ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

Segmentation and Analysis of Lumbar Spine MRI Scans for Vertebral Body Measurements

Helen Schneider[1†*], David Biesner[1†*], Akash Ashokan[1], Maximilian Broß[1], Rebecca Kador[1], Sandra Halscheidt[1], Gábor Bagyó[2], Peter Dankerl[2], Haissam Ragab[3], Jin Yamamura[2], Christoph Labisch[2], Rafet Sifa[1]

1- Fraunhofer IAIS
Sankt Augustin

2- evidia GmbH
Dortmund

3- Diagnostic and Interventional Radiology
University Medical Center Hamburg-Eppendorf

[†] first authors, equal contribution

Abstract.
This paper investigates a data- and knowledge-driven approach to automatically analyze lumbar MRI scans. The dataset used is an in-house dataset of 142 sagital lumbar spine images from German radiology practices of the evidia GmbH. We implement state-of-the-art deep learning methods to segment the individual vertebral bodies. Overall, a very accurate segmentation performance of 97% Dice Score was achieved. Based on this segmentation, pathologically relevant distances are calculated using rule-based computer vision methods. We focus on the anterior, posterior and middle height of a vertebra and the anterior and posterior distances between two lumbar vertebrae. We demonstrate the clinical value of this approach through a quantitative and qualitative result analysis.

1 Introduction

Medical imaging plays a crucial role in the diagnosis of spinal abnormalities, injuries, and other causes of back pain and mobility impairment. Lumbar MRI (Magnetic Resonance Imaging) scans are a widely used imaging tool to analyze diseases and abnormalities in the lower back [1]. The interpretation of these scans can be a time-consuming task that requires expert knowledge by radiologists and orthopedists. Therefore, the increasing numbers of required medical imaging pose a challenge for clinics and private practices. In addition, a lack of universally accepted imaging-based grading system or diagnostic criteria for lumbar spine MRI interpretation leads to significant inter-reader variability even among specialists which can degrade the perceived value of their reporting [2]. AI (artificial intelligence) and CV (computer vision) tools can support radiologists with these challenges. Physicians can be aided in their diagnosis by AI

*This research has been funded by the Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the Lamarr-Institute for Machine Learning and Artificial Intelligence, LAMARR22B.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

tools to optimize the workflow without compromising the quality of analysis or quality of patient care [1, 3, 4].

Measuring the sizes of and distances between lumbar vertebrae is one of the most important aspects of the MRI interpretation to define pathological abnormalities. For example, the measurements are needed to instance detect vertebral fractures or the ever-increasing degenerative osteochondrosis of the lumbar spine. Especially in the western world, osteochondrosis plays a major role in driving up morbidity and healthcare costs. In addition, other common pathologies like hyperlordosis, scoliose, spondylolisthesis can be detected by this procedure. Radiologists measure relevant distances manually by eye or by using visualization software [5]. However, abnormalities can easily be missed by this approach if the relevant scans are not noted by the medical staff. This manual approach contributes, for example, to the high number (66%-85%) of underreporting of vertebral fractures among radiologists [5]. Therefore, we investigate the potential of deep learning-enhanced computer vision algorithms for the automatic measurement of distances in lumbar MRI scans based on vertebrae segmentation masks.

In [6] and [7] the authors investigate a similar approach to segment spinal MRI scans with subsequent measurement of diameters in spinal vertebrae with the goal of detection of spinal fractures. In contrast to their work, we apply a combined 3D and 2D model, which allows for analysis of higher-resolution scans, segment a larger number of distinct vertebrae classes and provide additional measurements for each vertebra, which allows for the detection of various further pathologies. Semi-automatic approaches to spinal measurement have also been investigated in [8], in which the segmentation is conducted manually by the physicians. In addition, there are public data sets with associated algorithmic solutions available for the segmentation of vertebra bodies, e.g. [7] and [9]. However, in contrast to the data set used by us, these are not recordings of a European patient cohort.

In summary in this work we

- evaluate the use of deep learning image segmentation models for the detection of spinal vertebrae in an in-house volumetric MRI dataset of German radiology practices
- develop algorithms for reliable distance measurements from segmented MRI images based on expert knowledge, and
- demonstrate the clinical impact of the developed algorithm on the basis of various sample examples.

2 Methods

In this section we present the algorithmic approach to the segmentation and measurement of the spinal vertebrae. We implement a segmentation $f : I \mapsto \hat{Y}$ with a segmentation prediction $\hat{Y} \in \mathbb{R}^{C,D,H,W}$, for a given MRI scan $I \in \mathbb{R}^{D,H,W}$,

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

where $C, D, H, W \in \mathbb{N}$ are the number of vertebrae classes, the depth, height and width of the processed scan I, respectively.

