



OPEN ACCESS

EDITED BY

Yan Wang,
Chinese PLA General Hospital, China

REVIEWED BY

Stathis Hadjidemetriou,
University of Limassol, Cyprus
Luca Ferrarini,
University of Limassol, Cyprus

*CORRESPONDENCE

Litai Ma,
✉ ma.litai@163.com
Yang Jiang,
✉ 5313324@qq.com

RECEIVED 22 August 2023

ACCEPTED 24 November 2023

PUBLISHED 03 January 2024

CITATION

Xie J, Yang Y, Jiang Z, Zhang K, Zhang X, Lin Y, Shen Y, Jia X, Liu H, Yang S, Jiang Y and Ma L (2024), MRI radiomics-based decision support tool for a personalized classification of cervical disc degeneration: a two-center study. *Front. Physiol.* 14:1281506. doi: 10.3389/fphys.2023.1281506

COPYRIGHT

© 2024 Xie, Yang, Jiang, Zhang, Zhang, Lin, Shen, Jia, Liu, Yang, Jiang and Ma. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

MRI radiomics-based decision support tool for a personalized classification of cervical disc degeneration: a two-center study

Jun Xie^{1,2}, Yi Yang³, Zekun Jiang^{4,5}, Kerui Zhang³, Xiang Zhang³, Yuheng Lin⁵, Yiwei Shen³, Xuehai Jia³, Hao Liu³, Shaofen Yang⁶, Yang Jiang^{7*} and Litai Ma^{3*}

¹Information Technology Center, West China Hospital of Sichuan University, Chengdu, China,

²Information Technology Center, Sanya People's Hospital, Sanya, China, ³Department of Orthopedics, Orthopedic Research Institute, West China Hospital, Sichuan University, Chengdu, Sichuan, China,

⁴College of Computer Science, Sichuan University, Chengdu, Sichuan, China, ⁵West China Biomedical Big Data Center, Sichuan University, Chengdu, Sichuan, China, ⁶Cadre Health Section, Hezhou People's Hospital, Hezhou, Guangxi, China, ⁷Department of Orthopedic Spine, The Second Affiliated Hospital of Chengdu Medical College (China National Nuclear Corporation 416 Hospital), Chengdu, Sichuan, China

Objectives: To develop and validate an MRI radiomics-based decision support tool for the automated grading of cervical disc degeneration.

Methods: The retrospective study included 2,610 cervical disc samples of 435 patients from two hospitals. The cervical magnetic resonance imaging (MRI) analysis of patients confirmed cervical disc degeneration grades using the Pfirrmann grading system. A training set (1,830 samples of 305 patients) and an independent test set (780 samples of 130 patients) were divided for the construction and validation of the machine learning model, respectively. We provided a fine-tuned MedSAM model for automated cervical disc segmentation. Then, we extracted 924 radiomic features from each segmented disc in T1 and T2 MRI modalities. All features were processed and selected using minimum redundancy maximum relevance (mRMR) and multiple machine learning algorithms. Meanwhile, the radiomics models of various machine learning algorithms and MRI images were constructed and compared. Finally, the combined radiomics model was constructed in the training set and validated in the test set. Radiomic feature mapping was provided for auxiliary diagnosis.

Results: Of the 2,610 cervical disc samples, 794 (30.4%) were classified as low grade and 1,816 (69.6%) were classified as high grade. The fine-tuned MedSAM model achieved good segmentation performance, with the mean Dice coefficient of 0.93. Higher-order texture features contributed to the dominant force in the diagnostic task (80%). Among various machine learning models, random forest performed better than the other algorithms ($p < 0.01$), and the T2 MRI radiomics model showed better results than T1 MRI in the diagnostic performance ($p < 0.05$).

Abbreviations: MRI, magnetic resonance imaging; ROC, receiver operating characteristic; AUC, area under the curve; AF, annulus fibrosus; NP, nucleus pulposus; CDDD, cervical degenerative disc disease; DL, deep learning; CT, computed tomography; glcm, gray-level co-occurrence matrix; glrlm, gray-level run-length matrix; glszm, gray-level size-zone matrix; gldm, gray-level dependence matrix; ngtdm, neighboring gray-tone difference matrix; LoG, Laplacian of Gaussian; mRMR, minimum redundancy maximum relevance; XGBoost, eXtreme Gradient Boosting; SVM, support vector machine.

The final combined radiomics model had an area under the receiver operating characteristic curve (AUC) of 0.95, an accuracy of 89.51%, a precision of 87.07%, a recall of 98.83%, and an F1 score of 0.93 in the test set, which were all better than those of other models ($p < 0.05$).

Conclusion: The radiomics-based decision support tool using T1 and T2 MRI modalities can be used for cervical disc degeneration grading, facilitating individualized management.

KEYWORDS

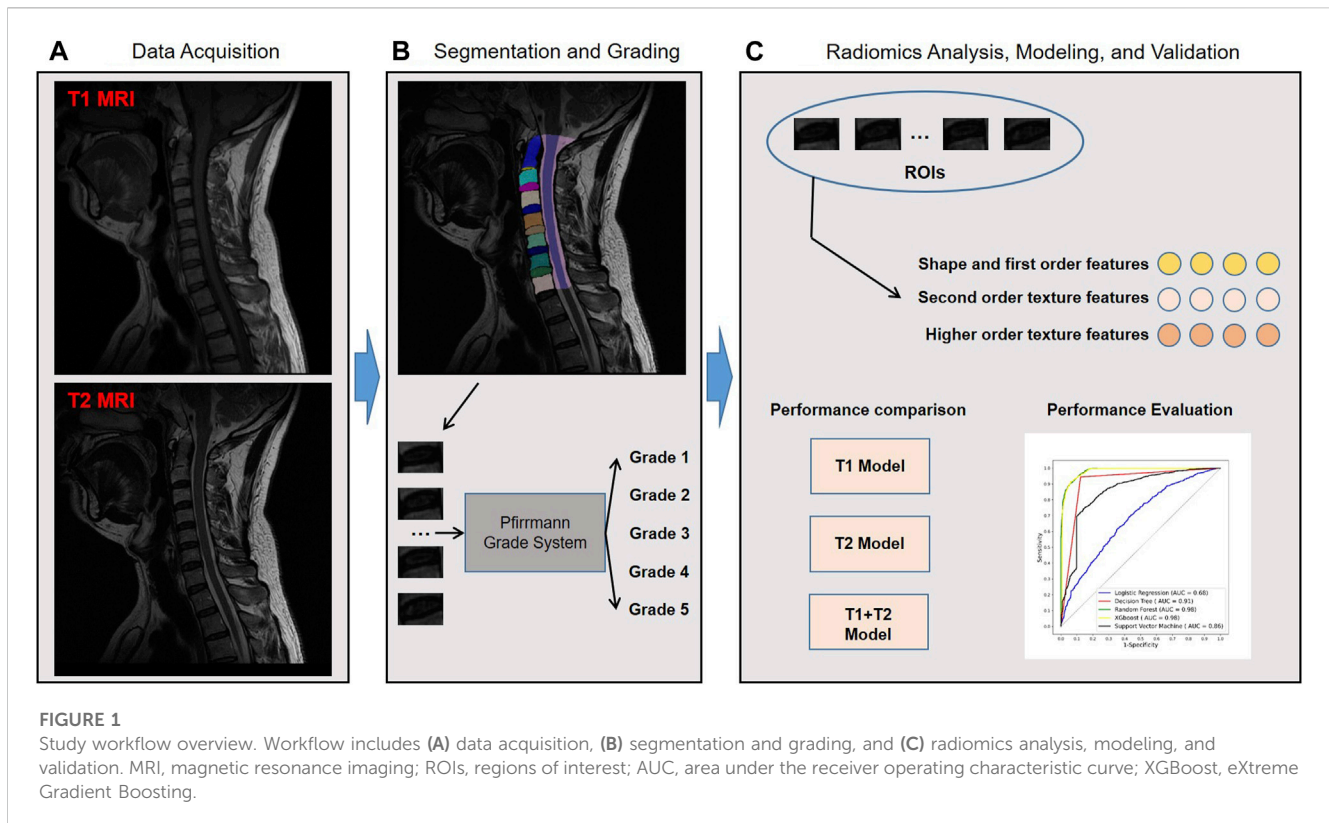
cervical disc degeneration, magnetic resonance imaging, radiomics, machine learning, quantitative image analysis

