Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 33, Num. 2 (2026) 226-233

RESEARCH ARTICLE

Automated Segmentation of the Spinal Column in MRI Volumes Using Deep Learning

Segmentação Automática da Coluna Vertebral em Volumes de Ressonância Magnética Utilizando Deep Learning

Renan Silva Carvalho[1], Leandro Henrique Furtado Pinto Silva[1], João Fernando Mari[1]

Abstract: Accurate segmentation of spinal structures in magnetic resonance imaging (MRI) is an important step for supporting the diagnosis of several diseases and related conditions. In this work, we present a systematic evaluation of three deep learning architectures, U-Net, Feature Pyramid Network (FPN), and SegFormer, combined with ResNet-50 and EfficientNet-B2 encoders for the semantic segmentation of the cervical spine (C1–C7) in MRI volumes from the VerSe 2020 dataset. The proposed pipeline includes dataset preparation, model training using Dice loss, Adam optimizer, and early stop strategy, and evaluation with standard metrics such as Intersection over Union (IoU), F1-Score, and accuracy. Results show that FPN with ResNet-50 achieved the best overall performance, reaching an IoU of 0.6696 and an F1-score of 0.8021, while EfficientNet-B2 provided more consistent results across different architectures. Data augmentation showed limited impact, with gains restricted to a few specific configurations. Qualitative analysis through 3D surface reconstruction further highlighted the limitations of 2D slice-based segmentation, particularly at the extremities of the vertebrae, suggesting the need for volumetric approaches. The contributions of this work include the development of a reproducible pipeline for spinal segmentation, a comparative evaluation of convolutional and Transformer-based models. Qualitative analysis suggests potential benefits in exploring 3D approaches in future work.
Keywords: spinal segmentation — magnetic resonance imaging (MRI) — deep learning — volumetric imaging

Resumo: A segmentação precisa das estruturas da coluna vertebral em imagens de ressonância magnética (MRI) é um passo importante para apoiar o diagnóstico de diversas doenças e condições relacionadas. Neste trabalho, apresentamos uma avaliação sistemática de três arquiteturas de deep learning, U-Net, Feature Pyramid Network (FPN) e SegFormer, combinadas com os encoders ResNet-50 e EfficientNet-B2 para a segmentação semântica da coluna cervical (C1–C7) em volumes de MRI do conjunto de dados VerSe 2020. O pipeline proposto inclui preparação do conjunto de dados, treinamento dos modelos utilizando a função de perda Dice, o otimizador Adam e estratégia de parada antecipada, além da avaliação com métricas padrão como Intersection over Union (IoU), F1-Score e acurácia. Os resultados mostram que a FPN com ResNet-50 obteve o melhor desempenho geral, alcançando IoU de 0,6696 e F1-score de 0,8021, enquanto a EfficientNet-B2 apresentou resultados mais consistentes entre diferentes arquiteturas. O aumento de dados apresentou impacto limitado, com ganhos restritos a poucas configurações específicas. A análise qualitativa por meio da reconstrução de superfícies 3D evidenciou ainda as limitações da segmentação baseada em fatias 2D, especialmente nas extremidades das vértebras, sugerindo a necessidade de abordagens volumétricas. As contribuições deste trabalho incluem o desenvolvimento de um pipeline reprodutível para segmentação da coluna, uma avaliação comparativa de modelos baseados em convoluções e em Transformers, e evidências empíricas que apoiam a adoção de abordagens 3D para melhorar a qualidade e a consistência das estruturas segmentadas.
Palavras-Chave: segmentação da coluna vertebral — ressonância magnética (RM) — deep learning — imagens volumétricas

[1] *Instituto de Ciências Exatas e Tecnológicas, Universidade Federal de Viçosa (UFV), Brazil*
*__Corresponding author__: joaofmari@gmail.com
DOI: http://dx.doi.org/10.22456/2175-2745.150902 • **Received:** 14/10/2025 • **Accepted:** 27/10/2025

1. Introduction

The spine plays a vital role in both supporting and protecting the human body, as well as mobilizing it. Thus, degenerative spinal diseases, such as spinal stenosis, are among the leading causes of low back pain, limiting quality of life and representing a significant clinical and economic burden. In addition to these conditions, spinal deformities such as scoliosis also require accurate diagnosis to support appropriate treatment [1]. Accurate segmentation of the vertebrae and associated structures in magnetic resonance imaging (MRI) is crucial for diagnosing problems in this region [2].

