ORIGINAL ARTICLE

Deep learning-based high-accuracy detection for lumbar and cervical degenerative disease on T2-weighted MR images

Wei Yi[1] · Jingwei Zhao[2] · Wen Tang[3] · Hongkun Yin[3] · Lifeng Yu[4] · Yaohui Wang[5] · Wei Tian[6]

Abstract

Purpose To develop and validate a deep learning (DL) model for detecting lumbar degenerative disease in both sagittal and axial views of T2-weighted MRI and evaluate its generalized performance in detecting cervical degenerative disease.

Methods T2-weighted MRI scans of 804 patients with symptoms of lumbar degenerative disease were retrospectively collected from three hospitals. The training dataset ($n=456$) and internal validation dataset ($n=134$) were randomly selected from the center I. Two external validation datasets comprising 100 and 114 patients were from center II and center III, respectively. A DL model based on 3D ResNet18 and transformer architecture was proposed to detect lumbar degenerative disease. In addition, a cervical MR image dataset comprising 200 patients from an independent hospital was used to evaluate the generalized performance of the DL model. The diagnostic performance was assessed by the free-response receiver operating characteristic (fROC) curve and precision–recall (PR) curve. Precision, recall, and F1-score were used to measure the DL model.

Results A total of 2497 three-dimension retrogression annotations were labeled for training ($n=1157$) and multicenter validation ($n=1340$). The DL model showed excellent detection efficiency in the internal validation dataset, with F1-score achieving 0.971 and 0.903 on the sagittal and axial MR images, respectively. Good performance was also observed in the external validation dataset I (F1-score, 0.768 on sagittal MR images and 0.837 on axial MR images) and external validation dataset II (F1-score, 0.787 on sagittal MR images and 0.770 on axial MR images). Furthermore, the robustness of the DL model was demonstrated via transfer learning and generalized performance evaluation on the external cervical dataset, with the F1-score yielding 0.931 and 0.919 on the sagittal and axial MR images, respectively.

Conclusion The proposed DL model can automatically detect lumbar and cervical degenerative disease on T2-weighted MR images with good performance, robustness, and feasibility in clinical practice.

Keywords Deep learning · Degenerative disc disease · Magnet resonance imaging · Spine

✉ Wei Yi
dr_weiyi@163.com

✉ Jingwei Zhao
kivist@163.com

Wen Tang
twen@infervision.com

Hongkun Yin
yhongkun@infervision.com

Lifeng Yu
57872016@qq.com

Yaohui Wang
13401041530@139.com

Wei Tian
realonly099@sina.com

[1] Department of Spine Surgery, Beijing Jishuitan Hospital, Beijing 100035, China

[2] Beijing Jishuitan Hospital, Research Unit of Intelligent Orthopedics, Chinese Academy of Medical Sciences, Beijing, China

[3] Institute of Advanced Research, Infervision Medical Technology Co., Ltd, Beijing, Beijing, China

[4] The Second Hospital of Zhangjiakou City, Zhangjiakou, Hebei Province, China

[5] Department of Trauma, Beijing Water Conservancy Hospital, Beijing 10036, China

[6] Beijing Jishuitan Hospital, Research Unit of Intelligent Orthopedics, Chinese Academy of Medical Sciences, Beijing, China

Introduction

Intervertebral disc degeneration was first described by Dexler in 1896 [1]. And now, it is a worldwide health problem related to enormous medical and social costs [2]. This degenerative progression occurs most commonly in the cervical and lumbar intervertebral discs. And in clinical practice, magnetic resonance imaging (MRI) is the best noninvasive assessment for investigating degenerative discs. However, there are several pitfalls in MR imaging. Some benign lesions and normal variations could be confused with more severe pathologies in clinics. Intervertebral disc prolapse sometimes mimics infective spondylitis or vertebral osteophytes. A sequestrated disc fragment could easily be mistaken for neurogenic tumors, epidural synovial cysts, epidural hematomas, and conjoined nerve roots [3].

In the last decade, there has been a massive increase in the use of artificial intelligence (AI) and machine learning (ML) technologies for auxiliary diagnosis clinically. It has excellent image recognition ability for the intervertebral disc. And it can improve diagnostic accuracy, mark critical information and reduce human error.

Previous reports show that deep learning approaches can achieve good performance in identifying, diagnosing, and grading lumbar degenerative diseases. The accuracy of some classification methods is comparable to that of radiologists [4–7]. However, most of these studies only use limited data from a single center. Thus, the robustness and generalization ability of the algorithms could be unreliable. In addition, many of those still require manual segmentations as input, which is expensive, time-consuming, and may also cause subjective bias.