1 Introduction

Neck pain is a highly prevalent musculoskeletal condition and the fourth leading cause of disability that has become a serious public health issue worldwide, imposing an enormous burden on patients, healthcare system, and the economic structure of countries (Cohen, 2015; Cohen and Hooten, 2017). According to the Global Burden of Disease 2017 study, the number of prevalent cases of neck pain was 288.7 million, and there has been a substantial increment in the past 3 decades (Safiri et al., 2020). A recent study reported that the annual cost for the treatment of neck pain and low back pain was estimated to be \$134.5 billion in the US, ranking first in terms of healthcare spending (Dieleman et al., 2020). Despite its high prevalence and the huge burden imposed on society, it receives relatively less research attention compared to low back pain (Cohen, 2015).

Similar to low back pain, a widely recognized contributor to neck pain is the degeneration of the cervical intervertebral disc (Fujimoto et al., 2012; Risbud and Shapiro, 2014; Theodore, 2020). The intervertebral disc comprises the peripherally located annulus fibrosus (AF), interior gel-like nucleus pulposus (NP), and cartilaginous endplates (CEPs), interposing between the two adjacent vertebral bodies and acting as the shock absorber of the spine. The most physiologically important degenerative changes in the intervertebral disc commence in NP and are usually characterized as decreased water content and loss of disc height, accompanied with the decreased yield strength of AF (Antoniou et al., 1996; Ferrara, 2012). These degenerative changes may alter biomechanical transfer and sensitize the nociceptive nerve fibers in the annulus and nucleus pulposus, which leads to disc herniation, nerve compression, and discogenic pain (Binch et al., 2015; Khan et al., 2017; Theodore, 2020). T2-weighted magnetic resonance imaging (MRI) is the most used imaging modality in the diagnosis of cervical degenerative disc disease (CDDD) due to its superiority in detecting the shape of the intervertebral disc and the water content of NP (Farshad-Amacker et al., 2015). Currently, the most widely used MRI classification system for intervertebral disc degeneration is based on the structure and signal intensity of the disc and the distinction of the nucleus and annulus as proposed by Pfirrmann et al. (2001). While each grade of disc degeneration is clearly defined in this system, it is still a laborious and time-consuming task and highly dependent on the expertise of radiologists and surgeons in clinical practice. The accurate and rapid automated classification of cervical disc degeneration on MRI remains a challenge.

Radiomics and deep learning (DL) approaches have proven to be effective methods for medical image analysis and obtained significant advances in the field of musculoskeletal system disease in recent years (Leung et al., 2020; von Schacky et al., 2020; Huang et al., 2020; Won et al., 2020; Gao et al., 2021; Goedmakers et al., 2021; Bayramoglu et al., 2021; Wang et al., 2022; Hallinan et al., 2021; Niemeyer et al., 2021; Zheng et al., 2022; Abdullah and Rajasekaran, 2022; Gebre et al., 2022). Swiecicki et al. (2021) proposed an automated DL model to evaluate the severity of knee osteoarthritis using knee radiographs according to the Kellgren–Lawrence grading system. Gebre et al. (2022) developed and compared DL models to detect hip osteoarthritis on clinical computer tomography (CT). Hallinan et al. (2021) developed a DL model for the automated detection and classification of central canal, lateral recess, and neural foraminal stenosis in the lumbar spine using sagittal and axial MRI. Several studies reported the feasibility of automated grading of lumbar disc degeneration based on MRI (Gao et al., 2021; Niemeyer et al., 2021; Zheng et al., 2022). Gao et al. (2021) proposed a push–pull regularization strategy to improve the convolutional neural network representation capability for intervertebral disc grading and demonstrated its superior performance. Niemeyer et al. (2021) presented a novel DL-based system for automatically evaluating lumbar disc degeneration according to Pfirrmann grading based on T2-weighted MRI slices, which achieved overall superior reproducibility compared with human interrater. Zheng et al. (2022) developed a segmentation network and a quantitation method to evaluate lumbar intervertebral disc degeneration and calculate the signal intensity and geometric features of disc degeneration. However, there is still a paucity of studies reporting the automated classification of cervical disc degeneration on MRI. The much smaller size of the cervical intervertebral disc compared to the lumbar disc and various anatomical morphologies and MRI signal intensities may increase the difficulty of automatic classification of cervical disc degeneration. The Segment Anything Model (SAM), as a vision foundation model, has shown significant advantages in zero-shot and few-shot segmentations in medical imaging (Cheng et al., 2023; Kirillov et al., 2023; Shi et al., 2023), and many studies have embedded it into the development process (Ma et al., 2023; Mazurowski et al., 2023). Zero-shot or few-shot learning implies that with no or a small amount of data, it is possible to use or fine-tune the foundation model to perform exceptionally well in specific downstream tasks. This addresses real-world challenges like limited clinical data or a lack of annotated data. Here, we can consider using the SAM-based DL model to achieve rapid cervical disc



segmentation and develop classification diagnosis algorithms on this basis.

Therefore, in this study, we aimed to develop and validate an MRI-based radiomics decision support tool for a personalized classification of cervical intervertebral disc degeneration according to the Pfirrmann scheme (Pfirrmann et al., 2001). To the best of our knowledge, this is the first study to develop the radiomics-based automated classification system for the intervertebral disc degeneration of the cervical spine. The imaging difference of discs with different degeneration grading scores may contribute to the understanding of the mechanisms underlying the onset and progression of disc diseases.

2 Materials and methods

2.1 Ethics and study design

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective study was approved by the Ethics Committee of Biomedical Research, West China Hospital (2021-1490). Written informed consent was waived owing to the retrospective nature of data collection (age/gender) and the use of de-identified MRI images.