For training and evaluation of the segmentation algorithm, we consider two approaches. First, we train a multiclass model M_{MC} for the segmentation of the images into $C = 10$ distinct classes. Second, we combine all vertebrae classes into a single positive class and train a binary classification model M_{SC}. We evaluate the binary model on the binary segmentation task and the multiclass model both on the multiclass segmentation task and the binary segmentation task. For the latter, we map each vertebra class of the prediction onto a single positive class.

The final distance measurements are rule-based algorithms on the basis of the predicted segmentation $\hat{Y}$.

Method	Dice Score	Jaccard Index
M_{SC}	97.4	94.4
M_{MC}	96.6	93.4
$\mathrm{M}_{\mathrm{MCBM}}$	97.3	94.6

Table 1: Dice Score and Jaccard Index in % on validation set.

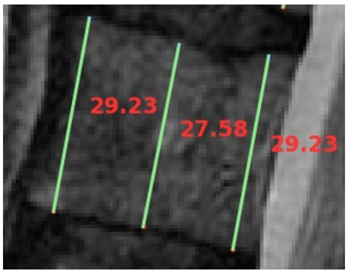


Fig. 1: Measured anterior, posterior and middle height of a lumbar vertebra (distances in mm).

2.1 Segmentation Algorithms

The segmentation approach of our work is based on [9], in which the authors apply a combined architecture of a 3D segmentation model and a 2D segmentation model for analysis of spinal MRI images. A down-scaled 3D scan is passed through a convolutional neural net (CNN) architecture called DeepLabV3 with a graph convolutional neural net (GCN) component. The GCN provides additional information on the relationship between label classes, such as the T11 and T12 thoracic vertebrae being closer to each other than the thoracic and sacral vertebrae. The high-resolution MRI scan is passed slice by slice to a 2D Residual UNet, along with the low-resolution segmentation slice of the 3D model, in order to calculate a final high-resolution segmentation of the vertebrae.

2.2 Algorithmic Distance Measurements

In consultation with the medical experts of the evidia GmbH the distance measurements of the vertebrae are based on six different anchor points per vertebra. To define these, the smallest bounding trapeze for each vertebra is determined based on the generated segmentation. The vertebra rotation is approximated as the angle of the bounding trapeze to the x-axis.

We measure the anterior and posterior height of a vertebra along two lines, perpendicular to the shorter trapeze side (top or bottom). These lines start in the corresponding vertices of the trapeze, shifted 10% inwards. We define the heights as the distance between the farthest segmented pixels (anchor points) along the lines. The middle height is calculated similarly along the parallel line centered between anterior and posterior height. We calculate the anterior and posterior distance between the lumbar vertebrae using the defined anchor points.

3 Data and Results

3.1 Data and Annotation

The data set used is an in-house data set provided by the evidia GmbH. A total of 142 volumetric T2-weighted sagittal lumbar spine MRI scans are available. Of these, 120 images are used for training and 22 for validation. Figure 2 shows exemplary slices of the 3D volume of different patients. Under the supervision of a radiology resident, two medical research assistants segmented the vertebrae of the scans. We consider 9 distinct vertebra in total (thoracic T11-T12, lumbar L1-L5 and sacral S1-S2) for a total of 10 labels for the segmentation task (including the negative class label).

In order to improve generalization of the models we apply random rotation (up to 5 degrees), random translation (up to 5%) and random scaling (up to 5%) to the training data. Additionally we crop each image horizontally by 25% left and right to remove empty space around the spine.

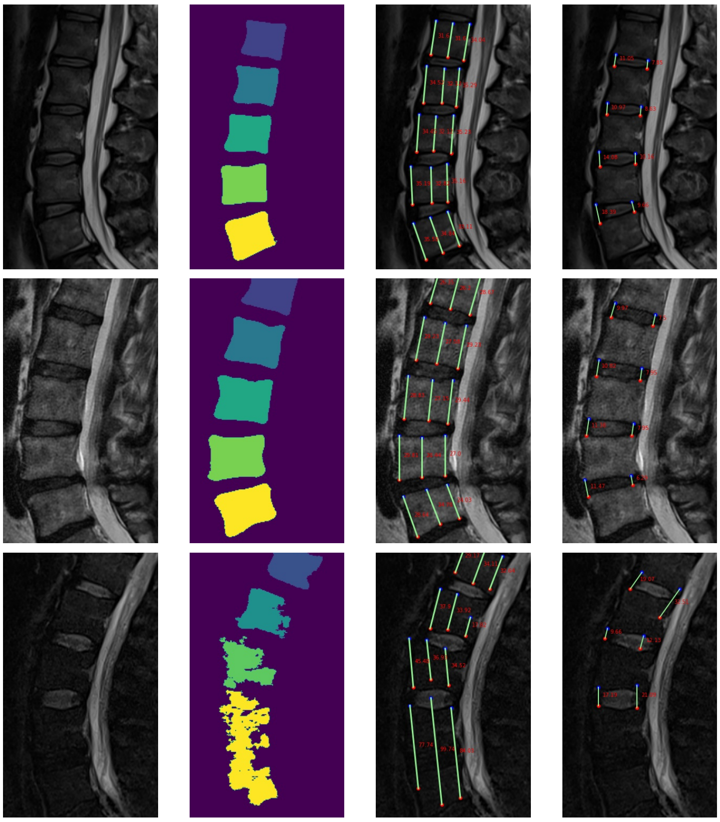
3.2 Quantitative and Qualitative Results

