中国临床研究 Chin J Clin Res, May 2024, Vol.37, No.5

Cite as: Ding ZM, Li HY, Chen L，et al. Fully automatic classification of lumbar disc degeneration based on deep learning [J]. Chin J Clin Res, 2024, 37(5):709-713.
DOI: 10.13429/j.cnki.cjcr.2024.05.039

Fully automatic classification of lumbar disc degeneration based on deep learning

DING Zhaoming*, LI Hongyan, CHEN Liang, CHEN Bing, SUN Hui, HOU Wentao, XIA Chunhua

**Imaging Center, The Third Affiliated Hospital of Anhui Medical University, Hefei, Anhui 230000, China*

Corresponding author: XIA Chunhua, E-mail: xiachunhua3775@sina.com

Abstract: Objective To investigate the feasibility of a deep learning model for the fully automatic classification of disc degeneration based on lumbar structures on sagittal T2WI images. **Methods** The lumbar T2WI image data of 94 patients who underwent lumbar spine MRI examination in the Third Affiliated Hospital of Anhui Medical University from August 2020 to June 2022 were retrospectively and continuously selected, and 466 discs were obtained. The lumbar intervertebral disk was manually annotated by 2 radiologists on sagittal T2WI images. The data were randomly divided into train set (n=300) , validation set (n=72), and test set (n=94). Firstly, a U-Net network was used to train the disc segmentation model. The evaluation indexes of the model included Dice coefficient and IoU score. Then, SpineNet network was used to train the disc segmentation model. The evaluation indexes of the model included Dice coefficient and IoU score. Then, SpineNet network was used to train the classification model, and the evaluation indexes of the model included accuracy (ACC), sensitivity (SEN), specificity (SPE), F1 score, and ROC curves. **Results** In the test set, the dice coefficient and IoU values of U-Net model for lumbar disc segmentation were 0.920 and 0.853, respectively. The ACC, SPE and SEN value of SpineNet classification models for lumbar disc degeneration were 0.913, 0.912 and 0.916, respectively. The ROC curve analysis showed that the AUC values for distinguishing mild to moderate, mild to serious, and moderate to serious lumbar disc degeneration were 0.89, 0.95, and 0.90, respectively. **Conclusion** It is feasible to realize the fully automatic classification of disc degeneration based on a deep learning network.

Keywords: Lumbar; Disc degeneration; Deep learning network ; T2WI sagittal image; U-Net model; Segmentation model; Classification model
Fund program: Anhui Provincial Health Commission Research Project (AHWJ2021b141)

Low back pain (LBP) is one of the most important health problems in modern society, mostly occurring in middle-aged and elderly people, even young people, with a lifetime incidence as high as 75% to 85%. Lumbar disc degeneration (LDD) is the leading cause of back pain[1] . Magnetic resonance imaging (MRI) is one of the most commonly used imaging techniques for the evaluation of LDD[2], and conventional T2-weighted imaging (T2WI) can effectively detect the disc morphology, height, and signal changes. The Pfirrmann grading system is one of the most accepted grading systems[3], which classifies LDD into 5 categories based on the signal intensity of the nucleus pulposus of the disc in the sagittal image of the T2WI, the demarcation between the nucleus pulposus and annulus fibrosus, and the disc height. This makes the assessment formal and comparable. However, it is highly dependent on the observer's level of expertise, with inevitably subjective errors, and is not consistent and efficient enough to be widely used. In recent years, network models based on deep learning have become very popular in various medical image segmentation and recognition tasks[4-5], and many researches have adopted this method for automatic classification of LDD[6-8] to reduce the subjectivity and instability brought by manual analysis and to improve work efficiency.

The aim of this study was to develop a fully automated deep learning model that automatically extracts intervertebral discs from routine T2WI sagittal images of the lumbar spine using a U-Net network, and then trains an LDD classifier based on a convolutional neural network (CNN) model (SpineNet)[9] to grade the degree of degeneration.

1 Materials and methods

1.1 General information

A retrospective consecutive sample of 94 patients admitted to the Third Affiliated Hospital of Anhui Medical University and underwent lumbar spine MRI examination from August 2020 to June 2022 was taken as the observation subjects. This is a retrospective study, only collecting imaging data, with no intervention in treatment; the informed consent exemption in this unit fully protects the rights and privacy of the subjects.

Inclusion criteria: lumbar spine MRI containing clear sagittal T2WI images.

Exclusion criteria: (1) history of lumbar spine slippage, deformity, and lumbar spine surgery; (2) non-disc diseases such as tumors, fractures, and infections; (3) poor image quality with metal artifacts. Of these, 47 were male and 47 were female, aged 22-83 years, with a mean age of 50 years. Four discs of poor quality were excluded and a total of 466 discs were included in the study. Of these, 300 discs were used as a training set, 72 discs as a tuning set and 94 discs as a test set.

1.2 Instruments and methods

A Siemens Trio-Tim 3.0 T magnetic resonance scanner with a spinal phased array coil was used and the scan sequences included conventional sagittal T1WI, T2WI and T2WI compression fat sequences and transverse T2WI sequences. Sagittal T2WI sequence parameters: TR=4000 ms, TE=96 ms, slice thickness=4 mm, slice spacing=0.4 mm, number of slices=13, field of view (FOV) =350 mm × 350 mm.

1.3 Image annotation

Image annotation, peer review and correction of the five lumbar discs were performed by two experienced diagnostic imaging physicians at the mid-sagittal level of the T2WI sequence, using 3D-Slicer annotation software and purple colour-coding of the lumbar discs.

1.4 Pfirrmann scoring of disc degeneration

The mid-sagittal level of the T2WI sequence was selected and the degree of disc degeneration was graded by a radiologist with 5 years of experience and a radiologist with 15 years of experience using the Pfirrmann scoring criteria[10]. If there was a discrepancy between the grading results of the two physicians, an agreement was reached after discussion. Following previous studies[11], the degree of disc degeneration was classified into three groups according to the Pfirrmann score: mild (grade I-II), moderate (grade III) and serious (grade IV-V).

