

## Research Paper

# A patch-based deep learning MRI segmentation model for improving efficiency and clinical examination of the spinal tumor

Weimin Chen<sup>a</sup>, Yong Han<sup>b,\*</sup>, Muhammad Awais Ashraf<sup>c</sup>, Junhan Liu<sup>b</sup>, Mu Zhang<sup>d</sup>, Feng Su<sup>d</sup>, Zhiguo Huang<sup>d</sup>, Kelvin K.L. Wong<sup>a,\*</sup>

<sup>a</sup> School of Information and Electronics, Hunan City University, Yiyang, Hunan 413000, China

<sup>b</sup> School of Design, Quanzhou University of Information Engineering, Quanzhou, Fujian 362000, China

<sup>c</sup> Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

<sup>d</sup> Department of Emergency, Xiangya Hospital, Central South University, Changsha, Hunan 410008, China

## HIGHLIGHTS

- Spine MRI image segmentation using a patch extraction (PE) based deep NN.
- PENN extracts discriminative features for precise representation of spine structures.
- PENN outperforms existing vertebrae and tumor segmentation methods.
- Achieves accuracy (93.2%) with clinical applications for diagnoses and treatment planning.
- Addresses limitations of traditional methods, providing accurate MRI image segmentation.

## ARTICLE INFO

## Keywords:

Gadolinium-enhanced T1 MRI  
Spinal tumors  
Full-automatic segmentation  
Patch model  
Deep learning

## ABSTRACT

**Background and objective:** Magnetic resonance imaging (MRI) plays a vital role in diagnosing spinal diseases, including different types of spinal tumors. However, conventional segmentation techniques are often labor-intensive and susceptible to variability. This study aims to propose a full-automatic segmentation method for spine MRI images, utilizing a convolutional-deconvolution neural network and patch-based deep learning. The objective is to improve segmentation efficiency, meeting clinical needs for accurate diagnoses and treatment planning.

**Methods:** The methodology involved the utilization of a convolutional neural network to automatically extract deep learning features from spine data. This allowed for the effective representation of anatomical structures. The network was trained to learn discriminative features necessary for accurate segmentation of the spine MRI data. Furthermore, a patch extraction (PE) based deep neural network was developed using a convolutional neural network to restore the feature maps to their original image size. To improve training efficiency, a combination of pre-training and an enhanced stochastic gradient descent method was utilized.

**Results:** The experimental results highlight the effectiveness of the proposed method for spine image segmentation using Gadolinium-enhanced T1 MRI. This approach not only delivers high accuracy but also offers real-time performance. The innovative model attained impressive metrics, achieving 90.6% precision, 91.1% recall, 93.2% accuracy, 91.3% F1-score, 83.8% Intersection over Union (IoU), and 91.1% Dice Coefficient (DC). These results indicate that the proposed method can accurately segment spine tumors CT images, addressing the limitations of traditional segmentation algorithms.

**Conclusion:** In conclusion, this study introduces a fully automated segmentation method for spine MRI images utilizing a convolutional neural network, enhanced by the application of the PE-module. By utilizing a patch extraction based neural network (PENN) deep learning techniques, the proposed method effectively addresses the deficiencies of traditional algorithms and achieves accurate and real-time spine MRI image segmentation.

\* Corresponding authors.

E-mail addresses: [U23092110161@cityu.edu.mo](mailto:U23092110161@cityu.edu.mo) (Y. Han), [kelvin.wong@hncu.edu.cn](mailto:kelvin.wong@hncu.edu.cn) (K.K.L. Wong).

<https://doi.org/10.1016/j.jbo.2024.100649>

Received 1 August 2023; Received in revised form 2 November 2024; Accepted 5 November 2024

Available online 16 November 2024

2212-1374/© 2024 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

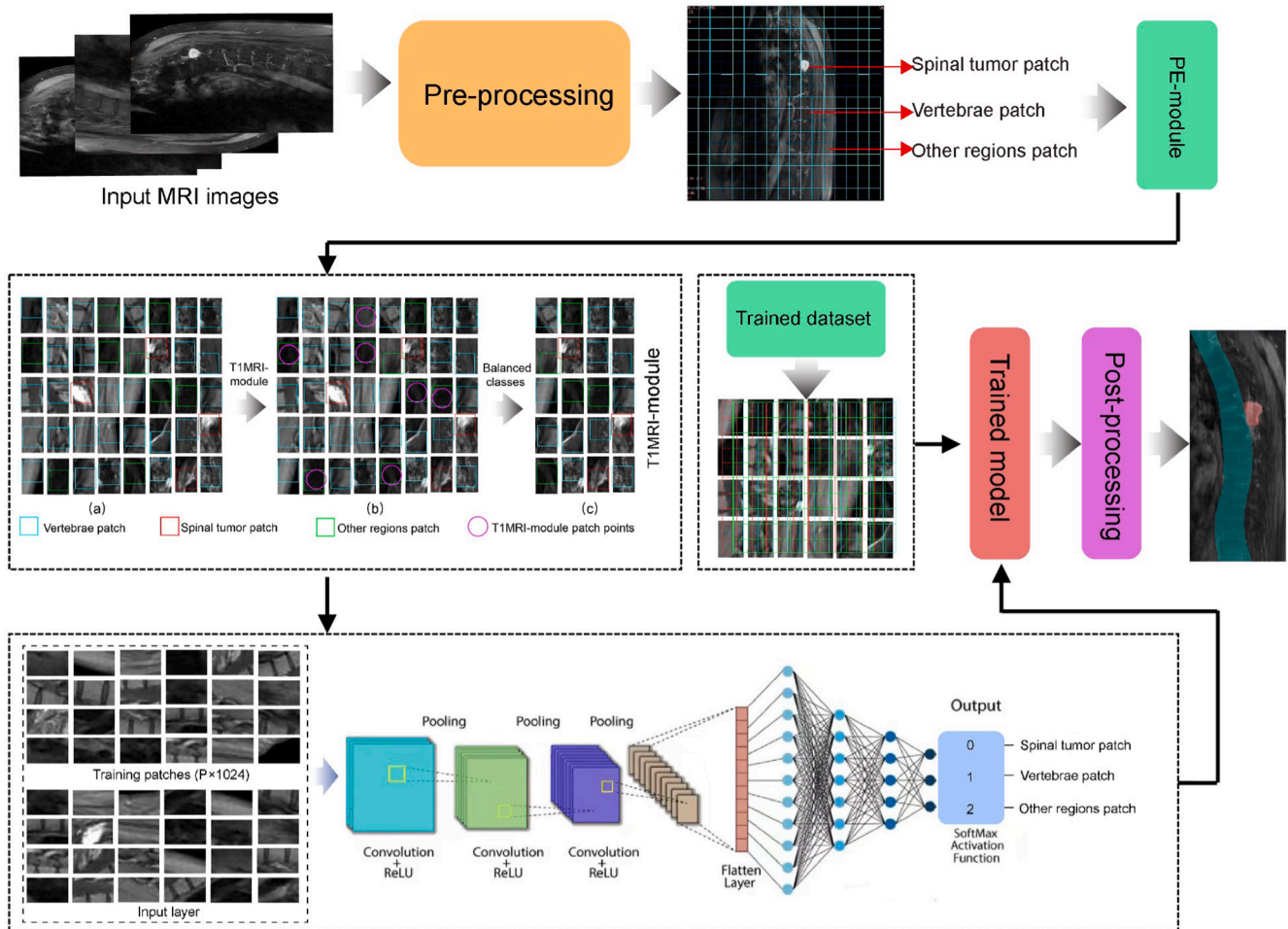
## 1. Introduction

The spine is a vital component of human anatomy, serving as the central support structure of the body. It connects the skull at the top, the ribs in the middle, and the hip bone at the bottom. As the backbone of the human body, it plays crucial roles in supporting the torso, protecting the spinal cord, and safeguarding internal organs. But in modern society, due to people's accelerated pace of life, heavy work pressure and other reasons, spine Patients with spondylosis continue to show a younger trend. According to statistics, at the age of 40. The number of people suffering from related diseases in the following groups accounted for more than 40 % [1]. Spinal tumors rank among the most prevalent tumors, with an incidence rate of 6 % to 10 % among all human bone tumors. Primary spinal tumors are quite rare, comprising no more than 5 % of cases, whereas metastatic tumors make up over 95 %. These metastatic tumors predominantly affect the thoracic spine (70 %), followed by [2], the lumbar spine (20 %) and the cervical spine (10 %). The spine is the most common site for malignant tumor metastasis, and about 30 % to 70 % of malignant tumor metastases will have spinal metastasis [3].