Manual segmentation is time-consuming and can be inconsistent between operators, and the challenge increases when working with volumetric images. This has encouraged the use of automated methods. In recent years, deep learning techniques have shown good results in semantic segmentation tasks, including spinal imaging. However, their performance depends on factors such as model architecture, encoder backbone, data variability, and augmentation strategies [3, 4]. Understanding how these choices affect segmentation quality can help build more reliable pipelines for clinical and research applications.

In this work, we evaluate three segmentation architectures, U-Net, FPN, and SegFormer, combined with both ResNet-50 and EfficientNet-B2 encoders. We then assess the performance of each combination in scenarios with and without data augmentation, using the VerSe 2020 dataset, specifically in the cervical spine (C1–C7). This evaluation helps understand the effect of different model choices on segmentation performance. Besides the quantitative evaluation, we also performed a qualitative analysis by reconstructing the segmented images as 3D polygonal surfaces.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the dataset, preprocessing steps, model architectures, and training procedure. Section 4 presents and discusses both quantitative and qualitative results. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Related works

Han et al. [5] proposed simultaneous segmentation and classification (normal and abnormal) of intervertebral discs, vertebrae, and neural foramina in magnetic resonance images. The authors highlighted a previously unprecedented finding, as no research had achieved the results presented. This novelity was due to three main factors: (i) the multitasking aspect, since segmenting multiple spinal structures is a much more complex task than segmenting individual structures; (ii) the multiple targets, since an average of 21 spinal structures per magnetic resonance imaging scan requires automated analysis, but such structures present great variety and variability; and (iii) weak spatial correlations with subtle differences between normal and abnormal structures generate indeterminacies. Thus, Han et al. present a framework based on a Recurrent Generative Adversarial Network (Spine-GAN) that addresses the aforementioned challenges. To evaluate the approach, the authors considered resonance images from 253 patients, which yielded a per-pixel accuracy of 96.2%, a Dice coefficient of 87.1%, a sensitivity of 89.1%, and a specificity of 86.0%.

Möller et al. [6] presented SPINEPS, a novel deep learning-based method for the semantic and instance segmentation of 14 spinal structures in sagittal T2-weighted turbo spin echo images of the entire body. The proposed model employs a sliding window approach with an auxiliary model to generate instance masks based on semantics. The authors evaluated the approach by combining a public dataset, a subset of the German National Cohort (NAKO), and an in-house dataset. The results demonstrated that for the public dataset, SPINEPS outperformed the nnUNet baseline in all structures considered and for all evaluation metrics. SPINEPS, trained with automated NAKO labeling and tested using in-house data, also achieved promising results, with a global average Dice score of 0.918. Thus, the approach proved to be robust for segmenting the 14 spinal structures considered, in addition to providing a semantic mask and an instance mask that separate the vertebrae and the intervertebral discs. This approach is the first to be publicly available that performs segmentation at this level of detail.

Saeed et al. [7] present a new multi-feature attention model for 3D spine segmentation, which modifies the MobileNet v3 framework by incorporating the Reverse Convolution Block Attention Module (RCBAM) and Feature Pyramid Pooling (FPP) modules, thereby enhancing feature extraction. The authors individually trained each MobileNet v3 module, which enables specific vision for axial, coronal, and sagittal areas of 3D images. Subsequently, the approach concatenates the individual features to form a single 3D feature map for input to the segmentation decoder. The results demonstrate that the approach proposed by Saeed et al. outperforms other models in terms of computational cost, with lower Dice Coefficient Score (DCS) and Intersection over Union (IoU) values for the VerSe 2019 and VerSe 2020 datasets.

These studies highlight significant advances in spinal segmentation. In contrast, our work focuses on a systematic comparison of widely used architectures and encoders under a unified experimental setup, establishing a clear and reproducible performance baseline for cervical spine MRI segmentation.

