bioengineering

MDPI

Article

MRI-Based Machine Learning and Radiomics Methods for Assessing Spinal Cord Function in Patients with Mild Cervical Spondylotic Myelopathy

He Wang [1,†], Kai Wang [1,†], Yutian Wang [1,†], Zhenlei Liu [1], Lei Zhang [1], Shanhang Jia [1], Kun He [2,3], Xiangyu Zhang [1] and Hao Wu [1,*]

[1] Department of Neurosurgery, Xuanwu Hospital, Capital Medical University, 45 Changchun St, Xicheng District, Beijing 100053, China; wanghe_15@126.com (H.W.); wangkaispine@xwh.ccmu.edu.cn (K.W.); wangyutian@mail.ccmu.edu.cn (Y.W.); zhenlei.liu@xwhosp.org (Z.L.); zhanglei@xwhosp.org (L.Z.); shanhangjia@126.com (S.J.); zhangxiangyu@mail.ccmu.edu.cn (X.Z.)
[2] Department of Neurosurgery, China-Japan Friendship School of Clinical Medicine, Peking University, No. 2 Yinghuayuan East Street, Chaoyang District, Beijing 100029, China; hekun39@foxmail.com
[3] Department of Neurosurgery, China-Japan Friendship Hospital, No. 2 Yinghuayuan East Street, Chaoyang District, Beijing 100029, China
* Correspondence: wuhaospine@xwh.ccmu.edu.cn
† These authors contributed equally to this work.

Abstract: (1) Background: Patients with mild cervical spondylotic myelopathy (CSM) who delay surgery risk progression. While PET evaluates spinal cord function, its cost and radiation limit its use. (2) Methods: In this prospective study, patients with mild cervical spondylosis underwent preoperative 18F-FDG PET-MRI. Narrowed spinal levels were classified based on whether SUV_{max} was decreased. Follow-up assessments were conducted. Two machine learning models using MRI T2-based radiomics were developed to identify stenotic levels and decreased SUV_{max}. (3) Results: Patients with normal SUV_{max} showed greater symptom improvement. The radiomics models performed well, with AUCs of 0.981/0.962 (training/testing) for stenosis detection and 0.830/0.812 for predicting SUV_{max} decline. The model outperformed clinicians in predicting SUV_{max} decline, improving the AUC by 10%. (4) Conclusion: Patients with preserved SUV_{max} have better outcomes. MRI-based radiomics shows potential for identifying stenosis and predicting spinal cord function changes for preoperative assessment, though larger studies are needed to validate its clinical utility.

Keywords: cervical spondylotic myelopathy; PET-MRI; radiomics; machine learning

check for updates

Academic Editors: Fabiano Bini and Pietro Aricò

Received: 24 April 2025
Revised: 5 June 2025
Accepted: 13 June 2025
Published: 17 June 2025

Citation: Wang, H.; Wang, K.; Wang, Y.; Liu, Z.; Zhang, L.; Jia, S.; He, K.; Zhang, X.; Wu, H. MRI-Based Machine Learning and Radiomics Methods for Assessing Spinal Cord Function in Patients with Mild Cervical Spondylotic Myelopathy. *Bioengineering* **2025**, *12*, 666. https://doi.org/10.3390/bioengineering12060666

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Degenerative cervical myelopathy (DCM) is the most common cause of spinal cord dysfunction and most importantly of patients with mild myelopathic symptoms [1]. According to AOSpine clinical practice guidelines for the management of DCM, the quality of evidence is low for providing surgical intervention for mild cervical spondylotic myelopathy (CSM) patients [2]. Mild CSM patients may initially refuse surgery due to mild symptoms; however, a significant proportion (20–62%) experience worsening conditions within 3–6 years, necessitating surgical intervention [3,4]. Therefore, the timely assessment of spinal cord function in mild CSM patients is crucial for determining surgical intervention and preventing disease progression.

Positron Emission Tomography (PET) based on 18F-FDG is currently an important method for evaluating cervical spinal cord function [5–8]. Previous studies have indicated

that cervical spondylosis patients can be classified into two types based on the maximum standardized uptake value (SUV_{max}): Type 1, with an increased or normal SUV_{max} in the compressed segment, characterized by a short medical history and rapid disease progression; and Type 2, with a decreased SUV_{max} in the compressed segment, characterized by a long medical history and chronic disease progression [9]. Previous research showed that patients with Type 1 cervical spondylosis typically experience significant improvement in symptoms after surgery, while symptoms in Type 2 patients show less noticeable improvement [10]. Therefore, conducting PET scans on patients can preoperatively assess spinal cord function, which is of significant importance for predicting patient prognosis. However, there is currently no research describing whether PET has evaluative value for spinal cord function in mild CSM patients. Furthermore, due to the radioactive nature of PET and its relatively high cost, conducting PET scans for all CSM patients is not advisable. Consequently, there is an urgent need for an alternative to PET for assessing spinal cord function.

To reduce radiation exposure and hospitalization costs, utilizing MRI to assess spinal cord function has become a main alternative to PET. Some studies reported a negative correlation between the high abnormal intramedullary signal on preoperative T2-weighted images and neurological improvement rates [11]. However, one study indicated that the signal on T2-weighted images does not correlate with prognosis [12]. Therefore, there remain challenges to objectively assess cervical spinal cord function using MRI.

Radiomics is a method of extracting numerous features from images and creating models based on these features to accurately analyze images [13]. Currently, numerous studies have employed radiomics to extract imaging features such as intracranial lesions and pathological changes, modeling predictions for aspects like drug treatment response and prognosis [14,15]. However, there is still a lack of reported research on spinal cord function-related studies based on radiomics.

In this study, we constructed a model based on radiomics to assess spinal cord function. The main contributions are as follows:

1. We prospectively collected a cohort of mild CSM patients and elucidated the relationship between decreased SUV_{max} and prognosis.
2. We constructed a model based on radiomics capable of identifying compressed cervical segments.
3. We developed a model based on radiomics capable of identifying segments with decreased SUV_{max}.
4. We conducted feature analysis on the models, yielding radiomic indicators with clinical relevance to guide clinical practice.

2. Methods

2.1. Study Population

This study prospectively enrolled mild CSM patients at our hospital, from January 2023 to June 2023. Although, according to AOSpine clinical practice guidelines, mild cervical spondylotic myelopathy (CSM) is defined by modified Japanese Orthopaedic Association (mJOA [16]) in the range of 15–17, to enroll more patients, we selected relatively mild CSM patients with mJOA $\geq$ 14 [2].

Inclusion criteria were (1) a radiological diagnosis of cervical spinal cord compression, (2) preoperative PET and MRI imaging examinations, and (3) relatively mild symptoms (mJOA $\geq$ 14). Exclusion criteria were (1) cervical spine tumors, (2) tuberculosis or other infectious diseases, and (3) a history of previous cervical spine surgery.

All patients were reassessed for mJOA scores 6 months after surgery, and the improvement in mJOA scores was calculated to evaluate treatment efficacy.

2.2. Image Scanning Process and Analysis

For all patients, preoperative PET-MRI scans were conducted using a scanner from UTH Manufacturer, with MRI sequences including coronal, sagittal, and axial T1- and T2-weighted imaging. The T2-weighted coronal parameters used for analysis were TR 4690 ms, TE 114 ms, pixel spacing of 0.4584 mm, and slice thickness of 3.3 mm. All patients underwent a minimum 12 h fasting period before PET studies. A dose of 370 MBq of 18F-FDG, with an effective radiation dose of 10 mSv, was injected one hour before the PET studies. The scanned PET images were limited to the cervical region, with a pixel spacing of 3.125 mm and a slice thickness of 2.78 mm.