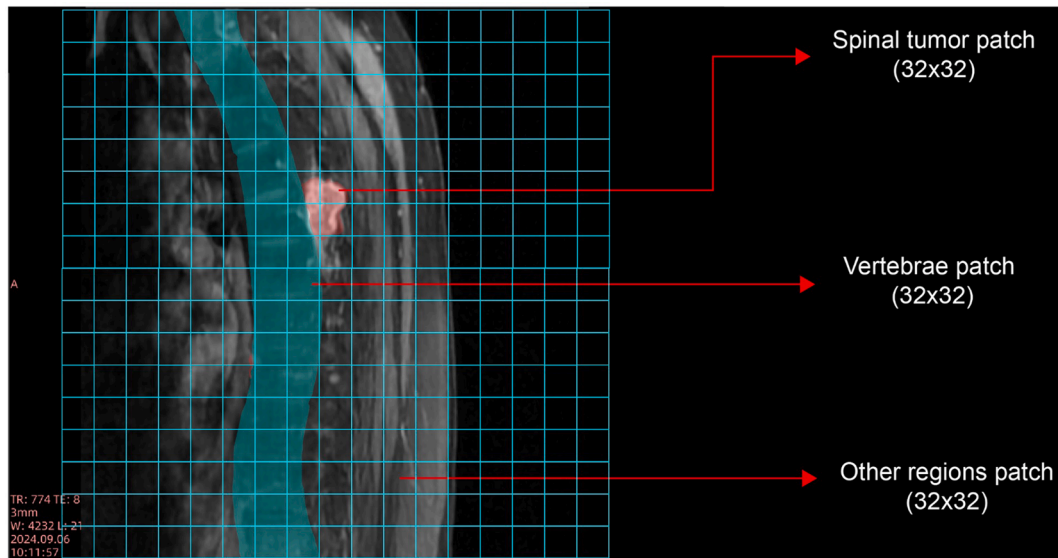
Previous experimental data show that the deep learning-based spinal image segmentation method can extract deep feature information, which significantly improves its accuracy, precision and segmentation efficiency compared with conventional approaches [4]. Leveraging the robust image feature extraction capabilities and high accuracy of deep learning technology, it has steadily emerged as the mainstream

approach for spinal tumor image segmentation. In spinal tumor diagnosis, MRI outperforms CT in differentiating soft tissues. It can clearly show the spinal cord, nerve roots, intervertebral discs and paravertebral soft tissues, and can accurately determine the relationship between the tumor and the surrounding soft tissues. CT has a weaker ability to distinguish the position, size, shape and relationship between the spinal cord and nerve roots in the spinal canal. In addition, MRI can be imaged in multiple directions, and can comprehensively observe the whole picture of the tumor, understand its range, growth direction and relationship with surrounding structures.

In the diagnostic process, it is often essential to segment the spinal region of interest in MRI scans using computer assistance. This segmentation, combined with three-dimensional reconstruction and visualization technologies, allows doctors to observe and analyze diseased areas more intuitively and clearly. Such capabilities significantly enhance the efficiency and success rate of disease treatment. However, during the imaging process of spine MRI images, the images are easily affected by equipment noise, so there is a certain amount of noise in the image, and there is no obvious boundary phenomenon between the spine boundary and other tissue boundaries, and the spine morphology changes irregularly and has a complex structure. etc. These all bring certain challenges to spine image segmentation. Aiming at the above problems, researchers have proposed many algorithms for spine image segmentation. For example, Kalidindi et al [5]. The proposed approach utilizes the super pixel method for segmenting spine images. The algorithm consists of several key steps: first, it preprocesses the images to



**Fig. 1.** An illustration of the architecture of our patch extraction based neural network (PENN) technique. This framework clearly depicts the processing of spinal MRI data from the scanning machine through a patch-based deep learning approach, culminating in final post-processing that delivers relevant segmentation results to enhance the efficiency of clinical examinations for spinal tumors.



**Fig. 2.** An illustration of the PE-module applied to a T1-weighted MRI scan of the spinal region, showcasing the corresponding segmented components and their labels.

eliminate noise and identify the candidate region of interest (ROI). Next, the superpixel method clusters the target and background within the ROI to achieve the segmentation result. Additionally, morphological closing and opening operations are applied to enhance the accuracy of the segmentation. Compared to manual segmentation of vertebral images, this method significantly increases the speed of the segmentation process. Suzani et al [6] utilizes the mean shape information of the spine obtained from the statistical shape model method to achieve semi-automatic segmentation. During the segmentation process, the initial step involves manually identifying the spine's position within the image. Subsequently, prior information, including the shapes and gradients of the models introduced, is employed as constraints to facilitate the segmentation. Pang et al [7] proposed an enhanced level set method (LSM) for segmentation. However, during practical implementation, it was found that the level set function (LSF) is particularly sensitive to image noise, resulting in difficulties in accurately segmenting the irregular boundaries of the spine.

The study utilizes the image ladder method, where the evolution of Level Set Functions (LSF) based on degree information can enhance accuracy. Aubert et al. [8] introduced statistical shape information of spine images to initialize the LSF, but this approach increased the computational complexity of the Level Set Method (LSM) despite improving segmentation accuracy. This research paper's main contribution is the proposal of a fully automated segmentation method for spine MRI images, employing a convolutional-deconvolutional neural network that leverages deep learning techniques. The paper presents an efficient and accurate model, namely patch-based deep learning (PDL). It requires no human intervention or prior information and addresses the limitations of traditional segmentation algorithms. Furthermore, we utilize a patch extraction based neural network (PENNN) to restore feature maps to their original image dimensions. This innovative combination of networks enables precise segmentation of spinal tumor MRIs, enhancing both efficiency and accuracy. To address the class imbalance issue during training, we implement the T1MRI module to sample an equal number of patches from both spinal tumor and non-spinal tumor categories. Ultimately, we showcase the effectiveness and efficiency of our approach using the tumor dataset from Xiangya Hospital of Central South University.

## 2. Methods

The methodology of this study involved using a convolutional neural network (CNN) to extract deep learning features from spinal images, providing a detailed representation of the anatomical structures. The CNN was carefully trained to identify the key features necessary for accurate segmentation of spinal MRI images. To tackle the issue of class imbalance, we sampled an equal number of patches from the T1 MRI module, including both spinal tumor patches and non-spinal tumor patches, during the training phase.