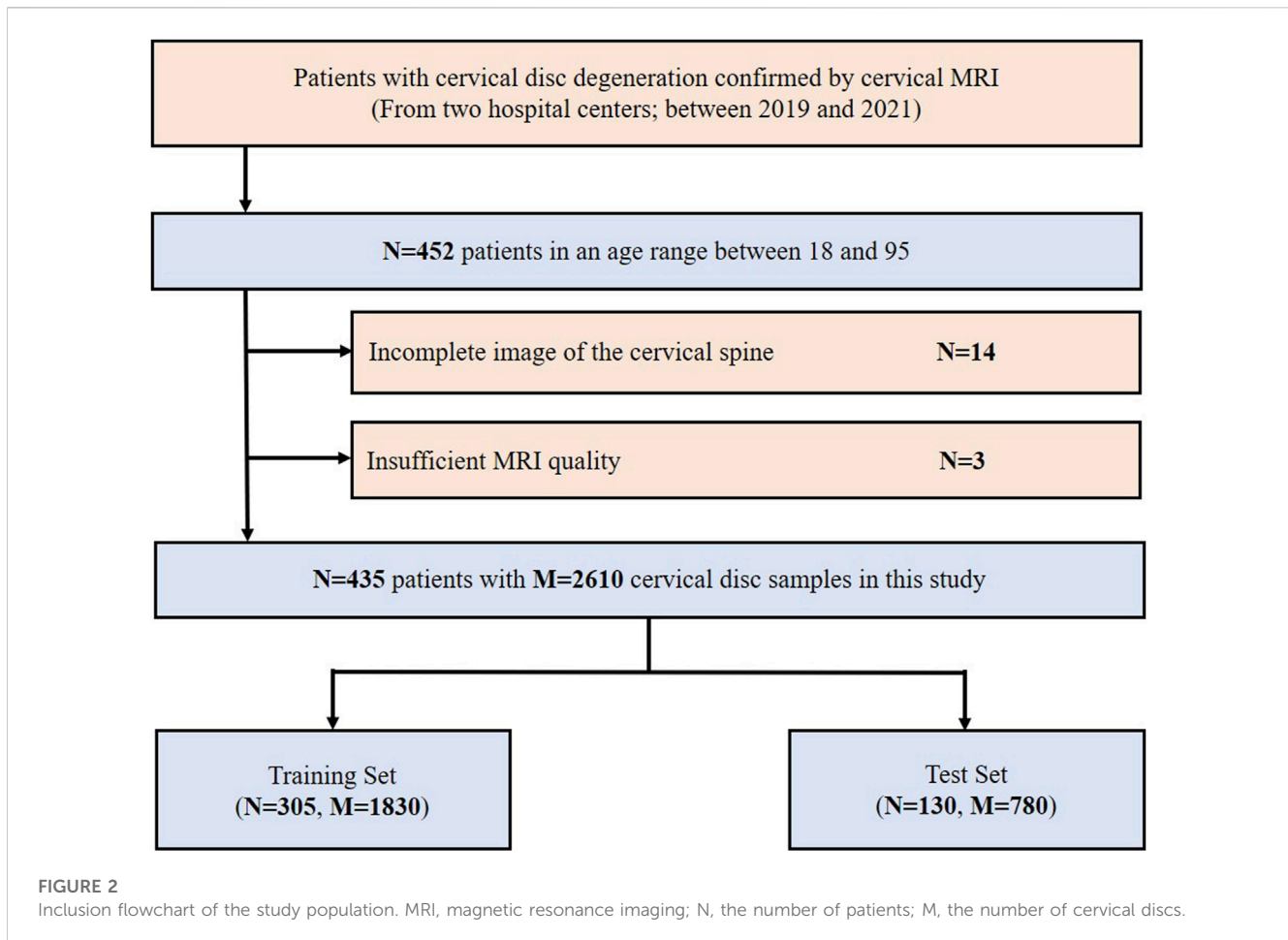
Figure 1 shows the workflow of the study. A total of 2,610 cervical disc samples of 435 patients were retrospectively analyzed. Each patient underwent both T1 and T2 MRI modalities. Each patient had six intercalated discs (C2/3, C3/4, C4/5, C5/6, C6/7, and C7/T1) on the image, which were all semi-automatically segmented using a deep learning-embedded

segmentation tool. Then, a fine-tuned MedSAM model was developed for automated disc segmentation. Meanwhile, the grading classification of cervical disc degeneration was performed using the Pfirrmann grading system. Based on the segmented regions of interest (ROIs), i.e., every disc, we extracted the high-throughput radiomic features including shape and first-order features, second-order texture features, and higher-order texture features. Then, we performed the diagnosis performance comparison between different machine learning algorithms and MRI images. Finally, through comparison, the optimal combined radiomics model was developed in the training set and validated in the test set.

2.2 Study population

A total of 452 consecutive patients aged between 18 and 95, for whom cervical MRI was prescribed for medical reasons, were scanned between 2019 and 2021 at the West China Hospital of Sichuan University and the First People's Hospital of Longquanyi District using either Siemens 3.0T scanners. Overall, 17 patients were excluded for the following reasons: 1) incomplete image of the cervical spine ($n = 14$) and 2) insufficient MRI quality ($n = 3$). Finally, 435 patients were retrospectively collected in this study. The inclusion and exclusion flowchart of the study population is shown in Figure 2.

The data were randomly divided into a training set (1,830 study samples of 305 patients) and a test set (780 study samples of 130 patients) according to the ratio of 7:3. The radiomics analysis, feature selection, and model development were implemented in the training set, and then the related radiomics models were validated in the test set.

**TABLE 1** Clinical characteristics of the study population.

Characteristic	Full (N = 435)	Training set (N = 305)	Test set (N = 130)
Age (years)			
Mean ±	47.66	47.63	47.74
SD	12.87	12.91	12.79
Gender			
Male	298 (68.5%)	213 (69.8%)	85 (65.4%)
Female	137 (31.5%)	92 (30.2%)	45 (34.6%)
Cervical disc samples	2,610	1,830	780
Degeneration grading			
Low grade	794 (30.4%)	527 (28.8%)	267 (34.2%)
High grade	1,816 (69.6%)	1,303 (71.2%)	513 (65.8%)

SD, standard deviation.

2.3 MRI data acquisition

All MRI examinations were performed on a 3.0T MRI system (MAGNETOM Skyra, Siemens, Germany and Discovery 750w, GE, United States) with a phased-array body surface coil. The imaging

protocol was the same for all patients. Images were acquired with a sagittal T2-weighted spin echo sequence with the following parameters: repetition time (TR), 3,064–6,220/3,675–4,200 msec; echo time (TE), 102–104/74 msec; matrix, 288 × 224/128 × 128; field of view (FOV), 256 × 256 mm²; slice thickness, 3.6/3.6 mm; and