3. Material and Methods

3.1 Dataset

The Vertebrae Segmentation (VerSe) dataset [8, 9, 10] was developed as a reference benchmark for automatic vertebra segmentation in three-dimensional medical imaging. It is publicly available and has been used in international benchmarking challenges. The dataset includes computed tomography (CT) and magnetic resonance imaging (MRI) scans of different spinal regions (cervical, thoracic, and lumbar), acquired from multiple clinical centers.

Manual annotations were produced by experts and validated by radiologists, covering up to 27 anatomical classes (cervical C1–C7, thoracic T1–T13, lumbar L1–L6, and background). The data are provided in volumetric format while preserving the original spatial resolution, and include scans with different orientations, such as axial and sagittal views. In some cases, axial volumes are isotropic, with voxels of equal size along the X, Y, and Z axes.

3.2 Dataset Preparation

The VerSe 2020 dataset is divided into three subsets: training, validation, and testing. It contains MRI volumes in both axial and sagittal orientations, with some axial scans being isotropic, meaning the voxel dimensions are equal along the Z, X, and Y axes. The original dataset comprises 27 anatomical classes; however, in this work, we focus exclusively on the cervical spine (C1–C7).

To adapt the dataset to our task, the segmentation masks were first relabeled by merging all classes corresponding to vertebrae C1 to C7 (labels 1 to 7) into a single class labeled as 1, while all other voxels were assigned label 0. The preprocessing pipeline then proceeded in two main steps. First, we discarded all volumes that did not contain any voxels labeled from 1 to 7, thereby removing cases without cervical spine representation. Second, we excluded all sagittal volumes, retaining only axial views, both isotropic and anisotropic.

Table 1 summarizes the number of volumes discarded and retained after each filtering step, as well as the final number of slices in each subset after preprocessing.

Table 1. Number of volumes and slices at each stage of dataset preparation. First, all volumes without any voxels labeled as cervical vertebrae C1–C7 were removed. Next, sagittal views were discarded, retaining only axial views (both isotropic and anisotropic). The final row reports the total number of slices for each split after preprocessing.

Dataset Preparation Step	Train Set	Validation Set	Test Set
Original VerSe 2020 dataset	61	80	73
After removing volumes without C1–C7	22	28	29
Final subset (axial views only)	17	21	25
Number of slices	7,844	11,719	17,109

3.3 Models and Encoders

We selected three segmentation architectures that reflect both classical and modern approaches (U-Net, FPN, and SegFormer). These models are widely adopted in medical image analysis and provide a solid basis for comparing segmentation performance.

U-Net is a widely recognized architecture for its robustness in medical image segmentation. This architecture has a symmetric encoder-decoder structure, where the encoder is responsible for reducing the spatial resolution of the image to highlight the most relevant features, and the decoder reverses the resolution reduction process to provide an output as a segmented image with the exact dimensions as the input image. U-Net also employs skip connections between corresponding encoder and decoder layers, which help preserve spatial information and improve the accuracy of the segmentation [11].

The Feature Pyramid Network (FPN) is a top-down architecture that utilizes upsampling from deeper layers, combined with corresponding shallower layers via lateral shortcut connections. This architecture enables object segmentation through more robust multiscale representations by combining the global semantic context present in deeper layers with fine spatial details present in shallower layers [12].

SegFormer is a Transformer-based semantic segmentation architecture that combines an encoder to capture hierarchical representations with a lightweight decoder. The encoder, known as a Mix Transformer (MiT), captures multiscale representations using Transformers that capture global dependencies. MiT applies progressive patch embedding operations to construct feature maps of different resolutions. The SegFormer decoder is based on Multi-Layer Perceptrons (MLP) layers that fuse the hierarchical features from the encoder to produce a segmentation map directly [13].

The previously described architectures were combined with two pre-trained encoders, ResNet-50 and EfficientNet-B2, to assess how different feature extraction backbones influence the segmentation performance.

Residual networks represented a significant advance in training deep networks by addressing the vanishing gradient problem. To achieve this, this type of architecture employs residual blocks through shortcut connections added to the outputs of convolutional blocks. Specifically, the ResNet-50 model used in this work has 50 trainable layers (with convolutions, normalizations, and activations), is pre-trained with ImageNet, and is widely used as an encoder for medical image segmentation applications [14].