The 3D-Slicer software (version 5.3.0) was employed to annotate the spinal cord in the five intervertebral disc segments (C2/3 to C6/7) by two neurosurgeons with 5 years of experience. The SUV_{max} of PET within these five segmental regions of interest (ROIs) was calculated [17]. These values represented the functional status of each cervical spinal cord segment.

2.3. Radiomic Feature Extraction

Based on segmented images, features were extracted from the axial T2-weighted imaging sequences using the pyradiomics software package (version 3.0) (https://github.com/Radiomics/pyradiomics, accessed on 1 January 2024). Due to the relatively small spinal cord area, a small sampling preprocessing step was designed to achieve finer resolution. Firstly, all images were resampled to isotropic voxels of $0.5 \times 0.5 \times 0.5$ mm^3. Voxel intensities were normalized and discretized at intervals of 5. Subsequently, four Laplacian of Gaussian (LoG) filters with sigma values ranging from 0.5 to 2.0 at intervals of 0.5 were used. Finally, 1197 radiomic features were extracted.

2.4. Clinical Task Setting

In this study, two tasks were defined for spinal cord segments. The first task involved determining whether the spinal cord segment was compressed, with two spinal cord imaging experts with over 10 years of experience making decisions to assess the degree of compression in each segment. The second task was to assess the function of the spinal cord segment, based on the SUV_{max} values from PET scans. If the SUV_{max} value in a level decreased by 10% compared to the nearest normal level, it indicated impaired spinal cord function. Segments with increased or unchanged SUV_{max} were considered to have intact spinal cord function.

Therefore, please note that in this study, levels with normal or increased SUV_{max} were both listed as normal SUV_{max} levels.

The training and testing sets were randomly divided in a 3:1 ratio at the patient level, with all five cervical spinal cord segments (C2/3 to C6/7) extracted for validation in task one (determining whether the cervical spinal cord is compressed). For task two (determining whether the cervical spinal cord function is impaired), only compressed segments were used for training and validation.

2.5. Construction of the Machine Learning Model

For both clinical tasks, the same procedure was employed to construct machine learning models. Over 1197 clinical features were extracted using the feature extraction method. To eliminate irrelevant and redundant data, intraclass correlation coefficient (ICC) values and hierarchical clustering were used to select features in an unsupervised learning manner [13]. To further select representative features, recursive feature elimination (RFE) was employed for supervised feature selection. Nine typical machine learning algorithms, including LinearSVC, Random Forest Classifier [18], Extra Trees [19], KNN, Decision

Trees [20], Gbdt [21], AdaBoost [22], MLP [23], and XGBoost [24], were chosen for model construction. The area under the curve (AUC) was used as the final evaluation metric. To objectively assess model quality, accuracy, sensitivity, specificity, PPV, and NPV indices were also calculated for each model.

Initially, the nine typical machine learning models were built based on the training set. Five-fold cross-validation in the training dataset was used to reduce overfitting and validate the robustness of the results. Subsequently, the trained models were further validated on the test set. The judgments of three clinical experts, including two with 3 years of spinal cord imaging experience and one with 10 years of experience, were compared with the constructed models in the test set.

2.6. Model Interpretation

Machine learning models with a large number of parameters perform well but lack interpretability, resembling a black box. To track the contribution of each input feature in the machine learning model, we introduced an explanation technique called Shapley Additive exPlanations (SHAPLEY) [25]. This technique calculates the marginal contribution when adding a feature to the model. In SHAP, the attributes of each feature (protective or risk factor) are thoroughly explained through marginal contributions.

For individual interpretation, we introduced the Local Interpretable Model-Agnostic Explanations (LIME) algorithm, which approximates the model's expression and explains the contribution of each feature to a specific patient [26].

2.7. Statistical Analysis

AUC calculations were performed using IBM SPSS Statistics 23.0 software (IBM Corporation). Image annotations were conducted using 3D-Slicer. Data processing was carried out with Python 3.7, and the models were constructed using the sklearn package. A two-sided *p*-value less than 0.05 was considered statistically significant.

3. Results

3.1. Patient Demographics

A total of 24 patients from our hospital were included in this study. Based on the SUV_{max}, patients were divided into two groups, with 10 patients having decreased SUV_{max} segments and 14 patients having normal (not decreased) SUV_{max} segments. There were no significant differences in age, gender, BMI, and disease course between the two groups (Table 1). Preoperative mJOA scores and follow-up mJOA scores at postoperative six months also showed no significant differences. However, in terms of mJOA improvement, the group with normal SUV_{max} exhibited a slightly greater improvement ($p = 0.043$).

Table 1. Clinical characteristics of patients' table.

	Decreased SUV_{max}	**Normal SUV_{max}**	***p*-Value**
Num	10	14	
Age (years)	56.9 ± 11.4	57.9 ± 8.3	0.801
Gender (M, %)	7 (70%)	7 (50%)	0.349
Course (months)	42 (6–114)	12 (5.5–66)	0.465
BMI	26.2 ± 1.5	25.1 ± 3.0	0.284
Pre-op mJOA	15 (14–16)	15.5 (14–16)	0.732
Post-op mJOA	16 (15.75–17)	17 (16–17)	0.24
mJOA improvement	1 (1–1.25)	2 (1–2)	0.043

Furthermore, the 24 patients were divided into a training set (n = 18) and a validation set (n = 6), totaling 116 segments. In task one, 18 patients with 87 segments were enrolled

in the training set, while 6 patients with 29 segments were enrolled in the test set. As described in Table 2, among the included patients, there were no compressed segments in the C2/3 level. The most commonly compressed segment was C5/6 (87.5%), followed by C4/5 (75%). In task two, a total of 56 segments were included, with 42 segments as the training set and 14 segments as the test set. Table 3 shows that the segments most prone to a decrease in PET uptake were C5/6 (42.86%) and C6/7 (40%). This suggests that the lower cervical spinal cord is more likely to experience a decrease in uptake when compressed, indicating a poorer prognosis.

Table 2. Datasets of identifying compressed levels.

Task 1	Training Dataset				Test Dataset			
Level	**Normal**	**Percent**	**Compressed**	**Percent**	**Normal**	**Percent**	**Compressed**	**Percent**
Total	45	51.72%	42	48.28%	15	51.72%	14	48.28%
C2/3	16	100.00%	0	0.00%	5	100.00%	0	0.00%
C3/4	12	66.67%	6	33.33%	5	83.33%	1	16.67%
C4/5	4	22.22%	14	77.78%	2	33.33%	4	66.67%
C5/6	2	11.11%	16	88.89%	1	16.67%	5	83.33%
C6/7	11	64.71%	6	35.29%	2	33.33%	4	66.67%

Table 3. Datasets of identifying decreased ^{18}F-FDG uptake levels.