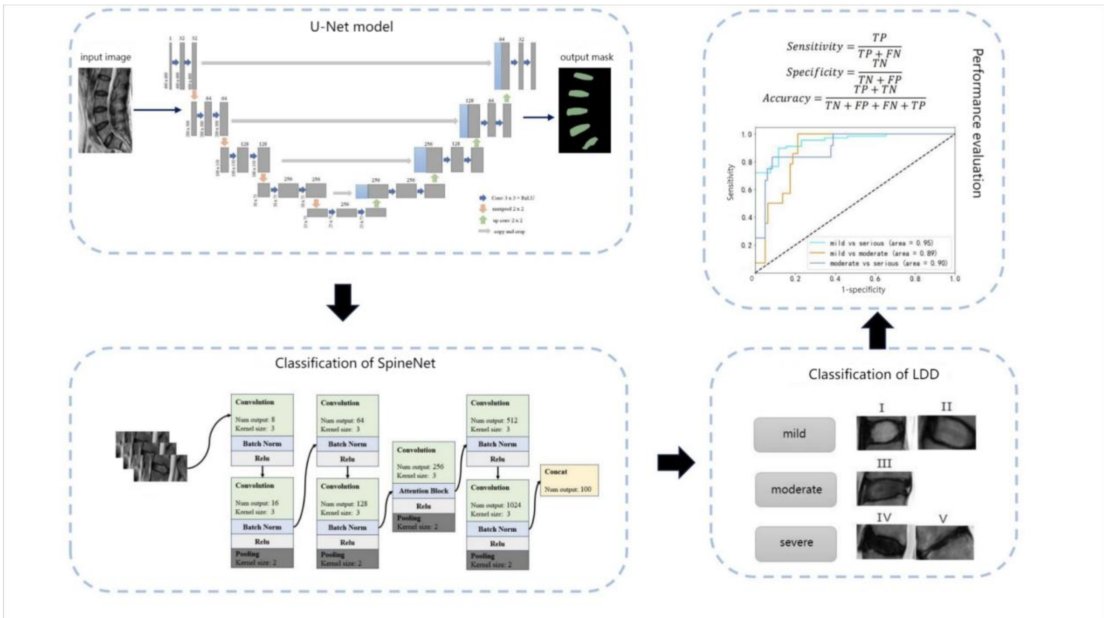
1.5 Model training

The hardware used for model training is a GPU Intel RTX 3090, 32 G. The software includes Python 3.8, Pytorch 1.11.0, and so on. Stochastic gradient descent (SGD) was used as the training optimizer.

Model training **[Figure 1]**: all data were randomly divided into training set, validation set and test set. In the first step, the segmentation model was trained, the input image was a T2WI image of the lumbar spine in the mid-sagittal plane, and the output data was the disc prediction region. Image pre-processing included automatic adjustment of window width and position, and the image size was set to 1×512×512. The U-Net model[11] was utilized for training, which consists of three parts: encoder, connection layer and decoder. The encoding path consists of four convolutional, pooling layers that continuously capture features at different scales through a convolutional kernel, while the decoding path has two parts: (1) gradually increasing the spatial dimensionality while decreasing the number of feature channels through the composition of four convolutional and inverse convolutional layers; and (2) combining the outputs of the decoder path with the outputs of the corresponding encoder paths through a jump connection. This helps to combine the underlying detailed spatial information with the high-level semantic information to improve the accuracy of segmentation. The primary purpose of the connection layer is to connect the encoder and decoder. In the second step to train the classification model, an attention mechanism-based CNN model (SpineNet)[9] is designed for training, the network architecture is based on VGG11, and the attention mechanism module is added to the network to improve the learning ability of the network. The input image is the intervertebral disc and its corresponding segmentation mask automatically segmented in the first step, and the size is set to 1×128×256. Firstly, the image is downsampled by a series of convolutions to extract the features of the image, and then the channel attention module is added to the high-level feature layer to perform further attentional learning for the deeper channels, aiming to extract more semantic features to help determine the grading of LDD. The output is then the degree of LDD (mild, moderate, serious). In addition, data enhancement techniques such as rotating, flipping, and adding Gaussian noise are added to the classification part to account for the small sample size. The results show that the introduction of the attention mechanism allows the CNN to achieve better results on the multi-classification problem compared to the traditional CNN. The main parameters of the network are as follows: batch_size=8, num_poch=300, and learning_rate=0.001.

1.6 Model evaluation

For the segmentation part, the model's efficacy was evaluated using the average Dice coefficient and average intersection over union (IoU) scores of the test set data. The average Dice coefficient takes values from 0 to 1, and the closer to 1, the better the model segmentation effect. For the classification part, the accuracy (ACC), sensitivity (SEN), specificity (SPE), F1 score and receiver operating characteristic curve (ROC) of the test set data were used to evaluate the model efficacy. The higher the ACC value, the better the classification effect.



Note: TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

Fig.1 The flowchart of the automatic classification of lumbar disc degeneration based on lumbar MR image

2 Results

2.1 General information

A total of 94 patients were enrolled, and 466 intervertebral discs were obtained finally, of which Grade I: 80, Grade II: 120, Grade III: 154, Grade IV: 71, and Grade V: 41. The discs were randomly divided into a training set (*n*=300), a tuning set (*n*=72), and a testing set (*n*=94).

2.2 Objective evaluation of model segmentation

The average dice coefficient values and average IoU scores of the segmentation model's segmentation results for lumbar discs in the tuning set as well as in the test set were shown in **Table 1**. In the tuning set, the average dice coefficient values and IoU values were 0.914 and 0.854, respectively. In the test set, the average dice coefficient values and IoU values were 0.920 and 0.853, respectively. **Figure 2** showed the MRI of lumbar disc with manual labeling and an automatic segmentation mask.

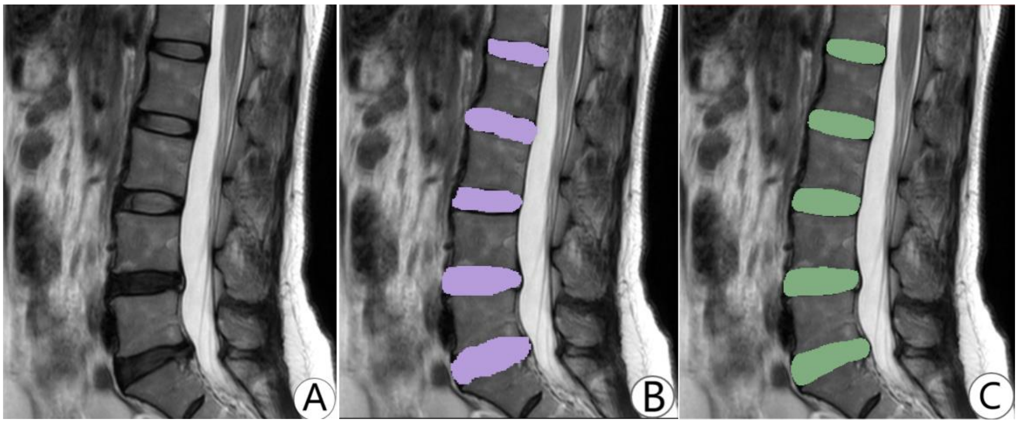
2.3 Objective evaluation of model LDD classification