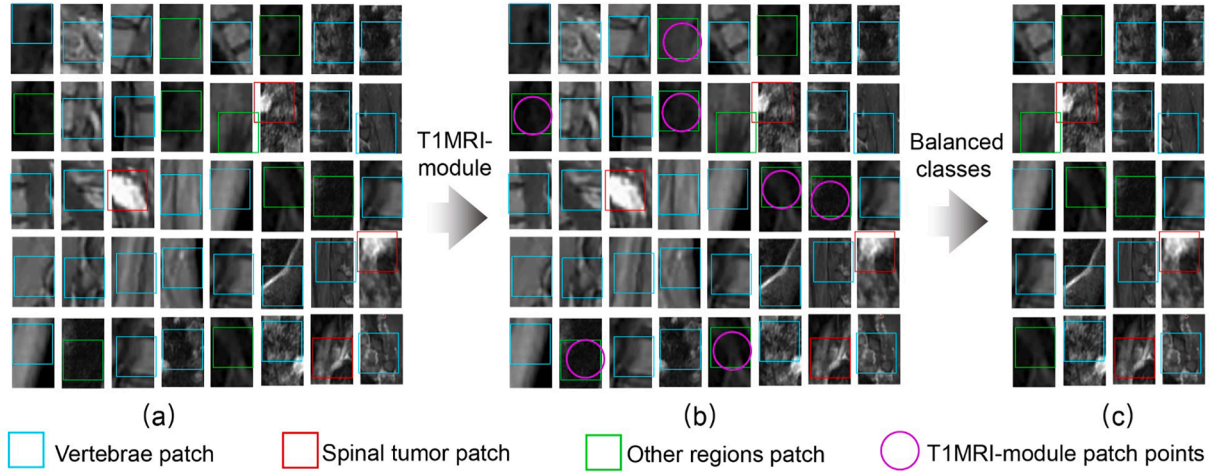
The pre-training phase involved initializing the CNN with weights acquired from a distinct but related task or dataset, enabling the network to gain valuable preliminary knowledge before fine-tuning on the specific spine MRI dataset. Subsequently, the improved SGD method was implemented to further refine the network's performance by proficiently updating the network's weights during the training process, thereby steering it towards an optimal solution. The research process is shown in Fig. 1. By harnessing the potential of patch-based deep learning, the aforementioned methodology aspired to augment the automated analysis and segmentation of spinal tumor images, thereby fostering more accurate diagnostic procedures and enhancing the quality of tumor patient care.

### 2.1. Preprocessing

To enhance the distinction between vertebrae and surrounding tissues, the primary goal of the preprocessing step is to identify bone pixels and eliminate noise from the image. We opted to employ the thresholding method to effectively remove noise artifacts from the entire MRI image. The input MRI data of the spine is volumetric data and needs to be processed slice by slice. Compared to other tissues, vertebrae can be distinguished from soft tissues by applying thresholding due to the high pixel intensity of the vertebrae in the MRI scan.

To differentiate between vertebrae and other structures in MRI images, we utilize deep learning models. The MRI images are then smoothed using a Gaussian filter with a sigma value of 2 to make sure that there are no intensity singularities and that the image gradients are well defined. The data is normalized to 0, 1, and 2. We use the PE module to extract overlapping patches of size  $n \times n$  from the input MRI data at a certain pixel spacing. A  $(32 \times 32)$  size patch image contains a total of 1024 pixels. If the total number of pixels within the patch is





**Fig. 3.** The category distribution in the plaque extraction reflects the sampling height imbalance. The category balancing of the image patches will be realized by the T1MRI module. The number of patches are deliberately reduced in this figure to illustrate the fundamental concept.

equal to or greater than 60 %, it is labeled as 1 (spine tumor); otherwise, it is labeled as 0 (vertebrae patch) and (other regions), respectively. The PE module constructs overlapping patches on 2D slices with a specific pixel step size, which is applied to the sliding window. As an image segmentation module, the PE module has been successfully applied to network training of patch-based deep learning models with improved categorization accuracy.

## 2.2. Magnetic resonance imaging

Because the spine's area in the MRI data is too tiny in relation to the backdrop, the number of image plaques following the PE module exhibits an imbalance. Due to the majority of patches being labeled as 0, the classifier may be biased in the background. There is some value in great sensitivity from a medical perspective. However, a high false-negative rate is inappropriate from a practical standpoint [9]. Therefore, in order to resolve and overcome this problem, it is essential to balance the size of the positive and negative training image patches. The T1MRI-module selects a portion of the majority class (background patches) in order to balance the training data.

Random image fragments in the majority class are removed in this module. (Fig. 2). The majority class is represented as class  $B$ , while the minority class is labeled as class  $F$ . The ratio of the sizes of these classes is defined as  $r$ . We conducted T1 MRI scans on class  $B$  to achieve a balanced ratio of  $r$ . After applying the T1MRI-module, the balanced ratios were  $r_{\text{non-spinal tumor patch}} = 0.6$  and  $r_{\text{spinal tumor patch}} = 0.4$ , compared to the imbalanced ratios of 0.94 and 0.06 prior to the T1MRI-module. This adjustment led to improvements in both the accuracy and convergence speed of the network during model training [10]. However,

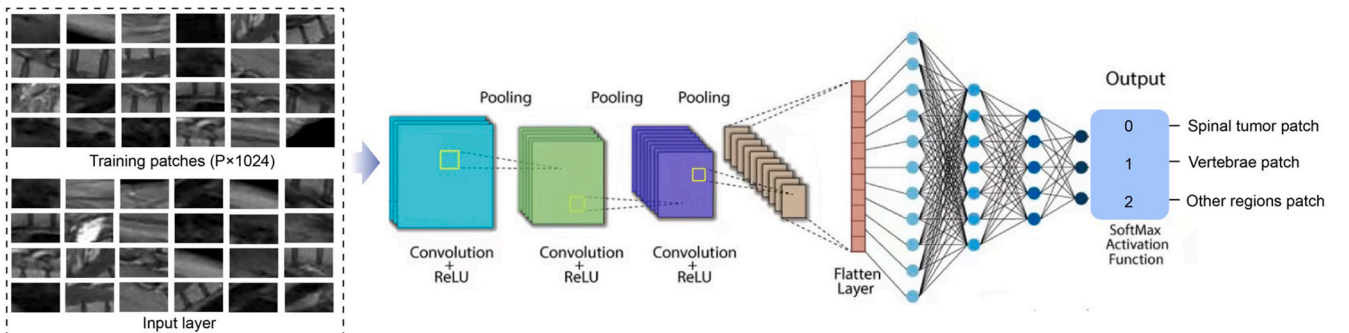
the balanced image patch categories were not part of the testing phase.

A key element of convolutional neural networks (CNNs) is the auto encoder (AE), which consists of three layers: an input layer, a hidden layer, and an output layer. Each layer in an AE is fully connected to the nodes in the other layers. By stacking multiple AEs, we can create a multilayer neural network. In our work, we enhance the three-layer CNN by stacking three AEs and pretraining the model using a greedy layer-wise approach specific to CNNs. Notably, this pertaining is conducted in an unsupervised manner, meaning that label (ground truth) information is not utilized.

We consider that  $x = (x_1, x_2, \dots, x_n)$  expresses the autoencoder input vector,  $y = (y_1, y_2, \dots, y_n)$  expresses the recon-structed representation vector of  $x$ , and the equation is  $z = (z_1, z_2, \dots, z_n)$  expresses  $k$  hidden node activation vector. Because the autoencoder uses the intermediate hidden layer to recreate input features on the output layer, it uses the weights  $w$  and bias  $b_1$  to encode the input  $x$  to  $= f(\omega_1 x + b_1)$ . Using the decoding weights  $\omega$  and bias  $b$ , the activation vector  $y$  in the hidden layer decodes the output. Then,  $y$  transfers the hidden layer latent representation to the output  $z$  using the formula  $y = f(\omega_2 y + b_2)$ . We used the following cost function to build a  $L_2$  regularized sparse autoencoder:

$$E(W, b) = \frac{1}{P} \sum_{j=1}^P \sum_{i=1}^n (y_{ij} - x_{ij})^2 + \lambda \Omega_{\text{weights}} + \beta \Omega_{\text{sparsity}} \quad (1)$$

Keep in mind that the AE network parameters are  $W$  (weights) and  $b$  (bias), that the first portion of the equation is mean squared error (MSE), and that the training data sample size is  $p$ . The  $L_2$  regularization on the encoding weights is the second part of the cost function, and  $\lambda$  represents the penalty coefficient of the  $L_2$  regularization term. The third part of the cost function is sparsity regularization, where  $\beta$  is the sparsity



**Fig. 4.** Integrating CNNs with sigmoid regression mapping for classification of tumor plaques, vertebrae, and other regions.