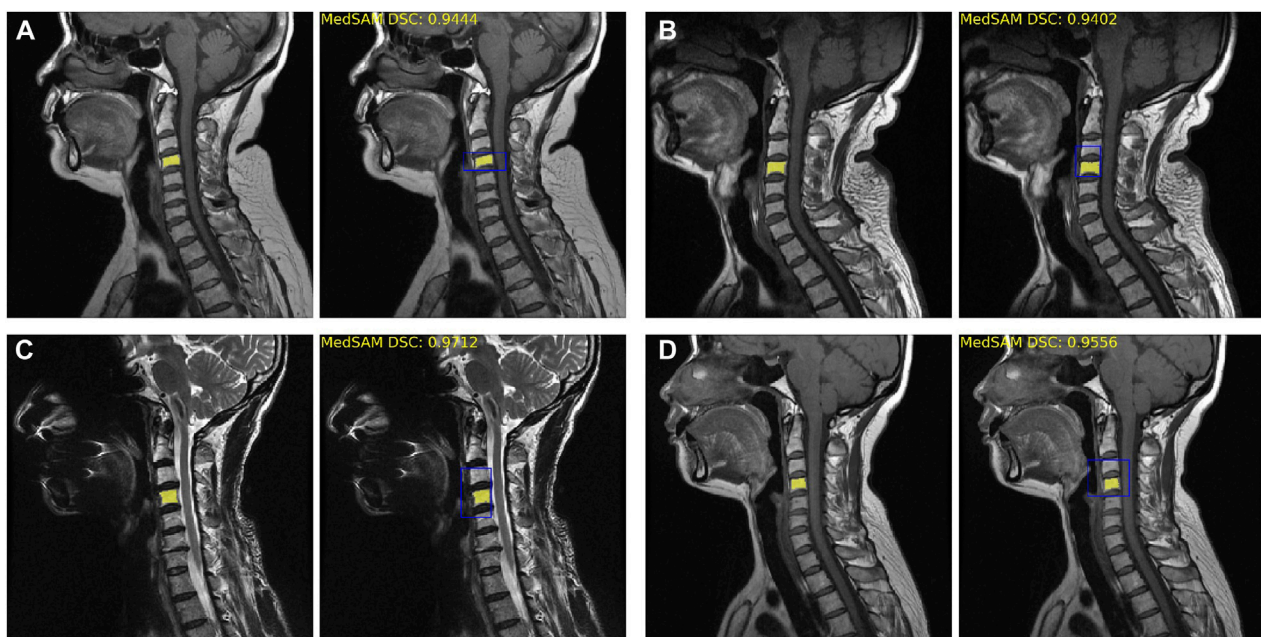


FIGURE 3

Fine-tuned MedSAM segmentation results. (A–D) Four disc segmentation examples. The left side is the real segmentation manually labeled by the doctors, and the right side is the prediction results obtained using MedSAM. The blue bounding box indicates the prompt input, and the yellow area indicates the segmentation results. DSC, Dice similarity coefficient.

intersection gap, 2 mm. T1-weighted MRI was acquired with the following parameters: an inversion time of approximately 400 ms, resolution at $0.8 \times 0.8 \times 0.4$ mm, 25 slices in 3 min 20 s, and single-shot mode with an echo train length (ETL) of 230.

2.4 Fine-tuned MedSAM segmentation model

We segmented the cervical discs from T1 and T2 MRI modalities separately. Six discs (C2/3, C3/4, C4/5, C5/6, C6/7, and C7/T1) were segmented as the independent study samples. First, ROI segmentation was performed by two orthopedic radiologists with approximately 5 years of experience by using Pair software (<https://www.aipair.com.cn/en/>, Version 2.7, RayShape, Shenzhen, China), which embedded deep learning algorithms and trained segmentation models inside (Liang et al., 2022). The final segmentation results were checked and modified by a third radiologist with more than 10 years of experience. By using the DL-based segmentation tool, the work efficiency had been significantly improved.

For achieving fast automated segmentation of discs, we provided a disc segmentation model by fine-tuning the MedSAM model (Ma et al., 2023) based on real results. The image encoder and box prompt encoder were frozen, and the mask encoder was re-trained for the task. Here, we only fine-tuned the MedSAM model on 64 MRI data with real segmentation, and the fine-tuned MedSAM model was evaluated in other datasets. The Dice coefficient was used to evaluate the segmentation performance.

2.5 Disc degeneration grade assessment

The disc degeneration grade assessment was assessed by an experienced orthopedic radiologist, according to the Pfirrmann guideline, conducted by Pfirrmann et al. (2001). The grading system was performed on T2 MRI, and all the discs were classified into five grades (grade 1–5). To facilitate clinical portable use and promote clinical decision-making, we classified grades 1 and 2 as low-grade disc degeneration and grades 3, 4, and 5 as high-grade (Pfirrmann et al., 2001). Based on the gold standard grading results, the radiomics models were developed using supervised learning in the subsequent experiments.

2.6 Radiomics analysis, modeling, and validation

Radiomics analysis was implemented using the PyRadiomics library (<https://www.radiomics.io/pyradiomics.html>, version 3.0.1), which was a commonly used tool for radiomics development (van Griethuysen JJM et al., 2017). The image preprocessing setting followed the previous work (Dong et al., 2022). Then, a total of 924 radiomic features were quantified, including shape and first-order features ($n = 32$), second-order texture features ($n = 73$), and higher-order texture features ($n = 819$). The second-order texture features were calculated using the gray-level co-occurrence matrix (glcm), gray-level run-length matrix (glrlm), gray-level size-zone matrix (glszm), gray-level dependence matrix (gldm), and neighboring gray-tone difference matrix (ngtdm). The higher-order texture features were quantified using the Laplacian of Gaussian (LoG) and wavelet transformation. The details of radiomic features can be seen in PyRadiomics documentation (<https://>

TABLE 2 Top 10 radiomic features on T1 and T2 MRI modalities.

MRI	Transformation	Feature type	Radiomic feature
T1	log-sigma-5-0-mm	First-order	Kurtosis
	wavelet-HH	ngtdm	Busyness
	original	First-order	10 percentile
	wavelet-HH	glcm	IMC1
	log-sigma-5-0-mm	First-order	Minimum
	log-sigma-1-0-mm	First-order	Skewness
	log-sigma-4-0-mm	First-order	10 percentile
	original	Shape	Maximum 2D Diameter Row
	wavelet-HH	glrlm	Run variance
	wavelet-LL	First-order	Kurtosis
T2	log-sigma-5-0-mm	First-order	Kurtosis
	log-sigma-2-0-mm	glszm	Gray-level non-uniformity
	wavelet-LL	glcm	Cluster prominence
	original	glszm	Large area low gray-level emphasis
	wavelet-HL	glcm	IMC1
	log-sigma-2-0-mm	First-order	Skewness
	original	Shape	Maximum 2D diameter column
	wavelet-LL	glcm	Correlation
	log-sigma-3-0-mm	glrlm	Long run low gray-level emphasis
	log-sigma-5-0-mm	gldm	Dependence variance

Log, Laplacian of Gaussian; ngtdm, neighboring gray-tone difference matrix; glcm, gray-level co-occurrence matrix; glrlm, gray-level run-length matrix; glszm, gray-level size-zone matrix; gldm, gray-level dependence matrix; IMC1, informational measure of correlation 1.

pyradiomics.readthedocs.io/en/latest/index.html). Most of them follow the image biomarker standardization initiative (IBSI).

For the important feature selection, minimum redundancy maximum relevance (mRMR) was performed, which was a minimal-optimal feature selection method for finding the smallest relevant feature subset (Zhao et al., 2019). mRMR was used for feature selection in many radiomics studies in recent years (Xie et al., 2021; Hou et al., 2022; Yang et al., 2022).

Here, we defined a machine learning pipeline for model construction and selection. Through mRMR feature selection, the top 10 features were selected for further modeling. For imbalance processing, an adaptive synthetic (ADASYN) algorithm, as a valuable oversampled method in radiomics (Han et al., 2023), was performed to balance the training data. Then, the logistic regression, decision tree, random forest, eXtreme Gradient Boosting (XGBoost), and support vector machine (SVM) were implemented for model construction and comparison. Here, all the machine learning hyperparameter optimization (HPO) was performed using Bayesian optimization, which was the state-of-the-art HPO algorithm (Yang and Shami, 2020). All the models were built in the training set and evaluated in the test set. The performance difference between different machine learning models was evaluated in all sets.

By using the machine learning pipeline, the optimal radiomics models were confirmed on T1 and T2 MRI modalities. Finally, we

combined the selected T1 and T2 radiomic features and built the combined radiomics model in the training set. The final radiomics model was validated in the test set.

2.7 Statistical analysis

All statistical analyses and machine learning algorithms were performed using SPSS (version 25; IBM Corporation) and Python (version 3.8). The Mann–Whitney U test and chi-squared test were implemented for continuous and count variables, respectively. The diagnostic performance was evaluated using the receiver operating characteristic (ROC) curve and AUC. The difference between machine learning models was evaluated using the DeLong test. p -value < 0.05 was considered significantly different.