The values of each parameter of the SpineNet classification model for automatic LDD classification results in the tuning set as well as in the test set are shown in **Table 2**. In the tuning set, the values of ACC, SPE, and SNE for LDD classification by the SpineNet model are 0.922, 0.923, and 0.944, respectively. In the test set, the values of ACC, SPE, and SNE for LDD classification by the SpineNet model are 0.913, 0.912, and 0.916, respectively. Net model's ACC, SPE, and SNE values for LDD classification are 0.913, 0.912, and 0.916, respectively.The ROC curve analysis results showed that the AUC values for distinguishing mild *vs* moderate, mild *vs* serious, and moderate *vs* serious LDD are 0.89, 0.95, and 0.90, respectively. **[Figure 3]**

Tab.1 Performance parameters of U-Net model for automatic segmentation of lumbar intervertebral discs [$M(Q_L, Q_U)$]

Type of set	Number of intervertebral discs	Performance indicators (2D UNet)	
		Dice coefficient	IoU Score
Tunning set	72	0.914 (0.781, 0.959)	0.854 (0.682, 0.891)
Test set	94	0.920 (0.796, 0.935)	0.853 (0.661, 0.877)

Note: $M(Q_L, Q_U)$, median (lower quartile, upper quartile).



Note: A, lumbar T2WI mid-sagittal position; B, lumbar intervertebral disc manual labeling results; C, lumbar intervertebral disc automatic segmentation mask.

Fig.2 Manually annotate and automatic segmentation mask of of lumbar intervertebral disc

Tab. 2 Diagnostic Performance parameters of SpineNet model for automatic classification of LDDs

Type of set	Number of intervertebral discs	Performance indicators (SpineNet)			
		ACC	SEN	SPE	F1 score
Tunning set	72	0.922	0.944	0.923	0.769
Test set	94	0.913	0.916	0.912	0.759

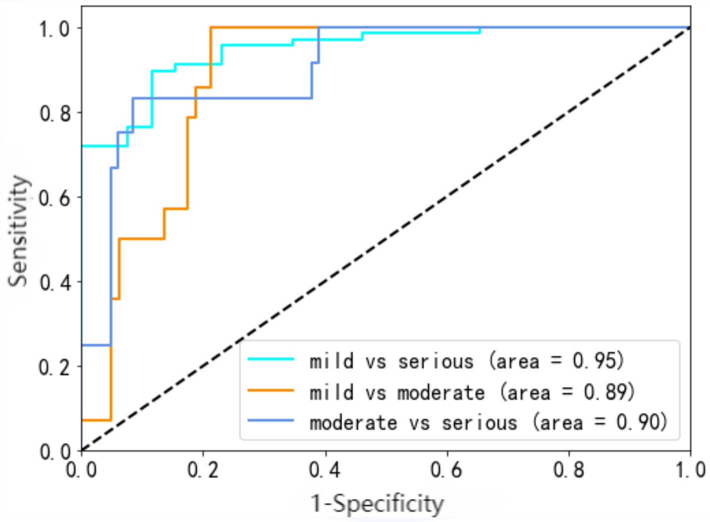


Fig.3 ROC curves of SpineNet model for automatic classification diagnosis of LDD

3 Discussion

In this study, we developed a deep learning system that can be applied to automatically segment intervertebral discs and grade degeneration in T2WI sagittal images. The system integrates the published basic network architectures of U-net and CNN. The U-net network is first used for automatic segmentation of intervertebral discs, and after evaluating the segmentation accuracy of the model, a CNN model with an added attention mechanism is further proposed for LDD classification. The developed system can reliably classify the Pfirrmann score of LDD in a fully automated manner. The average Dice coefficient of segmentation is 92.0% and the classification accuracy is 91.3%, which is satisfactory. In addition, the method in this study demonstrates the possibility of applying deep learning to small datasets, changes the stereotype that deep learning algorithms require a large number of training samples, and provides a framework to support the full automation of LDD classification.

In recent years, many studies have adapted machine learning methods to automatic LDD classification, including traditional methods based on manual feature extraction and methods based on deep learning. Deep learning, especially CNNs, has led to significant progress in this task. Compared to traditional machine learning methods, CNN does not rely on shallow rule-based image features, but automatically extracts features at different levels of abstraction from the input image through a series of convolutional operations, while using both perceptible and imperceptible image features for classification prediction, which is currently widely used[6,12-13]. However, only a few studies have achieved simultaneous segmentation and hierarchical diagnosis of intervertebral discs, which is directly generated by a network of radiologist-level diagnostic results, which is very helpful for the analysis of lumbar spine diseases. Jamaludin *et al.*[14] developed an automated diagnostic system for lumbar spine degeneration image features based on traditional automated segmentation methods in 2017, and the diagnostic accuracy of Pifrrmann grading was only 71%. Cheung *et al.* [15] constructed a new model to predict the progression of LDD by using unsupervised supervised DL model segmentation of vertebral body to predict the intervertebral disc region in 2022, and then extracted features from the intervertebral disc region to predict the degeneration grading through the basic CNN framework, and the results showed that the prediction accuracy of Pfirrmann grading progression reached 89. 9%, but the accuracy of the model for disc segmentation and categorization was not mentioned in the paper. Compared with previous works, this study uses a deep learning approach to complete the sequential task of intervertebral disc segmentation from segmentation to classification, realizes fully automatic classification of LDD, and provides highly accurate segmentation and diagnostic results.