In this multicenter study, we proposed a deep learning model for detecting degenerative discs in both sagittal and axial views of T2-weighted MRI. As an exploration of the application of deep learning for detecting lumbar degenerative disease, this work also assesses the robustness and generalization ability in detecting cervical degenerative disease on an independent dataset via a transfer learning approach.

Materials and methods

Patient enrollment

Patients who underwent MR scans from Jan 2019 to Jan 2020 in Center I hospital, July 2021 to Oct 2021 in center II hospital, and May 2021 to July 2021 in center III hospital were retrospectively reviewed. The inclusion criterion was that patients had been diagnosed with lumbar intervertebral disc prolapse. Patients were excluded if (1) they had received any treatment (surgery or chemoradiation) before the MR scan; (2) they had been diagnosed with neurogenic tumors, epidural synovial cysts, epidural hematomas, or infective spondylitis; (3) MR images could not be obtained or interpreted. Ultimately, 590 patients from center I, 100 patients from center II, and 114 from center III were enrolled in our study. The patients from the center I were randomly divided into a training dataset ($n = 456$) and an internal validation dataset ($n = 34$). The patients from center II to center III were used as independent external validation datasets.

MR protocol

All patients who underwent MR examination in Center I Hospital were performed on one 3.0 T MR scanner (Magnetom Vida, Siemens Healthcare GmbH, Germany) using the original standard spine coil (Spine 32, Siemens Healthcare GmbH, Germany). An institutional protocol was used for imaging consisting of the following three standard sequences: Sagittal T1-weighted (TR/TE 360/8.4 ms, the field of view (FOV) 280×280 mm, slice thickness and increment (SL) 4 mm, acceleration mode (AM) GRAPPA, acceleration factor (AF) 2, number of excitations (NEX) 2, acquisition time (AT) 112 s); sagittal T2-weighted (TR/TE 3000/96 ms, FOV 280×280 mm, SL 4 mm, AM None, NEX 1, AT 84 s); Axial T2-weighted (TR/TE 3000/91 ms, FOV 180×180 mm, SL 3.5 mm, AM GRAPPA, AF 2, NEX 3, AT 86 s).

The scanning parameters of MR for the patients from different hospitals are listed in Table 1.

Table 1 Summary of scanning parameters

Hospital	Center II hospital	Center III hospital
MR scanner	1.5 T Philips, Philips Medical Suzhou Co., Ltd, China	1.5 T Siemens Healthcare GmbH Henkestr. 127 91,052 Erlangen Germany
Sagittal T1	TR/TE 400/15.0 ms; Thickness 4 mm	TR/TE 580/9.9 ms; Thickness 4 mm
Sagittal T2	TR/TE 3000/100 ms; Thickness 4 mm	TR/TE 3440/121 ms; Thickness 4 mm
Axial T2	TR/TE 4338/120 ms; Thickness 4 mm	TR/TE 3920/122 ms; Thickness 4 mm

Pre-processing of MR images

The axial and sagittal views of MR images were collected from the picture archiving and communication system (PACS) and stored in a workstation as digital imaging and communications in medicine (DICOM) metadata for annotation and further analysis. The low-frequency intensity non-uniformity present in MRI images was corrected with the N4 bias field correction algorithm [8]. The lumbar degenerative discs were manually labeled as ground truth by two radiologists with more than ten years of experience in consensus. When there were disputes between the two radiologists, a senior expert with more than 20 years of experience in musculoskeletal imaging was consulted for the final decision. The pixel interpolation approach was used to transform the original sagittal-view and axial-view MR images to the voxel size of 12*256*256 and 24*256*256, respectively. 8630 annotations of 2497 lumbar degenerative discs from 804 patients were labeled as ground truth. To enhance the training and robustness of the convolutional neural network, sophisticated data augmentation techniques, including random flipping, random rotation, perturbations to contrast, Gaussian random noise, and nonlinear brightness transformation, were also performed [9].

Development of the deep learning model

A deep learning (DL) model was proposed for lumbar degenerative disc detection. The schematic illustration of the DL model is shown in Fig. 1. The DL model was based on the U-Net encoder-decoder architecture. Two modified 3D Resnet18 networks without parameter sharing were applied separately as encoders for the two different modalities (sagittal-view MR image and axial-view MR image). We modified the original Resnet18 network into a 3D version according to our data characteristics, and the channels between each layer were reduced to 16-32-64-128-192 due to the parameter redundancy in 3D networks. Transformer-based multi-modality cross attention was also applied to enhance the interaction of two MR modalities and better investigate multi-modal paired attention. The head number of multi-head attention was set to eight [10].