3 Results

3.1 Clinical characteristics

The clinical characteristics of the study population in the training and test sets are shown in Table 1. There was no significant difference in the two datasets. In the training set,

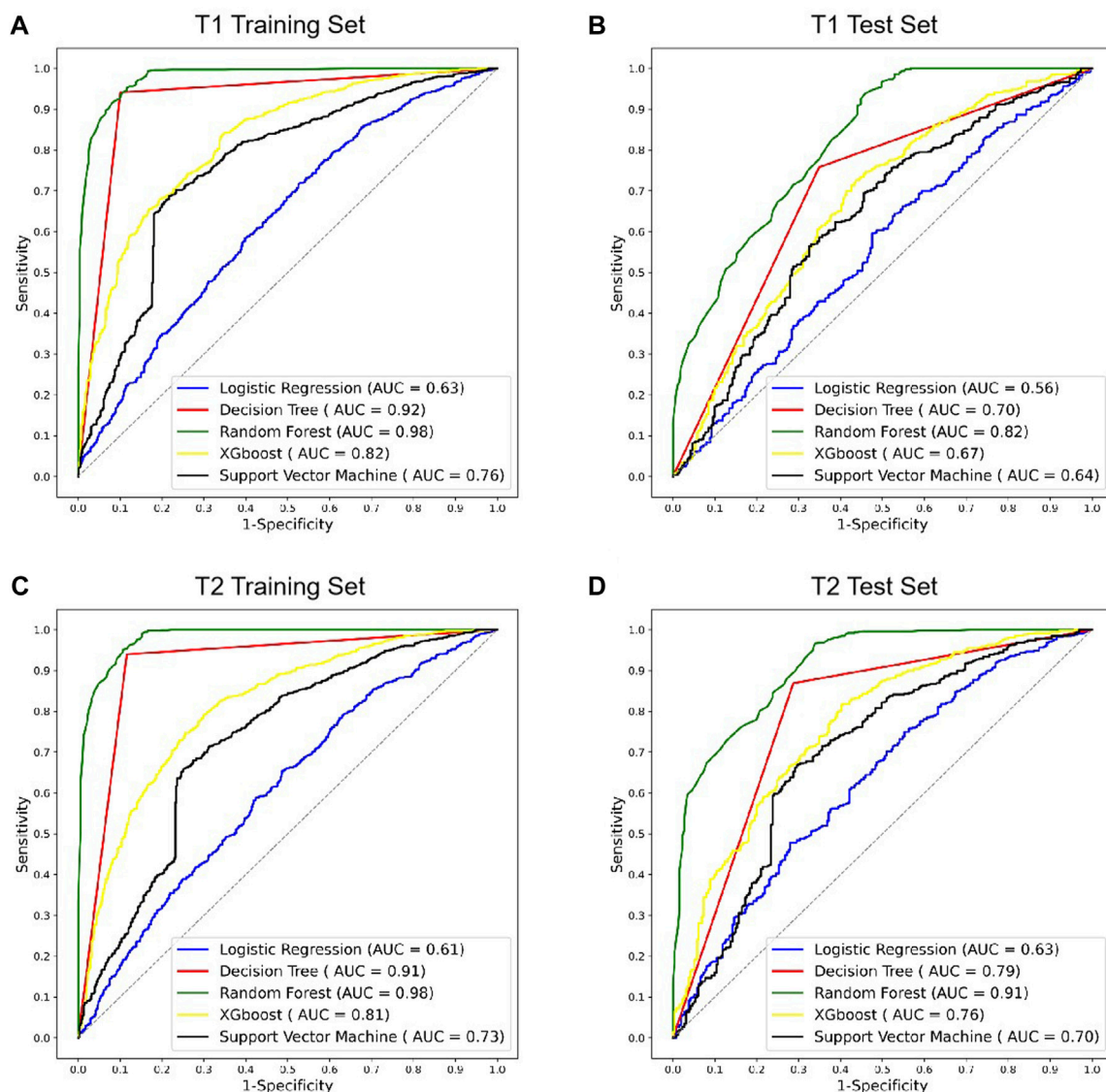


FIGURE 4

ROC curves of different radiomics models in T1 and T2 MRI modalities. (A,B) ROC curves of radiomics models using T1 MRI in the training and test sets. (C,D) ROC curves of radiomics models using T2 MRI in the training and test sets. ROC, receiver operating characteristic; AUC, area under the ROC curve; XGBoost, eXtreme Gradient Boosting.

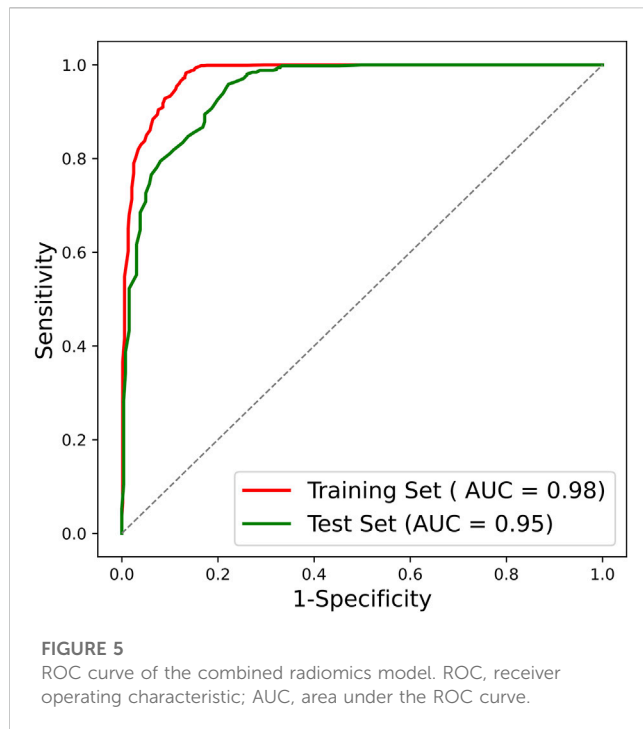
1,303 (71.2%) were defined as the high-grade disc degeneration. Of the test set, 513 (65.8%) disc samples were defined as high grade.

3.2 MedSAM segmentation

The fine-tuned MedSAM count achieved the mean Dice coefficient of 0.93 ± 0.04 . Figure 3 provides four examples of MedSAM segmentation, which all showed the good segmentation performance. When using the fine-tuned MedSAM model, the doctor only needs to roughly give a bounding box (like the blue box), and our model can achieve accurate disc segmentation (the yellow area).

3.3 Radiomic feature discovery

Table 2 shows the top 10 radiomic features in the T1 and T2 MRI modalities. All the features were selected using mRMR. Most of them were higher-order texture features ($N = 16$, 80%), including LoG and wavelet transform features. Of the feature type, first-order ($N = 8$, 40%) and glcm ($N = 4$, 20%) were the dominant factors. Compared with T1 and T2 radiomic features, we observe that the key features were relatively similar. In particular, the kurtosis feature appeared three times, and the kurtosis of log-sigma-5.0-mm first-order feature was the highest ranked radiomic feature in both MRI modalities.



3.4 Diagnostic performance across T1 and T2 MRI modalities

The radiomics models of various machine learning algorithms and MRI images were constructed and compared (Figure 4). In all the modes, random forest obtained the higher diagnostic performance than other machine learning algorithms (AUC = 0.82, $p < 0.01$, T1 test set; AUC = 0.91, $p < 0.01$, T2 test set). The performance of T2 MRI was significantly higher than T1 MRI ($p < 0.05$).