Accurate segmentation of lumbar spine structures is the basis for diagnosis and treatment of lumbar spine diseases[16]. The segmentation part of this study uses the more commonly used U-Net grid[17], which is a model proposed by Ronneberger *et al.*[12] in 2015. The framework overcomes the problem of low data volume by using elastic enhancement, but it requires pixel-level supervised learning, and achieves a better segmentation result in lumbar spine structure, with the segmentation accuracy reaching more than 90%[18-20]. Similar to previous studies, the average Dice coefficient of segmentation in this study reaches 92.0%, which is a more satisfactory segmentation result and lays the foundation for further accurate classification based on automatic segmentation. In the LDD classification part, this study based on VGG11, combines the disc image features and reduces the number of convolutional layers to adapt to the application scenario of a small sample size. At the same time, the addition of the attention mechanism module and data enhancement technology to the network improves the learning ability of the network, and the results show that the classification accuracy reaches 91.3% in the case of a small sample size, and the introduction of the attention mechanism makes the CNN achieve better results in the multi-classification problem. The introduction of the attention mechanism enables the CNN to achieve better results in multi-categorization problems. In the future, we need to increase the sample size to distinguish more detailed classes.

In addition, considering that the Pifrrmann score is evaluated based on the sagittal position in lumbar spine T2WI, the 2D model was used for training in this study, and although the 3D model can provide more complete information, the 2D model is closer to the actual situation of the Pifrrmann score. In fact, Jamaludin *et al.*[14] showed that the classification performance of the 3D model for the Pifrrmann score was not improved or even slightly reduced compared to the 2D model.

This study has several limitations. First, it is a retrospective, single-centre study with a small sample size and a study population of patients presenting to the hospital with selection bias. Future clinical applications require prospective multicentre studies with larger sample sizes. Secondly, the cases did not include patients with spinal deformities, slips and other conditions, which limits the applicability of the network architecture. Deep learning requires sufficient training data for each category, and the imbalance of data will affect the training results. Further collection of training data is needed to compensate for this deficiency in the future. Finally, image labelling in the model is highly supervised, and manual image labelling is the most accurate way to train the model, but the process is tedious and time-consuming, which limits the number of samples available for training. In addition, to further adapt to real clinical scenarios, more quantitative and functional sequences should be included in the future to explore the possibility of improving grading performance. The aim of this study is not to replace medical staff, but to support further research by providing a basic framework.

In conclusion, this study develops a high-precision deep learning network that can fully automate the identification of LDD grades on small datasets, providing a framework to support the full automation of LDD assessment.

Conflict of interest: None

References

[1] Liang XQ, Xie RY, Hou BW, et al. Feasibility study for evaluating lumbar intervertebral disc degeneration using histogram analysis of T2* values[J]. Eur Spine J, 2020, 29(10): 2600-2608.
[2] Zheng HD, Sun YL, Kong DW, et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI[J]. Nat Commun, 2022, 13(1): 841.
[3] Huang JW, Shen HT, Wu JL, et al. 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[J]. Spine J, 2020, 20(4): 590-599.
[4] Suri A, Jones BC, Ng G, et al. Vertebral deformity measurements at MRI, CT, and radiography using deep learning[J]. Radiol Artif Intell, 2022, 4(1): e210015.
[5] Lim DSW, Makmur A, Zhu L, et al. Improved productivity using deep learning-assisted reporting for lumbar spine MRI[J]. Radiology, 2022, 305(1): 160-166.

[6] Niemeyer F, Galbusera F, Tao YP, et al. A deep learning model for the accurate and reliable classification of disc degeneration based on MRI data[J]. Invest Radiol, 2021, 56(2): 78-85.
[7] Grob A, Loibl M, Jamaludin A, et al. External validation of the deep learning system "SpineNet" for grading radiological features of degeneration on MRIs of the lumbar spine[J]. Eur Spine J, 2022, 31(8): 2137-2148.
[8] Editors TPO. Retraction: the application of key feature extraction algorithm based on Gabor wavelet transformation in the diagnosis of lumbar intervertebral disc degenerative changes[J]. PLoS One, 2023, 18(9): e0292208.
[9] Qu XL, Lu HY, Tang WZ, et al. A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images[J]. Med Phys,2022 ,49(9):5787-5798.
[10] Pfirrmann CWA, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration[J]. Spine, 2001, 26(17): 1873-1878.
[11] Shakeri A, Shakeri M, Ojaghzadeh Behrooz M, et al. Infrarenal aortic diameter, aortoiliac bifurcation level and lumbar disc degenerative changes. A cross-sectional MR study[J]. Eur Spine J,2018, 27(5):1096-1104.
[12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[G/OL].International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241, (2015-05-18). http://lmb.informatik.uni-freiburg.de/ people/ronneber/u-net.
[13] Gao F, Liu S, Zhang X, et al. Automated grading of lumbar disc degeneration using a push-pull regularization network based on MRI [J] J Magn Reson Imaging,2021, 53(3):799-806.
[14] Jamaludin A, Lootus M, Kadir T,et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist[J]. Eur Spine J,2017,26(5):1374-1383.
[15] Cheung JPY, Kuang XH, Lai MKL, et al. Learning-based fully automated prediction of lumbar disc degeneration progression with specified clinical parameters and preliminary validation[J]. Eur Spine J, 2022, 31(8): 1960-1968.
[16] Liu JH, Cui ZM, Desrosiers C, et al. Grayscale self-adjusting network with weak feature enhancement for 3D lumbar anatomy segmentation[J]. Med Image Anal, 2022, 81: 102567.
[17] Zhao K, Guo L, Zhang XD, et al. The research status and progress of artificial intelligence in lumbar spine imaging[J]. Chin J Radiol, 2023, 57(4): 424-427. **[In Chinese]**
[18] Zhou JM, Damasceno PF, Chachad R, et al. Automatic vertebral body segmentation based on deep learning of Dixon images for bone marrow fat fraction quantification[J]. Front Endocrinol (Lausanne), 2020, 11: 612.
[19] Guo L, Zhao K, Zhu YF, et al. U-Net deep learning network for automatic segmentation of lumbar structures on sagittal T2 WI images[J]. Radiol Pract, 2022, 37(2): 224-229. **[In Chinese]**
[20] Cheng YK, Lin CL, Huang YC, et al. Automatic segmentation of specific intervertebral discs through a two-stage MultiResUNet model [J]. J Clin Med,. 2021,10(20):4760.