Task 2	Training Dataset				Test Dataset			
Level	**Normal**	**Percent**	**Decreased**	**Percent**	**Normal**	**Percent**	**Decreased**	**Percent**
Total	32	76.19%	10	23.81%	8	57.14%	6	42.86%
C2/3	0	/	0	/	0	/	0	/
C3/4	6	100.00%	0	0.00%	1	100.00%	0	0.00%
C4/5	12	85.71%	2	14.29%	3	75.00%	1	25.00%
C5/6	10	62.50%	6	37.50%	2	40.00%	3	60.00%
C6/7	4	66.67%	2	33.33%	2	50.00%	2	50.00%

3.2. Machine Learning Assessment of Spinal Cord Compression

Figure 1 shows the results of identifying compressed segments. After training and five-fold cross-validation, the LinearSVC model performed the best with an AUC of 0.981 in the training dataset. It achieved an accuracy of 0.93, sensitivity of 0.96, and specificity of 0.91. In the test set, the radiomics model obtained an AUC of 0.962 and an accuracy of 0.96. The AUC values from three clinical experts were 0.964, 0.950, and 1, with accuracies of 0.96, 0.96, and 1, respectively. These data suggest that the physicians involved were skilled at identifying spinal cord compression, and the machine learning models did not significantly outperform their assessments (p = 0.423).

Through SHAPLEY feature analysis (Figure 1e), we found that the T2 low signal (radiomics feature: firstorder Median, which represents the median signal of the ROI) in radiomics is a prominent parameter in determining compression. However, we considered that the exclusion of disc compression and surrounding cerebrospinal fluid in radiomics is a limitation resulting in low accuracy, as cerebrospinal fluid signal around the spinal cord is essential for clinical assessments. The focus on only the ROI rather than the entire image might be one reason why machine learning algorithms do not outperform clinical experts. Additionally, factors such as limited sample size and variability in physician expertise may also influence the outcomes.

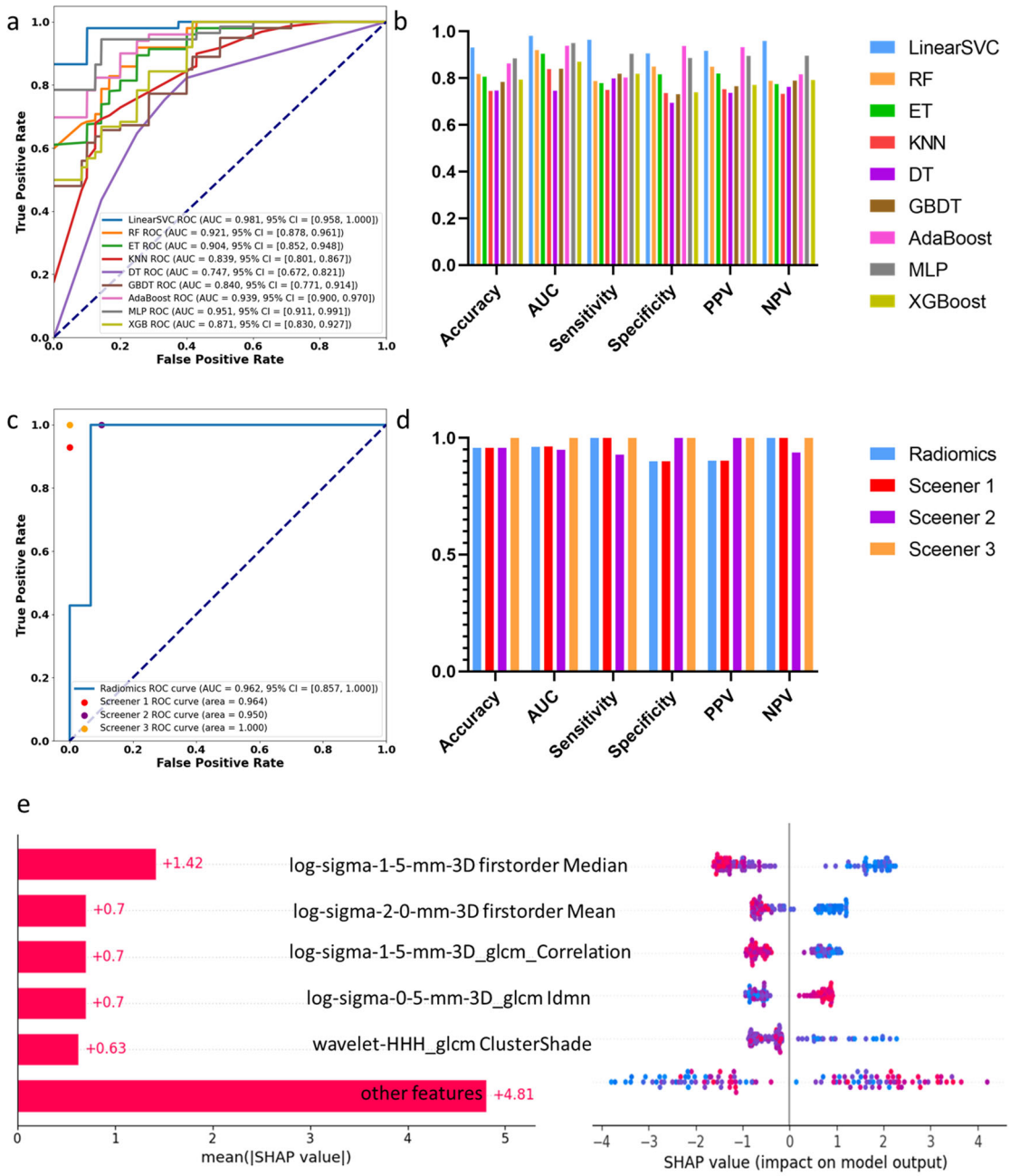

Figure 1. Results of identifying compression levels: (**a**) AUC curve of 9 machine learning algorithms on the training dataset; (**b**) statistic comparison of 9 machine learning algorithms on the training dataset; (**c**) AUC curve of proposed machine learning algorithm and 3 screeners on the test dataset; (**d**) statistic comparison of proposed machine learning algorithm and 3 screeners on the test dataset; (**e**) SHAPLEY explanations of the proposed model.

3.3. Machine Learning Assessment of Segments with Decreased SUV_{max}

Figure 2 shows the results of identifying decreased 18FDG uptake segments. Based on training and five-fold cross-validation results, the LinearSVC model performed the best with an AUC of 0.830 and an accuracy of 0.69. In the test set, the radiomics model obtained an AUC of 0.812 and an accuracy of 0.86. The AUC values from three clinical experts were 0.625, 0.604, and 0.708, with accuracies of 0.64, 0.57, and 0.71. The radiomics-based machine learning model significantly outperformed clinical doctors in detecting segments with decreased SUV_{max} with higher accuracy ($p = 0.035$).