**Fig. 5.** Sagittal examples of the spinal tumor dataset using Gd-e T1 MRI (T1 sequence scanning parameters: slices:15; FOV red: 260 mm; FOV phase: 235.6 mm; Slice thickness: 3 mm; TR: 774; TE: 8.7; Averages: 2; concatenations: 2; Flip angle: 160; acq matrix: 244×384; Base resolution: 384; TA: 1.25 min). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regularization coefficient and  $\Omega_{sparsity}$  is the Kullback–Leibler (KL) divergence, which is expressed as follows:

$$\Omega_{sparsity} = \sum_{i=1}^k \rho \log \frac{p_i}{p} + (1 - \rho) \log \frac{1 - p_i}{1 - p} \quad (2)$$

To balance the training dataset, a subset of the majority class (background patches) was chosen for the T1MRI module. This module removes random image patches from the majority class, as illustrated in Fig. 3. Given that the CNN employs an unsupervised learning approach, unlabeled data was utilized to train each layer of the network. Input reconstruction was generated using a feature vector, which the classifier employs to categorize the input data of the stacked sparse auto encoder. To distinguish between spinal tumor and non-spinal tumor patches, we employed a sigmoid regression layer (Fig. 4). Other classifiers that can be used in place of the sigmoid layer are MLP and SVM. The over-fitting issue causes the MLP, a feed-forward neural network with multiple layers and many nodes in each layer, to become trapped in local minima.

By reconstructing the score vector, SVM classifiers, on the other hand, require a great deal of generalization to generate a probability image. They do this by determining whether a pixel belongs to the target or background class based on its posterior probability value. But the sigmoid layer allows the joint to fine-tune the deep framework as a whole. One method for binary classification in supervised learning is sigmoid regression. By lowering the cost function, the output probabilities determine the likelihood coefficient vector for each class label.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

The output sigmoid function is denoted as  $\sigma$ , while  $x$  represents the input. During the supervised learning phase, the pre trained CNN and the sigmoid layer are combined into a unified classification model. To fine-tune the entire model, each iteration of the scaled gradient descent technique simultaneously adjusts the weights of all CNN layers along with the parameters of the sigmoid layer. [13].

### 3. Model training and evaluations

#### 3.1. Dataset

We utilized the spinal tumor MRI dataset sourced from Xiangya Hospital of Central South University, covering the years 2010 to 2024. Patients underwent examinations using a Siemens SKYRA 3.0 T magnetic resonance imaging system, which included both plain spinal MRI scans and gadolinium-enhanced imaging. Examples of images from the

dataset are presented in Fig. 5. The inclusion criteria are as follows: (1) The subjects' examination data are complete; (2) The subjects' general information and examination data are complete; (3) There is at least one spinal cord tumor lesion; (4) No treatment has been received. Exclusion criteria: (1) The scanning sequence is incomplete; (2) The image has artifacts. The cancer center has MRI images of the lumbar and cervical spine of 68 adults aged between 16 and 42 years old. Specifically, the MRI (Gd-e T1 sag) scanning parameters are: Ex: 32123; Se: 13; Im: 1/18; TR: 774; TE: 8.7; 3 mm; W: 2762; L: 1381. The lumbar spine MRI dataset contains 10 images of 50 lumbar vertebrae and the associated manual reference ground truth. The slice thicknesses is 3 mm. Each of the spinal tumor images was manually segmented to create a binary mask based on the vertebrae, spinal tumor, and the other tissue regions [14,15].

#### 3.2. Performance evaluation

The effectiveness of vertebral tumor segmentation using PENN is compared with other current methods using evaluation metrics [16]. These metrics are commonly used and well-known in medical image analysis. The quantitative assessment metrics for segmentation performance evaluation in this article are precision, recall, accuracy, F1-score, intersection over union (IoU), and Dice coefficient (DC). By contrasting the ground truth images with the anticipated segmented images, we assessed true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The calculation methods of these indicators are shown in Equations (4) to (9).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

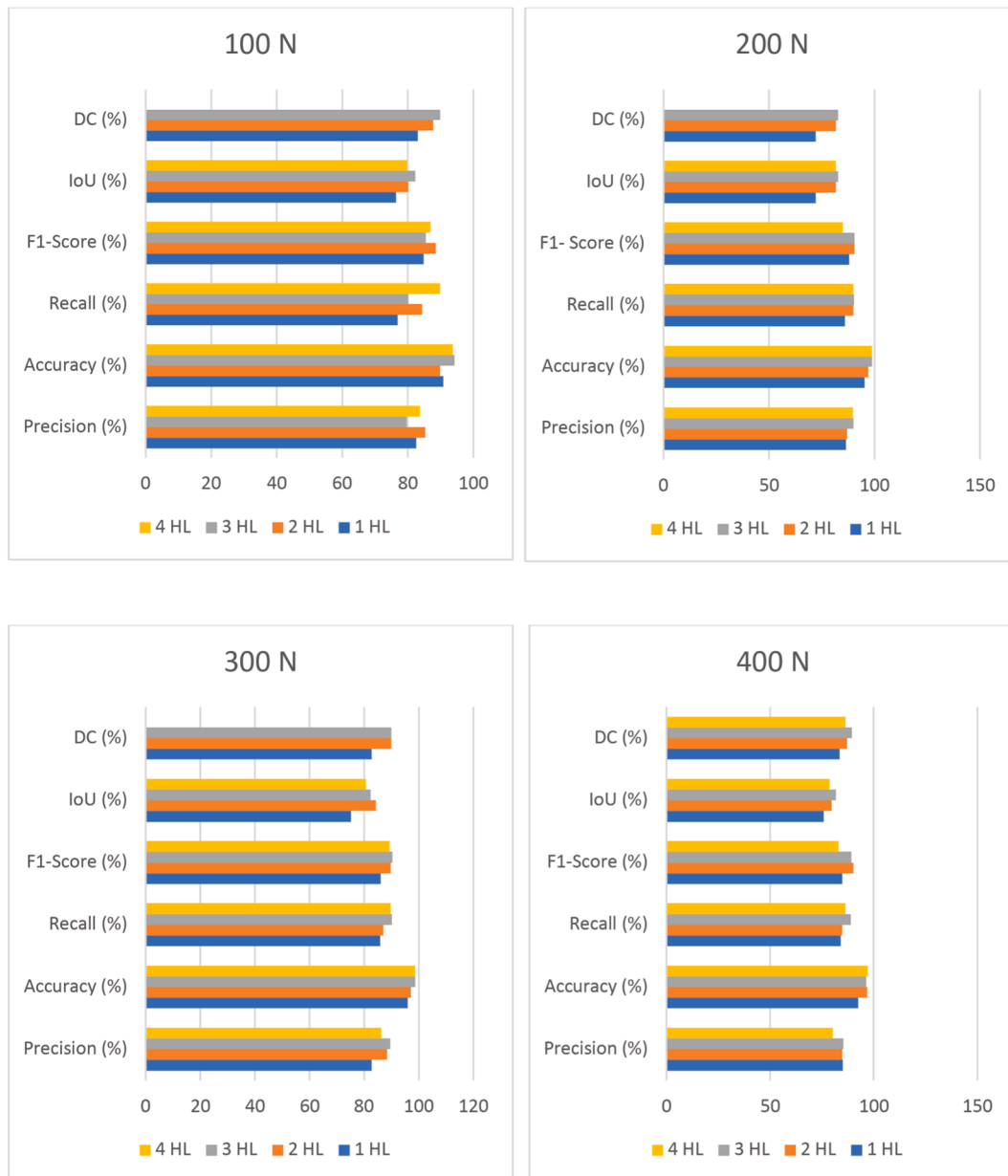
$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

$$\text{F1-score} = \frac{2 \times TP}{(2 \times TP) + FP + FN} \quad (7)$$

$$\text{Intersection over Union} = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \quad (8)$$

$$\text{DC} = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN} \quad (9)$$



**Fig. 6.** Performance of the suggested method with various architectures in terms of precision, accuracy, recall, F1-score, IoU, and DC. Each layer has hidden nodes (N) and a number of hidden layers (HL).