Submission received:2023-09-26 / Revised:2023-11-02

·论　著·

基于深度学习的腰椎间盘退变全自动分级

丁兆明[1]，　李鸿燕[2]，　陈亮[3]，　陈兵[1]，　孙辉[1]，　候文韬[1]，　夏春华[1]

1. 安徽医科大学第三附属医院影像中心，安徽 合肥 230000；2. 中国科学技术大学，安徽 合肥 230000；
3. 安徽医科大学，安徽 合肥 230000

摘要：目的　探讨深度学习模型在腰椎磁共振 T2 加权成像(T2WI)矢状图像上全自动识别腰椎间盘退变程度的可行性。**方法**　回顾性抽取 2020 年 8 月至 2022 年 6 月于安徽医科大学第三附属医院就诊并行腰椎 MRI 检查的 94 例患者的腰椎 T2WI 图像数据,共获得 466 个椎间盘,由两名放射科医生手动标注腰椎间盘,将数据随机分为训练集(300 个)、调优集(72 个)和测试集(94 个),首先使用 U-Net 网络训练椎间盘分割模型,模型评价指标包括 Dice 系数和交并比(IoU)分数;然后利用 SpineNet 网络训练分类模型进行评价,评价指标包括准确度、敏感度、特异度、F1 分数及 ROC 曲线。**结果**　测试集中 U-Net 模型对腰椎间盘分割的平均 Dice 系数值及 IoU 分数分别为 0.920、0.853;SpineNet 分类模型对腰椎间盘退变分类诊断的准确度、特异度、敏感度分别为 0.913、0.912、0.916,ROC 曲线分析示,该模型区分腰椎间盘退变轻度 *vs* 中度、轻度 *vs* 重度、中度 *vs* 重度的 AUC 值分别为0.89、0.95、0.90。**结论**　深度学习网络对腰椎间盘退变程度的全自动分类是可行的。

关键词：腰椎；椎间盘退变；深度学习网络；T2WI 矢状图像；U-Net 模型；分割模型；分类模型

中图分类号：R445.2　**文献标识码：**A　**文章编号：**1674-8182(2024)05-0709-05

Fully automatic classification of lumbar disc degeneration based on deep learning

DING Zhaoming*, LI Hongyan, CHEN Liang, CHEN Bing, SUN Hui, HOU Wentao, XIA Chunhua

** Imaging Center, The Third Affiliated Hospital of Anhui Medical University, Hefei, Anhui 230000, China*

Corresponding author: XIA Chunhua, E-mail: xiachunhua3775@sina.com

Abstract: Objective　To investigate the feasibility of a deep learning model for the fully automatic classification of disc degeneration based on lumbar structures on sagittal T2WI images. **Methods**　The lumbar T2WI image data of 94 patients who underwent lumbar spine MRI examination in the Third Affiliated Hospital of Anhui Medical University from August 2020 to June 2022 were retrospectively selected, and 466 discs were obtained. The lumbar intervertebral disc were manually annotated by 2 radiologists on sagittal T2WI images. The data were randomly divided into train set ($n=300$), validation set ($n=72$), and test set ($n=94$). Firstly, a U-Net network was used to train the disc segmentation model. The evaluation indexes of the model included Dice coefficient and IoU score. Then, SpineNet network was used to train the classification model, and the evaluation indexes of the model included accuracy, sensitivity, specificity, F1 score, and ROC curves. **Results**　In the test set, the dice coefficient and IoU values of U-Net model for lumbar disc segmentation were 0.920 and 0.853, respectively. The accuracy, specificity and sensitivity value of SpineNet classification models for lumbar disc degeneration were 0.913, 0.912 and 0.916, respectively. The ROC curve analysis showed that the AUC values for distinguishing mild to moderate, mild to serious, and moderate to serious lumbar disc degeneration were 0.89, 0.95, and 0.90, respectively. **Conclusion**　It is feasible to realize the fully automatic classification of disc degeneration based on deep learning network.

Keywords: Lumbar; Disc degeneration; Deep learning network; T2WI sagittal image; U-Net model; Segmentation model; Classification model

Fund program: Anhui Provincial Health Commission Research Project (AHWJ2021b141)

DOI: 10.13429/j.cnki.cjcr.2024.05.039

基金项目：安徽省卫生健康委科研项目（AHWJ2021b141）

通信作者：夏春华，E-mail：xiachunhua3775@sina.com

出版日期：2024-05-20

QR code for English version

下腰痛(low back pain, LBP)是现代社会常见的健康问题之一,多发于中老年人群,甚至年轻人,终生发病率高达 75% ~ 85%,腰椎间盘退变(lumbar disc degeneration, LDD)是引起腰痛的主要原因[1]。磁共振成像(MRI)是评估 LDD 最常用的影像学检查技术之一[2],常规 T2 加权成像(T2WI)可以很好地检测椎间盘形态、高度及信号变化。Pfirrmann 分级系统是目前最被接受的评分系统之一[3],它根据 T2WI 矢状图像上椎间盘髓核的信号强度、髓核与纤维环的分界以及椎间盘高度,将 LDD 分为 5 个等级,使评估形式化且具有可比性。然而,它高度依赖于观察者的专业知识水平,不可避免地存在主观误差,一致性和效率不高,无法广泛应用。近年来,基于深度学习(deep learning)的网络模型在各类医学图像分割和识别任务中广泛应用[4-5],许多研究将该方法用于 LDD 自动分类[6-8],以减少人工分析带来的主观性和不稳定性,提高工作效率。

本研究旨在开发一种全自动深度学习模型,该模型使用 U-Net 网络从腰椎常规 T2WI 矢状位图像中自动提取椎间盘,然后训练一个基于卷积神经网络(CNN)模型(SpineNet)[9]的 LDD 分类器进行退变程度分级。

1　材料与方法

1.1　一般资料　回顾性抽取 2020 年 8 月至 2022 年 6 月于安徽医科大学第三附属医院就诊并行腰椎 MRI 检查的患者 94 例为观察对象。本研究为回顾性研究,仅收集影像学资料,不干预治疗,研究对象权利和隐私得到充分保护,符合本单位豁免知情同意的情况。纳入标准:腰椎 MRI 检查包含清晰的矢状面 T2WI 图像。排除标准:(1) 腰椎滑脱、畸形及腰椎手术史;(2) 非椎间盘疾病如肿瘤、骨折、感染等;(3) 图像质量不佳,存在金属伪影。其中,男 47 例,女 47 例,年龄 22~83 岁,平均 50 岁。排除 4 个质量不佳的椎间盘,共计 466 个椎间盘纳入研究。其中 300 个椎间盘作为训练集,72 个椎间盘作为调优集,94 个椎间盘作为测试集。

