Int. j. adv. multidisc. res. stud. 2025; 5(6):954-961

Received: 08-10-2025
Accepted: 18-11-2025

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

YOLOv11 Models: Enhancing CT-Based Diagnosis of Cervical Foraminal Stenosis Through Deep Learning

[1] Nadeer M. Gharaibeh, [2] Liuye Yang, [3] Zaid M. Gharaibeh, [4] Fawwaz Al-Smadi, [5] Areej H. Al-Sarairah, [6] Huang Zhongyichen, [7] Omar Allasasmeh, [8] Anas O. K. Alrawashdeh, [9] Gang Wu, [10] Xiaoming Li

[1, 6, 9, 10] Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[2] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, 430022, China

[3] Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

[4] Department of Orthopedic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

[5] Diagnostic Imaging and Radiotherapy Program, Faculty of Health Sciences, School of Diagnostic and Applied Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur 50300, Malaysia

[5] Department of Radiology, Princess Aisha Bint Al Hussein College of Nursing and Health Sciences, Al Hussein Bin Talal University, Ma'an, 71111, Jordan

[7] Centre of Diagnostic, Therapeutic and Investigative Studies (CODTIS), Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur 56000, Malaysia

[8] Division of Cardiothoracic and Vascular Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430030, China

DOI: https://doi.org/10.62225/2583049X.2025.5.6.5291
Corresponding Author: **Xiaoming Li, Gang Wu**

Abstract

This study aims to construct a deep learning (DL) system from the YOLOv11 family by using computed tomography (CT) to overcome the scarcity of magnetic resonance imaging (MRI) availability and the standard imaging assessments to diagnose cervical foramina stenosis (CFS). This retrospective study involved 1,437 cervical spine CT scans from three centers (Center A, B, and C) of a local hospital. The training and validation were done using scans from Centers A and B, with Center C used as the external test set. A stenosis detection model, based on the YOLOv11 architecture with Multiplanar Reconstruction (MPR) technology, was employed. The model was trained on 5-fold cross-validated images (640×640 pixels) in five variants (n, s, m, l, x) that were labeled. Five radiologists analyzed 215 images from Center C, and performance metrics were determined. Statistical analysis was conducted using paired t-tests and McNemar tests, with Bonferroni and FDR corrections. In external testing (n = 215), the YOLOv11-x model had higher diagnostic performance at detecting cervical foraminal stenosis: sensitivity was 95.6% ($p = 0.005$), specificity was 97.4% ($p < 0.001$), and accuracy was 96.7% ($p < 0.001$). YOLOv11-x has demonstrated high performance in standardized CT-based diagnosis of CFS, and lightweight versions could be used in real-time clinical applications. This technique presents a deep learning modality, which is effective as an alternative diagnostic tool in settings with limited MRI resources. Future validation and correlation with clinical outcomes should verify the generalizability of treatment responses and symptom severity results.

Keywords: Cervical Foraminal Stenosis, Deep Learning, Computed Tomography, YOLOv11, Diagnostic Accuracy, Multiplanar Reconstruction

Introduction

The narrowing of the intervertebral foramina in the cervical spine is known as cervical foraminal stenosis (CFS), which can compress the exiting spinal nerve roots [1, 2]. The clinical manifestations of nerve root compression are characterized by radiculopathy, with symptoms such as paresthesia, muscle weakness, and pain that is experienced in the distribution of the compressed nerve root, as well as pain in the neck and shoulders [3, 4]. Facet hyperplasia, ligamentum flavum hypertrophy,

degenerative bone spurs (osteophytes), and lateral disc herniation are the principal factors contributing to neural foramina constriction and mechanical compression of the cervical nerve roots [5, 6]. Previous studies have shown that CFS occurs in approximately 10–25% of the adult population; therefore, an accurate diagnosis of CFS and its associated symptoms in the upper limbs becomes crucial for the development of effective treatment strategies, which may involve surgery [7-9].

The diagnosis of cervical radiculopathy is complicated due to the similarity between the symptoms and other syndromes, such as thoracic outlet syndrome, ulnar nerve neuropathy, and carpal tunnel syndrome, which have different etiologies [10]. CT and/or MRI cervical spine imaging assists in confirming CFS, but the cost and accessibility of MRI limit its use [11, 12]. While MRI is the primary modality for visualizing soft tissues, such as intervertebral discs and spinal cord pathology, CT is also useful for evaluating osseous structures, including the morphology of facet joints and bony stenosis, especially when MRI is not available [13]. CT scans of the cervical spine are also faster to obtain, enabling quicker patient throughput, particularly in the acute trauma setting [14]. Moreover, CT is more precise than MRI in assessing facet joint morphology (facet tropism) and diagnosing facet joint arthritis [15]. Therefore, CT serves as a useful second-line imaging modality for CFS in the absence of MRI and complements MRI, especially for measuring bony foraminal stenosis in emergency or resource-restricted settings.

