and nnU-Net.

Abbreviations: IVDs, Intervertebral Discs; IIS, Iterative Instance Segmentation; ASD, Average absolute Surface Distance; SD, Standard Deviation.

		IIS		nnU-Net	
		Mean	SD	Mean	SD
Vertebrae	Dice	0.93	0.05	0.92	0.05
	ASD (mm)	0.48	0.95	0.43	0.78
	Detection%	99.8%		100.0%	
IVDs	Dice	0.84	0.10	0.86	0.09
	ASD (mm)	0.54	0.45	0.54	1.04
	Detection%	98.7%		100.0%	
Spinal canal	Dice	0.92	0.04	0.92	0.03
	ASD (mm)	0.39	0.45	0.37	0.22
	Detection%	100.0%		100.0%	

exist for CT vertebra segmentation, such as VerSe which is the largest available vertebra segmentation dataset[23], currently no public datasets for MRI spine segmentation are available. Our dataset is of similar size to VerSe[23] and provides full segmentation of all relevant spinal structures on MR images. This allows for wider participation and collaboration in the field of spine segmentation, as it can be used to train and evaluate algorithms, as well as to compare to other datasets. The presented algorithms are the baseline results to which other algorithms can be compared.

Usage Notes

All training and validation data can be found at <https://doi.org/10.5281/zenodo.8009680> and are available under the CC-BY 4.0 license.

In order to allow for a fair comparison between different algorithms, including both baseline algorithms, a public segmentation challenge is hosted on the grand-challenge.org platform. The training and validation sets are made publicly available for everyone to develop and train their AI algorithms on. The test set will remain hidden on the Grand Challenge platform to avoid overfitting on it and enable a fair comparison. The test set consists of 39 lumbar MRI studies of unique patients, which includes 15 out of the 20 fully manually annotated studies. The remaining studies originate from the same four hospitals in a similar distribution as the presented dataset.

Participants are invited to submit a trained algorithm to the platform, which automatically executes the algorithm and determines its performance on the hidden test set. The challenge can be accessed on <https://spider.grand-challenge.org/>.

Code Availability

The original code of the IIS baseline algorithm is publicly available at: <https://github.com/DIAGNijmegen/SPIDER-Baseline-IIS>. The nnU-Net baseline algorithm from Isensee et al. can be found here: <https://github.com/MIC-DKFZ/nnUNet>. Both trained algorithms can also be used on the grand challenge platform:

1. <https://grand-challenge.org/algorithms/spider-baseline-iis/>
2. <https://grand-challenge.org/algorithms/spider-baseline-nnu-net/>

Acknowledgments

This study was funded by Radboud AI for Health (ICAI)

Author contributions

Jasper W. van der Graaf created the segmentation dataset, developed and trained the presented segmentation algorithms, and wrote the manuscript. Miranda L. van Hooff,

Marinus de Kleuver, and Bram van Ginneken provided oversight for the project and revised the manuscript. Constantinus F. M. Buckens assisted with the MRI segmentation. Matthieu Rutten, Job van Susante, and Robert Jan Kroeze provided MRI data from their respective hospitals and revised the manuscript. Nikolas Lessmann helped with the development and training of the presented segmentation algorithms, provided overall project management, and revised the manuscript.

Competing interests

Nikolas Lessmann is an employee at Stryker. Bram van Ginneken is CSO at Thirona.

References

- [1] Damian Hoy, Lyn March, Peter Brooks, Fiona Blyth, Anthony Woolf, Christopher Bain, Gail Williams, Emma Smith, Theo Vos, Jan Barendregt, Chris Murray, Roy Burstein, and Rachelle Buchbinder. The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases*, 73(6):968–974, June 2014.
- [2] Elin Kjelle, Eivind Richter Andersen, Arne Magnus Krokeide, Lesley J. J. Soril, Leti van Bodegom-Vos, Fiona M. Clement, and Bjørn Morten Hofmann. Characterizing and quantifying low-value diagnostic imaging internationally: A scoping review. *BMC Medical Imaging*, 22(1):73, December 2022.
- [3] Nathan M. Wnuk, Tarik K. Alkasab, and Daniel I. Rosenthal. Magnetic resonance imaging of the lumbar spine: Determining clinical impact and potential harm from overuse. *The Spine Journal*, 18(9):1653–1658, September 2018.
- [4] Maximilian T. Löffler, Anjany Sekuboyina, Alina Jacob, Anna-Lena Grau, Andreas Scharr, Malek El Hussein, Mareike Kallweit, Claus Zimmer, Thomas Baum, and Jan S. Kirschke. A Vertebral Segmentation Dataset with Fracture Grading. *Radiology: Artificial Intelligence*, 2(4):e190138, 2020.
- [5] Yang Deng, Ce Wang, Yuan Hui, Qian Li, Jun Li, Shiwei Luo, Mengke Sun, Quan Quan, Shuxin Yang, You Hao, Pengbo Liu, Honghu Xiao, Chungpeng Zhao, Xinbao Wu, and S. Kevin Zhou. CTSpine1K: A Large-Scale Dataset for Spinal Vertebrae Segmentation in Computed Tomography, July 2021.
- [6] Dženan Zukić, Aleš Vlasák, Jan Egger, Daniel Hořínek, Christopher Nimsy, and Andreas Kolb. Robust Detection and Segmentation for Diagnosis of Vertebral Diseases Using Routine MR Images: Robust Detection and Segmentation for Diagnosis of Vertebral Diseases Using Routine MR Images. *Computer Graphics Forum*, 33(6):190–204, September 2014.