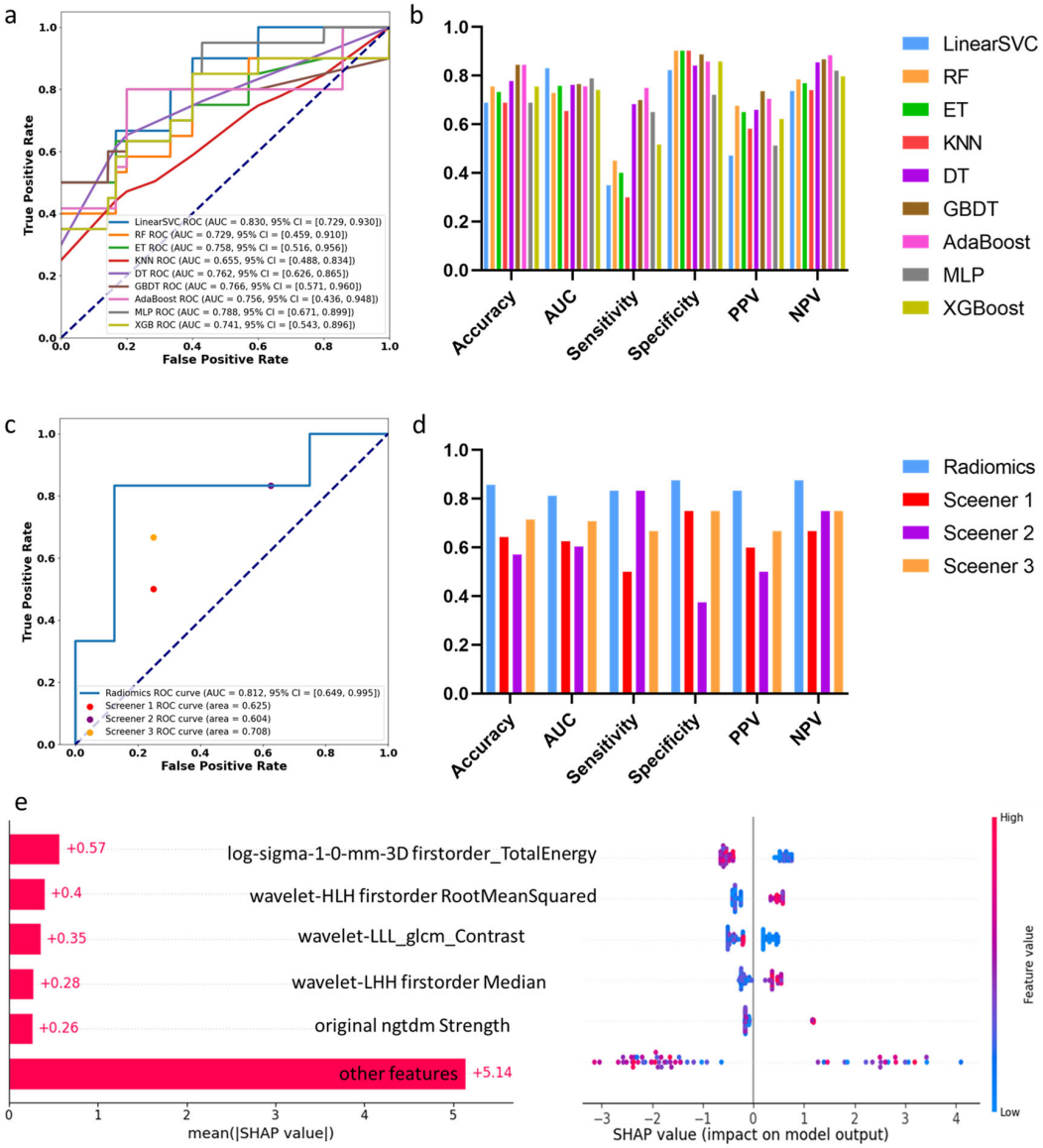

Figure 2. Results of identifying decreased ^{18}F-FDG uptake levels: (**a**) AUC curve of 9 machine learning algorithms on the training dataset; (**b**) statistic comparison of 9 machine learning algorithms on the training dataset; (**c**) AUC curve of proposed machine learning algorithm and 3 screeners on the test dataset; (**d**) statistic comparison of proposed machine learning algorithm and 3 screeners on the test dataset; (**e**) SHAPLEY explanations of the proposed model.

SHAPLEY feature analysis revealed that TotalEnergy, related to area and signal, had a high contribution. TotalEnergy indicated that segments with decreased SUV_{max} had a smaller cross-sectional area, signifying more severe compression. RootMeanSquared, associated with signal disorder within the ROI, indicated that levels with decreased SUV_{max} had more chaotic T2 signals. Additionally, a higher median T2 signal value was significantly correlated with decreased SUV_{max}. Radiomics-based machine learning, focusing on ROI's abnormal signals, far surpassed clinical doctors in accuracy, with AUC values exceeding 0.8 in both training and test sets, indicating the potential to assess spinal cord function based on T2-weighted imaging.

3.4. Individual Feature Interpretation Based on LIME

To better understand the reasons behind predicting decreased SUV_{max}, LIME was employed for feature interpretation of the constructed model.

Example 1 (Figure 3): A patient with cervical compression in C3 to C7 segments, where C3/4 showed unchanged 18FDG uptake, but C4/5, C5/6, and C6/7 exhibited decreased 18FDG uptake. The model indicated a 0.54 probability of normal SUV_{max} in C3/4, primarily due to its lack of significant signal elevation (Median = −0.14). However, other features leaned towards predicting decreased uptake, resulting in a more neutral probability. In C5/6, the model predicted a 0.93 probability of decreased SUV_{max}, associated with small TotalEnergy, chaotic signals, and a higher median T2 signal.