The diagnosis of CFS is a challenge, as CT and MRI usually visualize the foramina indirectly, and the diagnosis often depends on the physician's experience [16-18]. The exact levels of stenosis in the spine are difficult to determine accurately using routine methods [16]. Deep learning (DL) is a subset of machine learning that has recently received attention in medical imaging and has the potential to overcome these limitations. Convolutional neural networks (CNNs) and other forms of DL models can identify hard-to-detect patterns and features that are difficult to manually classify, making them beneficial for automated detection and classification tasks [19-21].

Despite the growing use of deep learning in spinal imaging studies, most have focused on conditions such as vertebral fractures or lumbar stenosis [22]. In contrast, CFS remains less studied due to the specificity of its diagnosis, which cannot be fully resolved through current methods. Most spinal examination procedures have used single-plane CT imaging, such as axial-only evaluation of lumbar stenosis or sagittal-only evaluation of fractures, or models based on MRI testing of cervical spine scans in general [23, 24]. These methods are inherently limited because they lack integrated multi-planar information to represent the complex cervical neural foramina anatomy in 3D.

To fill these gaps, we propose a diagnostic framework with three contributions. First, we present what we believe is one of the first medical imaging applications of the YOLOv11 architecture (introduced in 2025), an advanced network that builds upon the widely used YOLOv5–v8 models [25]. Second, our study focuses on cervical foraminal stenosis (CFS), a topic with a notable gap in the literature. Third, and most importantly, we introduce a highly integrated multiplanar reconstruction (MPR) pipeline that processes both axial and sagittal CT reconstructions. This design provides more anatomical context and reduces dependence on single-plane views, enabling a more accurate and robust CT-based diagnosis of CFS. Finally, this study evaluates the performance of YOLOv11 models with MPR for detecting CFS on CT images and aims to provide rapid, practical diagnosis when MRI is not readily available, as well as identify the optimal model for clinical practice.

Material and Methods

Patient Selection and Image Preparation

The study was approved by the Institutional Review Board (IRB) of Huazhong University of Science and Technology (Approval No. TJ-IRB202504051, April 28, 2025). This retrospective included patients with cervical intervertebral foramen stenosis diagnosed using CT. Inclusion criteria consisted of: (a) adults (age ≥18 years) who had diagnostic non-contrast cervical CT covering C1–C7 with slice thickness not exceeding 1.0 mm; (b) axial and sagittal multiplanar reconstructions available; (c) the clinical index test defined as the first eligible cervical CT exam per patient; and (d) reference-standard labels based on consensus reading by at least two radiologists. The exclusion criteria included: (a) low-quality CT images with significant artifacts, slice thickness > 1.0 mm, motion blur, or scanning range not covering the C1–C7 intervertebral foramen; (b) previous history of cervical spine surgery, combined cervical spondylosis lesions such as tumors, fractures, or infections that interfere with the assessment; and (c) incomplete clinical data or CT reconstruction sequences. The final included cases were independently evaluated by five radiologists with 5–10 years of experience using a double-blind method.

Although the reader panel offers a decent evaluation, larger multi-national center studies could enhance validation and generalizability. For each patient, three axial/sagittal CT images containing the most significant lesion segments were selected for quantitative analysis and model input.

Data Collection

A total of 1,437 cervical spine CT scans of cervical intervertebral stenosis patients were gathered from a local hospital (including three regions: Center A, Center B, and Center C) from March 2015 to March 2025. The scans were allocated as follows: Center A supplied 625 scans, Center B supplied 597 scans, and Center C supplied 215 scans. A step was taken to ensure the inclusion of diverse patient characteristics, including age, sex, and the degree of stenosis (mild, moderate, severe), as considered in the primary radiological reports.

The dataset included scans from three separate regions, each using different models of CT scanners and imaging settings. This natural heterogeneity in the parameters of the acquisition method increases the model's exposure to natural heterogeneity in real-world acquisitions compared to typical single-center, single-scanner research. To further assess performance, a 5-fold cross-validation was applied to the training and validation sets from Centers A and B. The findings were consistent, with an average accuracy of 93.2% ± 0.8%.

A total of 1,222 scans were contributed by Centers A and B, the majority of which were used in the training and validation sets for model development. Center C, in turn, provided the external test set, comprising 215 images: 123 from patients with intervertebral foramen stenosis and 92 from patients without. Each patient provided three MPR

images, typically including the most important axial and sagittal views, resulting in an equal number of cases. Most importantly, Center C was completely kept apart from both training and validation data, allowing it to be tested without any bias on the model's ability to perform in the real world.

Model Architecture and Processing

The YOLOv11 model family is characterized by identical architecture, which comprises three primary parts: the feature extractor backbone, the feature fusion neck (e.g., PAN, BiFPN), and the object localization and classification detection head. Although they all share the same fundamental design (YOLOv11n, s, m, l, x), these variants differ in their multiplicative scaling strategy, modifying depth (number of layers), width (number of channels), and input resolution [26-28]. This scaling allows predictable trade-offs between accuracy and computational cost across models. YOLOv11n is designed for edge devices and real-time applications (smallest/fastest, lowest accuracy). YOLOv11s, when deployed generally, balances speed and accuracy. YOLOv11m/l focuses on greater accuracy at the cost of higher computational requirements. YOLOv11x is the most accurate when inference is performed on cloud/offline, being the most computationally intensive [26, 29]. Therefore, YOLOv11 provides a versatile family of architectures that can be applied in various settings, including resource-constrained edge devices and powerful cloud networks by co-scaling depth, width, and resolution [30].