The proposed DL model was trained based on the unified loss function generalizing both dice loss and focal loss. The weights of hidden layers were randomly initialized, and the batch size and initial learning rate were set to 4 and 0.0001, respectively. Adam W, an algorithm that modifies the typical implementation of weight decay in Adam, was used as the optimizer in the training stage owing to its fast convergence and improved implementation of weight decay [11]. The training was terminated when the loss in the validation set stopped decreasing, and the max number of epochs was set to 300.

The supervised training process of the DL model was performed on a computer with a Core i7-6700 K 4.00-GHz central processing unit (Intel, Santa Clara, Calif), 32 GB memory, and a GeForce GTX 2070 graphics processing unit (NVIDIA, Santa Clara, Calif). The DL model was developed
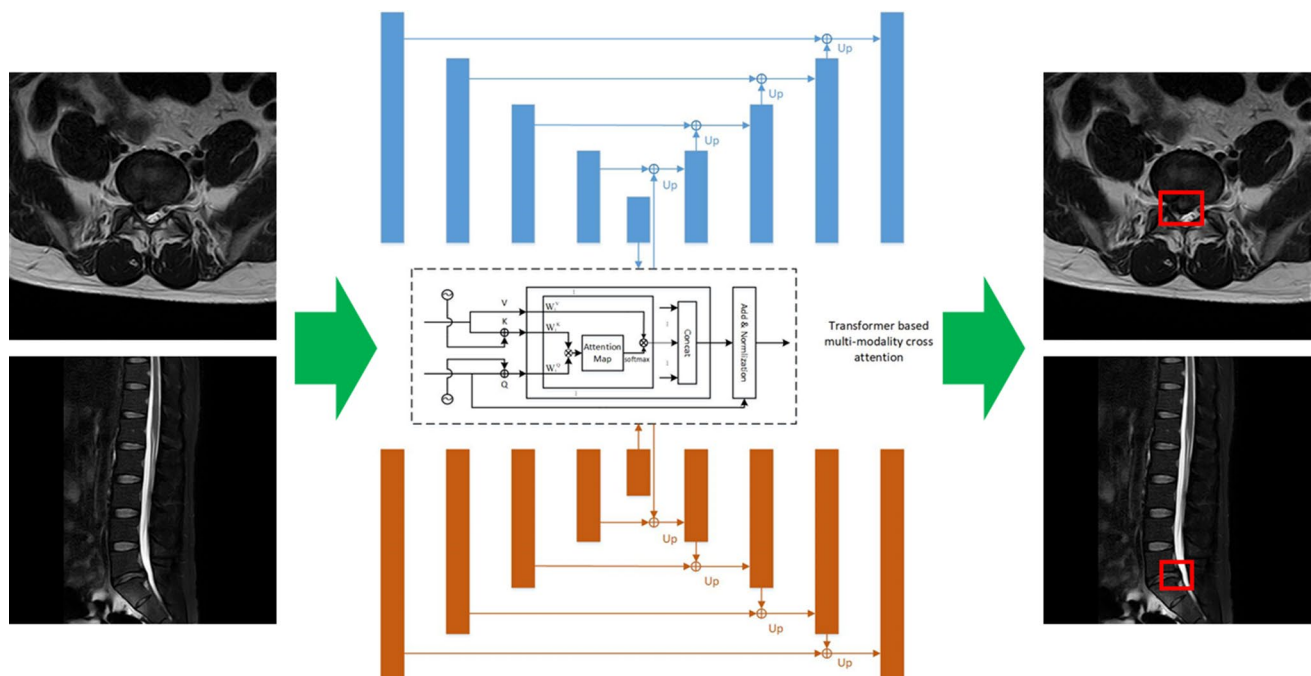

Fig. 1 Schematic illustration and detection flowchart of the 3D ResNet18 and transformer architecture of the proposed deep learning model

and validated with InferScholar platform version 3.5. Python 3.6.8 (https://www.python.org/) and the framework for neural networks of MXNet 1.5.0 (https://mxnet.incubator. apache.org/) were applied for model construction.