1.2　仪器与方法　采用 Siemens Trio-Tim 3.0 T 磁共振扫描仪,脊柱相控阵线圈,扫描序列包括常规矢状位 T1WI、T2WI 及 T2WI 压脂序列、横断位 T2WI 序列。矢状位 T2WI 序列参数:TR 4 000 ms,TE 96 ms,层厚 4 mm,层距 0.4 mm,视野(FOV)350 mm × 350 mm。

1.3　图像标注　由两位有经验的影像诊断医师在 T2WI 序列正中矢状位层面对腰椎 5 个椎间盘进行图像标注并相互检查校正,标注软件为 3D-Slicer,腰椎间盘的颜色标注为紫色。

1.4　椎间盘退变 Pfirrmann 评分　选择 T2WI 序列正中矢状位层面,由 1 名 5 年和 1 名 15 年工作经验的放射科医师,参考 Pfirrmann 评分标准[10]对椎间盘退变程度进行分级,当两名医师的评估结果存在差异时,经讨论达成一致。参考既往研究[11],根据 pfirrmann 评分将椎间盘退变程度分为三组:轻度(Ⅰ~Ⅱ级)、中度(Ⅲ级)、重度(Ⅳ~Ⅴ级)。

1.5　模型训练　模型训练的硬件为 GPU intel RTX 3090,32 G,软件包括 python3.8,pytorch1.11.0 等。使用随机梯度下降(SGD)作为训练优化器。

模型训练(如图 1 所示):将所有数据随机分为训练集、调优集和测试集。第一步训练分割模型,输入图像为正中矢状面腰椎 T2WI 图像,输出数据为椎间盘预测区域。图像预处理包括自动窗宽窗位的调整,图像大小设置为 1×512×512。利用 U-Net 模型[12]进行训练,该网络由编码器、连接层和解码器三个部分组成。编码路径由 4 个卷积层、池化层组成,不断通过卷积核捕获不同尺度的特征;解码路径则有两个部分组成:(1) 通过 4 个卷积和反卷积层组成,逐渐增加空间维度,同时减少特征通道数;(2) 将解码器路径的输出与相应的编码器路径的输出通过跳跃连接相结合。这有助于将底层的详细空间信息与高层次的语义信息结合起来,提高分割的准确性。连接层的主要目的是将编码器和解码器连接起来。第二步训练分类模型,设计了一个基于注意力机制的 CNN 模型(SpineNet)[9]进行训练,网络架构在基于 VGG11 的基础上,在网络中添加注意力机制模块,以提高网络的学习能力。输入图像为第一步自动分割的椎间盘及其相邻两张矢状面对应的椎间盘,大小设置为 3×128×256。首先通过一系列卷积对图片降采样,提取图像的特征,再在高层特征层加入通道注意力模块对深层的通道进行进一步的注意力学习,旨在挖掘出更多的语义特征来帮助判断 LDD 的分级。输出则是 LDD 的程度(轻度、中度、重度)。此外,为适应小样本量,分类部分添加了数据增强技术如旋转、翻转、添加高斯噪声等。结果表明,与传统的 CNN 相比,引入注意力机制,使得 CNN 在多分类问题上取得了更好的结果。网络的主要参数如下:batch_size = 8,num_poch = 300,learning_rate = 0.001。

1.6　模型评价　分割部分,用测试集数据的平均 Dice 系数和平均交并比(intersection over union, IoU)分数评价模型的效能。平均 Dice 系数取值为 0 ~ 1,

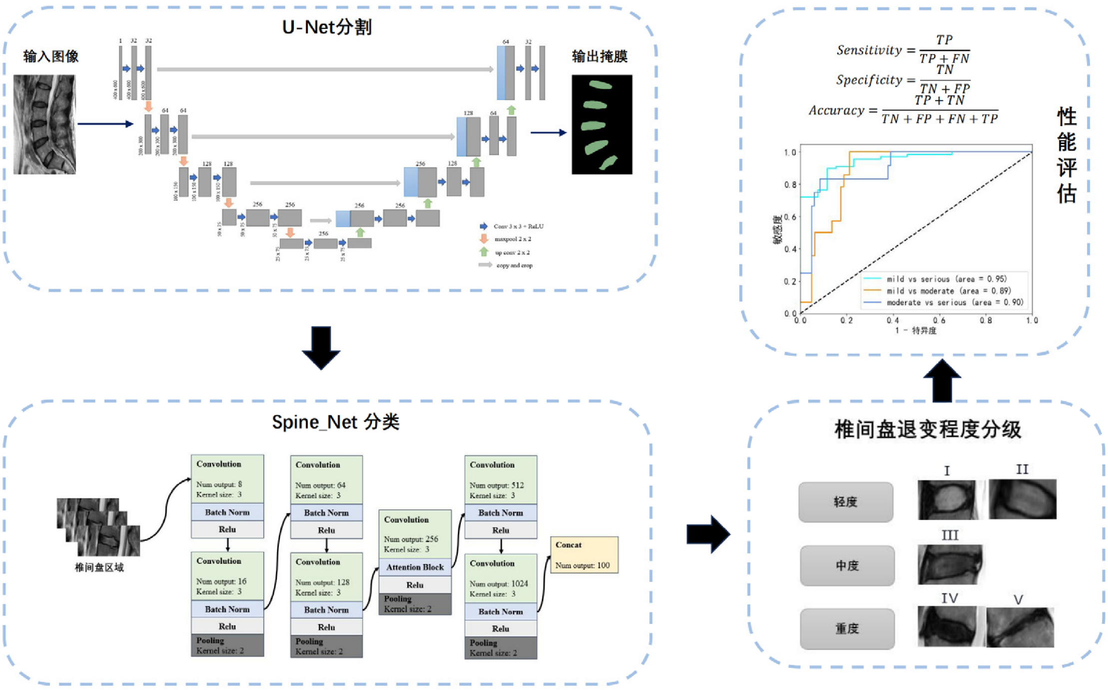
越接近 1 代表模型分割效果越好。分类部分,用测试集数据的准确度(accuracy, ACC),敏感度(sensitivity, SEN),特异度(specificity, SPE),F1 分数(F1 score)及受试者工作特征曲线(ROC)评价模型效能。ACC 值越高,分类效果越好。

2 结　果

2.1　一般资料　共入组 94 例患者,最终获得 466 个椎间盘,其中Ⅰ级 80 个,Ⅱ级 120 个,Ⅲ级 154 个,Ⅳ级 71 个,Ⅴ级 41 个,将其随机分为训练集($n=300$),调优集($n=72$),和测试集($n=94$)。