Additionally, we have integrated Multiplanar Reconstruction (MPR) technology into the model's input pipeline. In MPR, sagittal and coronal reformations are derived from the underlying axial CT data, providing a complementary assessment of the intervertebral foramina. This is crucial because axial slices alone can poorly assess the complex 3D anatomy of the cervical foramina. MPR enhances the visual data available to the model, which may contribute to improved stenosis detection.

The proposed study aims to analyze the application of YOLOv11 variants with MPR to detect cervical intervertebral foramen stenosis in CT images and identify the most efficient variant. We hypothesize that the scalable structure of these models, combined with the extended MPR views, can enhance diagnostic accuracy without compromising the clinical usability of the algorithms, thereby overcoming the trade-off between diagnostic accuracy and computational speed.

Model Training

A schematic illustration of the proposed model is shown in Figure 1. The five YOLOv11 variants (n, s, m, l, and x) were implemented and trained via Python scripts and command-line tools taken from the official Ultralytics YOLOv11 repository. All input images adhered to a standard 640×640-pixel size. A total of 1,227 labeled images from Centers A and B were randomly partitioned into a training set and validation set (8:2 ratio). Model training was performed for 100 epochs on an NVIDIA GeForce RTX 3060 GPU with 12 GB of memory. Batch size, optimizer (SGD with momentum), initial learning rate, weight decay, and other hyperparameters followed standard YOLOv11 practices and were fine-tuned based on validation performance. To mitigate overfitting, early stopping was applied if validation loss did not improve for 10 epochs, and dropout was incorporated into the neck module. Hyperparameters were fine-tuned using validation performance, with standard defaults such as a learning rate of 0.01 and a batch size of 16 selected as optimal. Final model weights were approximately 6.46 MB (varying by variant).

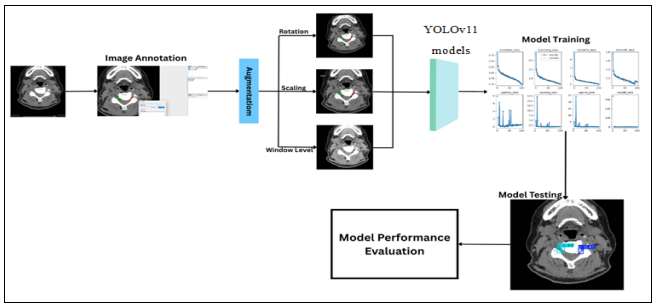

Fig 1: Schematic illustration of the YOLOv11 model architecture, showing image annotation, data augmentation techniques (rotation, scaling, window level adjustments), model training, and performance evaluation.

Comparison of Diagnostic Performance Between YOLOv11 Models and Radiologists

A total of 215 images from Center C were used for external validation. Five radiologists with 5–10 years of experience independently evaluated the CT scans, blinded to patient information, and annotated bounding boxes of all intervertebral foramen stenosis regions using the LabelMe annotation tool. True positives (TP), false positives (FP), true negatives (TN), false negatives (FN), and the time of diagnosing each image were obtained. The outcome of predictions by five variations of YOLOv11 was compared to radiologist annotations. Multiple comparison corrections and bootstrapping were used to ensure statistical reliability across metrics, particularly due to minor differences in sensitivity and specificity.

Metrics for Performance Evaluation

Specificity = TN / (TN + FP)
Recall (Sensitivity) = TP / (TP + FN)
Precision = TP / (TP + FP)
Accuracy = (TP + TN) / (TP + FP + TN + FN)
F1 score = (2 × Precision × Recall) / (Precision + Recall)

Statistical Analysis

All statistical analyses were performed using SPSS software version 26.0 (SPSS Inc., Chicago, Illinois, USA). For continuous variables, paired t-tests were used. CNN models and radiologists were compared in terms of diagnostic performance using the McNemar test. To minimize Type I errors, multiple comparisons were corrected using Bonferroni and False Discovery Rate (FDR) adjustments. A p-value of 0.05 was considered statistically significant, as is customary. It should be noted that a preliminary power analysis indicated that the sample size of n = 215 was sufficient to achieve a power of more than 80% to detect a 5% difference in accuracy, with an alpha level of 0.05 and an effect size of 0.3.

Results

In this study, 1,437 CT images of normal, mild, moderate, and severe stenosis were used. The images were divided into a training set (n = 978), validation set (n = 244), and

external test set (n = 215). The variants of YOLOv11 were compared to the diagnostic performance and efficiency of five radiologists with an average experience of 5–10 years. All YOLOv11 variants were successfully trained, as demonstrated by the gradual reduction of key loss functions, including bounding box loss (localization error), classification loss, and distribution focal loss (classification refinement) over the 100 epochs of both training and validation cohorts (Figure 2). This indicates effective learning and convergence without high overfitting. Figure 3 provides illustrative examples of stenosis detection using the trained YOLOv11 variant models on sample test images.