### 3.3. Model training

A total of 167,330 image patches were acquired for the model's training, of which 412,030 were spinal tumor patches and 112,080 were non-spinal tumor (vertebra) patches. We chose 161,024 image patches for validation and 80 % for training at random. The training patches are organized into 10,064 mini-batches, while the validation patches are divided into 2,516 mini-batches, each with a size of 64 to facilitate effective training. The proposed method consists of five network layers: an input layer with 1,024 neurons, three hidden layers each containing 200 neurons, and a final sigmoid layer with two neurons corresponding to the two output classes. This approach was implemented using MATLAB 2018 on a system equipped with a 1.80 GHz i7 CPU, 32 GB of RAM, and an NVIDIA GeForce MX250 GPU. The segmentation time of a 512×512 image took the algorithm about 23.42 h and 9 s to train.

A visualization model underpins the five-layered CNN, illustrating the feature representations from the first, second, and third hidden

layers. The first hidden layer, comprising 200 nodes, captures the learned feature representations of the vertebrae and other structures. The second and third hidden layers, each also containing 200 nodes, represent higher-level feature learning derived from image patches. The squares in the visualization indicate the weights connecting the original image's pixels to the hidden nodes. White pixels in the weight matrix indicate positive values, while gray pixels indicate negative values.

Various architectures may be selected based on the application's requirements. Based on DC requirements, we selected a CNN architecture with three layers, each of which has 200 hidden nodes. The training and randomly chosen test case outcomes are displayed individually in the confusion matrices in Fig. 6. displays the performance metrics that were computed for each conceivable combination of hidden layers and nodes.

**Table 1**  
Comparison of Performance Metrics (Precision, Recall, Accuracy, F1-Score, IoU, DC).

Methods	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)	IoU (%)	DC (%)
MultiResuNet	86.8	85.7	90.2	88.2	80.6	88.4
Our method	<b>90.6</b>	<b>91.1</b>	<b>93.2</b>	<b>91.3</b>	<b>83.8</b>	<b>91.1</b>

**Note** The bold indicates the optimal indicator.

**Table 2**  
The segmentation performance comparison (F1-score, IoU, DC).

Methods	F1-score(%)	IoU (%)	DC (%)
UNet	79.8	70.4	82.5
D-TVNet	78.3	74.7	85.2
MultiResuNet	88.2	80.6	88.4
SAU-Net	87.6	81.8	88.3
OP-convNet	90.2	82.3	89.9
Our method	<b>91.3</b>	<b>83.8</b>	<b>91.1</b>

**Note** The bold indicates the optimal indicator.

4. Results and discussion

Our method for segmenting vertebral tumors is based on PENN. The model we constructed demonstrated 90.6 % precision, 91.1 % recall, 93.2 % accuracy, 91.3 % F1-score, 83.8 % IoU, and 91.1 % DC in trials.

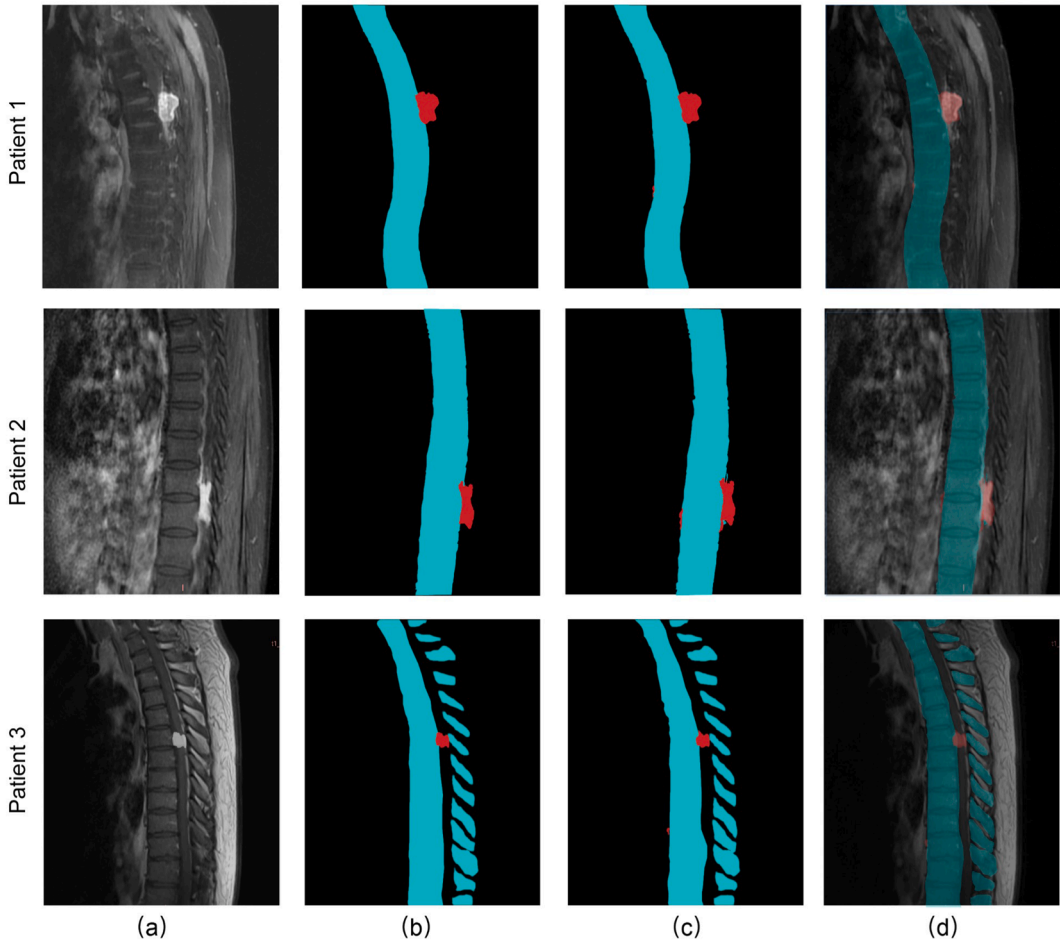
Our approach is contrasted with the well-known MultiResuNet model to demonstrate the efficacy of  $L_2$  regularized CNN.

The second and third parts of the cost function in Equation (1) become a three-layered stacked AE (MultiResuNet) model if the penalty coefficient  $\lambda$  and the sparsity regularization coefficient  $\beta$ , of the  $L_2$  regularization term are set to zero. The precision, recall, accuracy, F1-score, IoU, and DC means of our method and the MultiResuNet comparison model are shown in Table 1. The results show the significance of the  $L_2$  regularized CNN of our method that is in superior performance compared to MultiResuNet in all metrics.

Additionally, a number of deep learning models are compared with the suggested PENN method. Table 2 compares our method with popular segmentation techniques such as U-Net [22], D-TVNet [23,24], Multi-ResuNet [25], SAU-Net [26], and OP-convNet [27]. The results indicate that our proposed model outperformed all other models regarding F1-score, IoU, and DC in (Table 2). It's important to note that we used the same experimental setup for the MultiResuNet results, except for the sparsity regularization parameter  $\lambda$  in the hidden layers. Our model achieved higher F1-score%, IoU%, and DC% compared to the classical UNet, D-TVNet, MultiResuNet, SAU-Net, and OP-convNet..