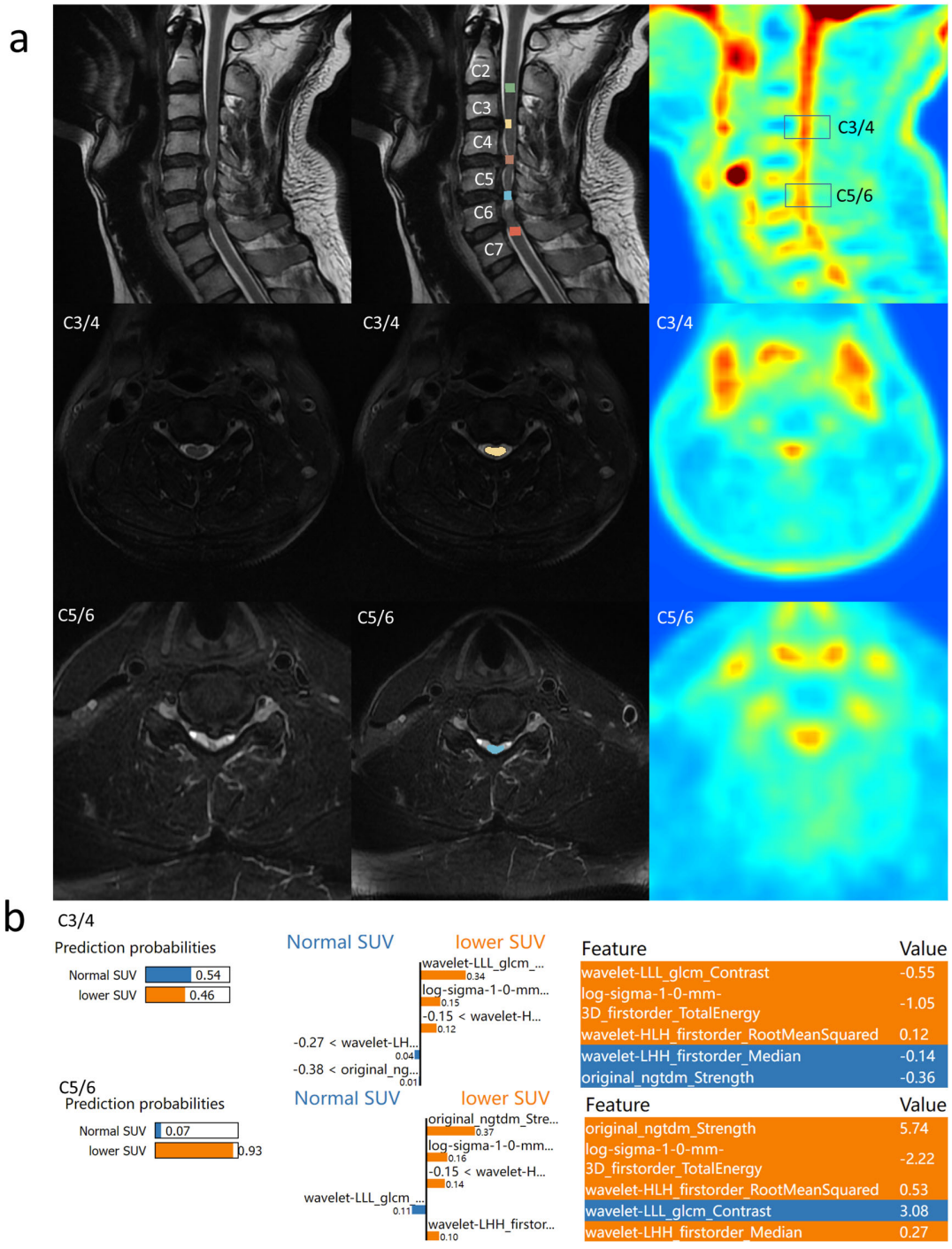

Figure 3. Prediction result of Example 1: (**a**) the PET/MR results of Example 1 are shown. Cervical spinal cord segments from C2 to C7 and five intervertebral disc segments were labeled. SUV_{max} was normal at the C3/4 level and SUV_{max} was decreased at the C5/6 level; (**b**) the LIME results of the proposed model are shown.

Example 2 (Figure 4): Another patient with compression in C5/6 and C6/7, where C5/6 showed decreased SUV_{max}, but C6/7 had a normal SUV_{max}. In C5/6, the model predicted a 0.97 probability of normal SUV_{max}, considering the absence of high median signals and non-chaotic signals. In C6/7, the model predicted a 0.90 probability of normal SUV_{max} due to the absence of high median signals and non-chaotic signals, resulting in successful prediction. This indicates that although the selected features are representative, they are not sufficient to fully replace PET in reflecting cervical cord function.

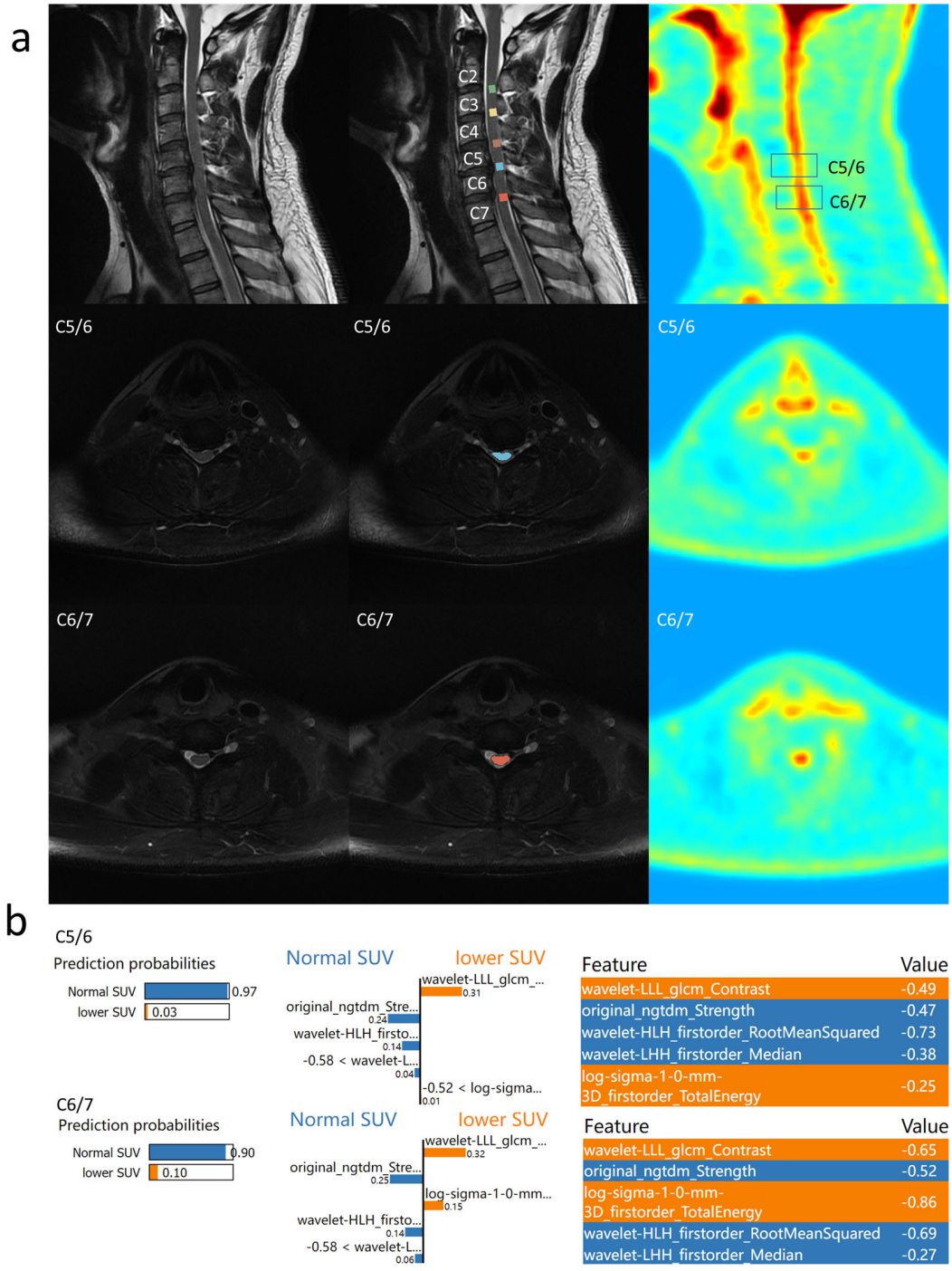

Figure 4. Prediction result of Example 2: (**a**) the PET/MR results of Example 2 are shown. Cervical spinal cord segments from C2 to C7 and five intervertebral disc segments were labeled. SUV_{max} decreased at the C5/6 level and SUV_{max} was normal at the C6/7 level; (**b**) the LIME results of the proposed model are shown. The proposed methods predicted that both SUV_{max} were normal at the two levels.

4. Discussion

In this study, a prospective collection of a cohort of mild CSM patients was conducted, followed by the development of a machine learning algorithm based on radiomics to automatically identify compressed segments and predict SUV_{max} reduction in the compressed segments. These two tasks achieved AUC values of 0.981 and 0.830 in the training set and 0.962 and 0.812 in the test set, respectively.

