BRITISH JOURNAL OF HOSPITAL MEDICINE

ARTICLE

A Deep Learning-Based Fully Automated Vertebra Segmentation and Labeling Workflow

Hongjiang Lu[1], Miao Liu[1], Kun Yu[2], Yuan Fang[1,*], Jing Zhao[3], Yang Shi[4,*]

[1]Department of Radiology, The 903rd Hospital of PLA Joint Logistics Support Force (Xihu Hospital Affiliated with Hangzhou Medical College), Hangzhou, Zhejiang, China
[2]Department of Head and Neck Surgery, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou, Zhejiang, China
[3]Department of Critical Care Medicine, The 903rd Hospital of PLA Joint Logistics Support Force (Xihu Hospital Affiliated with Hangzhou Medical College), Hangzhou, Zhejiang, China
[4]Center for Rehabilitation Medicine, Department of Radiology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou, Zhejiang, China
*Correspondence: fangyuan_0630@163.com (Yuan Fang); shiyang.911026@163.com (Yang Shi)

Abstract

Aims/Background Spinal disorders, such as herniated discs and scoliosis, are highly prevalent conditions with rising incidence in the aging global population. Accurate analysis of spinal anatomical structures is a critical prerequisite for achieving high-precision positioning with surgical navigation robots. However, traditional manual segmentation methods are limited by issues such as low efficiency and poor consistency. This work aims to develop a fully automated deep learning-based vertebral segmentation and labeling workflow to provide efficient and accurate preoperative analysis support for spine surgery navigation robots.

Methods In the localization stage, the You Only Look Once version 7 (YOLOv7) network was utilized to predict the bounding boxes of individual vertebrae on computed tomography (CT) sagittal slices, transforming the 3D localization problem into a 2D one. Subsequently, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm was employed to aggregate the 2D detection results into 3D vertebral centers. This approach significantly reduces inference time and enhances localization accuracy. In the segmentation stage, a 3D U-Net model integrated with an attention mechanism was trained using the region of interest (ROI) based on the vertebral center as input, effectively extracting the 3D structural features of vertebrae to achieve precise segmentation. In the labeling stage, a vertebra labeling network was trained by combining deep learning architectures—ResNet and Transformer, which are capable of extracting rich intervertebral features, to obtain the final labeling results through post-processing based on positional logic analysis. To verify the effectiveness of this workflow, experiments were conducted on a dataset comprising 106 spinal CT datasets sourced from various devices, covering a wide range of clinical scenarios.

Results The results demonstrate that the method performed excellently in the three key tasks of localization, segmentation, and labeling, with a Mean Localization Error (MLE) of 1.42 mm. The segmentation accuracy metrics included a Dice Similarity Coefficient (DSC) of 0.968 $\pm$ 0.014, Intersection over Union (IoU) of 0.879 $\pm$ 0.018, Pixel Accuracy (PA) of 0.988 $\pm$ 0.005, mean symmetric distance (MSD) of 1.09 $\pm$ 0.19 mm, and Hausdorff Distance (HD) of 5.42 $\pm$ 2.05 mm. The degree of classification accuracy reached up to 94.36%.

Conclusion These quantitative assessments and visualizations confirm the effectiveness of our method (vertebra localization, vertebra segmentation and vertebra labeling), indicating its potential for deployment in spinal surgery navigation robots to provide accurate and efficient preoperative analysis and navigation support for spinal surgeries.

Key words: spine; intervertebral disc displacement; scoliosis; X-ray computed tomography; deep learning; image segmentation

Submitted: 13 May 2025 Revised: 20 August 2025 Accepted: 22 August 2025

How to cite this article:
Lu H, Liu M, Yu K, Fang Y, Zhao J, Shi Y. A Deep Learning-Based Fully Automated Vertebra Segmentation and Labeling Workflow. Br J Hosp Med. 2025. https://doi.org/10.12968/hmed.2025.0443

Introduction

As a key structure in the human skeletal system, the spine has the important functions of enabling upright posture and protecting the spinal cord and nerves (Durbas et al, 2025; Kawsar and Chowdhury, 2024). Herniated disc, scoliosis, fractures, and spinal tumors are some of the prevalent spinal disorders worldwide, with their incidence continuing to rise in conjunction with the accelerated aging of the global population (Huang et al, 2024; Wirth et al, 2023). These conditions not only affect the quality of daily life for patients but may also lead to severe neurological damage and even life-threatening situations. Therefore, early detection and accurate diagnosis of spinal diseases are crucial for clinical treatment (Abdou et al, 2025).

Surgery navigation robots are a form of advanced technology that can greatly improve surgical accuracy and safety through automatic analyses of imaging data such as X-rays, computed tomography (CT) scans, and magnetic resonance imaging (MRI) scans, providing high-precision real-time positioning and operation guidance (Devito et al, 2010; Kantelhardt et al, 2011). In the process of utilizing these robots, rapid extraction of accurate spinal anatomical information from a patient is a prerequisite for precise vertebral registration and navigation in subsequent steps.

Manual segmentation is a common method of vertebra segmentation, which requires experienced radiologists or orthopedic experts to conduct detailed analysis of scan images. Results from manual segmentation are generally accurate, but the process is labor-intensive and time-consuming, while the interpretation may be affected by subjective bias. When dealing with complex spinal lesions or cases with large anatomical variations, the consistency of manual segmentation is often difficult to maintain. Traditional segmentation methods, including region growing methods (Fu et al, 2018), regression forests (Glocker et al, 2012), and statistical models (Neubert et al, 2012), can shorten the duration of analysis, but it is highly challenging to achieve high-accuracy segmentation with these methods, especially in complex cases or in situations requiring the utilization of varying imaging instruments. Various deep learning approaches have been investigated and applied for spinal segmentation (Galbusera et al, 2019). By training on large annotated datasets, deep learning models can autonomously learn the anatomical structural characteristics of the spine and achieve highly robust segmentation despite variations in image quality. Therefore, developing a fully automated spinal vertebral segmentation and labeling workflow based on deep learning, which can provide precise and rapid anatomical information for surgical navigation robots, is an important research direction in the fields of medical image processing and robot-assisted surgery.

In recent years, the adoption of deep learning has made significant strides in the field of medical image analysis. In the task of spinal segmentation, convolutional neural networks (CNNs) and their variants, such as U-Net and 3D U-Net (Çiçek et al, 2016), have become mainstream methods (Masuzawa et al, 2020; Chmelik et al, 2018; Cheng et al, 2021). These methods demonstrate excellent performance in the automated analysis and segmentation of spinal images and can efficiently and accurately extract the spine and vertebral regions. However, although deep

learning models can provide relatively accurate segmentation results, several challenges remain in practical applications. First, spinal image datasets are typically limited in size, and the manual annotation process is both time-consuming and dependent on expert knowledge, making it difficult to ensure the quality and consistency of data labeling. These limitations present significant challenges for training a deep learning model. To address this issue, researchers have employed various data augmentation techniques (Chaitanya et al, 2021) to mitigate the problem of data scarcity. Secondly, morphological variations in vertebrae are apparent among patients, especially in the presence of lesions and deformities, where these changes could present a higher degree of complexity for segmentation. Accurately segmenting vertebrae within such complex and variable anatomical structures remains one of the challenges in current research. To enhance the robustness of segmentation models, researchers have begun to explore multi-modal learning, integrating information from different imaging modalities such as CT, MRI, and X-rays to improve model performance in the context of varying image qualities and anatomical variations (Yan et al, 2024). Meanwhile, an increasing number of studies have begun to focus on deep learning-based multi-task learning frameworks (Tran et al, 2020), such as simultaneously performing vertebra segmentation and labeling, or combining vertebra segmentation with lesion detection tasks, to build more comprehensive and intelligent models. In addition, technologies such as attention mechanisms and reinforcement learning have also been introduced to improve the model's focus on key anatomical regions and improve its overall robustness (Wang et al, 2022; Zhang et al, 2021). However, existing methods still face several limitations and often struggle to efficiently perform the localization, segmentation, and labeling of the spine, especially in achieving high accuracy and generalization in the segmentation and labeling of pathological vertebrae.

In light of the aforementioned challenges, this study proposes a deep learning-based workflow for vertebra analysis in spinal surgery navigation robots, aiming to enhance efficiency and accuracy in preoperative analysis and navigation support. This workflow, which consists of three cascaded deep learning models, addresses the key tasks of spinal localization, segmentation, and labeling.