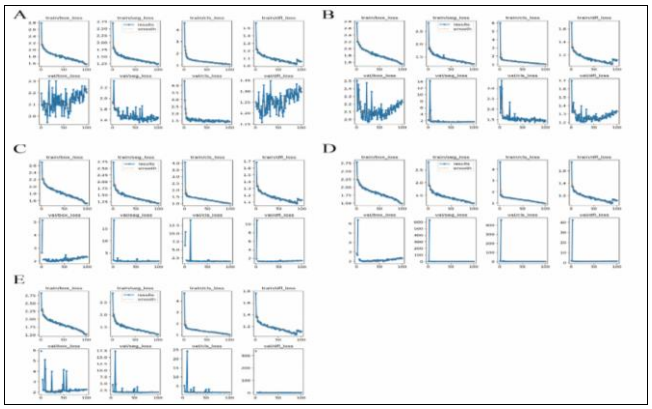

Fig 2: Training performance of YOLOv11 variant models. The graphs display reductions in box loss, classification loss, and distribution focal loss across training epochs, with upper lines representing the training cohort and lower lines representing the validation cohort. (A) YOLOv11-n; (B) YOLOv11-s; (C) YOLOv11-m; (D) YOLOv11-l; (E) YOLOv11-x.

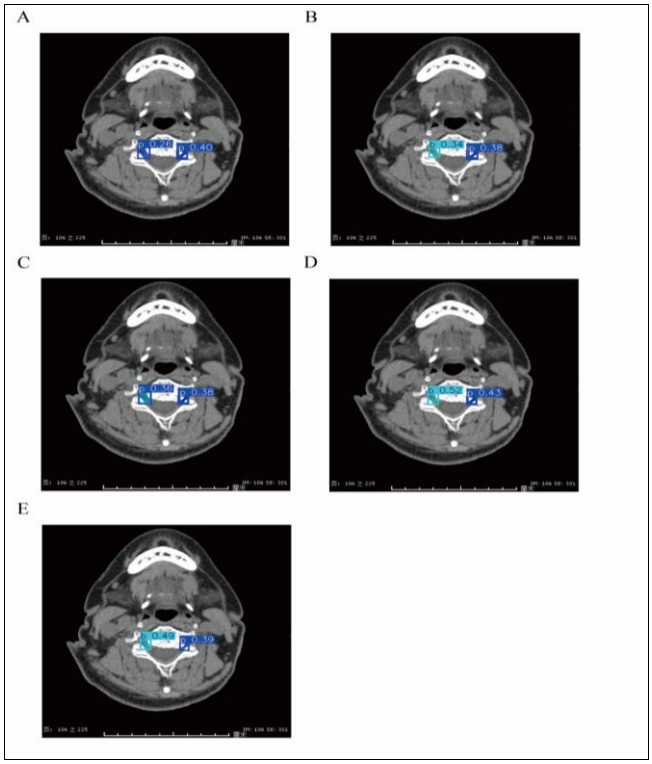

Fig 3: Prediction results of YOLOv11 variant models on sample CT images. Each subfigure presents the output of a different variant: (A) YOLOv11-n; (B) YOLOv11-s; (C) YOLOv11-m; (D) YOLOv11-l; (E) YOLOv11-x.

Diagnostic Performance: External Test Set (n=215)

The diagnostic performance metrics of the five YOLOv11 variants and the average performance of the five radiologists on the external test set are summarized in Table 1. Bonferroni correction and FDR adjustments were used to adjust the p-values of each comparison to account for multiple comparisons. To quantify uncertainty and increase statistical reliability, bootstrap resampling (1,000 iterations) was used to compute 95% confidence intervals. The level of inter-reader agreement between the radiologists was high (Cohen's $\kappa > 0.82$). YOLOv11-x and YOLOv11-l were the models with the highest overall performance, while the lighter variants, YOLOv11-n and YOLOv11-s, exhibited trade-offs: YOLOv11-s had a lower specificity of 86.8% compared to 85.7% for the radiologists, and YOLOv11-n had a higher specificity of 91.2% compared to 85.7%.

Table 1: Diagnostic performance of YOLOv11 series models versus radiologists (mean experience: 5–10 years) on the external test set (n = 215). Sensitivity, specificity, accuracy, and F1 scores are reported. Statistical comparisons against radiologists were performed using McNemar tests, with significance markers (p-values) indicating differences in sensitivity, specificity, and accuracy. Bonferroni and FDR adjusted p-values are reported to account for multiple comparisons