4.1. Deep Learning and Machine Learning Applications for CSM Patients

For the automatic identification of compressed segments, a review indicates that the diagnostic accuracy of disc herniation or spinal stenosis based on deep learning is already above 0.95, similar to clinical doctors' accuracy [27]. However, diagnostic performance varies across imaging modalities. Deep learning models developed for degenerative cervical myelopathy have demonstrated superior inter-rater agreement in diagnosing spinal canal stenosis on MRI ($\kappa = 0.78$, $p < 0.001$ vs. κ range = 0.57–0.70 for human readers) and classifying foraminal stenosis ($\kappa = 0.80$, $p < 0.001$ vs. κ range = 0.63–0.69), with diagnostic accuracy comparable to manual interpretation (DL: 92.3%, readers: 92.3–100.0%) [28]. In contrast, deep learning approaches based on X-ray imaging have shown significantly improved accuracy, far surpassing human interpretation (89.7% vs. 68.3%) [29].

Regarding prognosis, machine learning models have achieved an AUC of 0.777 in predicting the length of hospital stay following cervical spine surgery [30]. Another study reported that machine learning models attained average AUC values of 0.776, 0.846, 0.775, and 0.747, predicting a prolonged length of stay, non-home discharges, 30-day remissions, and major complications, respectively [31].

Deep learning in image recognition comprehensively understands semantic information throughout the entire image, whereas radiomics-based machine learning understands only the texture features in the annotated ROI [32]. In the task of identifying spinal cord compression, changes in morphological structure are crucial, and radiomics cannot integrate this feature. SHAPLEY analysis revealed that spinal cord compression not only involves morphological changes but also often includes a reduction in the intramedullary T2 signal in the compressed segment.

4.2. Association Between PET/MRI and Prognosis

Current research indicates that PET is a crucial examination for assessing spinal cord functional reserve and predicting prognosis [33]. Frank's studies demonstrate a significant correlation between SUV_{max} values in PET and prognosis [10]. Segments with increased SUV_{max} values show significant postoperative symptom improvement ($p = 0.001$), while those with decreased SUV_{max} values show no significant improvement postoperatively ($p > 0.05$). Similarly, Kenzo found a high correlation between SUV ratio (SUVR) and neurological improvement ($R = 0.837$, $p = 0.001$) [12]. Therefore, predicting whether SUV_{max} decreases has significant implications for prognosis, indirectly representing spinal cord function. In our prospective cohort study, we similarly found more significant symptom improvement in the normal SUV_{max} group ($p = 0.043$). Additionally, we observed that the lower cervical cord is prone to decreased SUV_{max} when compressed, consistent with Frank's findings that SUV_{max} decreased in the C7 segment of CSM patients [9].

Predicting the increase or decrease in SUV_{max} values based on MRI is an important basis for predicting surgical outcomes. Kenzo showed a negative correlation between the abnormal low intramedullary signal on T1-weighted images and the high intramedullary signal on T2-weighted images with a neurological improvement rate ($p < 0.05$) [11]. Furthermore, the SIR (Signal Intensity Ratio), indicating the ratio of increased lesion signal intensity to C7/T1 disc level signal intensity, demonstrated a negative correlation between

neurological improvement rate and SIR on T1-weighted images (R = 0.617, $p < 0.01$), but its correlation with SIR on T2-weighted images was not significant (R = −0.256) [11]. However, in the radiomics model constructed in this study, we found a relationship between high T2 signal intensity and decreased SUV_{max}. Additionally, in this study, through radiomics feature analysis, we discovered that features such as chaotic signals and smaller areas might also contribute to decreased SUV_{max}.

4.3. Treatment Options for Patients with Cervical Spondylotic Myelopathy

Currently, there is still controversy over whether surgery should be actively performed on CSM patients with different severity degrees. Frank's study suggests that in moderate to severe CSM patients, those exhibiting decreased 18FDG uptake did not show significant improvement postoperatively (pre-op mJOA = 11.6, post-op mJOA = 12.0, $p > 0.05$), while those exhibiting increased 18FDG uptake showed significant improvement (pre-op mJOA = 9.5, post-op mJOA = 13.6, $p = 0.001$) [10]. This indicates that for moderate to severe CSM patients, timely surgery can significantly improve symptoms before 18FDG uptake decreases. However, there are still questions regarding whether surgery should be actively performed for mild CSM patients. To explore the prognosis of mild CSM, the 24 patients enrolled in this study had relatively mild symptoms (median mJOA of 15 points). Follow-up results showed that for mild CSM patients, regardless of whether PET showed reduced uptake, active surgery could improve patient symptoms (SUV_{max} decrease group: pre-op mJOA = 15, post-op mJOA = 16, $p = 0.012$; normal SUV_{max} group: pre-op mJOA = 15.5, post-op mJOA = 17, $p = 0.001$). However, for patients without decreased SUV_{max}, their symptom improvement was more significant ($p = 0.043$). Combined with the previous literature, we believe that patients with normal SUV_{max} can benefit from surgery regardless of the severity of the disease, although complete recovery is difficult for moderately severe patients. For patients with decreased SUV_{max} in mild CSM, symptoms can still improve after surgery, but for moderate to severe CSM patients, symptoms do not significantly improve.

Therefore, we built a treatment flowchart for CSM patients based on the MRI and PET findings, as Figure 5 shows. For mild CSM patients with normal SUV_{max}, the decision to undergo surgery or long-term follow-up could be based on the patient's willingness. For mild CSM patients with decreased SUV_{max} and moderate to severe CSM patients with normal SUV_{max}, we recommended timely surgical treatment to prevent progression. For moderate to severe CSM patients with decreased SUV_{max}, due to poor surgical outcomes, we recommend fully informing patients about the poor prognosis before surgery.

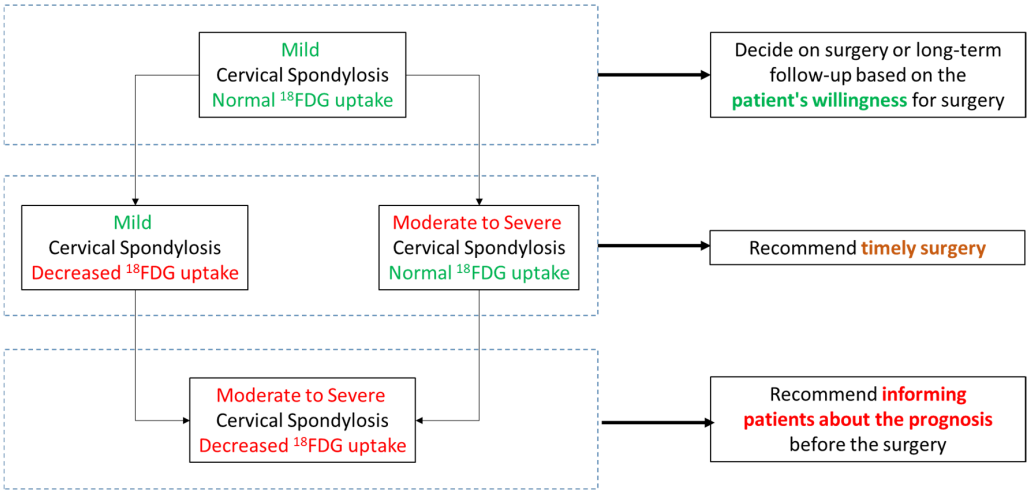

Figure 5. Proposed flowchart of the treatment of cervical spondylotic myelopathy.

4.4. Other Methods for the Assessment of Cervical Spondylotic Myelopathy

Diffusion Tensor Imaging (DTI), an advanced neuroimaging technique for visualizing the integrity of white matter tracts, has gained considerable attention in the study of CSM in recent years. Studies have demonstrated that patients with mild CSM exhibit significantly higher fractional anisotropy (FA) values compared to those with moderate to severe disease, whereas the differences in FA between moderate and severe cases are not statistically significant [34]. Furthermore, DTI can assist in identifying the responsible segment of spinal cord compression based on localized FA reductions [35].