- [7] Chengwen Chu, Daniel L. Belavý, Gabriele Armbrrecht, Martin Bansmann, Dieter Felsenberg, and Guoyan Zheng. Fully Automatic Localization and Segmentation of 3D Vertebral Bodies from CT/MR Images via a Learning-Based Method. *PLOS ONE*, 10(11):e0143327, November 2015.
- [8] Jiawei Huang, Haotian Shen, Jialong Wu, Xiaojian Hu, Zhiwei Zhu, Xiaoqiang Lv, Yong Liu, and Yue 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. *The Spine Journal*, 20(4):590–599, April 2020.
- [9] Abhinav Suri, Brandon C. Jones, Grace Ng, Nancy Anabaraonye, Patrick Beyrer, Albi Domi, Grace Choi, Sisi Tang, Ashley Terry, Thomas Lechner, Iman Fathali, Nikita Bastin, Helene Chesnais, and Chamith S. Rajapakse. A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. *Bone*, 149:115972, August 2021.
- [10] Wuke Peng, Liang Li, Libin Liang, Hui Ding, Lei Zang, Shuo Yuan, and Guangzhi Wang. A convenient and stable vertebrae instance segmentation method for transforaminal endoscopic surgery planning. *International Journal of Computer Assisted Radiology and Surgery*, June 2021.
- [11] Anjany Sekuboyina, Jan Kukačka, Jan S. Kirschke, Bjoern H. Menze, and Alexander Valentinitich. Attention-Driven Deep Learning for Pathological Spine Segmentation. In Ben Glocker, Jianhua Yao, Tomaž Vrtovec, Alejandro Frangi, and Guoyan Zheng, editors, *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, volume 10734, pages 108–119. Springer International Publishing, Cham, 2018.
- [12] Naoto Masuzawa, Yoshiro Kitamura, Keigo Nakamura, Satoshi Iizuka, and Edgar Simo-Serra. Automatic segmentation, localization, and identification of vertebrae in 3D CT images using cascaded convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 681–690. Springer, 2020.
- [13] Anjany Sekuboyina, Alexander Valentinitich, Jan S. Kirschke, and Bjoern H. Menze. A Localisation-Segmentation Approach for Multi-label Annotation of Lumbar Vertebrae using Deep Nets. *arXiv:1703.04347 [cs]*, March 2017.
- [14] Nikolas Lessmann, Bram van Ginneken, Pim A de Jong, and Ivana Išgum. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical Image Analysis*, page 14, 2019.
- [15] Hao Chen, Qi Dou, Xi Wang, Jing Qin, Jack C. Y. Cheng, and Pheng-Ann Heng. 3D Fully Convolutional Networks for Intervertebral Disc Localization and Segmentation. In Guoyan Zheng, Hongen Liao, Pierre Jannin, Philippe Cattin, and Su-Lin Lee, editors, *Medical Imaging and Augmented Reality*, volume 9805, pages 375–382. Springer International Publishing, Cham, 2016.
- [16] Xing Ji, Guoyan Zheng, Daniel Belavy, and Dong Ni. Automated Intervertebral Disc Segmentation Using Deep Convolutional Neural Networks. In Jianhua Yao, Tomaž Vrtovec, Guoyan Zheng, Alejandro Frangi, Ben Glocker, and Shuo Li, editors, *Computational Methods and Clinical Applications for Spine Imaging*, volume 10182, pages 38–48. Springer International Publishing, Cham, 2016.
- [17] Robert Korez, Bulat Ibragimov, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec. Intervertebral disc segmentation in MR images with 3D convolutional networks. In Martin A. Styner and Elsa D. Angelini, editors, *SPIE Medical Imaging*, page 1013306, Orlando, Florida, United States, February 2017.
- [18] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, and Milan Sonka. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.
- [19] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- [20] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A. Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M. Summers, Bram Van Ginneken, Michel Bilello, Patrick Bilic, Patrick F. Christ, Richard K. G. Do, Marc J. Gollub, Stephan H. Heckers, Henkjan Huisman, William R. Jarnagin, Maureen K. McHugo, Sandy Napel, Jennifer S. Golia Pernicka, Kawal Rhode, Catalina Tobon-Gomez, Eugene Vorontsov, James A. Meakin, Sebastien Ourselin, Manuel Wiesenfarth, Pablo Arbeláez, Byeonguk Bae, Sihong Chen, Laura Daza, Jianjiang Feng, Baochun He, Fabian Isensee, Yuanfeng Ji, Fucang Jia, Ildoo Kim, Klaus Maier-Hein, Dorit Merhof, Akshay Pai, Beomhee Park, Mathias Perslev, Ramin Rezaiifar, Oliver Rippel, Ignacio Sarasua, Wei Shen, Jaemin Son, Christian Wachinger, Liansheng Wang, Yan Wang, Yingda Xia, Daguang Xu, Zhanwei Xu, Yefeng Zheng, Amber L. Simpson, Lena Maier-Hein, and M. Jorge Cardoso. The Medical Segmentation Decathlon. *Nature Communications*, 13(1):4128, July 2022.
- [21] Shumao Pang, Chunlan Pang, Lei Zhao, Yangfan Chen, Zhihai Su, Yujia Zhou, Meiyang Huang, Wei Yang, Hai Lu, and Qianjin Feng. SpineParseNet: Spine Parsing