Table 1 describes the final evaluation scores on the validation dataset regarding the micro-averaged Dice Score and Jaccard Index. It can be seen that both the multiclass model $\mathrm{M_{MC}}$ and the binary model $\mathrm{M_{SC}}$ provide good segmentation results. Overall, the binary model $\mathrm{M_{SC}}$ achieves a slightly increased segmentation performance, up to 1% improvement compared to the multiclass evaluation of the $\mathrm{M_{MC}}$ model. However, if we compare both models with respect to the binary evaluation, we can see that similarly strong performance of apporximatly 97% Dice Score is achieved, despite the more complex training of $\mathrm{M_{MC}}$.

We observe that differentiating the distinct vertebrae only comes with a small loss of segmentation performance, which is why we focus the following evaluations on the multiclass segmentation model $\mathrm{M_{MC}}$.

Figure 2 illustrates the generated distance measurements for three different patients with the corresponding segmentation. We show a single 2D slice for each patient. We see in the upper two samples that with suitable segmentation the relevant distances for the lumbar vertebrae can be calculated well with the introduced rule-based methods. This concerns both the different height measurements of a vertebral body and the distances between the lumbar vertebrae. In addition, the presented segmentations underlines the good Dice Scores discussed in 1. The lumbar vertebrae L1-L5 can be separated and their shapes

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.



(a) (b) (c) (d)

Fig. 2: Original scan (a), predicted segmentation (b), measurements of vertebrae heights (c) and distances between vertebrae (d). The segmentation of the first two scans is predicted cleanly which results in very accurate distance measurements. The third scan is less clear and more difficult to segment, which results in inaccurate distance measurements.

are well recognized. The third example illustrates that if the image quality is too low, the segmentation algorithm is no longer able to recognize and separate the individual vertebra. If the vertebral contours are not segmented correctly, the rule-based distance measurements can no longer calculate the height of the individual vertebral bodies and the distances between them. We show a more detailed example of the calculated measurements for a single vertebra in Figure 1. Overall, the examples presented indicate that the introduced approach is suitable to automatically calculate the relevant distances of the lumbar vertebral bodies in MRI scans.

ESANN 2023 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 4-6 October 2023, i6doc.com publ., ISBN 978-2-87587-088-9. Available from http://www.i6doc.com/en/.

4 Conclusion and Outlook

This paper evaluated the feasibility of a data- and knowledge-driven lumbar spine MRI analysis approach. A very accurate volumetric segmentation of the vertebrae with up to a Dice Score of 97% is achieved. Based on this segmentation, rule-based distance measurements were implemented from expert knowledge, and their high clinical relevance was discussed in a qualitative analysis. We plan to use the developed approach to implement patient cohort comparisons in the future. Additionally, the algorithm will be extended to the analysis of intervertebral discs.

References

[1] Desmond Shi Wei Lim and Andrew Makmur et al. Improved Productivity Using Deep Learning-assisted Reporting for Lumbar Spine MRI. *Radiology*, 305(1):160–166, 2022.

[2] Michael C Fu et al. Interrater and intrarater agreements of magnetic resonance imaging findings in the lumbar spine: significant variability across degenerative conditions. *The Spine Journal*, 14(10):2442–2448, 2014.

[3] Helen Schneider et al. Towards Symmetry-Aware Pneumonia Detection on Chest X-Rays. In *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 543–550. IEEE, 12 2022.

[4] Helen Schneider and David Biesner et al. Improving Intensive Care Chest X-Ray Classification by Transfer Learning and Automatic Label Generation. In *European Symposium on Artificial Neural Networks (ESANN)*, pages 457–462, 01 2022.

[5] Abhinav Suri et al. Vertebral Deformity Measurements at MRI, CT, and Radiography Using Deep Learning. *Radiology: Artificial intelligencee*, 4(1):e210015, 2022.

[6] Christoph Germann et al. Performance of a deep convolutional neural network for MRI-based vertebral body measurements and insufficiency fracture detection. *European Radiology*, 33, 12 2022.

[7] Rao Farhat Masood et al. Deep Learning based Vertebral Body Segmentation with Extraction of Spinal Measurements and Disorder Disease Classification. *Biomedical Signal Processing and Control*, 71:103230, 2022.

[8] Natalia Friska et al. A semi-automated method to measure the height of lumbar intervertebral discs. In *2022 2nd International Conference on Computers and Automation (CompAuto)*, pages 32–35. IEEE, 2022.

[9] Shumao Pang et al. 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, 2021.