Transfer learning for the detection of cervical degenerative disease

To investigate the generalizability of our DL model in the detection of common vertebral lesions, we also applied a transfer learning framework to the detection of cervical degenerative disease. In detail, the convolutional layers of the DL model were frozen and transferred into the new model, while the fully connected layers (softmax layer) were retrained with randomly initialized parameters on the top of the transferred convolutional layers. Then, the newly initialized DL model took the original image bottlenecks as input and retrained to detect degenerative spine lesions. We collected an external cervical MR image dataset containing the sagittal-view and axial-view MR images of 200 patients with the cervical degenerative disease in Center I hospital from Oct 2021 to Jan 2022. The inclusion criterion was that patients had been diagnosed with cervical intervertebral disc prolapse. Patients were excluded if (1) they had received any treatment (surgery or chemoradiation) before the MR scan; (2) they had been diagnosed with neurogenic tumors, epidural synovial cysts, epidural hematomas, or infective spondylitis; (3) MR images could not be obtained or interpreted. The annotation and pre-processing of MR images were the same as those of the lumbar degenerative patients, which had been described before.

Model performance evaluation

To evaluate the capacity of the DL model for detecting the lumbar degenerative disc in the validation dataset, a precision-recall curve was plotted to show the precision-recall pairs for different probability thresholds. The precision is the ratio of $TP/(TP+FP)$, and the recall is the ratio of $TP/(TP+FN)$, where TP, FP, and FN are the numbers of true positives, false positives, and false negatives, respectively. A TP was defined as a correct detection of the lumbar degenerative disc. An FP was defined as a wrong prediction of the lumbar degenerative disc. An FN was defined as a missed detection of the lumbar degenerative disc. Since there were no true negatives in the lumbar degenerative disc detection task, the ROC curve and specificity were not applicable in this study. Therefore, the free-response receiver operating characteristic (fROC) curve was used to evaluate the comprehensive performance of the DL model. F1 score was also used to measure the weighted average of precision and recall of our DL model, which was defined as follows: $F1 = 2 \times Precision \times Recall/(Precision + Recall)$.

Statistical analysis

The Mann–Whitney U test was used to evaluate the differences in the numerical variables across different categories. Differences in the dichotomous variables were calculated with the Chi-squared test. A two-sided p value less than 0.05 was considered statistically significant. All analyses were performed using SPSS for Windows (version 26.0, IBM).

Results

Study population characteristics

The flowchart of patient enrollment is presented in Table 2, and a total of 268 males and 188 females were divided into the training dataset with an average age of 43.4 years (range 12–86 years). There were 67 males and 67 females in the internal validation dataset with an average age of 48.0 years (range 12–79 years), 51 males and 49 females in the external validation dataset I with an average age of 52.7 years (range 16–84 years), and 55 males and 59 females in the external validation dataset II with an average age of 59.5 years (range 17–88 years), respectively. The total number of labeled lumbar degenerative discs in the training dataset, internal validation dataset, external validation dataset I, and external validation dataset II was 1157, 395, 194, and 144, respectively. There were 607 labeled cervical degenerative regions from 128 males to 72 females with an average age of 53.9 years (range 23–82 years) in the external cervical MR image dataset.

Model performance for detecting lumbar degenerative disc

The DL model showed favorable performance in lumbar degenerative disc detection in the validation dataset, with the F1 score and the areas under the fROC curve achieving 0.971 (95% CI 0.951–0.987) and 0.968 (95% CI 0.943–0.994) for the sagittal-view MR images, and 0.903 (95% CI 0.870–0.935) and 0.896 (95% CI 0.870–0.932)

Table 2 Clinical characteristics of the enrolled patients

Dataset	Male	Female	Age (range)
Lumbar discs			
Training dataset	268	188	43.4 (12–86)
Internal validation dataset	67	67	48.0 (12–79)
External validation dataset I	51	49	52.7 (16–84)
External validation dataset II	55	59	59.5 (17–88)
Cervical discs			
External cervical dataset	128	72	53.9 (23–82)

for the axial-view MR images, respectively. The DL model also showed promising detective capability in the external validation dataset I, and the F1 score and areas under the fROC curve were 0.768 (95% CI 0.693–0.838) and 0.764 (95% CI 0.691–0.837) for the sagittal-view MR images as well as 0.837 (95% CI 0.762–0.895) and 0.808 (95% CI 0.636–0.880) for the axial-view MR images. Similar performance was observed in the external validation dataset II with the F1 score and areas under the fROC curve yielding 0.787 (95% CI 0.723–0.855) and 0.732 (95% CI 0.648–0.817) for the sagittal-view MR images, and 0.770 (95% CI 0.678–0.844) and 0.721 (95% CI 0.640–0.802) for the axial-view MR images, respectively. The detailed precision and recall of the DL model in these multicenter validation datasets are summarized in Table 3. The fROC curve is shown in Fig. 2, and the precision-recall curve is shown in Fig. 3.