3.5 Diagnostic performance of the combined radiomics model

After confirming the selected radiomic features and random forest modeling method, the final combined radiomics model was constructed and evaluated. Figure 5 shows the ROC curves of the combined model. The AUC of the training ROC was 0.98 and the test AUC was 0.95. Table 3 shows the diagnostic performance between the final T1, T2, and combined radiomics models. The combined model had the accuracy of 89.51%, the precision of

87.07%, the recall of 98.83%, and the F1 score of 0.93 in the test set, which was better than those of other models ($p < 0.05$).

3.6 Radiomic feature mapping

Here, in order to provide the visualization tools that are easy to use clinically, the radiomic feature maps are provided (Figure 6). Only the top three feature maps of the two modalities were used to aid in diagnosis, enabling some mode differences to be seen on these virtual medical imaging through the comparison of low- and high-grade disc degeneration.

4 Discussion

There is still a paucity of studies reporting the automated grading of cervical disc degeneration on MRI. In clinical practice, the automated decision support tool to intelligently grade disc degeneration through MRI will greatly assist physicians in individualized patient management. In this study, we developed and validated an MRI radiomics-based cervical disc degeneration grading method, which can automatically segment the disc ROIs, extract valuable radiomic features, and predict the degeneration grades, showing a high diagnostic performance (AUC = 0.98 in the training set; AUC = 0.95 in the test set).

We found that both T1 and T2 MRI modalities showed good diagnostic results, although T2 showed higher performance than T1. To the best of our knowledge, T2 MRI is the most used MRI modality in the diagnosis of cervical degenerative disc disease (Farshad-Amacker et al., 2015), and many studies developed AI models only based on T2 MRI (Gao et al., 2021; Zheng et al., 2022). Our study showed that even though T1 MRI was macroscopically difficult to use directly for disc grading, it was still possible to construct the diagnostic model with good performance through radiomics and machine learning methods. By integrating the valuable information in T1 and T2, the performance of the combined model will significantly be improved, which may have some positive hints for clinical practice.

For the valuable radiomic features, higher-order texture features showed the dominant force, which was consistent with the findings of previous studies (Jiang et al., 2022a; Jiang et al., 2022b). Texture features can adequately characterize the heterogeneous information between tumors or inflammation (Alobaidli et al., 2014; Song et al., 2023), and the higher-order transformations (wavelet, LoG, convolutional neural network, etc.) may have the potential to further enhance the expression of this heterogeneity, contributing

TABLE 3 Diagnosis performance of the final radiomics models.

Model	Training set				Test set			
	Accuracy (%)	Precision (%)	Recall (%)	F1	Accuracy (%)	Precision (%)	Recall (%)	F1
T1 MRI	94.95	94.07	99.15	0.96	80.72	79.13	96.28	0.87
T2 MRI	93.96	93.73	98.07	0.95	85.89	83.75	97.66	0.90
Combined	94.94	94.38	98.76	0.97	89.51	87.07	98.83	0.93

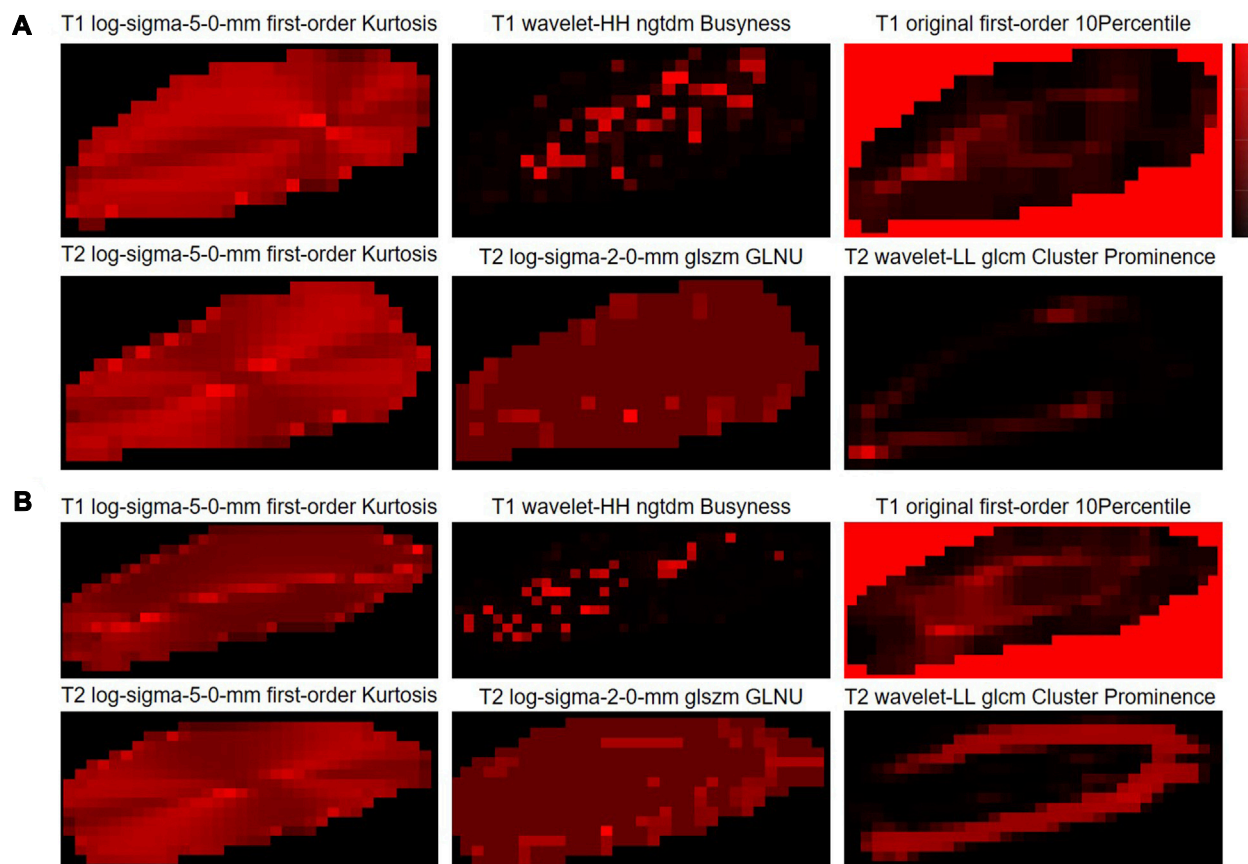


FIGURE 6

Radiomic feature maps for the classification of cervical disc degeneration. (A) Low-grade cervical disc and (B) high-grade cervical disc. Log, Laplacian of Gaussian; ngtdm, neighboring gray-tone difference matrix; glszm, gray-level size-zone matrix; GLNU, gray-level non-uniformity; glcm, gray-level co-occurrence matrix.

to the diagnostic performance (Jiang et al., 2022a; Jiang et al., 2022b). The mRMR, ADASYN, and Bayesian optimization algorithms also showed the excellent selection performance in radiomics analysis, which is consistent with the findings of Xie et al. (2021), Hou et al. (2022), Wang et al. (2023), and Wu et al. (2023).

Here, we also provided the radiomic feature maps for the classification of cervical disc degeneration (Figure 5). We can still visualize relatively well the differences in patterns between low- and high-grade degeneration, especially the log-sigma-5.0-mm first-order kurtosis. However, in fact, the difference in this virtual imaging was not very significant, and it was still difficult to be used as an independent imaging biomarker that gives clinicians a direct and significant indication. However, to a certain extent, the imaging difference of discs with different degeneration grading scores may help understand the mechanisms underlying the onset and progression of disc diseases.