EfficientNet B2 employs Compound Scaling operations, enabling the network to scale its capacity in a balanced manner. This balance allows it to simultaneously process image width, depth, and resolution without significantly increasing computational cost. This approach also employs Mobile Inverted Bottleneck Convolution, utilizing shortcut connections, each of which is fed by a Squeeze-and-Excitation block to recalibrate the channels, thereby giving greater weight to the most relevant features [15].

3.4 Training Procedure

As described previously, we trained models by combining different architectures with pre-trained encoders. This section details the training and validation procedures, including the use of data augmentation to improve generalization. The baseline transformation, applied to the validation and test sets and to training without augmentation, consists of resizing images to 512×512, normalizing them using ImageNet statistics, and converting them to tensors. The data augmentation strategy extends this baseline with horizontal and vertical flips, random rotations of 90°, affine transformations (translation of

5%, scaling from 90% to 110%, and rotations between $-20°$ and $20°$), brightness and contrast adjustments (limit factor of 0.2), and Gaussian noise. All transformations are applied with a probability of 0.5, except for Gaussian noise we used 0.3.

All experiments were carried out using the PyTorch framework and the Segmentation Models Pytorch (SMP)[16][1]. The models were trained with the Adam optimizer, an initial learning rate of 1×10^{-4}. A batch size of 8 was used consistently. To control the learning rate, we employed a plateau-based learning rate scheduler, which decreases the learning rate when the validation loss stops improving over 10 epochs. Training was performed for a maximum of 400 epochs, with early stopping based on validation loss, defined by a patience parameter to halt training if no improvement was observed for 21 consecutive epochs.

As the task is binary segmentation, ground-truth masks were preprocessed into tensors of shape $(B,1,H,W)$. The Dice loss was used as the optimization criterion, given its effectiveness in handling class imbalance and its direct optimization of overlap between predicted and reference masks. Model outputs were passed through a sigmoid activation and thresholded at 0.5 to generate binary predictions.

Performance was monitored after each epoch on both training and validation sets. We employed the SMP metrics library to compute mean Intersection over Union (mIoU), F1-score, and accuracy, all with macro averaging. The model with the lowest validation loss was saved as a checkpoint for later inference and evaluation.

3.5 Model Evaluation

All metrics were computed using the SMP library, adopting macro reductions based on the aggregated counts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). We report three standard measures widely employed in medical image segmentation.

The Intersection over Union (IoU), or Jaccard coefficient, quantifies the overlap between the predicted and ground-truth masks:

$$\text{IoU} = \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}}.$$

IoU penalizes both over- and under-segmentation and ranges from 0 to 1, with higher values indicating greater spatial agreement.

The F1-Score, or Dice coefficient, is the harmonic mean of precision and recall, reflecting both accuracy and completeness of the segmentation:

$$\text{F1-Score} = \frac{2 \times \text{TP}}{2 \times \text{TP}+\text{FP}+\text{FN}}.$$

This metric is particularly robust under class imbalance, as it mitigates the influence of the background class.

[1] <https://github.com/qubvel-org/segmentation_models.pytorch>

The accuracy measures the proportion of correctly classified pixels:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}}.$$

Although intuitive, accuracy may be misleading in medical image segmentation, where the background dominates the image. High accuracy can, therefore, mask poor segmentation quality of structures of interest.

3.6 Surface Reconstruction

The 2D segmented slices were used to reconstruct the 3D surface of the spinal structures. For this purpose, we employed the classical marching cubes algorithm [17], which iterates over the volumetric image by examining each cube formed by eight neighboring vertices and determines how the isosurface intersects the cube, thereby selecting the most appropriate triangulation that separates the object from the background. This procedure results in a polygonal mesh that approximates the surface of the structures of interest. We adopted marching cubes to enable a qualitative assessment of the segmentations produced by our models.

3.7 Computational Resources

The experiments were performed on a desktop workstation featuring an Intel Core i5 CPU running at 3.0 GHz, 32 GB of RAM, and an NVIDIA Titan Xp GPU with 12 GB of dedicated memory. The software stack consisted of Python 3.12 and PyTorch 2.7.1 compiled with CUDA 12.6 support. For model implementation, we employed the Segmentation Models PyTorch (SMP) framework, version 0.5, in combination with the Albumentations library (version 2.0.8) to handle data augmentation procedures. Data preprocessing, visualization, and metric computation relied on auxiliary packages, including scikit-learn 1.7, scikit-image 0.25.2, Pillow 11.0, and Matplotlib 3.10.3. For surface reconstruction, we employed PyMCubes[2], and MeshLab[3] [18] for 3D visualization.