Methods

Vertebra Localization

As the first step in the vertebral analysis workflow, vertebra localization was used to locate the central position of each vertebra in the spine based on the input CT images. In this step, we applied the You Only Look Once version 7 (YOLOv7) model (Wang et al, 2023), which offers high accuracy and real-time performance.

The original CT image data dimensions were normalized to H' $\times$ 512 $\times$ 512 (where H' depended on the original size and spacing). The normalized image had the same spacing in three directions. The 3D CT images were sliced along the sagittal plane, and the dimensions of each slice were normalized to 640 $\times$ 640 pixels. These 2D slices were utilized as model inputs, with bounding box annotations delineated around the larger vertebrae within the slices to provide training labels.

We implemented a fully automated vertebral analysis workflow for a spinal surgery navigation robot, encompassing vertebral localization, segmentation, and labeling, by concatenating three deep learning models (Fig. 1).

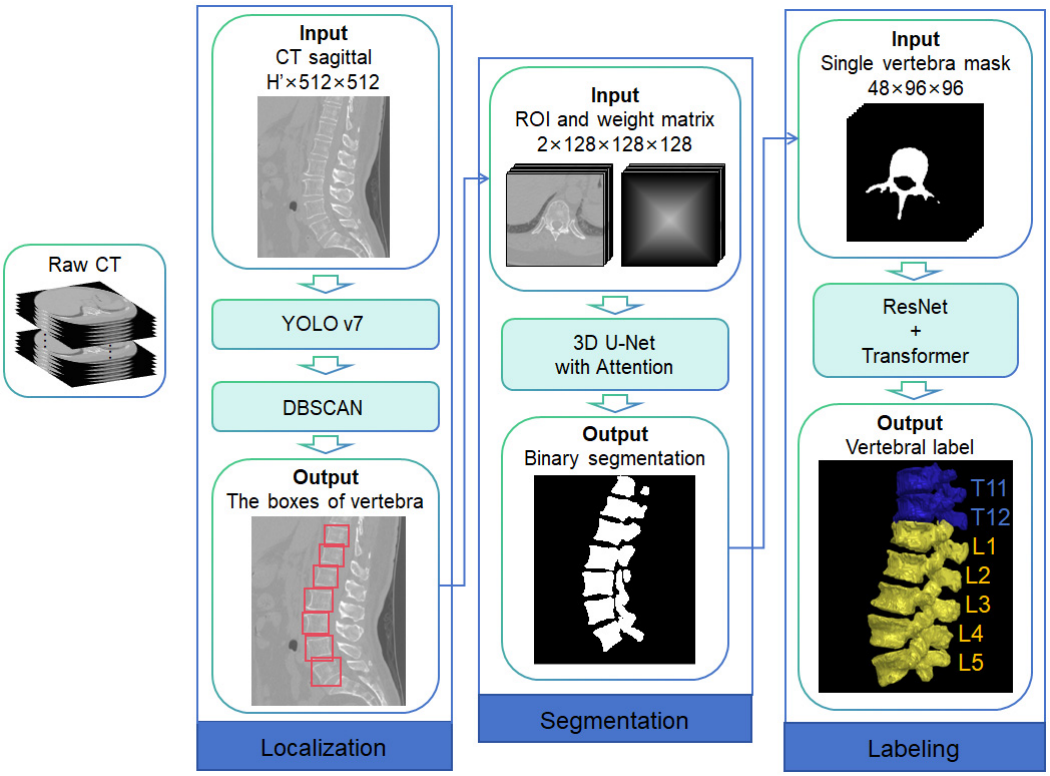

Fig. 1. Fully automated vertebra segmentation and labeling workflow based on deep learning. CT, computed tomography; YOLOv7, You Only Look Once version 7; DBSCAN, Density-Based Spatial Clustering of Applications with Noise; ROI, region of interest. Fig. 1 was created using Microsoft Visio software (version 2022, Microsoft Corporation, Hangzhou, China).

YOLOv7 primarily consists of three main modules. The Backbone module is mainly used for extracting image features, employing an improved Cross Stage Partial (CSP) Net as the backbone network to further optimize efficiency and expressive ability. The Neck module is utilized for fusing features at different scales, using the Path Aggregation Network (PAN) structure to achieve the transmission and enhancement of multi-scale features. The head module is utilized for generating the final detection results. Additionally, YOLOv7 incorporates innovations in modules such as the Extended Efficient Layer Aggregation Network (E-ELAN) and the Spatial Pyramid Pooling Fast Cross Stage Partial Connections with Convolution (SPPFCSPC), which further enhance the model's performance. The YOLOv7 model was applied to all CT sagittal slices to predict the vertebral bounding boxes.

The 2D detection results obtained from YOLOv7, including the bounding boxes and their centers, were projected onto a single plane. Subsequently, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method was employed to perform a two-stage clustering on these projected bounding boxes (Ester et al, 1996). Clustering at the first stage was performed at the 3D position of the

centers of all bounding boxes. The bounding boxes of the same vertebra distributed on different sagittal slices were clustered into one class. The number of classes was the number of vertebrae contained in the CT image. Clustering at the second stage was performed on the length and width of the bounding boxes in each class, and the class with the most members was used as the main class. By excluding bounding boxes of the remaining classes, a set of 2D slice bounding boxes of each vertebra was obtained. Additional logical processing was applied to the clustering results, where the number of 2D bounding boxes for each vertebra was counted, and vertebra classes with a bounding box count of less than 0.2 times the maximum number were excluded. A region of interest (ROI) was defined for each of the N vertebrae identified in the CT image, with each ROI centered on the respective vertebra's center point. The ROI was then enlarged and padded with background values to ensure uniformity, resulting in a standardized image matrix with a size of 128 × 128 × 128 for each vertebra.

Vertebra Segmentation

After obtaining the number of vertebrae and their respective center points as described in section "Vertebra Localization" the next step is to perform segmentation on each vertebra to achieve a mask of each segment within the CT images. Since precise segmentation of vertebra is a pixel-level classification task, we thus introduced the 3D U-Net network combined with an attention mechanism.

A region of interest (ROI) was defined for each of the N vertebrae identified in the CT image, with each ROI centered on the respective vertebra's center point. The ROI was then enlarged and padded with background values to ensure uniformity, resulting in a standardized image matrix with a size of 128 × 128 × 128 for each vertebra. In addition, a weight matrix was constructed to achieve a larger weight at the center, with the weight in the surrounding being smaller. The image matrix and the weight matrix were concatenated in high dimensions to form an input tensor with a size of 2 × 128 × 128 × 128. This process was repeated for all N vertebrae, yielding N distinct input tensors.

3D U-Net consists of an encoder and decoder structure, where the encoder extracts low-level and high-level features, and the decoder restores the original image size by gradually upsampling. To further enhance segmentation accuracy, we incorporated an attention mechanism into the skip connections between each corresponding encoder and decoder pair. This attention mechanism dynamically adjusts the weights of each position in the feature map by learning a spatial attention map from the input feature map, enabling the network to adaptively focus on the spinal vertebral region while ignoring background and noise information. This enhances the network's focus on the target region. The output layer utilizes a sigmoid activation function, enabling the network to generate a predicted value between 0 and 1 for each voxel.

Voxels with predicted values greater than 0.5 were marked as 1, whereas those with predicted values less than or equal to 0.5 were labeled as 0; therefore, the binary segmentation mask of each vertebra was obtained. Based on the displacement between the center of the currently processed vertebra and the original CT image,

the vertebra mask was mapped back to the original CT image space. The central coordinates of the vertebra were recalculated based on the pixel-level segmentation results of each vertebra.

To further improve segmentation accuracy, we incorporated an attention mechanism into the skip connections between each corresponding encoder and decoder pair. This enables the network to adaptively focus on the vertebral region at each spatial location while disregarding background and noise information. The output layer uses a sigmoid activation function, allowing the network to yield a predicted value in the range of 0 to 1 for each voxel. The details on the 3D U-Net architecture used for single vertebra segmentation and how the self-attention mechanism is integrated are presented in Fig. 2.