Transfer learning performance on cervical degenerative disease

In the model robustness and generalized evaluation on the external cervical MR image dataset, the DL model had achieved an F1 score of 0.931 (95% CI 0.907–0.955), with a precision of 0.974 (95% CI 0.953–0.991) and a recall of 0.893 (95% CI 0.857–0.929) for the sagittal MR images, and an F1 score of 0.919 (95% CI 0.891–0.944), with a precision of 0.942 (95% CI 0.909–0.973) and a recall of 0.897 (95% CI 0.858–0.934) for the axial MR images, respectively. The areas under the fROC curve were 0.911 (95% CI 0.873–0.950) for the sagittal MR images and 0.882 (95% CI 0.839–0.925) for the axial MR images. These results demonstrated the generalization capability of our DL model, which had good overall accuracy for cervical degenerative disease detection. The fROC curve and the precision-recall curve of the DL model in the external cervical MR image dataset are shown in Fig. 4.

Discussion

The aging of society has resulted in a significant increase in spinal images over the past decades, for both the growing burden of spinal disease and the more popularized application of MR. Artificial intelligence (AI) and machine learning (ML) technologies are playing a more critical role

Table 3 Detailed performance of the deep learning model in the internal validation dataset, external validation dataset I, and external validation dataset II for the lumbar degenerative disc detection

| Dataset	Precision (95% CI)		Recall (95% CI)		F1-score (95% CI)	
	Sagittal MRI	Axial MRI	Sagittal MRI	Axial MRI	Sagittal MRI	Axial MRI
Internal validation	0.971 (0.943, 0.995)	0.948 (0.898, 0.983)	0.971 (0.947, 0.990)	0.862 (0.818, 0.909)	0.971 (0.951, 0.987)	0.903 (0.870, 0.935)
External validation I	0.785 (0.699, 0.869)	0.885 (0.809, 0.952)	0.753 (0.670, 0.837)	0.794 (0.710, 0.865)	0.768 (0.693, 0.838)	0.837 (0.762, 0.895)
External validation II	0.737 (0.655, 0.829)	0.670 (0.565, 0.758)	0.843 (0.761, 0.924)	0.905 (0.827, 0.972)	0.787 (0.723, 0.855)	0.770 (0.678, 0.844)
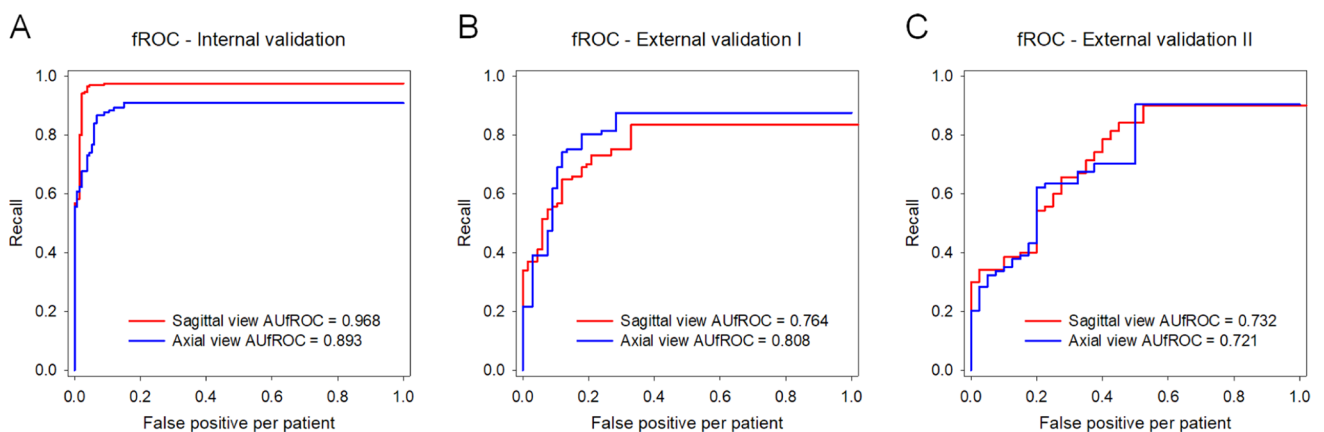
Fig. 2 Free-response receiver operating characteristic (fROC) curve analysis of the deep learning model in the internal validation dataset (**A**), external validation dataset I (**B**) and external validation dataset II (**C**)
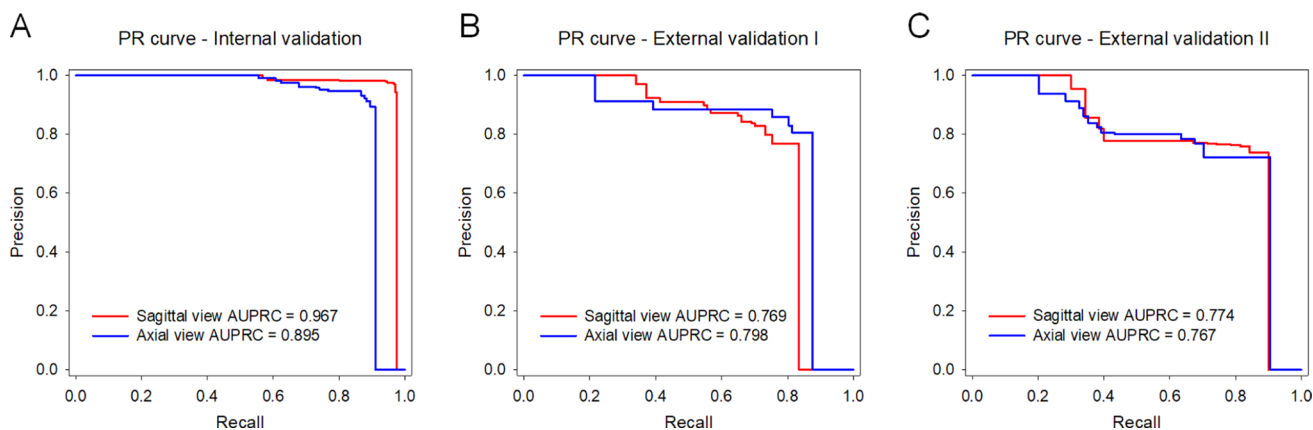
Fig. 3 Precision–recall (PR) curve analysis of the deep learning model in the internal validation dataset (**A**), external validation dataset I (**B**) and external validation dataset II (**C**)