- for Volumetric MR Image by a Two-Stage Segmentation Framework With Semantic Image Representation. *IEEE Transactions on Medical Imaging*, 40(1):262–273, January 2021.
- [22] Guoyan Zheng, Chengwen Chu, Daniel L. Belavý, Bulat Ibragimov, Robert Korez, Tomaz Vrtovec, Hugo Hutt, Richard Everson, Judith Meakin, Isabel López Andrade, Ben Glocker, Hao Chen, Qi Dou, Pheng-Ann Heng, Chunliang Wang, Daniel Forsberg, Aleš Neubert, Jurgen Fripp, Martin Urschler, Darko Stern, Maria Wimmer, Alexey A. Novikov, Hui Cheng, Gabriele Ambrecht, Dieter Felsenberg, and Shuo Li. Evaluation and comparison of 3D intervertebral disc localization and segmentation methods for 3D T2 MR data: A grand challenge. *Medical Image Analysis*, 35:327–344, January 2017.
- [23] Anjany Sekuboyina, Malek E. Hussein, Amirhossein Bayat, Maximilian Löffler, Hans Liebl, Hongwei Li, Giles Tetteh, Jan Kukačka, Christian Payer, Darko Štern, Martin Urschler, Maodong Chen, Dalong Cheng, Nikolas Lessmann, Yujin Hu, Tianfu Wang, Dong Yang, Daguang Xu, Felix Ambellan, Tamaz Amiranashvili, Moritz Ehlke, Hans Lamecker, Sebastian Lehnert, Marilia Lirio, Nicolás Pérez de Olaguer, Heiko Ramm, Manish Sahu, Alexander Tack, Stefan Zachow, Tao Jiang, Xinjun Ma, Christoph Angerman, Xin Wang, Kevin Brown, Alexandre Kirszenberg, Élodie Puybureau, Di Chen, Yiwei Bai, Brandon H. Rapazzo, Timyoas Yeah, Amber Zhang, Shangliang Xu, Feng Hou, Zhiqiang He, Chan Zeng, Zheng Xiangshang, Xu Liming, Tucker J. Netherton, Raymond P. Mumme, Laurence E. Court, Zixun Huang, Chenhang He, Li-Wen Wang, Sai Ho Ling, Lê Duy Huynh, Nicolas Boutry, Roman Jakubicek, Jiri Chmelik, Supriti Mulay, Mohanasankar Sivaprakasam, Johannes C. Paetzold, Suprosanna Shit, Ivan Ezhov, Benedikt Wiestler, Ben Glocker, Alexander Valentinitich, Markus Rempfler, Björn H. Menze, and Jan S. Kirschke. VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images. *Medical Image Analysis*, 73:102166, October 2021.