2.2　模型分割结果的客观评价　调优集以及测试集中,分割模型对腰椎间盘分割结果的平均 Dice 系数值及平均 IoU 分数见表 1。在调优集中,平均 Dice 系数值及 IoU 值分别为 0.914、0.854。在测试集中,平均 Dice 系数值及 IoU 值分别为 0.920、0.853。图 2 为腰椎间盘手动标注和自动分割掩膜的 MR 影像。

2.3　模型 LDD 分类结果的客观评价　调优集以及测试集中,SpineNet 分类模型对 LDD 自动分类结果的各参数值见表 2。在调优集中,SpineNet 模型对 LDD 分类的 ACC、SPE、SEN 值分别为 0.922、0.923、0.944。在测试集中,SpineNet 模型对 LDD 分类的 ACC、SPE、SEN 值分别为 0.913、0.912、0.916。ROC 曲线分析结果显示,Spine Net 模型区分轻度 *vs* 中度、轻度 *vs* 重度、中度 *vs* 重度 LDD 的 AUC 值分别为0.89、0.95、0.90。见图 3。



注:TP,真阳性;TN,真阴性;FP,假阳性;FN,假阴性。

图 1　腰椎 MR 椎间盘退变全自动分类模型训练流程

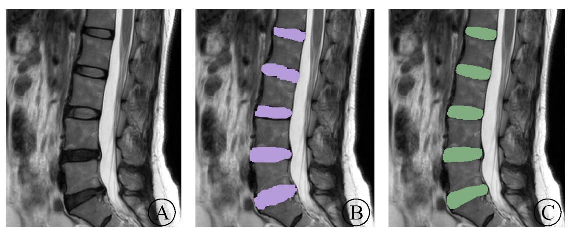
Fig. 1　The flowchart of the automatic classification of lumbar disc degeneration based on lumbar MR image

表 1　U-Net 模型对腰椎间盘自动分割的性能参数

Tab. 1　Performance parameters of U-Net model for automatic segmentation of lumbar intervertebral discs

集别	椎间盘个数	性能指标(2D UNet)	
		Dice 系数	IoU 分数
调优集	72	0.914(0.781, 0.959)[a]	0.854(0.682, 0.891)[a]
测试集	94	0.920(0.796, 0.935)[a]	0.853(0.661, 0.877)[a]

注:[a] 为平均数(最小值,最大值)。



注:A,腰椎 T2WI 正中矢状位;B,腰椎间盘手动标注结果;C,腰椎间盘自动分割掩膜。

图 2　腰椎间盘手动标注和自动分割掩膜的 MR 影像

Fig. 2　Manually annotate and automatic segmentation mask of of lumbar intervertebral disc

表 2　SpineNet 模型对 LDD 自动分类的诊断性能参数
Tab. 2　Diagnostic performance parameters of SpineNet model for automatic classification of LDD

集别	椎间盘个数	性能指标(SpineNet)			
		ACC	SEN	SPE	F1 score
调优集	72	0.922	0.944	0.923	0.769
测试集	94	0.913	0.916	0.912	0.759

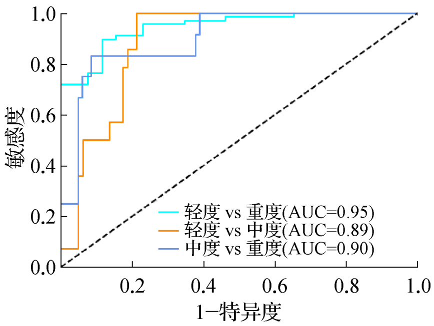


图 3　SpineNet 模型对 LDD 自动分类诊断的 ROC 曲线
Fig. 3　ROC curves of SpineNet model for automatic classification diagnosis of LDD

3　讨　论

本研究开发了一种可应用于 T2WI 矢状位图像自动分割椎间盘并进行退变分级的深度学习系统。该系统集成了已发布的 U-net 和 CNN 的基本网络架构。首先利用 U-net 网络对椎间盘进行自动分割，在评估了模型的分割精度后，进一步提出添加注意力机制的 CNN 模型进行 LDD 分类。开发的系统可以可靠地对 LDD 的 Pfirrmann 评分进行全自动分级，分割平均 Dice 系数为 92.0%，分类准确度达 91.3%，结果令人满意。此外，本研究的方法证明了在小规模数据集上使用深度学习的可能性，改变了深度学习算法需要大量训练样本的刻板印象，为 LDD 分级完全自动化提供了框架支持。

近年来，许多研究将机器学习方法用于 LDD 自动分类，包括传统的基于手工提取特征的方法和基于深度学习的方法。深度学习，尤其是 CNN 使这一任务取得了显著的进展。与传统的机器学习方法相比，CNN 不依赖于基于规则的浅图像特征，而是通过一系列卷积运算，自动从输入图像中提取不同层次的抽象特征，同时利用可感知和不可感知的图像特征进行分类预测，目前被广泛采用[6,12-13]，但只有少数研究实现了对椎间盘的同步分割和分级诊断，通过网络直接生成放射科医生级别的诊断结果，对腰椎疾病的分析有很大帮助。2017 年，Jamaludin 等[14]基于传统自动分割方法开发了一个腰椎退变影像特征的自动诊断系统，Pifrrmann 分级的诊断准确率仅为 71%。2022 年，Cheung 等[15]构建一种预测 LDD 进展的新模型，采用无监督 DL 模型分割椎体的方法来预测椎间盘区域，然后通过 CNN 基本框架从椎间盘区域提取特征预测退变分级，结果表明，Pfirrmann 分级进展预测准确率达到 89.9%，但文中未提及模型对椎间盘分割及分类的准确性。与之前的工作相比，本研究使用深度学习方法完成了椎间盘从分割到分类的序贯任务，实现对 LDD 的全自动分类，并提供了高精度的分割及诊断结果。