In terms of prognostic evaluation, dynamic DTI studies have shown that preoperative FA values measured in different cervical spine positions are significantly associated with postoperative functional recovery ($p < 0.029$) [36]. Beyond DTI, other multimodal approaches have also shown promise in the diagnosis and prognostication of CSM. For example, a deep learning model based on electroencephalography (EEG) achieved an accuracy of 92.5% in binary classification tasks for CSM diagnosis [37]. Additionally, intraoperative ultrasound monitoring has been found to correlate closely with JOA scores and may hold potential prognostic value [38]. In summary, these multimodal imaging and physiological signal analysis techniques offer novel, multidimensional approaches to the comprehensive evaluation of spinal cord function in CSM patients.

4.5. Limitations

This study has certain limitations. Firstly, the sample size is relatively small. As a prospective study, it included 24 patients with a total of 116 segments, with only 56 segments available in task two. Although this number surpasses or is comparable to previous related studies [9,10], for machine learning algorithms based on radiomics, further expansion of the sample size is necessary. Secondly, this study only analyzed T2 data, while prior studies have indicated that low signals on T1 also hold prognostic significance [11]. Therefore, future research will incorporate multimodal radiomics to construct machine learning models, thereby enhancing predictive accuracy. Thirdly, multi-center studies are needed to validate the stability and robustness of the model.

4.6. Future Directions

Future research should focus on expanding the cohort, particularly by including more moderate to severe CSM cases for model training and validation to enhance generalizability. In addition, combining radiomics with other biomarkers, such as physiological measurements or clinical assessments, may further improve the model's predictive performance. These steps aim to advance toward a comprehensive decision-support system that can aid personalized treatment planning in CSM before surgery. The proposed method has the potential to preoperatively evaluate prognosis-related risk, especially in mild CSM patients, for whom surgical decision-making is often challenging.

5. Conclusions

This prospective study included a cohort of mild cervical spondylotic myelopathy (CSM) patients with follow-up assessments. Follow-up results indicated that the improvement in symptoms for the not-decreased SUV_{max} group was higher than that for the decreased SUV_{max} group. We recommend timely surgical treatment for mild cervical spondylosis patients with decreased SUV_{max} to prevent symptom worsening. Meanwhile, the machine learning algorithm based on radiomics completed two tasks: the automatic identification of compressed segments and the automatic identification of decreased 18FDG uptake segments. The results showed that the radiomics-based machine learning algorithm outperformed clinical doctors by 10% in predicting whether 18FDG uptake would

decrease. This study also identified several imaging features related to decreased 18FDG uptake, including a smaller cross-sectional area, chaotic signals, and higher intramedullary signals, which provided an alternative to PET examinations. Radiomics-based prediction of SUV_{max} changes could potentially reflect spinal cord function and could be worth further exploration for its prognostic and clinical relevance.

Author Contributions: H.W. (He Wang), K.W. and Y.W. contributed equally to the present study. H.W. (Hao Wu) initiated the study. Y.W., Z.L., L.Z. and S.J. collected PET/MR images. H.W. (He Wang) and K.H. performed segmentation. Y.W. and X.Z. helped to collect clinical data and follow-up data. H.W. (He Wang) and K.W. established the model. H.W. (He Wang) and K.W. wrote the draft. All authors revised the manuscript critically. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2023YFC2507703).

Institutional Review Board Statement: The study was approved by the Xuanwu Hospital ethics committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The data and code used to support the findings of this study are available from the corresponding author on request.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Abbreviations

CSM	Cervical spondylotic myelopathy
mJOA	Modified Japanese Orthopaedic Association scale
SHAPLEY	Shapley Additive exPlanations
LIME	Local Interpretable Model-Agnostic Explanations
SUV_{max}	The maximum standardized uptake value

References

1. Badhiwala, J.H.; Ahuja, C.S.; Akbar, M.A.; Witiw, C.D.; Nassiri, F.; Furlan, J.C.; Curt, A.; Wilson, J.R.; Fehlings, M.G. Degenerative cervical myelopathy—Update and future directions. *Nat. Rev. Neurol.* **2020**, *16*, 108–124. [CrossRef] [PubMed]
2. Fehlings, M.G.; Kwon, B.K.; Tetreault, L.A. Guidelines for the management of degenerative cervical myelopathy and spinal cord injury: An introduction to a focus issue. *Glob. Spine J.* **2017**, *7*, 6S–7S. [CrossRef] [PubMed]
3. Karadimas, S.K.; Erwin, W.M.; Ely, C.G.; Dettori, J.R.; Fehlings, M.G. Pathophysiology and natural history of cervical spondylotic myelopathy. *Spine* **2013**, *38*, S21–S36. [CrossRef]
4. Rhee, J.; Tetreault, L.A.; Chapman, J.R.; Wilson, J.R.; Smith, J.S.; Martin, A.R.; Dettori, J.R.; Fehlings, M.G. Nonoperative versus operative management for the treatment degenerative cervical myelopathy: An updated systematic review. *Glob. Spine J.* **2017**, *7* (Suppl. 3), 35S–41S. [CrossRef]
5. Aiello, M.; Alfano, V.; Salvatore, E.; Cavaliere, C.; Picardi, M.; Della Pepa, R.; Nicolai, E.; Soricelli, A.; Vella, A.; Salvatore, M.; et al. [^{18}F] FDG uptake of the normal spinal cord in PET/MR imaging: Comparison with PET/CT imaging. *EJNMMI Res.* **2020**, *10*, 91. [CrossRef]
6. Lam, M.; Burke, C.J.; Walter, W.R.J.D.; Radiology, I. Correlation of 18F-FDG PET/CT uptake with severity of MRI findings and epidural steroid injection sites in patients with symptomatic degenerative disease of the lumbar spine: A retrospective study. *Diagn. Interv. Radiol.* **2021**, *27*, 580–586. [CrossRef]
7. Tan, X.; Li, D.; Wu, X.; Yang, Y.; Hou, Q.; He, L.; Jiang, L. Physiologically intense FDG uptake of distal spinal cord on total-body PET/CT. *Ann. Nucl. Med.* **2022**, *36*, 643–650. [CrossRef] [PubMed]
8. Brancato, V.; Borrelli, P.; Alfano, V.; Picardi, M.; Mascalchi, M.; Nicolai, E.; Salvatore, M.; Aiello, M. The impact of MR-based attenuation correction in spinal cord FDG-PET/MR imaging for neurological studies. *Med Phys.* **2021**, *48*, 5924–5934. [CrossRef]