Fig. 4 The free-response receiver operating characteristic (fROC) curve analysis (**A**) and precision-recall (PR) curve analysis (**B**) of the deep learning model in the external cervical MR image dataset
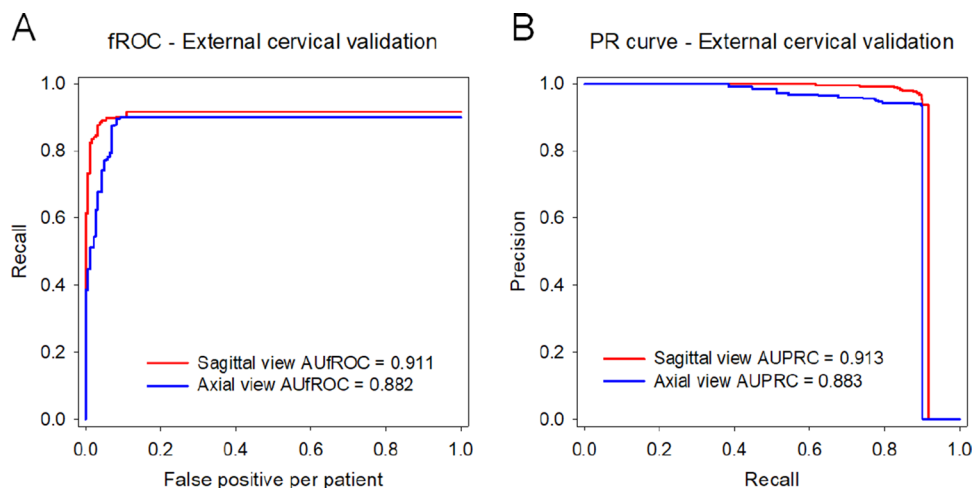
in the diagnosis of spinal disorders [12]. Researchers and engineers are working together to develop AI-assisted diagnostic systems to improve the accuracy of diagnosis and reduce the disease burden on society. In recent years, machine learning techniques have been applied in the diagnosis of spinal disorders, such as spinal degenerative disease, trauma, oncology, and deformity. Recent literature also shows the potential of deep learning-based approaches for reliable quantifications of the vertebrae and discs and for facilitating decision-making in the treatment of lumbar disc herniations. However, most of the ML studies are based on limited data in a single center. It is essential to develop optimized algorithms that allow AI programs have the ability to analyze all kinds of images precisely from different medical centers with various scanning parameters. Furthermore, it is also crucial for an ideal AI system to give both qualitative and quantitative descriptions of the lesions. For example, in the case of intervertebral disc prolapse, the AI system should not only tell surgeons the level of the prolapsed disc, but also exhibit the borderline

of the lesion and its relationship with other critical organs and tissues.