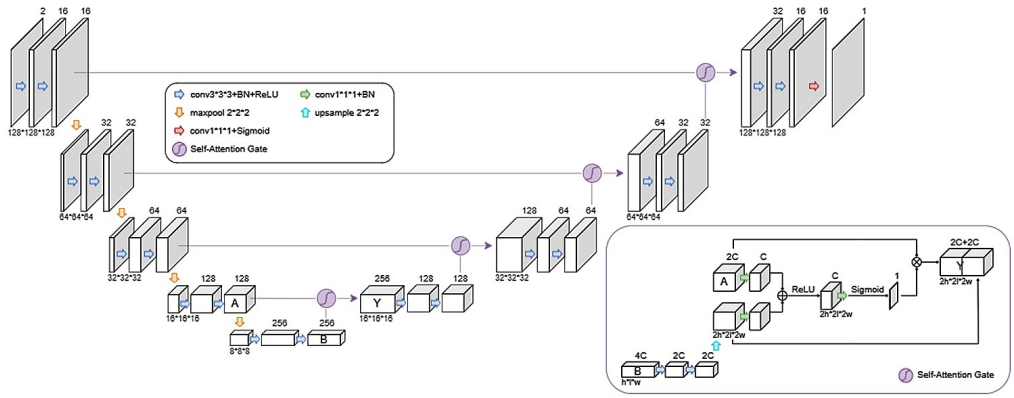

Fig. 2. The 3D U-Net architecture for single vertebra segmentation, which incorporates a self-attention mechanism. The calculation process of the self-attention module is shown in the inset at the bottom-right corner. BN, Batch Normalization; ReLU, Rectified Linear Unit. The asterisk in the diagram indicates multiplication.

Vertebra Labeling

For each vertebra, labeling was performed with the goal of obtaining the actual label of each vertebra. This is a crucial step in the fully automated vertebral analysis workflow for surgical navigation robots. In this step, we combined the deep learning architectures, namely ResNet (He et al, 2016) and Transformer (Rao et al, 2021).

The binary mask of each individual vertebra, generated by the segmentation model described in section "Vertebra Segmentation", was used as the input data and resized to form an input matrix with a size of 48 × 96 × 96.

ResNet solves the gradient vanishing problem in deep learning-based training through residual connections and can effectively capture local features in images. Transformer networks can handle long-distance dependency problems through multi-head self-attention mechanisms, enhance the model's understanding of global information, and can be used to analyze the spatial structure of vertebrae. The ResNet34 network was used to extract the local features of the vertebra, which were then integrated into the Transformer network. The specific network structure is shown in the

Fig. 3. One-hot encoding was used in the output layer to represent the vertebra labels, with the predicted label corresponding to the class with the highest probability score.

Global logic processing was performed on the model outputs to address some possible recognition errors. First, the identification results of each vertebra were assembled into a sequence according to their spatial order from top to bottom. The entire sequence was then examined, with particular attention to the junctions between the cervical-thoracic and thoracic-lumbar regions. If there were duplicate or missing numbers at the junction, the distance between the vertebrae centers was used to determine whether it was an identification error or a missed segmentation. The labels of each vertebra were corrected according to the spatial order. Finally, the results of vertebra identification and labeling were obtained.

ResNet addresses the vanishing gradient problem in deep learning training through residual connections and effectively captures local features in images. The Transformer network, on the other hand, handles long-range dependencies via a multi-head self-attention mechanism, enhancing the model's ability to understand global information and making it suitable for analyzing the spatial structure of vertebrae. The specific network architectures are depicted in Fig. 3. For each vertebral class, one-hot encoding was used in the output layer to represent the vertebral labels, with the predicted label corresponding to the class with the highest probability score.

The accuracy assessment follows a two-level hierarchy:

(1) Single-vertebra classification accuracy (Table 4) reflects independent recognition capability; (2) Sequence accuracy (Table 3) evaluates global label logic constrained by inter-vertebral spatial relationships.

Post-processing with positional logic correction can improve the sequence accuracy by 5% (Table 5).

Localization errors were calculated as the Euclidean distance between predicted and ground truth vertebral centers. The 90.72% error threshold was determined by means of percentile analysis of 500 clinical cases, with 3.0 mm set as the safety margin for surgical navigation.

Model Training

This study was approved by the Medical Research Ethics Committee of the 903rd Hospital of PLA Joint Logistics Support Force (Xihu Hospital Affiliated with Hangzhou Medical College) (approval no. 20220722/08/02/019) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all patients.

Patients meeting the following criteria were included in the present study: (1) age $\geq$18 years; (2) complete thoracic/lumbar CT scans; and (3) no prior spinal surgery. Exclusion criteria include: (1) severe artifacts (including metal-induced artifacts with Hounsfield Unit [HU] $>$3000); (2) congenital spinal deformities; (3) metastatic spinal lesions; (4) incomplete coverage of T1-L5 vertebrae; and (5) motion artifacts affecting $\geq$3 consecutive slices. The CT images used in this paper were obtained from 116 patients, including 564 complete lumbar vertebrae and 837

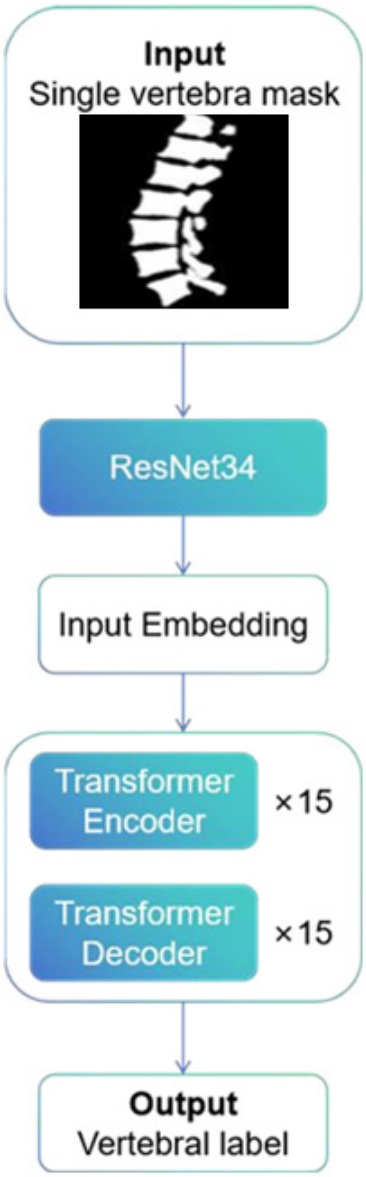

Fig. 3. Single vertebra labeling using deep learning architectures, namely ResNet and Transformer.

complete thoracic vertebrae. The dataset contained normal vertebrae and vertebrae in different pathological states (such as fractures, degeneration, etc.). A total of 100 cases were randomly selected from among the original cohort as the training set, and the remaining 16 cases were regarded as the test set.

For the vertebra localization task, the bounding box coordinates and center position of each vertebra were annotated as training labels. The bounding box was used to represent the area where the vertebra was located, and the center coordinates were used as a reference for positioning. A multi-task loss function was used, including the average position error, i.e., the average distance between the center of the vertebra and the center label, and the target confidence error, i.e., the coincidence of the bounding box. For this task, the stochastic gradient descent (SGD)

optimizer was used, the learning rate was set to 0.001, and 50 epochs were trained, with each batch size set to 16.

For the vertebra segmentation task, all CT images were normalized to 128 × 128 × 128, and the vertebral regions were manually annotated. Binary masks of the same size were constructed as training labels, where the vertebral region was marked as 1 and other regions were marked as 0. The loss function included weighted cross-entropy loss and Dice loss. For this task, the Adam optimizer was used, the learning rate was set to 0.0001, and 150 epochs were trained, with each batch size set to 4.

For the vertebra labeling task, the binary mask used in vertebra segmentation was annotated with the corresponding vertebral category (such as L1, T12, etc.) as the training label. For this task, the cross-entropy loss and Adam optimizer were used; the learning rate was set to 0.0001, and the training was performed for 60 epochs, with each batch size set to 16.

A flowchart depicting the patient inclusion process is shown in Fig. 4.

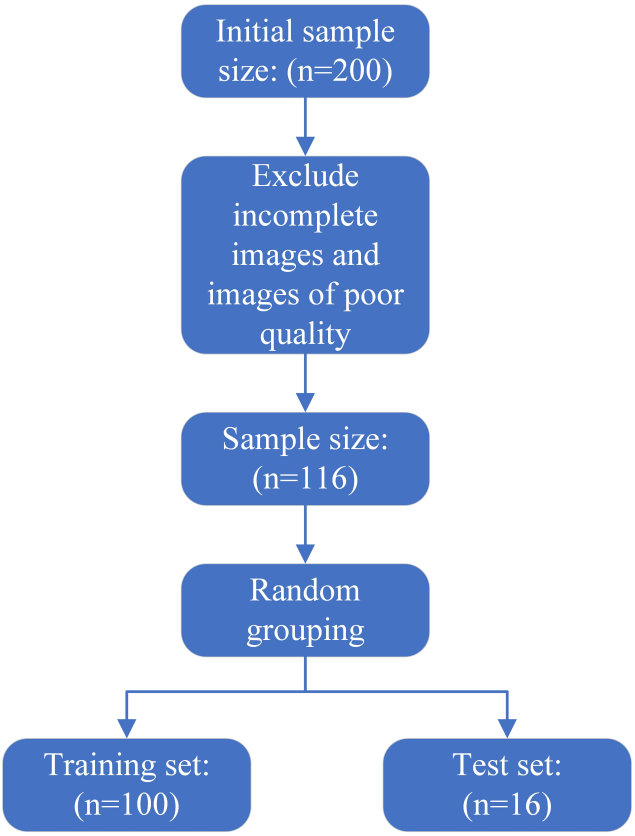

Fig. 4. Flowchart depicting patient inclusion (n = 116).