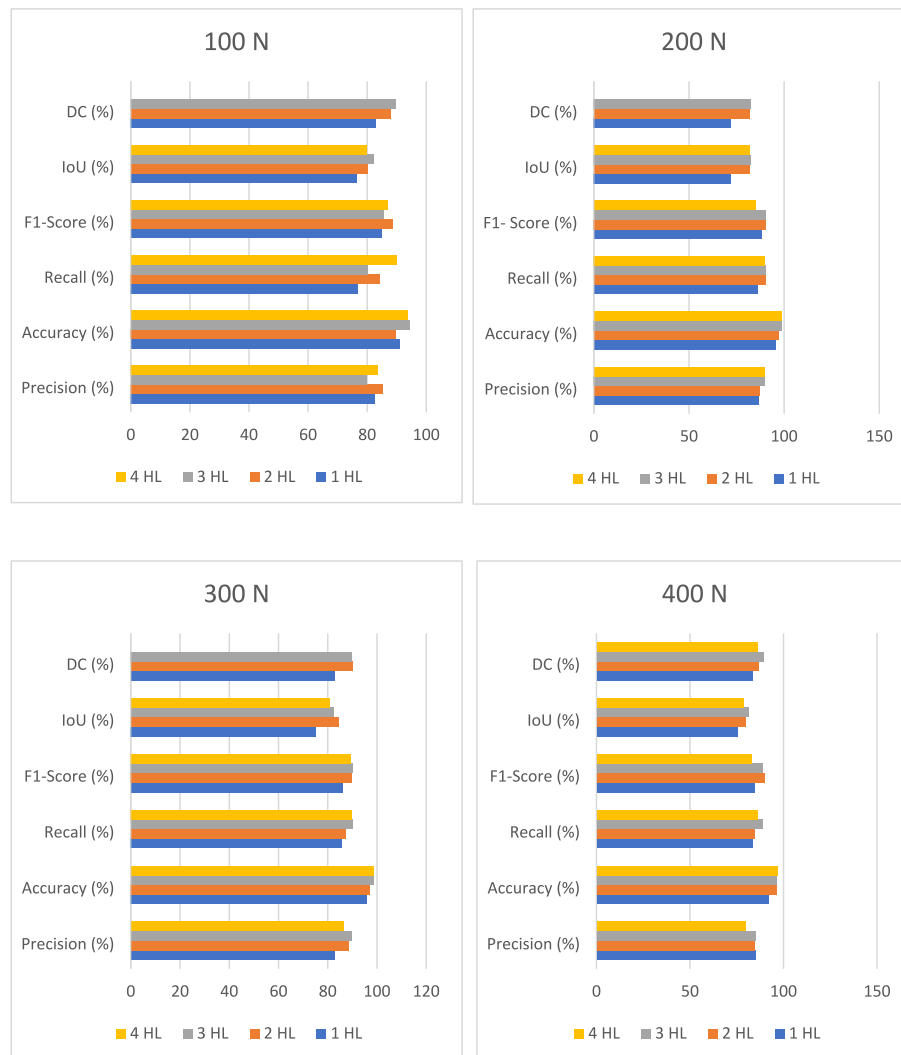
The original photos, along with their labeled counterparts, predicted segmented images, and the predicted segments overlaid on the original images, are presented in rows as qualitative examples of segmentation results for 2D axial images (Fig. 7). The results demonstrate that our proposed method achieved effective segmentation with high performance.

The performance comparison outlined in Figs. 8 and 9 highlights the



**Fig. 7.** Examples demonstrating the segmentation results: (a) axial plane images; (b) ground truth; (c) our predicted segmented images; and (d) prediction images overlaid on the original images. We conducted experiments on two samples (Patient 1 and Patient 2) using MRI scans of tumors that are slightly off-center from the spine, while Patient 3's scan aligns directly with the center of the spinal cord, revealing the hidden vertebrae in the back.





**Fig. 8.** Performance Comparison of Various Methods for Image Segmentation, Highlighting Precision, Recall, Accuracy, F1-Score, IoU, and Dice Coefficient (DC).

efficacy of our proposed method against the established MultiResU-Net model in thyroid nodule segmentation. Our method demonstrates superior performance across all evaluated metrics, with precision reaching 90.6 %, a significant improvement over the 86.8 % achieved by MultiResU-Net. This increase in precision is complemented by enhanced recall (91.1 % vs. 85.7 %), indicating not only a reduction in false positives but also an ability to accurately identify more true positive cases. The higher accuracy of our method (93.2 %) further reflects its robustness in correctly classifying the samples, resulting in an F1-score of 91.3 %, which underscores a balanced trade-off between precision and recall.

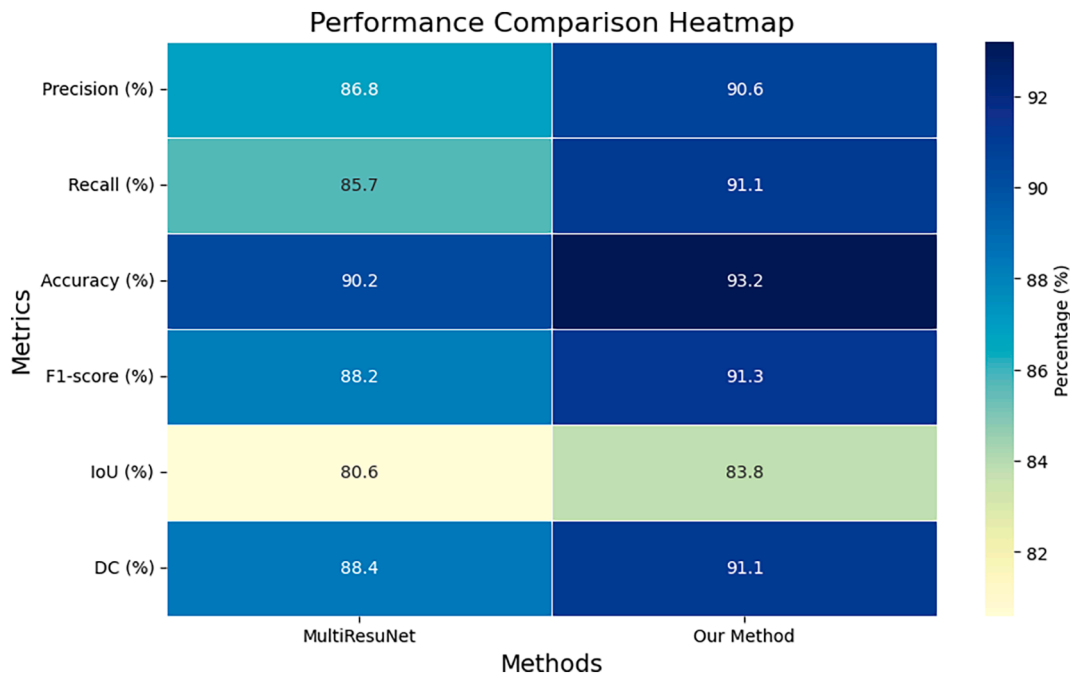
Moreover, our method outperforms MultiResU-Net in both the Intersection over Union (IoU) and Dice Coefficient (DC), with values of 83.8 % and 91.1 %, respectively, compared to 80.6 % and 88.4 % for the baseline model. The improvements in these metrics illustrate our method's effectiveness in capturing the spatial overlap between the predicted and actual segmentation masks, which is crucial for precise nodule delineation. The results underscore the potential of our approach to enhance diagnostic accuracy in clinical settings, ultimately contributing to better patient outcomes by providing more reliable segmentation of thyroid nodules in ultrasound images.