This study included a wide range of patients aged 12–86 years to add a relatively wide variety of disc morphology and hydration to the Deep Learning database. And it provided more reasonable results. The major findings of our study reveal that the deep learning algorithm has achieved good performance and is highly consistent with the radiologist's expert reading for detecting lumbar degenerative disease. Several previous researches have reported automated detection of intervertebral disc degeneration using deep learning approaches [13]. However, these studies only use limited data from a single center to validate the performance. As an important step for evaluating a deep learning-based model, external validation is necessary for assessing the model's robustness and generalization, which could avoid the overestimation of model performance caused by overfitting [14]. Various studies have demonstrated that the diagnostic performance of deep learning models can vary across different datasets [15, 16]. Our proposed model

showed high detective accuracy on both sagittal and axial views of T2-weighted MRI in the internal validation dataset. The model also had good robustness, achieving favorable performance in two external validation datasets.

Quantitative analysis provides valuable information and can help clinicians make better decisions, and some deep learning-based lumbar disc degeneration-related quantitative models have been developed in recent years [2, 17, 18]. However, the reliability and validity of quantitative segmentation and measurement of disease regions still need further validation in clinical practice. Moreover, the diseased area of the degenerative cervical spine is commonly irregular, with unclear boundaries in the MRI image, which makes the accurate annotation of the disease region difficult. Meanwhile, the quality of MR images is mainly dependent on the scanning parameters, which are usually changed across different institutions or even varied in each MR scanner in the same center [19]. Therefore, the subjective bias of assessment and segmentation of lesion areas in the MR images seems inevitable, even between experienced radiologists [20]. This, in turn will have a substantial impact on the training and validation of the segmentation neural networks. Therefore, in this case, we choose to use detection networks to solve this problem rather than the segmentation networks such as U-Net.

An important distinction of our work from previous studies is the use of transfer learning to detect cervical disc herniation. Cervical and lumbar disc herniations are both degenerative spinal diseases in their pathophysiological mechanism, and the degenerative discs in the cervical and lumbar regions show similar characteristics in MR imaging. For these reasons, our research further explored the possibility of transferring the learning technique generated from the lumbar dataset to evaluate cervical disks. Although it was only a preliminary study, the results were quite inspiring. As a highly effective technique, transfer learning is suitable for dealing with medical images, particularly when faced with limited data [21, 22]. By using the transfer learning technique, Kermany et al. [23] built a generalized platform that could both classify diabetic macular edema and age-related macular degeneration on optical coherence tomography images and diagnose pediatric pneumonia on chest X-ray images. In this study, we also investigated the effectiveness of transfer learning in detecting cervical degenerative disease on limited T2-weighted MR images from 200 patients. We found that our model retained high detective accuracy, thereby illustrating the generalization of our proposed deep learning algorithm and the power of transfer learning to achieve good performance even with a small dataset.

There are several limitations of our study. First, although a relatively larger cohort from several institutions was collected for model development and validation, we are still aware that the sample size is not big enough for a deep learning algorithm that contains millions of parameters. More patients from different hospitals should be collected to improve the accuracy of lumbar and cervical degenerative disease detection. Second, our study focused on the analysis of T2-weight MR images, which were most commonly used to investigate degenerative discs. However, other modalities such as radiographs, CT, and lumbar discography could also infer anatomic changes and help to exclude other diagnoses [24]. The fusion of multi-modality images such as CT and MR could be a promising area of future investigation. Third, although our model had achieved high accuracy in detecting degenerative discs, there were some technical difficulties in the accurate identification of the boundary of a herniated disc, especially the floor of the lesion in MRI. Further quantitative analysis, including characterization of different degeneration grades, would lead to more critical in clinical practice, which would also be our future work.

Conclusion

The proposed DL model can automatically detect lumbar and cervical degenerative disease on T2-weighted MR images with good performance, robustness, and feasibility in clinical practice.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Fenn J, Olby NJ, Canine Spinal Cord Injury Consortium (CANSORT-SCI) (2020) Classification of intervertebral disc disease. Front Vet Sci 6(7):579025. https://doi.org/10.3389/fvets.2020.579025

2. Huang J, Shen H, Wu J, Hu X, Zhu Z, Lv X, Liu Y, Wang Y (2020) Spine Explorer: a deep learning based fully automated

program for efficient and reliable quantifications of the vertebrae and discs on sagittal lumbar spine MR images. Spine J 20(4):590–599. https://doi.org/10.1016/j.spinee.2019.11.010