In the study, we used each cervical disc as an independent study sample, and all six different cervical discs in each person were mixed to perform the modeling analysis. The better diagnostic performance showed that the different cervical discs could be identified basically using the same mode. In addition, the gold standard for Pfirrmann assessment usually classifies discs into five categories, and for the purpose of clinical decision-making we have used only two classifications: low grade (grades 1 and 2)

and high grade (grades 3, 4, and 5). In future studies, we will further expand the amount of data to build clinical tools for automatic segmentation and five-category diagnosis. Moreover, we will also consider other clinical grading criteria (Adams and Dolan, 2012; Wang, 2018) and compare the performance differences between machine learning models built under different criteria.

There are also some limitations to the study. First, as a retrospective study, there was no clinical information enrolled. Perhaps adding the broader range of clinical factors could further enhance the performance of the model. Second, although it was a two-center study, there was still a need to further expand the validation of independent center data for the decision support tool to be further promoted and validated. Therefore, a larger multi-center study is needed.

5 Conclusion

In conclusion, we demonstrated that the radiomics-based decision support tool by integrating T1 and T2 MRI modalities can be used for a personalized classification of cervical disc degeneration, showing the robust diagnostic performance, and may aid in clinical decision-making and individualized management.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Biomedical Research, West China Hospital (2021-1490). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

JX: conceptualization, data curation, formal analysis, validation, writing–original draft, and writing–review and editing. YY: conceptualization, data curation, funding acquisition, investigation, project administration, resources, validation, writing–original draft, and writing–review and editing. ZJ: conceptualization, methodology, software, validation, visualization, writing–original draft, and writing–review and editing. KZ: data curation, formal analysis, investigation, and writing–original draft. XZ: data curation, formal analysis, investigation, and writing–original draft. YS: formal analysis, investigation, and writing–review and editing. XJ: data curation, investigation, software, and writing–review and editing. HL:

supervision, validation, and writing–review and editing. SY: formal analysis, investigation, and writing–review and editing. YJ: conceptualization, formal analysis, project administration, writing–original draft, and writing–review and editing. LM: conceptualization, funding acquisition, project administration, resources, supervision, writing–original draft, and writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Health Commission of Sichuan Province, Project of China (No. 21PJ037), and the Sichuan Science and Technology Program (No. 2023YFG0126).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdullah, S. S., and Rajasekaran, M. P. (2022). Automatic detection and classification of knee osteoarthritis using deep learning approach. *Radiol. Med.* 127 (4), 398–406. doi:10.1007/s11547-022-01476-7
- Adams, M. A., and Dolan, P. (2012). Intervertebral disc degeneration: evidence for two distinct phenotypes. *J. Anat.* 221, 497–506. doi:10.1111/j.1469-7580.2012.01551.x
- Alobaidli, S., McQuaid, S., South, C., Prakash, V., Evans, P., and Nisbet, A. (2014). The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning. *Br. J. Radiol.* 87 (1042), 20140369. doi:10.1259/bjr.20140369
- Antoniou, J., Steffen, T., Nelson, F., Winterbottom, N., Hollander, A. P., Poole, R. A., et al. (1996). The human lumbar intervertebral disc: evidence for changes in the biosynthesis and denaturation of the extracellular matrix with growth, maturation, ageing, and degeneration. *J. Clin. Invest.* 98 (4), 996–1003. doi:10.1172/JCI118884
- Bayramoglu, N., Nieminen, M. T., and Saarakkala, S. (2021). Automated detection of patellofemoral osteoarthritis from knee lateral view radiographs using deep learning: data from the Multicenter Osteoarthritis Study (MOST). *Osteoarthr. Cartil.* 29 (10), 1432–1447. doi:10.1016/j.joca.2021.06.011
- Binch, A. L., Cole, A. A., Breakwell, L. M., Michael, A. L., Chiverton, N., Creemers, L. B., et al. (2015). Nerves are more abundant than blood vessels in the degenerate human intervertebral disc. *Arthritis Res. Ther.* 17, 370. doi:10.1186/s13075-015-0889-6
- Cheng, D., Qin, Z., Jiang, Z., Zhang, S., Lao, Q., and Kang, L. (2023). Sam on medical images: a comprehensive study on three prompt modes. <https://arxiv.org/abs/2305.00035>.
- Cohen, S. P. (2015). Epidemiology, diagnosis, and treatment of neck pain. *Mayo Clin. Proc.* 90 (2), 284–299. doi:10.1016/j.mayocp.2014.09.008
- Cohen, S. P., and Hooten, W. M. (2017). Advances in the diagnosis and management of neck pain. *Bmj* 358, j3221. doi:10.1136/bmj.j3221
- Dieleman, J. L., Cao, J., Chapin, A., Chen, C., Li, Z., Liu, A., et al. (2020). US health care spending by payer and health condition, 1996–2016. *Jama* 323 (9), 863–884. doi:10.1001/jama.2020.0734
- Dong, Y., Jiang, Z., Li, C., Dong, S., Zhang, S., Lv, Y., et al. (2022). Development and validation of novel radiomics-based nomograms for the prediction of EGFR mutations and Ki-67 proliferation index in non-small cell lung cancer. *Quant. Imaging Med. Surg.* 12 (5), 2658–2671. doi:10.21037/qims-21-980
- Farshad-Amacker, N. A., Farshad, M., Winklehner, A., and Andreisek, G. (2015). MR imaging of degenerative disc disease. *Eur. J. Radiol.* 84 (9), 1768–1776. doi:10.1016/j.ejrad.2015.04.002
- Ferrara, L. A. (2012). The biomechanics of cervical spondylosis. *Adv. Orthop.* 2012, 493605. doi:10.1155/2012/493605
- Fujimoto, K., Miyagi, M., Ishikawa, T., Inoue, G., Eguchi, Y., Kamoda, H., et al. (2012). Sensory and autonomic innervation of the cervical intervertebral disc in rats: the pathomechanics of chronic discogenic neck pain. *Spine (Phila Pa 1976)* 37 (16), 1357–1362. doi:10.1097/BRS.0b013e31824ba710
- Gao, F., Liu, S., Zhang, X., Wang, X., and 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. doi:10.1002/jmri.27400
- Gebr, R. K., Hirvasniemi, J., van der Heijden, R. A., Lantto, I., Saarakkala, S., Leppilahti, J., et al. (2022). Detecting hip osteoarthritis on clinical CT: a deep learning application based on 2-D summation images derived from CT. *Osteoporos. Int.* 33 (2), 355–365. doi:10.1007/s00198-021-06130-y
- Goedmakers, C. M. W., Lak, A. M., Duey, A. H., Senko, A. W., Arnaout, O., Groff, M. W., et al. (2021). Deep learning for adjacent segment disease at preoperative MRI for cervical radiculopathy. *Radiology* 301 (3), 664–671. doi:10.1148/radiol.2021204731
- Hallinan, JTPD, Zhu, L., Yang, K., Makmur, A., Algazwi, D. A. R., Thian, Y. L., et al. (2021). Deep learning model for automated detection and classification of central canal,

- lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology* 300 (1), 130–138. doi:10.1148/radiol.2021204289
- Han, P. L., Jiang, Z. K., Gu, R., Huang, S., Jiang, Y., Yang, Z. G., et al. (2023). Prognostic prediction of left ventricular myocardial noncompaction using machine learning and cardiac magnetic resonance radiomics. *Quant. Imaging Med. Surg.* 13 (10), 6468–6481. doi:10.21037/qims-23-372
- Hou, J., Li, H., Zeng, B., Pang, P., Ai, Z., Li, F., et al. (2022). MRI-based radiomics nomogram for predicting temporal lobe injury after radiotherapy in nasopharyngeal carcinoma. *Eur. Radiol.* 32 (2), 1106–1114. doi:10.1007/s00330-021-08254-5
- Huang, J., Shen, H., Wu, J., Hu, X., Zhu, Z., Lv, X., et al. (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. doi:10.1016/j.spinee.2019.11.010
- Jiang, Z., Wang, B., Han, X., Zhao, P., Gao, M., Zhang, Y., et al. (2022a). Multimodality MRI-based radiomics approach to predict the posttreatment response of lung cancer brain metastases to gamma knife radiosurgery. *Eur. Radiol.* 32 (4), 2266–2276. doi:10.1007/s00330-021-08368-w
- Jiang, Z., Yin, J., Han, P., Chen, N., Kang, Q., Qiu, Y., et al. (2022b). Wavelet transformation can enhance computed tomography texture features: a multicenter radiomics study for grade assessment of COVID-19 pulmonary lesions. *Quant. Imaging Med. Surg.* 12 (10), 4758–4770. doi:10.21037/qims-22-252
- Khan, A. N., Jacobsen, H. E., Khan, J., Filippi, C. G., Levine, M., Lehman, R. A., Jr, et al. (2017). Inflammatory biomarkers of low back pain and disc degeneration: a review. *Ann. N. Y. Acad. Sci.* 1410 (1), 68–84. doi:10.1111/nyas.13551
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). Segment anything. Available at: <https://arxiv.org/abs/2304.02643>.
- Leung, K., Zhang, B., Tan, J., Shen, Y., Geras, K. J., Babb, J. S., et al. (2020). Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the osteoarthritis initiative. *Radiology* 296 (3), 584–593. doi:10.1148/radiol.2020192091
- Liang, J., Yang, X., Huang, Y., Li, H., He, S., Hu, X., et al. (2022). Sketch guided and progressive growing GAN for realistic and editable ultrasound image synthesis. *Med. Image Anal.* 79, 102461. doi:10.1016/j.media.2022.102461
- Ma, J., He, Y., Li, F., You, C., and Wang, B. (2023). Segment anything in medical images. Available at: <https://arxiv.org/abs/2304.12306>.
- Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., and Zhang, Y. (2023). Segment anything model for medical image analysis: an experimental study. *Med. Image Anal.* 89, 102918. doi:10.1016/j.media.2023.102918
- Niemeyer, F., Galbusera, F., Tao, Y., Kienle, A., Beer, M., and Wilke, H. J. (2021). A deep learning model for the accurate and reliable classification of disc degeneration based on MRI data. *Invest. Radiol.* 56 (2), 78–85. doi:10.1097/RLI.0000000000000709
- Pfirrmann, C. W., Metzendorf, A., Zanetti, M., Hodler, J., and Boos, N. (2001). Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)* 26 (17), 1873–1878. doi:10.1097/00007632-200109010-00011
- Risbud, M. V., and Shapiro, I. M. (2014). Role of cytokines in intervertebral disc degeneration: pain and disc content. *Nat. Rev. Rheumatol.* 10 (1), 44–56. doi:10.1038/nrrheum.2013.160
- Safiri, S., Kolahi, A. A., Hoy, D., Buchbinder, R., Mansournia, M. A., Bettampadi, D., et al. (2020). Global, regional, and national burden of neck pain in the general population, 1990–2017: systematic analysis of the Global Burden of Disease Study 2017. *Bmj* 368, m791. doi:10.1136/bmj.m791
- Shi, P., Qiu, J., Abaxi, S. M. D., Wei, H., Lo, F. P., and Yuan, W. (2023). Generalist vision foundation models for medical imaging: a case study of segment anything model on zero-shot medical segmentation. *Diagn. (Basel)* 13 (11), 1947. doi:10.3390/diagnostics13111947
- Song, M. X., Yang, H., Yang, H. Q., Li, S. S., Qin, J., and Xiao, Q. (2023). MR imaging radiomics analysis based on lumbar soft tissue to evaluate lumbar fascia changes in patients with low back pain. *Acad. Radiol.* 30 (23), 2450–2457. doi:10.1016/j.acra.2023.02.038
- Swiecicki, A., Li, N., O'Donnell, J., Said, N., Yang, J., Mather, R. C., et al. (2021). Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists. *Comput. Biol. Med.* 133, 104334. doi:10.1016/j.cmpbiomed.2021.104334
- Theodore, N. (2020). Degenerative cervical spondylosis. *N. Engl. J. Med.* 383 (2), 159–168. doi:10.1056/NEJMra2003558
- van Griethuysen Jjm, Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77 (21), e104–e107. doi:10.1158/0008-5472.CAN-17-0339
- von Schacky, C. E., Sohn, J. H., Liu, F., Ozhinsky, E., Jungmann, P. M., Nardo, L., et al. (2020). Development and validation of a multitask deep learning model for severity grading of hip osteoarthritis features on radiographs. *Radiology* 295 (1), 136–145. doi:10.1148/radiol.2020190925
- Wang, L., Wu, X., Tian, R., Ma, H., Jiang, Z., Zhao, W., et al. (2023). MRI-based pre-Radiomics and delta-Radiomics models accurately predict the post-treatment response of rectal adenocarcinoma to neoadjuvant chemoradiotherapy. *Front. Oncol.* 13, 1133008. doi:10.3389/fonc.2023.1133008
- Wang, Y., Bi, Z., Xie, Y., Wu, T., Zeng, X., Chen, S., et al. (2022). Learning from highly confident samples for automatic knee osteoarthritis severity assessment: data from the osteoarthritis initiative. *IEEE J. Biomed. Health Inf.* 26 (3), 1239–1250. doi:10.1109/JBHI.2021.3102090
- Wang, Y. X. J. (2018). Senile osteoporosis is associated with disc degeneration. *Quant. Imaging Med. Surg.* 8 (6), 551–556. doi:10.21037/qims.2018.07.04
- Won, D., Lee, H. J., Lee, S. J., and Park, S. H. (2020). Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. *Spine (Phila Pa 1976)* 45 (12), 804–812. doi:10.1097/BRS.00000000000003377
- Wu, K., Miu, X., Wang, H., and Li, X. (2023). A Bayesian optimization tuning integrated multi-stacking classifier framework for the prediction of radiodermatitis from 4D-CT of patients underwent breast cancer radiotherapy. *Front. Oncol.* 13, 1152020. doi:10.3389/fonc.2023.1152020
- Xie, Y., Zhao, H., Guo, Y., Meng, F., Liu, X., Zhang, Y., et al. (2021). A PET/CT nomogram incorporating SUVmax and CT radiomics for preoperative nodal staging in non-small cell lung cancer. *Eur. Radiol.* 31 (8), 6030–6038. doi:10.1007/s00330-020-07624-9
- Yang, F., Zhang, J., Zhou, L., Xia, W., Zhang, R., Wei, H., et al. (2022). CT-based radiomics signatures can predict the tumor response of non-small cell lung cancer patients treated with first-line chemotherapy and targeted therapy. *Eur. Radiol.* 32 (3), 1538–1547. doi:10.1007/s00330-021-08277-y
- Yang, L., and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415, 295–316. doi:10.1016/j.neucom.2020.07.061
- Zhao, Z., Anand, R., and Wang, M. (2019). “Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform,” in Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Washington, DC, USA, October 2019 (IEEE), 442–452. doi:10.1109/DSAA.2019.00059
- Zheng, H. D., Sun, Y. L., Kong, D. W., Yin, M. C., Chen, J., Lin, Y. P., et al. (2022). Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. *Nat. Commun.* 13 (1), 841. doi:10.1038/s41467-022-28387-5