腰椎结构的精准分割是腰椎疾病诊断和治疗的基础[16]。本研究分割部分采用目前较为常用的 U-Net 网格[17]，该模型由 Ronneberger 等[12]在 2015 年提出，框架通过使用弹性增强克服了数据量少的问题，但需要像素级监督学习，在腰椎结构中取得了较好的分割效果，分割精度达 90% 以上[18-20]。与既往研究类似，本研究分割平均 Dice 系数达 92.0%，分割结果较为满意，为进一步在自动分割基础上进行精准分级奠定了基础。在 LDD 分类部分，本研究在基于 VGG11 的基础上，结合椎间盘图像特征，减少卷积层数，以适应样本量小的应用场景，同时在网络中添加注意力机制模块及数据增强技术，以提高网络的学习能力，结果显示在小样本量的情况下分类准确度达到 91.3%，引入注意力机制使得 CNN 在多分类问题上能够取得更好的结果。未来需增加样本量，以区分更详细的等级。

此外，考虑到 Pifrrmann 评分是基于腰椎 T2WI 中矢状位进行评估的，本研究采用 2D 模型进行训练，虽然 3D 模型能够提供更完整的信息，但 2D 模型更接近 Pifrrmann 评分的真实情况。Jamaludin 等[14]的研究表明，相比 2D 模型，3D 模型对 Pifrrmann 评分的分类性能并无改进甚至略有降低。

本研究有一些局限性。首先，是一项回顾性的单中心研究，样本量较小，研究群体为来院就诊患者，存在选择偏倚。未来的临床应用需要进行更大样本的前瞻性多中心研究。其次，病例没有包含脊柱畸形、滑脱等疾病的患者，限制了网络架构的适用性。深度学习需要为每个类别提供足够的训练数据，数据的不平衡会对训练结果产生影响，未来需进一步收集训练数据，弥补这种不足。最后，模型中的图像标记是高度监督的，手动标注图像是训练模型最准确的方法，但过程繁琐、耗时，限制了可用于训练的样本数量。此外，为了进一步适应真实的临床场景，未来应包括更多的定量和功能序列，以探索分级性能提高的可能性。本研究的目的并非取代医务人员，而是通过提供

基础框架,为进一步的研究提供支持。

综上所述,本研究开发了一种高精度的深度学习网络,可以在小规模数据集上全自动识别 LDD 程度,为 LDD 评估的完全自动化提供了框架支持。

利益冲突　无

参考文献

[1] Liang XQ, Xie RY, Hou BW, et al. Feasibility study for evaluating lumbar intervertebral disc degeneration using histogram analysis of T2* values[J]. Eur Spine J, 2020, 29(10): 2600-2608.

[2] Zheng HD, Sun YL, Kong DW, et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI[J]. Nat Commun, 2022, 13(1): 841.

[3] Huang JW, Shen HT, Wu JL, et al. 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[J]. Spine J, 2020, 20(4): 590-599.

[4] Suri A, Jones BC, Ng G, et al. Vertebral deformity measurements at MRI, CT, and radiography using deep learning[J]. Radiol Artif Intell, 2022, 4(1): e210015.

[5] Lim DSW, Makmur A, Zhu L, et al. Improved productivity using deep learning-assisted reporting for lumbar spine MRI [J]. Radiology, 2022, 305(1): 160-166.

[6] Niemeyer F, Galbusera F, Tao YP, et al. A deep learning model for the accurate and reliable classification of disc degeneration based on MRI data[J]. Invest Radiol, 2021, 56(2): 78-85.

[7] Grob A, Loibl M, Jamaludin A, et al. External validation of the deep learning system "SpineNet" for grading radiological features of degeneration on MRIs of the lumbar spine[J]. Eur Spine J, 2022, 31(8): 2137-2148.

[8] Editors TPO. Retraction: The application of key feature extraction algorithm based on Gabor wavelet transformation in the diagnosis of lumbar intervertebral disc degenerative changes [J]. PLoS One, 2023, 18(9): e0292208.

[9] Qu XL, Lu HY, Tang WZ, et al. A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images[J]. Med Phys,2022 ,49(9):5787-5798.

[10] Pfirrmann CWA, Metzdorf A, Zanetti M, et al. Magnetic resonance classification of lumbar intervertebral disc degeneration[J]. Spine, 2001, 26(17): 1873-1878.

[11] Shakeri A, Shakeri M, Ojaghzadeh Behrooz M, et al. Infrarenal aortic diameter, aortoiliac bifurcation level and lumbar disc degenerative changes: a cross-sectional MR study[J]. Eur Spine J, 2018,27(5):1096-1104.

[12] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[G/OL].International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer,2015: 234-241 (2015-05-18)[2023-08-23]. http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.

[13] Gao F, Liu S, Zhang X, et al. Automated grading of lumbar disc degeneration using a push-pull regularization network based on MRI [J]. J Magn Reson Imaging,2021, 53(3):799-806.

[14] Jamaludin A, Lootus M, Kadir T, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist[J]. Eur Spine J, 2017, 26(5): 1374-1383.

[15] Cheung JPY, Kuang XH, Lai MKL, et al. Learning-based fully automated prediction of lumbar disc degeneration progression with specified clinical parameters and preliminary validation [J]. Eur Spine J, 2022, 31(8): 1960-1968.

[16] Liu JH, Cui ZM, Desrosiers C, et al. Grayscale self-adjusting network with weak feature enhancement for 3D lumbar anatomy segmentation[J]. Med Image Anal, 2022, 81: 102567.

[17] 赵凯,郭丽,张晓东,等.人工智能在腰椎影像中的研究现状和进展[J].中华放射学杂志,2023,57(4):424-427.
Zhao K, Guo L, Zhang XD, et al. The research status and progress of artificial intelligence in lumbar spine imaging[J]. Chin J Radiol, 2023, 57(4): 424-427.

[18] Zhou JM, Damasceno PF, Chachad R, et al. Automatic vertebral body segmentation based on deep learning of Dixon images for bone marrow fat fraction quantification [J]. Front Endocrinol (Lausanne), 2020, 11: 612.

[19] 郭丽,赵凯,朱逸峰,等.使用 U-Net 深度学习网络对腰椎矢状 T_2WI 图像自动分割的可行性研究[J].放射学实践,2022,37(2):224-229.
Guo L, Zhao K, Zhu YF, et al. U-Net deep learning network for automatic segmentation of lumbar structures on sagittal T_2WI images [J]. Radiol Pract, 2022, 37(2): 224-229.

[20] Cheng YK, Lin CL, Huang YC, et al. Automatic segmentation of specific intervertebral discs through a two-stage MultiResUNet model [J]. J Clin Med, 2021, 10(20): 4760.

收稿日期:2023-09-26　**修回日期**:2023-11-02　**编辑**:石嘉莹