3. Shikhare SN, Singh DR, Peh WC (2014) Variants and pitfalls in MR imaging of the spine. Semin Musculoskelet Radiol 18(1):23–35. https://doi.org/10.1055/s-0034-1365832

4. Castro-Mateos I, Hua R, Pozo JM, Lazary A, Frangi AF (2016) Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images. Eur Spine J 25(9):2721–2727. https://doi.org/10.1007/s00586-016-4654-6

5. Beulah A, Sharmila TS, Pramod VK (2022) Degenerative disc disease diagnosis from lumbar MR images using hybrid features. Vis Comput 38:2771–2783. https://doi.org/10.1007/s00371-021-02154-x

6. Lehnen NC, Haase R, Faber J, Rüber T, Vatter H, Radbruch A, Schmeel FC (2021) Detection of degenerative changes on mr images of the lumbar spine with a convolutional neural network: a feasibility study. Diagnostics (Basel) 11(5):902. https://doi.org/10.3390/diagnostics11050902

7. Gao F, Liu S, Zhang X, Wang X, Zhang J (2021) Automated grading of lumbar disc degeneration using a push-pull regularization network based on MRI. J Magn Reson Imaging 53(3):799–806. https://doi.org/10.1002/jmri.27400

8. Beare R, Lowekamp B, Yaniv Z (2018) image segmentation, registration and characterization in r with simple ITK. J Stat Softw 86:8. https://doi.org/10.18637/jss.v086.i08

9. Zhang L, Wang X, Yang D et al (2020) Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE Trans Med Imaging 39(7):2531–2540

10. Zhuang X, Liu F, Hou J, Hao J, Cai X (2022) transformer-based interactive multi-modal attention network for video sentiment detection. Neural Process Lett 54:1943–1960. https://doi.org/10.1007/s11063-021-10713-5

11. Llugsi R, El Yacoubi S, Fontaine A, Lupera P (2021) Comparison between adam, adamax and adam w optimizers to implement a weather forecast based on neural networks for the Andean city of quito. IEEE Fifth Ecuad Tech Chapters Meet. https://doi.org/10.1109/ETCM53643.2021.9590681

12. Merali ZA, Colak E, Wilson JR (2021) Applications of machine learning to imaging of spinal disorders: current status and future directions. Global Spine J 11(1_suppl):23S-29S. https://doi.org/10.1177/2192568220961353

13. Ma S, Huang Y, Che X, Gu R (2020) Faster RCNN-based detection of cervical spinal cord injury and disc degeneration. J Appl Clin Med Phys 21(9):235–243. https://doi.org/10.1002/acm2.13001

14. Park SH, Han K (2018) Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 286(3):800–809. https://doi.org/10.1148/radiol.2017171920

15. Kim DW, Jang HY, Kim KW, Shin Y, Park SH (2019) Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. Korean J Radiol 20(3):405–410. https://doi.org/10.3348/kjr.2019.0025

16. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK (2018) Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 15(11):e1002683. https://doi.org/10.1371/journal.pmed.1002683

17. Pang S, Su Z, Leung S, Nachum IB, Chen B, Feng Q, Li S (2019) Direct automated quantitative measurement of spine by cascade amplifier regression network with manifold regularization. Med Image Anal 55:103–115. https://doi.org/10.1016/j.media.2019.04.012

18. Zheng HD, Sun YL, Kong DW, Yin MC, Chen J, Lin YP, Ma XF, Wang HS, Yuan GJ, Yao M, Cui XJ, Tian YZ, Wang YJ (2022) Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. Nat Commun 13(1):841. https://doi.org/10.1038/s41467-022-28387-5

19. Kruggel F, Turner J, Muftuler LT (2020) Alzheimer's disease neuroimaging. Initiative impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49(3):2123–2133. https://doi.org/10.1016/j.neuroimage.2009.11.006

20. Obuchowicz R, Oszust M, Piorkowski A (2020) Interobserver variability in quality assessment of magnetic resonance images. BMC Med Imaging 20(1):109. https://doi.org/10.1186/s12880-020-00505-z.

21. Morid MA, Borjali A, Del Fiol G (2021) A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 128:104115. https://doi.org/10.1016/j.compbiomed.2020.104115

22. Karimi D, Warfield SK, Gholipour A (2021) Transfer learning in medical image segmentation: new insights from analysis of the dynamics of model parameters and learned representations. Artif Intell Med 116:102078. https://doi.org/10.1016/j.artmed.2021.102078

23. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

24. Hasz MW (2012) Diagnostic testing for degenerative disc disease. Adv Orthop 2012:413913. https://doi.org/10.1155/2012/413913