9. Floeth, F.W.; Stoffels, G.; Herdmann, J.; Jansen, P.; Meyer, W.; Steiger, H.-J.; Langen, K.-J. Regional impairment of 18 F-FDG uptake in the cervical spinal cord in patients with monosegmental chronic cervical myelopathy. *Eur. Radiol.* **2010**, *20*, 2925–2932. [CrossRef]
10. Floeth, F.W.; Stoffels, G.; Herdmann, J.; Eicker, S.; Galldiks, N.; Steiger, H.-J.; Langen, K.-J. Prognostic value of 18F-FDG PET in monosegmental stenosis and myelopathy of the cervical spinal cord. *J. Nucl. Med.* **2011**, *52*, 1385–1391. [CrossRef]
11. Uchida, K.; Nakajima, H.; Yayama, T.; Kobayashi, S.; Shimada, S.R.; Tsuchida, T.; Okazawa, H.; Mwaka, E.M.; Baba, H. High-resolution magnetic resonance imaging and 18FDG-PET findings of the cervical spinal cord before and after decompressive surgery in patients with compressive myelopathy. *Spine* **2009**, *34*, 1185–1191. [CrossRef] [PubMed]
12. Uchida, K.; Nakajima, H.; Okazawa, H.; Kimura, H.; Kudo, T.; Watanabe, S.; Yoshida, A.; Baba, H. Clinical significance of MRI/18 F-FDG PET fusion imaging of the spinal cord in patients with cervical compressive myelopathy. *Eur. J. Nucl. Med.* **2012**, *39*, 1528–1537. [CrossRef] [PubMed]
13. Gillies, R.J.; Kinahan, P.E.; Hricak, H.J.R. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [CrossRef]
14. Sun, Q.; Chen, Y.; Liang, C.; Zhao, Y.; Lv, X.; Zou, Y.; Yan, K.; Zheng, H.; Liang, D.; Li, Z.-C. Biologic pathways underlying prognostic radiomics phenotypes from paired MRI and RNA sequencing in glioblastoma. *Radiology* **2021**, *301*, 654–663. [CrossRef]
15. Ligero, M.; Garcia-Ruiz, A.; Viaplana, C.; Villacampa, G.; Raciti, M.V.; Landa, J.; Matos, I.; Martin-Liberal, J.; Ochoa-De-Olza, M.; Hierro, C.; et al. A CT-based radiomics signature is associated with response to immune checkpoint inhibitors in advanced solid tumors. *Radiology* **2021**, *299*, 109–119. [CrossRef]
16. Benzel, E.C.; Lancon, J.; Kesterson, L.; Hadden, T. Cervical laminectomy and dentate ligament section for cervical spondylotic myelopathy. *J. Spinal Disord.* **1991**, *4*, 286–295. [CrossRef]
17. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **2012**, *30*, 1323–1341. [CrossRef] [PubMed]
18. Breiman, L. Random forests. *Mach Learn.* **2001**, *45*, 5–32. [CrossRef]
19. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
20. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
21. Friedman, J.H.; Analysis, D. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
22. Schapire, R.E.; Singer, Y. BoosTexter: A Boosting-based System for Text Categorization. *Mach. Learn.* **2000**, *39*, 135–168. [CrossRef]
23. Gardner, M.W.; Dorling, S.R. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [CrossRef]
24. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
25. Rodríguez-Pérez, R.; Bajorath, J. Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *J. Comput. Aided Mol. Des.* **2020**, *34*, 1013–1026. [CrossRef]
26. de Sousa, I.P.; Maria Bernardes Rebuzzi Vellasco, M.; da Silva, E.C. Local interpretable model-agnostic explanations for classification of lymph node metastases. *Sensors* **2019**, *19*, 2969. [CrossRef]
27. Qu, B.; Cao, J.; Qian, C.; Wu, J.; Lin, J.; Wang, L.; Ou-Yang, L.; Chen, Y.; Yan, L.; Hong, Q.; et al. Current development and prospects of deep learning in spine image analysis: A literature review. *Quant. Imaging Med. Surg.* **2022**, *12*, 3454–3479. [CrossRef] [PubMed]
28. Lee, A.; Wu, J.; Liu, C.; Makmur, A.; Ting, Y.H.; Nor, F.E.M.; Tan, L.Y.; Ong, W.; Tan, W.C.; Lee, Y.J.; et al. Deep learning model for automated diagnosis of degenerative cervical spondylosis and altered spinal cord signal on MRI. *Spine J.* **2025**, *25*, 255–264. [CrossRef]
29. Xie, Y.; Nie, Y.; Lundgren, J.; Yang, M.; Zhang, Y.; Chen, Z. Cervical spondylosis diagnosis based on convolutional neural network with x-ray images. *Sensors* **2024**, *24*, 3428. [CrossRef]
30. Zhang, B.; Huang, S.; Zhou, C.; Zhu, J.; Chen, T.; Feng, S.; Huang, C.; Wang, Z.; Wu, S.; Liu, C.; et al. Prediction of additional hospital days in patients undergoing cervical spine surgery with machine learning methods. *Comput. Assist. Surg.* **2024**, *29*, 2345066. [CrossRef]
31. Karabacak, M.; Bhimani, A.D.; Schupper, A.J.; Carr, M.T.; Steinberger, J.; Margetis, K. Machine learning models on a web application to predict short-term postoperative outcomes following anterior cervical discectomy and fusion. *BMC Musculoskelet. Disord.* **2024**, *25*, 401. [CrossRef]
32. Niemeyer, F.; Galbusera, F.; Tao, Y.; Kienle, A.; Beer, M.; Wilke, H. A deep learning model for the accurate and reliable classification of disc degeneration based on MRI data. *Investig. Radiol.* **2020**, *56*, 78–85. [CrossRef]
33. Floeth, F.W.; Galldiks, N.; Eicker, S.; Stoffels, G.; Herdmann, J.; Steiger, H.-J.; Antoch, G.; Rhee, S.; Langen, K.-J. Hypermetabolism in 18F-FDG PET predicts favorable outcome following decompressive surgery in patients with degenerative cervical myelopathy. *J. Nucl. Med.* **2013**, *54*, 1577–1583. [CrossRef]

34. Cho, N.; Al-Shawwa, A.; Jacobs, W.B.; Evaniew, N.; Bouchard, J.; Casha, S.; Duplessis, S.; Lewkonia, P.; Nicholls, F.; Soroceanu, A.; et al. Spinal Cord Tract Integrity in Degenerative Cervical Myelopathy. *Neurosurgery* **2022**, *10*, 1227. [CrossRef] [PubMed]
35. Fang, Y.; Li, S.; Wang, J.; Zhang, Z.; Jiang, W.; Wang, C.; Jiang, Y.; Guo, H.; Han, X.; Tian, W. Diagnostic efficacy of tract-specific diffusion tensor imaging in cervical spondylotic myelopathy with electrophysiological examination validation. *Eur. Spine J.* **2024**, *33*, 1230–1244. [CrossRef] [PubMed]
36. Wang, X.; Tian, X.; Zhang, Y.; Zhao, B.; Wang, N.; Gao, T.; Zhang, L. Predictive value of dynamic diffusion tensor imaging for surgical outcomes in patients with cervical spondylotic myelopathy. *BMC Med. Imaging* **2024**, *24*, 260. [CrossRef] [PubMed]
37. Li, S.; Yang, B.; Dou, Y.; Wang, Y.; Ma, J.; Huang, C.; Zhang, Y.; Cao, P. Aided diagnosis of cervical spondylotic myelopathy using deep learning methods based on electroencephalography. *Med Eng. Phys.* **2023**, *121*, 104069. [CrossRef]
38. Liu, W.; Li, J.; Shu, T.; Ji, Q.; Wang, X.; Li, R.; Sui, Y.; He, D.; Xu, Z. Preliminarily exploring the intraoperative ultrasonography characteristics of patients with degenerative cervical myelopathy. *BMC Musculoskelet. Disord.* **2024**, *25*, 538. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.