The use of a CNN in patch-based classification for automatic vertebral segmentation utilizing three different MRI spine tumor datasets is unprecedented, as far as we are aware. Our method demonstrated CNN's powerful capacity to automatically partition vertebrae by leveraging

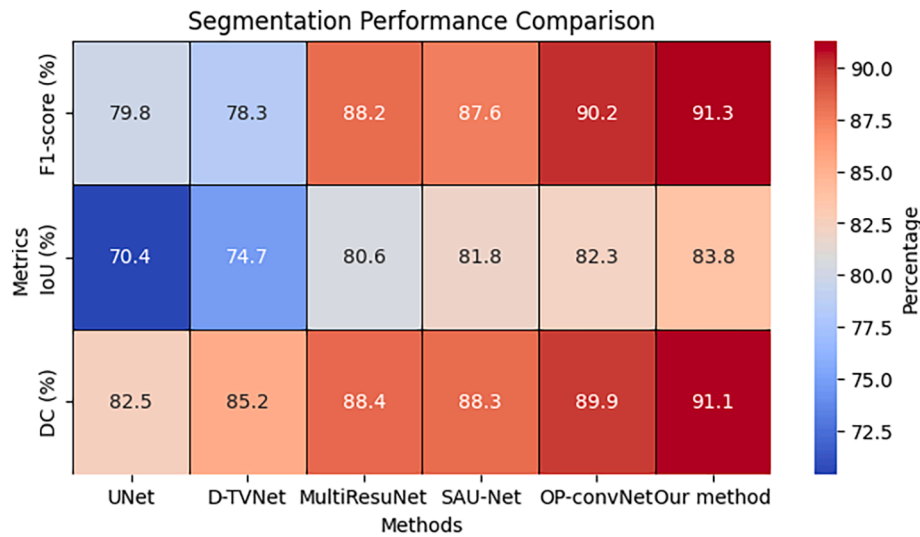
deep learning's superior data mining advantage on large data. With no human involvement, training, or analysis, our method might operate as fully automated CAD software; no handcrafted characteristics would need to be selected. In the hectic clinical settings of today, this is a crucial aspect of CAD [28,29].

The segmentation performance comparison presented in Fig. 10 highlights the effectiveness of various methods in accurately segmenting images, measured through the F1-score, Intersection over Union (IoU), and Dice Coefficient (DC). Among the methods evaluated, our proposed approach demonstrates superior performance, achieving an F1-score of 91.3 %, an IoU of 83.8 %, and a DC of 91.1 %. These results not only surpass those of established techniques such as UNet, D-TVNet, MultiResU-Net, SAU-Net, and OP-convNet but also indicate a significant enhancement in segmentation accuracy. Notably, OP-convNet, which previously held the highest metrics, shows a lower F1-score and DC compared to our method, suggesting that our approach effectively balances precision and recall better than existing models. The improvements observed with our method may be attributed to its advanced architecture and the incorporation of self-attention mechanisms, which allow for a more nuanced understanding of spatial features within the ultrasound images, ultimately leading to enhanced diagnostic precision in clinical applications.

In unsupervised learning, we utilized convolutional neural networks to extract high-level features from overlapping patches. Our method effectively classified vertebrae, tumor patches, and other regions based



**Fig. 9.** Visual Comparison of MultiResU-Net and Our Method Performance.



**Fig. 10.** Segmentation Performance Comparison Among Different Methods (F1-score, IoU, DC).

on these advanced criteria. To enhance classification accuracy, these high-level features were integrated into a sigmoid regression layer. The proposed approach demonstrates accuracy, reliability, and robustness, yielding strong overall performance in clinical applications. Given that CNNs are a type of neural network, the model's capability to classify image patches was significantly influenced by the convergence of the training process. Insufficient training epochs could result in a prematurely trained CNN, leading to suboptimal performance. Therefore, a convergence test is necessary to determine the optimal number of epochs in deep learning. In our experiments, we employed 500 epochs for pretraining, which helped prevent time waste and ensured convergence during training.

Neural network architecture is yet another important factor. As previously said, there are no universal standards for creating the architecture of a neural network. In fact, the complexity of the data being used determines the best neural network architecture [31]. In our case,

optimization studies could give an early indication of CNN architecture design. Additionally, we discovered that in order to construct deep feature representations that had a beneficial effect on the fine-tuning stage, sparse regularization was required during training. Sparsity forces the filters to extract finer details from image patches during training. When compared to other well-known models, the performance comparison demonstrated the efficacy and superior capabilities of our technique [20,32].

The test patches are successfully classified into tumor patches, vertebral patches, and other locations using our suggested method. The vertebral tumor is subsequently separated from the image patch reconstruction. Every level of the spine has a unique set of vertebral patterns. Nevertheless, certain images of spinal tumors are still not well segmented. In the spinal column, there are significant morphological differences between widely spaced vertebrae, such as the lower lumbar and upper thoracic vertebrae. This variability complicates the accurate

segmentation of vertebral tumor images [33]. Our proposed model faces challenges in segmenting tumors in the upper thoracic vertebrae (T1-T3) due to the presence of rib structures, and the L5 vertebra also demonstrated lower Dice Coefficient (DC) values compared to other vertebrae. In future implementation, we can utilize machine learning coupled with advanced modelling and optimization techniques [11,12,17–21] to develop an image retrieval system [30] for bone cancer [34] followed by segmentation of the critical regions for a more detailed oncological analysis [35].

## 5. Conclusion

This study introduces a patch-based deep learning (PDL) system for the automated segmentation of spinal tumors. We extracted overlapping patches from 2D image slices for model training. The proposed patch extraction based neural network (PENN) approach leverages these image patches to capture a high-level feature representation of pixel intensity in an unsupervised manner. A sigmoid layer utilizes these high-level features to accurately classify patches into vertebrae and non-vertebrae categories. After optimizing key parameters such as the number of hidden layers, the size of hidden nodes, and the number of epochs, our method achieved optimal performance. It is also shown that sparsity requirements on hidden layers are effective. Additionally, we show that our method is effective and feasible on the tumor dataset, which qualifies it for use in clinical settings. It was found that in order to construct feature representations that favorably impact the last supervised tuning stage, training with sparsity regularization is required. False positives are also reduced by the method of identifying vertebral areas using picture patches as opposed to individual pixels. The PENN approach shows great promise in addressing problems brought on by differences in vertebral morphology. Our PDL approach outperformed several state-of-the-art vertebral tumor segmentation techniques, including U-Net, D-TVNet, MultiResU-Net, SAU-Net, and OP-convNet, in terms of segmentation accuracy. We also undertook further research to identify the limitations of our method, especially concerning different types of spinal tumors. To strengthen our strategy in these situations, future research will create a deep neural network architecture that is more discriminative.

## Ethics Approval

Ethics was approved by the institutional review committee of our hospital.

## Data availability statement

Part of the data is contained within the paper and the rest are available upon request.

## CRediT authorship contribution statement

**Weimin Chen:** Conceptualization, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yong Han:** Funding acquisition, Investigation, Project administration, Resources, Software, Supervision, Validation. **Muhammad Awais Ashraf:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology. **Junhan Liu:** Investigation, Methodology, Resources, Software. **Mu Zhang:** . **Feng Su:** Validation, Visualization, Writing – original draft, Writing – review & editing. **Zhiguo Huang:** Conceptualization, Data curation, Formal analysis, Investigation. **Kelvin K.L. Wong:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication. In addition, special thanks to Xiangya Hospital of Central South University for providing medical data on spinal tumors for this study.