Model	Sensitivity	P-value (vs Radiologist)	Specificity	P-value (vs Radiologist)	Accuracy	P-value (vs Radiologist)	F1 Score	Bonferroni Adjusted P-value	FDR Adjusted P-value
YOLOv11-n	87.3%	0.434	91.2%	0.008	89.8%	0.023	0.9225	1.000	0.465
YOLOv11-s	89.6%	0.267	86.8%	0.776	90.1%	0.001	0.928	1.000	0.308
YOLOv11-m	89.3%	0.267	97.0%	p<0.001	92.1%	0.002	0.940	1.000	0.308
YOLOv11-l	94.2%	0.029	92.4%	0.047	94.9%	0.0014	0.9597	0.145	0.0435
YOLOv11-x	95.6%	0.005	97.4%	p<0.001	96.7%	p<0.001	0.974	0.025	0.0107
Radiologists	85.1%	N/A	85.7%	N/A	85.6%	N/A	0.909	N/A	N/A

In the external test set, the F1 Score, sensitivity, specificity, and accuracy of the improved YOLO model variants were as follows: YOLOv11-n achieved an F1 Score of 0.9225, sensitivity of 87.3%, specificity of 91.2%, and accuracy of 89.8%; YOLOv11-s achieved an F1 Score of 0.928, sensitivity of 89.6%, specificity of 86.8%, and accuracy of 90.1%; YOLOv11-l achieved an F1 Score of 0.9597, sensitivity of 94.2%, specificity of 92.4%, and accuracy of 94.9%; YOLOv11-m achieved an F1 Score of 0.940, sensitivity of 89.3%, specificity of 97.0%, and accuracy of 92.1%; and YOLOv11-x achieved an F1 Score of 0.974, sensitivity of 95.6%, specificity of 97.4%, accuracy of 96.7%, and a P-value of 0.005 for sensitivity compared to radiologists.

For comparison, the radiologists achieved a mean value of F1 Score of 0.909, sensitivity of 85.1%, specificity of 85.7%, and accuracy of 85.6%. When analyzed using the McNemar test, YOLOv11-x showed statistically significant improvements over the radiologists in sensitivity ($p = 0.005$), specificity ($p < 0.001$), and overall accuracy ($p < 0.001$). Meanwhile, YOLOv11-l achieved higher sensitivity (94.2% vs. 85.1%; $p = 0.029$), specificity (92.4% vs. 85.7%; $p = 0.047$), and accuracy (94.9% vs. 85.6%; $p = 0.0014$), all statistically significant. YOLOv11-m had significantly higher specificity (97.0% vs. 85.7%; $p < 0.001$), a non-significant difference in sensitivity (89.3% vs. 85.1%; $p = 0.267$), and significantly higher accuracy (92.1% vs. 85.6%; $p = 0.002$). YOLOv11-s showed a non-significant difference in sensitivity (89.6% vs. 85.1%; $p = 0.267$) and specificity (86.8% vs. 85.7%; $p = 0.776$), with higher accuracy (90.1% vs. 85.6%; $p = 0.001$). YOLOv11-n showed a non-significant difference in sensitivity (87.3% vs. 85.1%; $p = 0.434$) and higher specificity (91.2% vs. 85.7%; $p = 0.008$) and accuracy (89.8% vs. 85.6%; $p = 0.023$).

Diagnostic Efficiency

The processing times per image for the YOLOv11 models and the radiologists are detailed in Table 2. All the model variants were highly efficient compared to the radiologists (all $p < 0.001$ by paired t-test). The lightweight YOLOv11-n model achieved real-time processing, with a total time of 112.3 ms per image, mainly due to the rapid inference time (103.6 ms). The processing time of each model increased as the size and complexity grew: YOLOv11-s (231.3 ms), YOLOv11-m (523.2 ms), and YOLOv11-l (646.1 ms). YOLOv11-x is suited for use alongside radiologists, although it is slower than required, needing 1,174.1 ms per image. Despite being the largest model, it remains quicker than radiologists and usable in the clinic.

Table 2: Diagnostic efficiency comparison: The time per image to process the image in models of YOLOv11 and radiologists (average experience: 5-10 years). Time metrics (milliseconds) incorporate pre-processing, inference, and post-processing (milliseconds). Paired t-tests were used to establish the degree of statistical significance ($p < 0.001$) between the performance of each of the models and radiologists on the same image sets

Model	Pre-process (ms)	Inference (ms)	Post-process (ms)	Total time (ms)	P-value (vs Radiologist)
YOLOv11-n	4.4	103.6	4.3	112.3	<0.001
YOLOv11-s	4.4	222.6	4.3	231.3	<0.001
YOLOv11-m	4.1	515.4	3.7	523.2	<0.001
YOLOv11-l	4.1	637.7	4.3	646.1	<0.001
YOLOv11-x	4.0	1166.3	3.8	1174.1	<0.001
Radiologists	N/A	N/A	N/A	210000	N/A

Discussion

This study has proven that the YOLOv11-MPR framework has the capacity to enhance the CT-based detection of cervical foraminal stenosis (CFS). The results are especially impressive: in the external test set, YOLOv11-x scored 96.7% accuracy, 95.6% sensitivity, and 97.4% specificity (accuracy: $p < 0.001$; sensitivity: $p = 0.005$; specificity: $p < 0.001$ compared to radiologists), surpassing human readers and setting a new benchmark in automated detection (Table 1).