This work used a Windows 10 system, 64 G memory, NVIDIA GeForce RTX 3060 Ti GPU, CUDA 11.1. The Pytorch framework was used to build the network structure and trained under a Python 3.9 (TechGuru Software Solutions, Beijing, China) environment.

Evaluation Criteria

For the vertebra localization task, the Mean Localization Error (MLE) and Intersection over Union (IoU) were used for evaluation on sagittal 2D slices. MLE was used to evaluate the average distance between the vertebral center predicted by the model and the actual vertebral center (in millimeters), and it was measured using Eqn. 1:

$$\mathrm{MLE} = \frac{1}{N}\sum_{i=1}^{N}\left|\bar{P}_i - \bar{G}_i\right| \tag{1}$$

where $\bar{P}_i$ is the predicted center coordinate of the i-th vertebra, $\bar{G}_i$ is the ground-truth center coordinate of the i-th vertebra, and N is the number of vertebrae in the test set. IoU was used to evaluate the degree of overlap between the predicted bounding box and the ground-truth bounding box, and it was calculated using Eqn. 2:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

where A is the predicted bounding box and B is the ground-truth bounding box. After completing the DBSCAN clustering process to obtain the 3D center of each vertebra, the result was compared with the ground-truth center of each vertebra. MLE was used for evaluation and position accuracy of the center was calculated.

For the vertebra segmentation task, indicators such as Dice Similarity Coefficient (DSC), IoU, Pixel Accuracy (PA), False Positive Dice (FPD), False Negative Dice (FND), Mean Surface Distance (MSD), and Hausdorff Distance (HD) were used for evaluation. Similar to IoU, DSC was used to evaluate the overlap between the predicted segmentation result and the ground-truth segmentation label, and it was determined using Eqn. 3:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{3}$$

where A is the predicted segmentation result, while B is the actual annotated vertebral area. PA is defined as the ratio of the number of correctly segmented pixels to the total number of pixels, which is used to measure the degree of match between the predicted result and the ground-truth label at the pixel level. FPD was used to measure the degree to which the background area was incorrectly segmented as the foreground area, and FND was used to measure the degree to which the foreground area was missed. Eqns. 4,5,6 were used to calculate the PA, FPD, and FND, respectively:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$FPD = \frac{FP}{FP + TN} \tag{5}$$

$$FND = \frac{FN}{TP + FN} \tag{6}$$

where *TP* is True Positives, *TN* is True Negatives, *FP* is False Positive, *FN* is False Negatives. MSD is the average surface distance between the segmentation result and the ground-truth label. It is mainly used to evaluate the average difference between the boundaries of two-point sets and reflects the accuracy of the segmentation boundary. The formulas for determing HD are given in Eqns. 7,8,9:

$$HD = \frac{1}{2}\left(\frac{\sum_{i=1}^{|S_G|} d(p_i, S_P)}{|S_G|} + \frac{\sum_{j=1}^{|S_P|} d(q_j, S_G)}{|S_P|}\right) \tag{7}$$

$$d(p_i, S_P) = \min_{q_i \in S_P} d(p_i, q_i) \tag{8}$$

$$d(q_j, S_G) = \min_{p_i \in S_G} d(p_i, q_j) \tag{9}$$

where S_G is the surface point set of the ground-truth label; S_P is the surface point set of the predicted segmentation; and $d(p, q)$ is the Euclidean distance between points p and q. In simple terms, S_G and S_P can be understood as sizes of point sets P and G, respectively. HD refers to the maximum shortest distance between the boundaries of two-point sets, which is calculated by determining the maximum difference between the two-point sets, using Eqn. 10:

$$HD = \max\left(\max_{p \in G} \min_{q \in P} d(p, q), \max_{q \in P} \min_{p \in G} d(p, q)\right) \tag{10}$$

For the vertebra labeling task, the overall classification accuracy, as well as the recall and precision of vertebra labeling of each category, were calculated to evaluate the performance of the algorithm.

In addition to evaluating our internal dataset, we conducted comparative experiments on the public Vertebral Segmentation 2020 (VerSe2020) dataset to benchmark against state-of-the-art methods (Saeed et al, 2025; Xu et al, 2024). Evaluation metrics, such as Dice coefficient and mean IoU, were compatible with our internal evaluation protocol to ensure comparability.

Results

Experimental Results

The proposed vertebral analysis workflow consists of three modules: localization, segmentation, and labeling. Three deep learning models were used and post-processed, respectively. The outputs of the deep learning models and the post-processing results in the three modules were quantitatively evaluated and visualized.

Fig. 5 shows the vertebra localization results of a representative case during the intermediate process. Fig. 5a–c shows the results of the YOLOv7 network. The bounding boxes of all vertebrae were marked on all sagittal slices, and the centers of the bounding boxes were taken as the vertebral centers. When this step was being performed, we set the bounding box to be the minimum bounding rectangle