## References

- [1] A. Rezaei, et al., Mechanical testing setups affect spine segment fracture outcomes, *J. Mech. Behav. Biomed. Mater.* 100 (2019) 103399.
- [2] M.L. Goodwin, et al., Spinal tumors: diagnosis and treatment, *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 30 (17) (2022) e1106–e1121.
- [3] L.L. Mechtler, K. Nandigam, Spinal cord tumors: new views and future directions, *Neurol. Clin.* 31 (1) (2013) 241–268.
- [4] F. Jiang, et al., Medical image semantic segmentation based on deep learning, *Neural Comput. & Applic.* 29 (2018) 1257–1265.
- [5] K.K.V. Kalidindi, et al., Introduction of a novel “segmentation line” to analyze the variations in segmental lordosis, location of the lumbar apex, and their correlation with spinopelvic parameters in asymptomatic adults, *Asian Spine Journal* 16 (4) (2022) 502.
- [6] A. Suzani, A. Rasoulia, A. Seitel, S. Fels, R.N. Rohling, P. Abolmaesumi, “Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images,” in *Proceedings of the Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, Article ID 941514, Orlando, Florida, February 2015.
- [7] S. Pang, C. Pang, L. Zhao, et al., SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation, *IEEE Trans. Med. Imaging* 40 (1) (2021) 262–273.
- [8] B. Aubert, et al., Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting, *IEEE Trans. Med. Imaging* 38 (12) (2019) 2796–2806.
- [9] K.K.L. Wong, *Cybernetical intelligence: engineering cybernetics with machine intelligence*, John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN: 9781394217489, 2023.
- [10] Z. Tang, S. Wang, Y. Li, Dynamic NOX emission concentration prediction based on the combined feature selection algorithm and deep neural network[J], *Energy* 2024 (292) (2024) 130608.
- [11] Y. Shao, Y. Yin, S. Du, L. Xi, A surface connectivity based approach for leakage channel prediction in static sealing interface, *ASME Trans. J. Tribol.* 141 (2019), 062201-1-11.
- [12] K.K.L. Wong, A geometrical perspective for the bargaining problem, *PLoS One* 5 (4) (2010) e10331.
- [13] J. Yao, J.E. Burns, D. Forsberg, et al., A multi-center milestone study of clinical vertebral CT segmentation, *Comput. Med. Imaging Graph.* 49 (2016) 16–28.
- [14] A. Sekuboyina, A. Valentinitsch, J.S. Kirschke, B.H. Menze, A Localisation-Segmentation Approach for Multi-Label Annotation of Lumbar Vertebrae Using Deep Nets, 2017.
- [15] A. Sekuboyina, J. Kukacka, J.S. Kirschke, B.H. Menze, A. Valentinitsch, Attention-driven deep learning for pathological spine segmentation, in *Proceedings of the International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging*, Canada, September 2017, pp: 108-119.
- [16] R. Janssens, G. Zeng, G. Zheng, Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks, in *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging*, pp. 893–897, ISBI 2018, Washington, DC, USA, April 2018.
- [17] Du. Shichang, Xu. Rui, L. Li, Modeling and analysis of multiproduct multistage manufacturing system for quality improvement, *IEEE Transaction on Systems, Man, and Cybernetics: Systems.* 48 (5) (2018) 801–820.
- [18] K. Wang, G. Li, Du. Shichang, L. Xi, T. Xia, State space modelling of variation propagation in multistage machining processes for variable stiffness structure workpieces, *Int. J. Prod. Res.* 59 (13) (2021) 4033–4052.
- [19] G. Li, S. Du, B. Wang, J. Lv, Y. Deng, High definition metrology-based quality improvement of surface texture in face milling of workpieces with discontinuous surfaces, *ASME Transaction on Manufacturing Science and Engineering.* 2022, 144: 031001-1-18.
- [20] G. Li, S. Du, D. Huang, C. Zhao, Y. Deng, Dynamics modeling-based optimization of process parameters in face milling of workpieces with discontinuous surfaces, *ASME Transaction on Manufacturing Science and Engineering.* 2019. 141: 101009-1-15.



- [21] C. Zhao, J. Lv, Du. Shichang, Geometrical deviation modeling and monitoring of 3D surface based on multi-output gaussian process, *Measurement* 199 (2022) 111569.
- [22] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (October 2015) 234–241.
- [23] J. Lyu, X. Bi, S. Banerjee, Z. Huang, F.H.F. Leung, T.T.-Y. Lee, D.-D. Yang, Y.-P. Zheng, S.H. Ling, Dual-task ultrasound spine transverse vertebrae segmentation network with contour regularization, *Comput. Med. Imag. Graph.*, vol. 89, Apr. 2021, Art. no. 101896.
- [24] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 609–616, Montreal Quebec Canada, July 2009.
- [25] N. Ibtehaz, M.S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation, *Neural Netw.* 121 (Jan. 2020) 74–87.
- [26] Y. Zhang, L. Yuan, Y. Wang, J. Zhang, SAU-Net: efficient 3D spine MRI segmentation using inter-slice attention, *Medical Imaging With Deep Learning* (2020) 903–913.
- [27] S.F. Qadri, L. Shen, M. Ahmad, S. Qadri, S.S. Zareen, S. Khan, OP-convNet a patch classification-based framework for CT vertebrae segmentation, *IEEE Access* 9 (2021) 158227–158240.
- [28] P. Charilaou, R. Battat, Machine learning models and over-fitting considerations, *World J. Gastroenterol.* 28 (5) (2022) 605.
- [29] P. Bansal, K. Gehlot, A. Singhal, et al., Automatic detection of osteosarcoma based on integrated features and feature selection using binary arithmetic optimization algorithm, *Multimed. Tools Appl.* 81 (2022) 8807–8834.
- [30] T. Zeng, H. Lv, M.A. Ashraf, M. Ling, Y. Chen, Y. Liu, X. Chen, Y. Li, J. Huang, Management of sports injury treatment and radiological data analysis based on enhanced MRI image retrieval using autoencoder-based deep learning, *Journal of Radiation Research and Applied Sciences*, Volume 17, Issue 3, 2024, 101022, ISSN 1687-8507.
- [31] H. Ge, Z. Zhu, J. Ouyang, M.A. Ashraf, Z. Qiu, U.M. Ibrahim, Integration of manifold learning and density estimation for fine-tuned face recognition, *Symmetry* 16 (2024) 765, <https://doi.org/10.3390/sym16060765>.
- [32] X. Xue, Z. Liu, T. Xue, W. Chen, X. Chen, Machine learning for the prediction of acute kidney injury in patients after cardiac surgery, *Front. Surg.* 2022 (9) (2022) 946610.
- [33] S. Zhao, B. Chen, H. Chang, B. Chen, S. Li, Reasoning discriminative dictionary-embedded network for fully automatic vertebrae tumor diagnosis, *Med. Image Anal.* 79 (2022) 102456.
- [34] Taisheng Zeng, Yuguang Ye, Yusi Chen, Daxin Zhu, Yifeng Huang, Ying Huang, Yijie Chen, Jianshe Shi, Bijiao Ding, Jianlong Huang, Mengde Ling. Deep hashing and attention mechanism-based image retrieval of osteosarcoma scans for diagnosis of bone cancer, *J. Bone Oncol.* 49 (2024) 100645.
- [35] Ketong Zhao, Ping Dai, Ping Xiao, Yuhang Pan, Litao Liao, Junru Liu, Xuemei Yang, Zhenxing Li, Yanjun Ma, Jianxi Liu, Zhengbo Zhang, Shupeng Li, Hailong Zhang, Sheng Chen, Feiyue Cai, Zhen Tan, Automated segmentation and source prediction of bone tumors using ConvNeXtv2 Fusion based Mask R-CNN to identify lung cancer metastasis, *J. Bone Oncol.* 48 (2024) 100637.