In comparison to prior studies in the spinal region that used the YOLO algorithm, which have focused on lumbar canal stenosis or vertebral fractures, this study targets cervical foraminal stenosis, a more difficult complication to detect and an insufficiently explored field in deep learning research. The framework has a more detailed spatial comprehension of foraminal structure that combines multiplanar CT reconstructions with YOLOv11. To the best of our knowledge, it is one of the early applications of YOLOv11 to cervical foraminal stenosis and CT-only automated CFS detection.

To understand the decision-making of the model better, gradient-weighted class activation maps (Grad-CAM) were generated on the detection head. These visual replications consistently highlighted the edges of the foraminal areas involved in osteophytic invasion and facet-joint hypertrophy, which are key features considered by radiologists during CFS diagnosis. This alignment supports the fact that YOLOv11-MPR predictions can be biologically feasible and makes it more confident that the predictions can be interpreted with high clinical transparency.

The YOLOv11-MPR framework was found to be more effective, as the diagnosis could take less time compared to

manual reading. The performance of the model is proportional to its predicted complexity. The lighter models (n/s) have high speed, whereas the larger models (l/x) offer the highest accuracy. YOLOv11-n runs in real time (112.3 ms/image) and is more than 178 times faster than radiologists (Table 2).

The YOLOv11 scaling method achieves adaptable solutions for speed and accuracy based on clinical demands. YOLOv11-x, which achieves 96.7% accuracy, is suitable for situations requiring conclusive diagnoses, such as surgical planning. It clinically yielded better results in detecting mild stenosis and had a lower false-negative rate by 10.5% compared to radiologists, due to its high sensitivity of 95.6%. The simple and fast YOLOv11-n can be used in emergency triage or in areas where resources are limited, and real-time processing is needed, although it has lower sensitivity (87.3%). YOLOv11-l provides a balanced alternative, with 94.9% accuracy (compared to 85.6% for radiologists) and a computational cost of 646.1 ms per image, offering a reasonable trade-off.

The YOLOv11-MPR framework has high diagnostic accuracy and efficiency, making it suitable for multiple clinical settings. YOLOv11-x, with high specificity (97.4%), can be used as a screening tool in primary care or low-resource settings, providing initial assessments with minimal intervention, followed by additional testing (e.g., MRI) for more precise diagnoses. The model can also serve as a decision-assist tool in certain radiological procedures, improving consistency and reducing observational variability, especially in ambiguous cases of stenosis [8, 31]. Although it does not replace MRI, YOLOv11-MPR complements MRI by offering quicker assessments of osseous foraminal stenosis, while MRI remains necessary to evaluate soft-tissue and discogenic pathology [32].

The small size (e.g., YOLOv11-n at 6.46 MB) and fast inference times enable YOLOv11 to be run in PACS to retrieve, analyze, and report on a large number of images without interfering with workflow. The integration of multiplanar reconstruction (MPR) with YOLOv11 is a novel approach, using axial and sagittal views to better identify the problematic anatomy of the 3D foraminal region [33]. The scalable architecture enables deployment across edge devices (YOLOv11-n) and cloud servers (YOLOv11-x), and it is computationally light enough for portable devices.

Limitations include the inability to measure dynamic changes in stenosis during neck movement and the lower soft-tissue resolution of static CT, which cannot assess discogenic stenosis. Although the data were institutionally homogeneous, variability was introduced across several Centers of a local hospital in different regions, and the sample size, although quite large, may still limit generalizability. This has been mitigated by robust external validation, but a logical follow-up would be a large-scale, prospective, multi-national external validation study to ascertain that it can generalize to other populations and diverse healthcare systems. Patient symptoms were not correlated; clinical relevance could be further enhanced by adding electromyography (EMG) or functional biomarkers. Comparison was made to five radiologists with 5–10 years of experience; this gives a representative panel; however, larger multi-institutional reader studies would be more valid.

Conclusion

This study combined multiplanar CT reconstruction (MPR) with a scalable deep learning model to build a novel diagnostic framework for cervical foraminal stenosis (CFS). Several outcomes were achieved with this framework. It was highly diagnostic, with accuracy (96.7%), specificity (97.4%), and sensitivity (95.6%) significantly higher than those of radiologists, and it improved the detection of mild stenosis compared with radiologists. Secondly, it increased clinical efficiency; the lightweight YOLOv11-n is capable of real-time triage (112.3 ms/image), and YOLOv11-x can diagnose images 178 times faster than clinicians, enhancing workflow effectiveness. Third, it offers a robust CT-based option in areas with limited MRI availability, facilitating more equitable access to spinal disease diagnosis.

In general, the recommended framework will improve the quality and speed of routine radiologic assessments while retaining the necessary radiologist control and interpretive competence. The introduction of dynamic imaging facilities, along with clinical biomarkers, might be used in the future to create more precise diagnostic observations and enhance the capacity to personalize and customize therapeutic planning, depending on the particularities of each case.

Funding

This study has received funding by National Natural Scientific Foundation of China (Number: 82371938).

Institutional Review Board Statement

This study is retrospective and does not require informed consent. The study was approved by the Institutional Review Board (IRB) of Huazhong University of Science and Technology (Approval Number: TJ-IRB202504051).