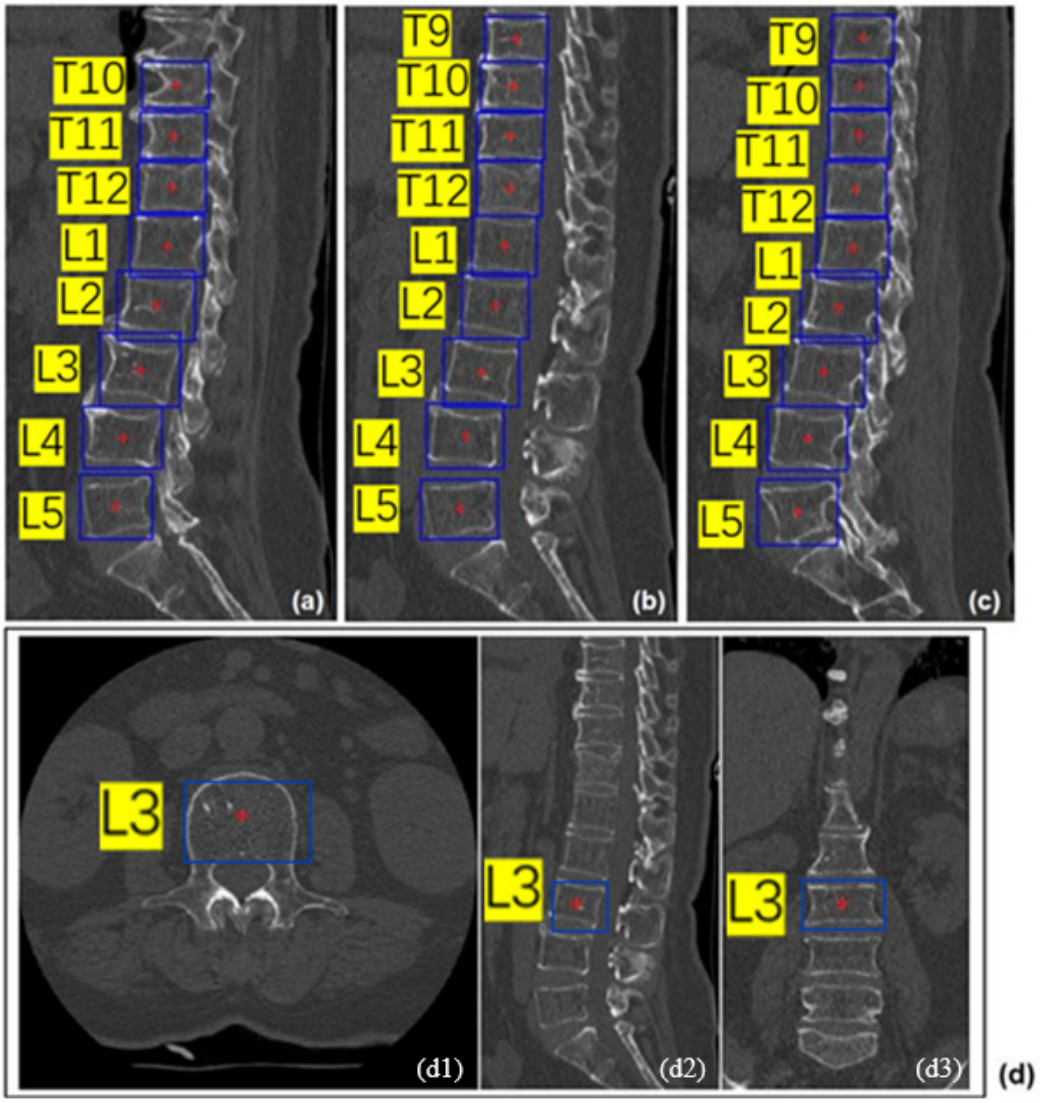

Fig. 5. A representative case of the vertebra localization task. (a–c) The vertebra bounding boxes (blue) and centers (red) at different sagittal plane locations, located at 1.21 mm, 11.06 mm, and 23.30 mm, respectively. (d) The L3 vertebra center location (blue boxes) identified after DBSCAN clustering. The d1 shows the precise localization of the L3 vertebral body center point (red marker) in the axial CT slice. This point is located within the central trabecular bone region of the vertebral body and aligns with the pedicle projection line; d2 demonstrates the position of the L3 vertebral body's center point in the sagittal plane, located at the middle third of the vertebral body and equidistant from the superior and inferior endplates, consistent with the characteristics of the vertebral body's biomechanical center; d3 demonstrates the symmetrical distribution of the L3 vertebral body center point in the coronal plane, positioned along the vertebral midline, with symmetrical projections of the bilateral pedicles. This validates the spatial accuracy of the DBSCAN clustering algorithm.

of the vertebra, and no rotation was applied to the rectangle. Therefore, the bounding box area of some tilted vertebrae appeared to be larger, but this did not affect subsequent calculations. In addition, when the sagittal slice was located at the edge of the vertebra, the vertebra may present in the shape of a polygon, resulting in the inability to identify the bounding box, such as T9 as shown in Fig. 5a. However,

Table 1. Evaluation results of the vertebra localization task.

	Part 1		Part 2			
	MLE (mm)	IoU	MLE (mm)	Acc1.5 (%)	Acc2.0 (%)	Acc3.0 (%)
T1	2.01 ± 0.71	0.825 ± 0.054	1.95	26.83	46.34	85.37
T2	1.87 ± 0.68	0.846 ± 0.049	1.65	35.71	61.90	88.10
T3	1.98 ± 0.85	0.841 ± 0.051	1.73	30.23	53.49	76.74
T4	1.51 ± 0.77	0.849 ± 0.035	1.43	40.91	65.91	86.36
T5	1.66 ± 0.75	0.828 ± 0.044	1.62	36.84	71.93	91.23
T6	1.83 ± 0.82	0.851 ± 0.046	1.78	39.66	56.90	93.10
T7	1.24 ± 0.92	0.866 ± 0.052	1.05	61.02	77.97	94.92
T8	1.33 ± 0.88	0.848 ± 0.033	1.24	65.17	79.78	95.51
T9	1.99 ± 0.69	0.852 ± 0.031	1.91	21.35	42.70	89.89
T10	1.97 ± 0.72	0.851 ± 0.056	1.85	28.57	49.45	89.01
T11	1.38 ± 0.81	0.847 ± 0.043	1.24	44.64	74.11	92.86
T12	1.61 ± 0.95	0.839 ± 0.047	1.49	49.11	66.96	90.18
L1	1.26 ± 0.88	0.853 ± 0.038	1.02	53.57	74.11	91.96
L2	1.19 ± 0.71	0.858 ± 0.041	0.86	62.28	79.82	94.74
L3	1.46 ± 0.72	0.855 ± 0.049	1.35	56.14	71.05	92.11
L4	1.15 ± 0.85	0.851 ± 0.032	0.89	62.28	84.21	96.49
L5	1.32 ± 0.83	0.849 ± 0.029	1.03	59.09	78.18	93.64
Mean	1.51 ± 0.82	0.849 ± 0.046	1.42	45.49	66.75	90.72

Notes: Part 1 shows the results on the 2D slice of the CT sagittal plane; Part 2 shows the results of the vertebral center in 3D space after clustering post-processing. Acc1.5, Acc2.0, and Acc3.0 represent the accuracy of MLE within 1.5 mm, 2 mm, and 3 mm, respectively.
Abbreviations: IoU, Intersection over Union; MLE, Mean Localization Error.

the bounding boxes of T10 and T11 in Fig. 5a were correctly recognized, indicating that the shape of the polygon would affect the recognition results. Following the post-processing with DBSCAN clustering, bounding boxes were divided into distinct clusters, each representing a vertebra. The average center of the bounding boxes within each cluster was then computed to determine the final vertebral center. As a result, the center positions of all vertebrae were obtained. Fig. 5d shows the position of the center of one vertebra (L3) in the axial, sagittal, and coronal planes. The number of centers corresponded to the number of vertebrae detected, forming the basis for subsequent calculations.

The quantitative evaluation results of the vertebral localization module are shown in Table 1. The mean MLE of the vertebral center output by the YOLOv7 network was around 2 mm, with a standard deviation of 1 mm, and the bounding box IoU exceeded 0.8, indicating good overlap between the predicted box and the ground-truth box and relatively accurate localization. After processing with DBSCAN clustering, the MLE of the vertebral center was further reduced because some abnormal detection boxes were excluded in the process, and the mean processing corrected some offsets, thereby yielding better results. In order to evaluate the grading accuracy of center offset, we set three levels: 1.5 mm, 2.0 mm, and

Table 2. Quantitative evaluation results of the vertebra segmentation task.

	DSC	IoU	PA	FPD	FND	MSD (mm)	HD (mm)
T1	0.923 ± 0.019	0.826 ± 0.012	0.984 ± 0.005	0.014	0.032	1.24 ± 0.23	7.46 ± 1.39
T2	0.949 ± 0.013	0.851 ± 0.016	0.987 ± 0.004	0.012	0.030	1.06 ± 0.20	5.69 ± 2.05
T3	0.961 ± 0.018	0.876 ± 0.019	0.989 ± 0.003	0.009	0.034	1.05 ± 0.19	7.68 ± 2.11
T4	0.965 ± 0.006	0.863 ± 0.009	0.988 ± 0.003	0.011	0.030	1.02 ± 0.15	6.36 ± 1.28
T5	0.977 ± 0.020	0.879 ± 0.030	0.989 ± 0.003	0.010	0.026	1.13 ± 0.23	6.30 ± 2.26
T6	0.939 ± 0.015	0.850 ± 0.015	0.986 ± 0.007	0.013	0.026	1.10 ± 0.25	5.40 ± 3.29
T7	0.963 ± 0.012	0.890 ± 0.013	0.989 ± 0.004	0.010	0.026	1.16 ± 0.23	3.65 ± 2.01
T8	0.979 ± 0.013	0.889 ± 0.011	0.989 ± 0.004	0.009	0.030	1.09 ± 0.16	4.89 ± 2.23
T9	0.973 ± 0.013	0.889 ± 0.028	0.989 ± 0.007	0.009	0.029	0.95 ± 0.15	6.83 ± 0.92
T10	0.963 ± 0.012	0.877 ± 0.013	0.987 ± 0.006	0.011	0.024	1.05 ± 0.16	5.02 ± 1.63
T11	0.979 ± 0.018	0.891 ± 0.011	0.989 ± 0.004	0.010	0.023	1.19 ± 0.19	4.21 ± 3.20
T12	0.969 ± 0.013	0.891 ± 0.022	0.989 ± 0.008	0.010	0.027	1.14 ± 0.13	4.27 ± 1.05
L1	0.979 ± 0.018	0.895 ± 0.016	0.989 ± 0.009	0.009	0.026	1.04 ± 0.25	6.00 ± 2.55
L2	0.979 ± 0.014	0.894 ± 0.023	0.989 ± 0.007	0.010	0.027	1.02 ± 0.13	3.99 ± 2.16
L3	0.971 ± 0.011	0.891 ± 0.033	0.988 ± 0.007	0.011	0.024	1.16 ± 0.16	5.19 ± 1.99
L4	0.960 ± 0.009	0.878 ± 0.018	0.987 ± 0.005	0.012	0.024	1.00 ± 0.19	4.86 ± 3.06
L5	0.972 ± 0.010	0.898 ± 0.024	0.989 ± 0.004	0.010	0.022	1.15 ± 0.22	4.41 ± 1.69
Mean	0.968 ± 0.014	0.879 ± 0.018	0.988 ± 0.005	0.011	0.027	1.09 ± 0.19	5.42 ± 2.05

Abbreviations: DSC, Dice Similarity Coefficient; FND, False Negative Dice; FPD, False Positive Dice; HD, Hausdorff Distance; IoU, Intersection of Union; MSD, mean symmetric distance; PA, Pixel Accuracy.

3.0 mm. At the 1.5 mm level, the performance of vertebral center offset of all categories was average, but the results of lumbar vertebrae were slightly better than those of the thoracic vertebrae. This difference may be attributed to the data distribution. The deep learning-based analysis of the thoracic vertebrae was relatively insufficient due to a higher number of lumbar vertebrae than thoracic vertebrae in our data set. The center offset of 90.72% of the vertebrae can be $\leq$3.0 mm, and the results of lumbar vertebrae were slightly better than those of the thoracic vertebrae. In the vertebra segmentation module, we divided a larger ROI based on the vertebral center for segmentation. Provided that the entire vertebra was included in the ROI, the calculation could be executed reliably. Therefore, the accuracy of the current vertebral center offset could meet the requirements for subsequent calculations.

The binary segmentation masks of each vertebra were output by the vertebral segmentation module, and the segmentation results were compared with the ground-truth labels. The quantitative evaluation results are presented in Table 2. The DSC was generally above 0.95, and the IoU was generally above 0.85, indicating a high degree of overlap between the segmentation results and the ground-truth labels. For results related to the thoracic vertebrae, the mean DSC ranged from 0.923 to 0.979, and the mean IoU ranged from 0.826 to 0.891. The lowest DSC and IoU values of the thoracic vertebrae were observed at T1, while the highest values were observed at T11 and T12; the DSC of lumbar vertebrae all exceeded 0.960. This phenomenon may be caused by fewer cases with T1 but more cases containing T11 to L5 in the

dataset. The greater the amount of data, the higher the learning degree of the 3D U-Net model. The mean values of PA, FPD, and FND were 0.988, 0.011, and 0.027, respectively, indicating that the segmentation results had good pixel-level accuracy. PA is the ratio of correctly segmented pixels to the total number of pixels, including both vertebral and background pixels; therefore, the PA values were high across all categories. FPD was lower than FND, but upon examining the data, we found that the low FPD values were primarily due to the significantly larger number of background pixels compared to vertebral pixels, resulting in a larger denominator in the FPD formula. Considering the subsequent requirements of spinal surgery navigation robots, we designed the segmentation module with a tendency to ensure that all pixels belonging to the vertebral region are segmented, resulting in more false positive pixels than false negative pixels. MSD is the average distance between the surfaces of the segmentation results and the ground truth, while HD is the maximum shortest distance between the surfaces of the segmentation results and the ground truth; both were used to measure discrepancies in surface distance. The mean MSD was 1.09, indicating good precision of the segmentation boundaries. The mean HD was 5.42, suggesting that the maximum discrepancy between the segmented surfaces and the ground-truth surfaces was relatively small. The highest MSD value occurred at T1, and the HD values of thoracic vertebrae were slightly worse than those of lumbar vertebrae, possibly due to the segmentation challenges at the junctions of thoracic vertebrae and ribs, leading to larger local surface discrepancies.

The visualization results of vertebral segmentation are displayed in Fig. 6, with panels (a–c) showing the segmentation results of three vertebral segments of the same representative case. Fig. 6 shows the segmentation boundaries overlaid on the original CT images, where it can be observed that the segmentation boundaries generally coincide well with the actual vertebra. Especially in the vertebral plate region, the segmentation boundary lines for the vertebral plate, pedicle, and vertebral foramen structures were very clear in all three views and generally aligned well with the edges of the vertebrae in the CT images. The boundary structures of the spinous and transverse processes showed a higher degree of complexity, making segmentation in this region more challenging than in the vertebral plate region. Fig. 6b1 shows issues such as indentation of the transverse process boundary and missing tips, whereas Fig. 6c1 depicts the shifting spinous process boundary. Since the upper and lower vertebrae are connected by the spinous processes, the upper and lower boundaries of the spinous processes are also prone to unclear segmentation, as shown in Fig. 6b2. Fig. 6a4–c4 display the 3D models of the T10, L1, and L3 after segmentation, showing that the segmented surfaces were smooth and continuous, and the structures such as the vertebral plate, foramen, pedicle, and spinous process were very clear. Both the quantitative evaluation and the visualization results demonstrated that our vertebra segmentation module performs effectively, with the 3D U-Net model combined with an attention mechanism proving highly suitable for vertebra segmentation tasks.

The binary results of vertebra segmentation were input into the vertebra labeling network, and the outcomes are presented in Table 3. Since the vertebral labeling is a multi-classification task, its overall classification performance was evaluated

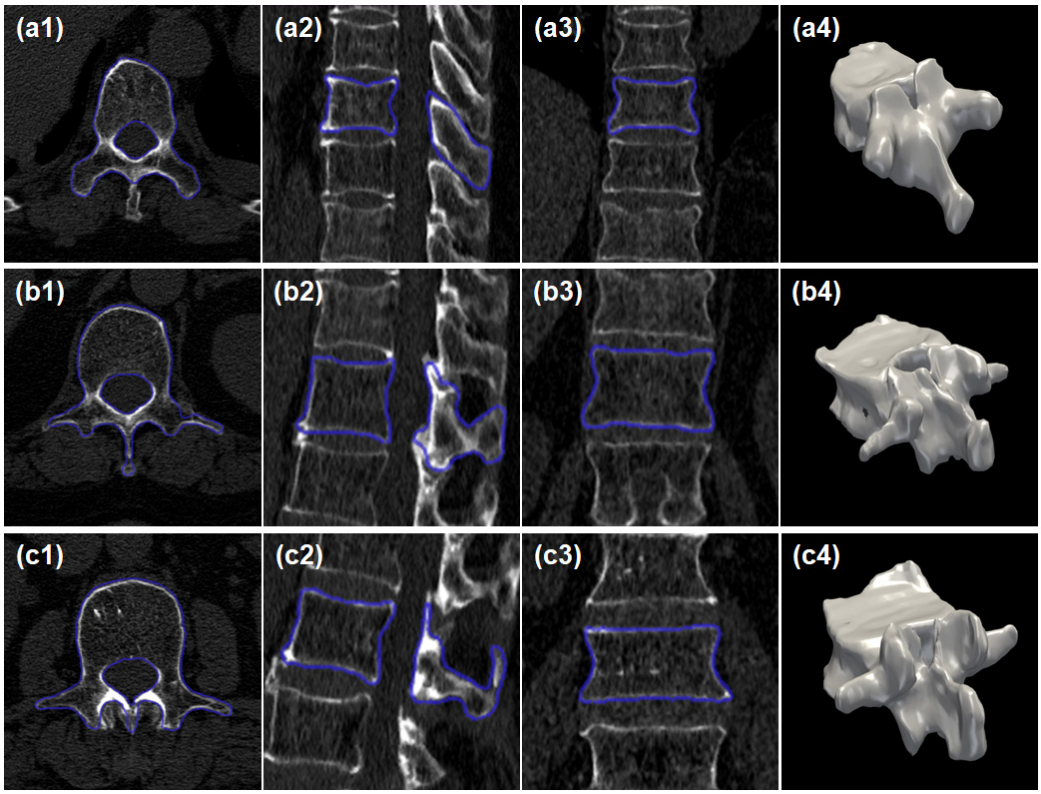

Fig. 6. Example results from a vertebra segmentation task. (a1–c1) Axial plane segmentation results for T10, L1, and L3, respectively. (a2–c2) Sagittal plane segmentation results for T10, L1, and L3, respectively. (a3–c3) Coronal plane segmentation results for T10, L1, and L3, respectively. (a4–c4) 3D model segmentation results for T10, L1, and L3, respectively.

in terms of classification accuracy. The overall classification accuracy of the vertebral labeling was 89.36%, with the accuracy rates reaching 88.05% and 91.31% for labeling thoracic and lumbar vertebrae, respectively. The recall evaluates the proportion of correctly identified instances that actually belong to a certain category, with an average recall of 88.49% for the thoracic vertebrae and 91.31% for the lumbar vertebrae. Precision measures the proportion of vertebrae predicted by the model to be of a certain category that actually belong to that category, with a precision of 87.14% for the thoracic vertebrae and 92.46% for the lumbar vertebrae. Table 4 displays the per-vertebra classification performance metrics, including accuracy (correct classification rate), recall (sensitivity), and precision (positive predictive value) for each vertebral level from T1 to L5. The thoracic vertebrae show an average accuracy of 98.7%, while the lumbar vertebrae achieve a rate of 98.8%. The evaluation results for the lumbar vertebrae were slightly higher than those for the thoracic vertebrae, indicating that the identification was better for the lumbar vertebrae than for the thoracic vertebrae. The recall and precision values for most vertebrae were found to be consistent. The difference in the evaluation results between the thoracic and lumbar vertebrae may stem from their variations in morphological characteristics. In some cases, the morphology of different thoracic vertebrae was similar, especially in the lower thoracic region (T7 to T12). These similar morphological traits in the thoracic vertebrae contribute to a higher

Table 3. Performance evaluation of vertebra labeling on regional basis.