Conflict of Interest

The authors declare no conflicts of interest.

Abbreviations

CFS	Cervical Foraminal Stenosis
YOLOv11	You Only Look Once version 11
DL	Deep Learning
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
MPR	Multiplanar Reconstruction
EMG	Electromyography
IRB	Institutional Review Board
FDR	False Discovery Rate

References

1. Straface JI, García-Ramos C, Conde-Espinoza R, Osuna-Espinoza MC, Barragan-Hermosillo JL, Acosta-Cortez JL, *et al*. Clinical correlation of oblique magnetic resonance imaging for cervical foraminal stenosis. Cureus, 2025. Doi: https://doi.org/10.7759/cureus.78130
2. Paul A, Lewis M, Saklatvala J, McCall I, Shadforth M, Croft P, *et al*. Cervical spine magnetic resonance imaging in primary care consulters with shoulder pain: A case control study. Ann Rheum Dis. 2007; 66(10):1363-1368. Doi: https://doi.org/10.1136/ard.2006.064881
3. Enomoto K, Eguchi Y, Sato T, Norimoto M, Inoue M, Watanabe A, *et al*. Usefulness of simultaneous magnetic resonance neurography and apparent T2 mapping for the diagnosis of cervical radiculopathy. Asian Spine J. 2022; 16(1):47-55. Doi: https://doi.org/10.31616/asj.2020.0668

4. Seo J, Suk KS, Kwon JW, Kim N, Lee BH, Moon SH, *et al*. Cervical foraminal stenosis as a risk factor for cervical kyphosis following cervical laminoplasty. Spine J. 2022; 22(7):1271-1280. Doi: https://doi.org/10.1016/j.spinee.2022.03.017
5. Abbed KM, Coumans JVC. Cervical radiculopathy: Pathophysiology, presentation, and clinical evaluation. Neurosurgery. 2007; 60(1 Suppl 1):S28-S34. Doi: https://doi.org/10.1227/01.NEU.0000249223.51871.C2
6. Engel G, Bender YY, Adams LC, Boker SM, Fahlenkamp UL, Wagner M, *et al*. Evaluation of osseous cervical foraminal stenosis in spinal radiculopathy using susceptibility-weighted magnetic resonance imaging. Eur Radiol. 2019; 29(4):1855-1862. Doi: https://doi.org/10.1007/s00330-018-5769-4
7. Ko S, Choi W, Lee J. The prevalence of cervical foraminal stenosis on computed tomography of a selected community-based Korean population. Clin Orthop Surg. 2018; 10(4):433-438. Doi: https://doi.org/10.4055/cios.2018.10.4.433
8. Strobel K, Pfirrmann CWA, Schmid M, Hodler J, Boos N, Zanetti M. Cervical nerve root blocks: Indications and role of MR imaging. Radiology. 2004; 233(1):87-92. Doi: https://doi.org/10.1148/radiol.2331030423
9. Finelli DA, Hurst GC, Karaman BA, Simon JE, Duerk JL, Bellon EM. Use of magnetization transfer for improved contrast on gradient-echo MR images of the cervical spine. Radiology. 1994; 193(1):165-171. Doi: https://doi.org/10.1148/radiology.193.1.8090886
10. Meacock J, Schramm M, Selvanathan S, Currie S, Stocken D, Jayne D, *et al*. Systematic review of radiological cervical foraminal grading systems. Neuroradiology. 2021; 63(3):305-316. Doi: https://doi.org/10.1007/s00234-020-02596-5
11. Kang KC, Lee HS, Lee JH. Cervical radiculopathy focus on characteristics and differential diagnosis. Asian Spine J. 2020; 14(6):921-930. Doi: https://doi.org/10.31616/asj.2020.0647
12. Lee HD, Jeon CH, Chung NS, Yoon HS, Chung HW. Is the severity of cervical foraminal stenosis related to the severity and sidedness of symptoms? Healthcare (Basel Switz.). 2021; 9(12):1743. Doi: https://doi.org/10.3390/healthcare9121743
13. Kim GU, Park WT, Chang MC, Lee GW. Diagnostic technology for spine pathology. Asian Spine J. 2022; 16(5):764-775. Doi: https://doi.org/10.31616/asj.2022.0374
14. Gonzalez-Beicos A, Nunez DB. Role of multidetector computed tomography in the assessment of cervical spine trauma. Semin Ultrasound CT MR. 2009; 30(3):159-167. Doi: https://doi.org/10.1053/j.sult.2009.02.001
15. Xu C, Ding ZH, Xu YK. Comparison of computed tomography and magnetic resonance imaging in the evaluation of facet tropism and facet arthrosis in degenerative cervical spondylolisthesis. Genet Mol Res. 2014; 13(4):4102-4109. Doi: https://doi.org/10.4238/2014.May.30.5
16. Schell A, Rhee JM, Holbrook J, Lenehan E, Park KY. Assessing Foraminal Stenosis in the Cervical Spine: A Comparison of Three-Dimensional Computed Tomographic Surface Reconstruction to Two-Dimensional Modalities. Glob Spine J. 2017; 7(3):266-271. Doi: https://doi.org/10.1177/2192568217699190
17. Van der Cruyssen F, Croonenborghs TM, Renton T, Hermans R, Politis C, Jacobs R, *et al*. Magnetic resonance neurography of the head and neck: State of the art, anatomy, pathology and future perspectives. Br J Radiol. 2021; 94(1122):20200798. Doi: https://doi.org/10.1259/bjr.20200798
18. Kim W, Ahn KS, Kang CH, Kang WY, Yang KS. Comparison of MRI grading for cervical neural foraminal stenosis based on axial and oblique sagittal images: Concordance and reliability study. Clin Imaging. 2017; 43:165-169. Doi: https://doi.org/10.1016/j.clinimag.2017.03.008
19. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, *et al*. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021; 8(1):53. Doi: https://doi.org/10.1186/s40537-021-00444-8
20. Park S, Kim JK, Chang MC, Park JJ, Yang JJ, Lee GW. Assessment of fusion after anterior cervical discectomy and fusion using convolutional neural network algorithm. Spine (Phila Pa 1976). 2022; 47(7):1645-1650. Doi: https://doi.org/10.1097/BRS.0000000000004439
21. Kim JK, Choo YJ, Shin H, Choi GS, Chang MC. Prediction of ambulatory outcome in patients with corona radiata infarction using deep learning. Sci Rep. 2021; 11:7989. Doi: https://doi.org/10.1038/s41598-021-87176-0
22. Hong N, Whittier DE, Glüer CC, Leslie WD. The potential role for artificial intelligence in fracture risk prediction. Lancet Diabetes Endocrinol. 2024; 12(9):596-600. Doi: https://doi.org/10.1016/S2213-8587(24)00153-0
23. Deep learning reconstruction for accelerated spine MRI: prospective analysis of interchangeability. Radiology. Available from: https://pubs.rsna.org/doi/10.1148/radiol.212922 Accessed October 23, 2025.
24. Foreman SC, Schinz D, El Husseini M, Goller SS, Weißinger J, Dietrich AS, *et al*. Deep learning to differentiate benign and malignant vertebral fractures at multidetector CT. Radiology. 2024; 310:e231429. Doi: https://doi.org/10.1148/radiol.231429
25. Wan Z, Kong W, Tang Y, Ma F, Ji Y, Peng Y, *et al*. A computer vision-based approach for high-throughput automated analysis of arabidopsis seedling phenotypes. Plant Physiol. 2025; 198(3):kiaf275. Doi: https://doi.org/10.1093/plphys/kiaf275
26. Vogelgesang M, Kaczmarek V, Lopes ACP, Li C, Ionescu E, Schebek L. Automated material flow characterization of WEEE in sorting plants using deep learning and regression models on RGB data. Waste Manag (NY). 2025; 204:114904. Doi: https://doi.org/10.1016/j.wasman.2025.114904
27. Wang J, Zhou C, Wang W, Zhang H, Zhang A, Cui D. A multimodal deep learning model for detecting endoscopic images of near-infrared fluorescence capsules. Biosens Bioelectron. 2025; 278:117251. Doi: https://doi.org/10.1016/j.bios.2025.117251
28. Park HJ, Kim SS, Lee SY, Park NH, Chung EC, Rho MH, *et al*. A practical MRI grading system for cervical foraminal stenosis based on oblique sagittal images. Br J Radiol. 2013; 86(1023):20120515. Doi: https://doi.org/10.1259/bjr.20120515