Region	Average accuracy (%)	Average recall (%)	Average precision (%)
Thoracic	88.05	88.49	87.14
Lumbar	91.31	91.31	92.46
T1–L5 spine	89.36	89.32	88.71

Table 4. Performance evaluation of vertebra labeling on a per-vertebra basis.

Vertebra	Average accuracy (%)	Average recall (%)	Average precision (%)
T1	98.40	90.24	88.10
T2	98.70	90.48	86.36
T3	98.90	86.05	88.10
T4	98.80	88.64	84.78
T5	98.90	92.98	86.89
T6	98.60	87.93	89.47
T7	98.90	93.22	82.09
T8	98.90	86.52	88.51
T9	98.90	84.27	88.24
T10	98.70	85.71	85.71
T11	98.90	88.39	89.19
T12	98.90	87.50	88.29
L1	98.90	89.29	90.09
L2	98.90	92.98	93.81
L3	98.80	90.35	91.96
L4	98.70	92.11	92.92
L5	98.90	91.82	93.52

tendency in vertebral misidentification. On the other hand, morphological differences across lumbar vertebrae were relatively more pronounced, minimizing the risk of vertebral misidentification.

With the independent identification results for each vertebra following vertebra labeling, logical errors can be identified upon global observation, especially for vertebrae from top to bottom of the spine covering T8, T9, T9, T11, and T12. In light of this, we designed a post-processing logic to correct such errors. For instance, the aforementioned error can be rectified by calculating the correct probabilities of the two "T9" segments based on positional relationships, combined with the information of "missing T10", thereby correcting the second "T9" to "T10". After post-processing, the evaluation metrics were recalculated, and the results are presented in Table 5. The overall classification accuracy increased by 5.00%, with the rates for thoracic and lumbar vertebrae improving by 5.38% and 4.43%, respectively. The average recall and precision also showed similar increases, indicating that post-processing can effectively address identification errors and significantly enhance the overall performance of the vertebra labeling task.

Upon completion of the vertebra localization, segmentation and labeling tasks, a vertebral mask with labels can be obtained. Fig. 7 shows a representative case,

Table 5. Overall performance evaluation of vertebra labeling after post-processing.

Region	Average accuracy (%)	Average recall (%)	Average precision (%)
Thoracic	93.43	94.08	93.87
Lumbar	95.74	95.74	95.58
T1–L5 spine	94.36	94.57	94.37

where the CT scan includes T9 to L5. Fig. 7a depicts the 3D segmentation mask of all vertebrae, in which the blue-colored regions represent the thoracic vertebrae, whereas the yellow-colored regions represent the lumbar vertebrae. Fig. 7b,c show the overlapping images of the vertebral segmentation mask and the original CT scan from different viewing angles; the mask aligns well with the underlying bony structures. Although some boundary ambiguity issues were observed in regions where the thoracic vertebrae interface with the ribs, the issue remains relatively minor.

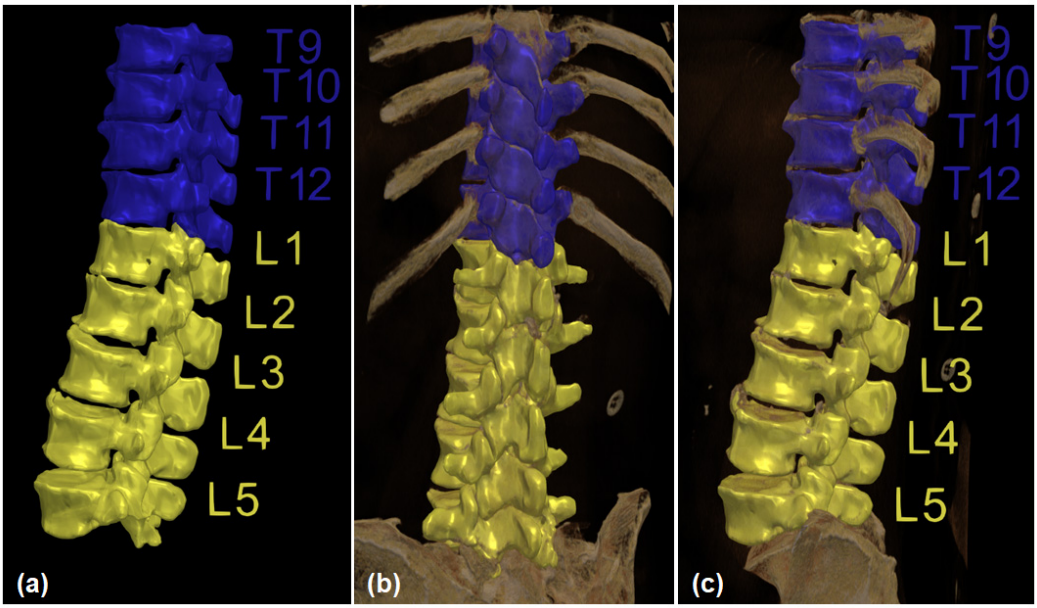

Fig. 7. Vertebra segmentation and labeling for a representative case. (a) 3D segmentation mask of all vertebrae. (b,c) Overlapping images of the vertebra segmentation mask and the original CT scan from different viewing angles.

To demonstrate the superiority of our method (vertebra localization, vertebra segmentation and vertebra labeling), we compared its performance on the VerSe2020 dataset with the methods of Saeed et al (2025) and Xu et al (2024). In terms of evaluation metrics such as Dice coefficient and mean IoU, our method significantly outperformed the reference methods (Table 6).

Discussion

This study introduces an intelligent, deep learning-driven navigation system for spinal surgery, which achieves fully automated vertebral analysis via an inno-

Table 6. Performance comparison on the VerSe2020 dataset with generic popular convolutional networks.

Method	Performance in VerSe2020 dataset	
	Dice coefficient	Mean IoU
3D-MFA (Saeed et al, 2025)	91.73	90.28
LE-NeXt (Xu et al, 2024)	80.6	89.8
Our method (vertebra localization, vertebra segmentation and vertebra labeling)	**96.80**	**91.18**

Abbreviations: VerSe2020, Vertebral Segmentation 2020; IoU, Intersection of Union; 3D-MFA, 3D Multi-Feature Attention; LE-NeXt, Lightweight Efficient NeXt. Bold values indicate the best results.

vative three-stage cascaded architecture. Demonstrating strong clinical feasibility, the system boasts a localization accuracy of 1.42 mm (with 90.72% errors $\leq$3.0 mm) and a DSC value of 0.968.

From a technical perspective, the YOLOv7+DBSCAN combination yields a 90.72% localization accuracy at a 3.0 mm threshold. The attention mechanism-enhanced 3D U-Net achieved an HD of 5.42 $\pm$ 2.05 mm, with 82.6% of vertebral samples having an HD below 5.42 mm, indicating that the model can attain accurate boundary segmentation in most cases. Furthermore, post-processing logic improves labeling accuracy from 89.36% to 94.36%, effectively correcting junctional errors like T9–T10 mislabeling through spatial logic.

The system offers substantial clinical benefits: localization errors of $<$3 mm ensures a safe screw placement; it's fully automated process reduces human-dependent errors; and the cross-institutional performance consistency (DSC 0.968 $\pm$ 0.014) ensures standardized output.