29. Bhagwat A, Dutta S, Saha D, Reddy MJB. An online 11 kV distribution system insulator defect detection approach with modified YOLOv11 and MobileNetV3. Sci Rep. 2025; 15:15691. Doi: https://doi.org/10.1038/s41598-025-99756-5
30. Yang Z, Lan X, Wang H. Comparative analysis of YOLO series algorithms for UAV-based highway distress inspection: Performance and application insights. Sens (Basel Switz.). 2025; 25(5):1475. Doi: https://doi.org/10.3390/s25051475
31. Fu MC, Webb ML, Buerba RA, Neway WE, Brown JE, Trivedi M, *et al*. Comparison of agreement of cervical spine degenerative pathology findings in magnetic resonance imaging studies. Spine J (Phila Pa 1976). 2016; 16(1):42-48. Doi: https://doi.org/10.1016/j.spinee.2015.08.026
32. Nevalainen MT, Vähä J, Räsänen L, Bode MK. Diagnostic utility of 3D MRI sequences in the assessment of central, recess, and foraminal stenoses of the spine: A systematic review. Skeletal Radiol. 2024; 53(12):2575-2584. Doi: https://doi.org/10.1007/s00256-024-04210-5
33. Badhiwala JH, Ahuja CS, Akbar MA, Witiw CD, Nassiri F, Furlan JC, *et al*. Degenerative cervical myelopathy - update and future directions. Nat Rev Neurol. 2020; 16(2):108-124. Doi: https://doi.org/10.1038/s41582-019-0303-0