The current system is not without any shortcomings. Among the limitations are a recognition rate of 84.27% for T9 vertebrae and the relatively limited size of thoracic variant data for validation. Optimization is also required for severe deformities (e.g., scoliosis with Cobb angle $>$50°). Future work will involve multicenter studies and the adoption of intraoperative fluoroscopy fusion to validate our findings so as to enhance the system's clinical applicability. Despite these limitations, our system has outperformed existing solutions in key areas such as segmentation accuracy and computational efficiency. In addition, the construction of this deep learning-based fully automated vertebra segmentation and labeling workflow aligns with the global trend in innovating intelligent navigation systems for spinal surgery, offering innovative solutions to the traditional challenges in vertebra segmentation and labeling.

Conclusion

This study presents an innovative multi-modal framework that achieves three major technological breakthroughs in spinal imaging analyses. Firstly, the localization module combines density clustering with deep learning, significantly enhancing the performance in determining vertebral center. Secondly, the segmentation

module employs an anatomy-optimized attention mechanism, effectively improving robustness in the segmentation of complex boundaries. Lastly, the classification module utilizes a hybrid architecture and intelligent post-processing in a collaborative design, enabling recognition of vertebral type at clinically acceptable accuracy levels. Evidenced by results from a robust single-center validation (with a DSC of 0.968 ± 0.014), the system demonstrates strong potential for clinical applications. At the current stage, multicenter validation is still underway and is part of future work.

Key Points

- This study presents a fully automated vertebral analysis workflow that achieves surgical navigation-level accuracy (Mean Localization Error of 1.42 mm, Dice Similarity Coefficient of 0.968), which reduces inter-observer variability and preoperative planning time and addresses crucial needs for emergency spinal cases by reducing inter-observer variability.
- The cascaded system demonstrates robust performance in vertebral analysis across challenging thoracic vertebrae (T1: DSC 0.923, L5: DSC 0.972), with a localization accuracy of 90.72% (with YOLOv7+DBSCAN), an average HD of 5.42 ± 2.05 mm in segmentation (with attention mechanism-enhanced 3D U-Net), and improved labeling accuracy in the range of 89.36% to 94.36%.
- Despite high overall accuracy, performance for T9 (84.27% recall) is limited by data scarcity; therefore, future work will focus on multicenter validation and pathology-specific modules for severe deformities.
- The system enhances surgical safety (<3 mm error margin for screw placement), workflow efficiency (40% time reduction), and standardization (DSC 0.968 ± 0.014), positioning it as a transformative tool for spine navigation.
- The three-stage cascaded architecture integrates localization, segmentation, and labeling with attention mechanisms, offering a balance between computational efficiency and clinical precision.

Availability of Data and Materials

All data included in this study are available from the corresponding authors upon reasonable request.

Author Contributions

HJL, ML, KY, YF, JZ and YS contributed to the conception or design of the work. HJL analyzed the data, drafted and revised the manuscript. All authors contributed to the important editorial changes of important content in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Medical Research Ethics Committee of the 903rd Hospital of PLA Joint Logistics Support Force (Xihu Hospital Affiliated with Hangzhou Medical College) (approval no. 20220722/08/02/019) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all patients.

Acknowledgement

Not applicable.

Funding

This study is supported by Zhejiang Province Medical and Health Science and Technology Plan General Project (2024KY235).

Conflict of Interest

The authors declares no conflict of interest.

References

Abdou A, Kades S, Masri-Zada T, Asim S, Bany-Mohammed M, Agrawal DK. Lumbar Spinal Stenosis: Pathophysiology, Biomechanics, and Innovations in Diagnosis and Management. Journal of Spine Research and Surgery. 2025; 7: 1–17. https://doi.org/10.26502/fjsrs0082

Chaitanya K, Karani N, Baumgartner CF, Erdil E, Becker A, Donati O, et al. Semi-supervised task-driven data augmentation for medical image segmentation. Medical Image Analysis. 2021; 68: 101934. https://doi.org/10.1016/j.media.2020.101934

Cheng P, Yang Y, Yu H, He Y. Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net. Scientific Reports. 2021; 11: 22156. https://doi.org/10.1038/s41598-021-01296-1

Chmelik J, Jakubicek R, Walek P, Jan J, Ourednicek P, Lambert L, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. Medical Image Analysis. 2018; 49: 76–88. https://doi.org/10.1016/j.media.2018.07.008

Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 (pp. 424–432). Athens, Greece. Springer International Publishing. 2016.

Devito DP, Kaplan L, Dietl R, Pfeiffer M, Horne D, Silberstein B, et al. Clinical acceptance and accuracy assessment of spinal implants guided with SpineAssist surgical robot: retrospective study. Spine. 2010; 35: 2109–2115. https://doi.org/10.1097/BRS.0b013e3181d323ab

Durbas A, Yilgor C, Alanay A. Functional Anatomy of the Spine. In Doral MN, Karlsson J, Nyland J, Bilge O, Hamrin Senorski E (eds.) Sports Injuries (pp. 787–801). Springer: Cham. 2025.

Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD. 1996; 96: 226–231.

Fu G, Lu H, Tan JK, Kim H, Zhu X, Lu J. Segmentation of spinal canal region in CT images using 3D region growing technique. In 2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT) (pp. 1–4). IEEE. 2018.

Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. JOR Spine. 2019; 2: e1044. https://doi.org/10.1002/jsp2.1044

Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E. Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference (pp. 590–598). Nice, France, 1–5 October 2012. Springer Berlin Heidelberg. 2012.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778). 2016.

Huang X, Cai Y, Chen K, Ren Q, Huang B, Wan G, et al. Risk factors and treatment strategies for adjacent segment disease following spinal fusion (Review). Molecular Medicine Reports. 2024; 31: 33. https://doi.org/10.3892/mmr.2024.13398

Kantelhardt SR, Martinez R, Baerwinkel S, Burger R, Giese A, Rohde V. Perioperative course and accuracy of screw positioning in conventional, open robotic-guided and percutaneous robotic-guided, pedicle screw placement. European Spine Journal. 2011; 20: 860–868. https://doi.org/10.1007/s00586-011-1729-2

Kawsar KA, Chowdhury FH. Spinal Anatomy, Mobility, Balance, and Deformity. Principles of Neurosurgery: A Concise Text. 2024; 431.

Masuzawa N, Kitamura Y, Nakamura K, Iizuka S, Simo-Serra E. Automatic segmentation, localization, and identification of vertebrae in 3D CT images using cascaded convolutional neural networks. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference (pp. 681–690). Lima, Peru, 4–8 October 2020. Springer International Publishing. 2020.

Neubert A, Fripp J, Engstrom C, Schwarz R, Lauer L, Salvado O, et al. Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models. Physics in Medicine and Biology. 2012; 57: 8357–8376. https://doi.org/10.1088/0031-9155/57/24/8357

Rao RM, Liu J, Verkuil R, Meier J, Canny J, Abbeel P, et al. 'MSA Transformer', Proceedings of the 38th International Conference on Machine Learning. 18–24 July 2021. Microtome Publishing: Brookline, MA, USA. 2021.

Saeed MU, Bin W, Sheng J, Saleem S. 3D MFA: An automated 3D Multi-Feature Attention based approach for spine segmentation using a multi-stage network pruning. Computers in Biology and Medicine. 2025; 185: 109526. https://doi.org/10.1016/j.compbiomed.2024.109526

Tran VL, Lin HY, Liu HW. MBNet: A multi-task deep neural network for semantic segmentation and lumbar vertebra inspection on X-ray images. In Proceedings of the Asian Conference on Computer Vision. 2020.

Wang CY, Bochkovskiy A, Liao HY. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464–7475). 2023.

Wang S, Jiang Z, Yang H, Li X, Yang Z. Automatic Segmentation of Lumbar Spine MRI Images Based on Improved Attention U-Net. Computational Intelligence and Neuroscience. 2022; 2022: 4259471. https://doi.org/10.1155/2022/4259471

Wirth F, Bergamaschi EC, Forti FD, Bergamaschi JP. Development of indications for endoscopic spine surgery: an overview. International Journal of Translational Medicine. 2023; 3: 321–333. https://doi.org/10.3390/ijtm3030023

Xu G, Wang C, Li Z, Zhai J, Wang S. Efficient spine segmentation network based on multi-scale feature extraction and multi-dimensional spatial attention. International Journal of Imaging Systems and Technology. 2024; 34: e23046. https://doi.org/10.1002/ima.23046

Yan H, Zhang G, Cui W, Yu Z. Multi-modality hierarchical fusion network for lumbar spine segmentation with magnetic resonance images. Control Theory and Technology. 2024; 22: 612–622. https://doi.org/10.1007/s11768-024-00231-9

Zhang D, Chen B, Li S. Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy. Medical Image Analysis. 2021; 67: 101861. https://doi.org/10.1016/j.media.2020.101861