4. Results and Discussion

Table 2 summarizes the segmentation performance of all evaluated models on the test set, reporting IoU, F1-Score, and Accuracy. All metrics were computed using macro reduction, where true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were first aggregated across all test images before metric computation. Because MRI volumes contain a large proportion of background pixels, accuracy values tend to be artificially high and thus are less informative for model comparison. Accuracy is still reported for completeness, but our analysis primarily focuses on IoU and F1-score. The table compares three segmentation architectures (U-Net, FPN, and SegFormer), each combined with two encoders (ResNet-50 and EfficientNet-B2), under two

[2] https://github.com/pmneila/PyMCubes
[3] https://www.meshlab.net/

Table 2. Comparison of segmentation performance across models, encoders, and data augmentation settings. For clarity, the best results of each model are underlined, while the overall best results across all models are highlighted in bold.

		No Data Augmentation			Data Augmentation		
Model	**Encoder**	IoU	F1-Score	Accuracy	IoU	F1-Score	Accuracy
Unet	ResNet-50	0.4880	0.6559	0.9982	0.4657	0.6355	0.9981
	EfficientNet-B2	0.6571	0.7931	0.9991	0.6638	0.7979	0.9991
FPN	ResNet-50	**0.6696**	**0.8021**	**0.9992**	0.5263	0.6896	0.9983
	EfficientNet-B2	0.6368	0.7781	0.9990	0.6581	0.7938	0.9990
SegFormer	ResNet-50	0.4323	0.6037	0.9978	0.5491	0.7089	0.9986
	EfficientNet-B2	0.6581	0.7938	0.9991	0.6165	0.7627	0.9989

training conditions: without data augmentation and with data augmentation. For clarity, the best results of each model are underlined per metric (IoU, F1-Score, and Accuracy), while the overall best results across all models are highlighted in bold.

The results in Table 2 show that the combination of FPN with ResNet-50 reached the highest values without data augmentation, with an IoU of 0.6696 and an F1-Score of 0.8021. The other models, U-Net and SegFormer, obtained their best results when paired with EfficientNet-B2. Overall, EfficientNet-B2 yielded better performance across most scenarios, except for the best overall configuration (FPN + ResNet-50 without augmentation). Considering the peak results of each model, SegFormer achieved the lowest scores among the evaluated architectures.

Data augmentation increased performance for EfficientNet-B2 only when combined with U-Net and FPN. For SegFormer, the trend was reversed; improvements were observed only with ResNet-50. Among the top results under augmentation, only the U-Net with EfficientNet-B2 reached the highest values (IoU = 0.6638, F1-Score = 0.7979). In contrast, the best performance of SegFormer occurred with EfficientNet-B2 in the scenario without augmentation (IoU = 0.6581, F1-Score = 0.7938).

Overall, these results highlight two main points: (i) FPN with ResNet-50 is the most effective configuration in this study, and (ii) the choice of encoder strongly impacts performance, with EfficientNet-B2 providing more consistent results across models. The limited gains from augmentation suggest that future work should explore more diverse strategies, including mix-based data augmentation methods [1]. Figure 3 shows six equally spaced axial slices from the CT volume gl108 (175 slices), sampled within the valid-voxel range, ie.e, from the first to the last slice containing at least one foreground voxel. Row (a) displays the original CT slices, (b) the ground-truth segmentation masks, (c) the predicted masks, and (d) the segmentation maps, where green highlights true positives (TP), red false positives (FP), orange false negatives (FN), and black true negatives (TN). A higher number of segmentation errors (false positives and false negatives) can be observed at the extremes of the valid-voxel range. We attribute this to the use of a 2D approach, where slices are segmented independently. To verify this hypothesis, future work should explore fully 3D deep learning models that incorporate volumetric context and relationships between adjacent slices.

Similarly, Figure 1 presents six equally spaced slices from the CT volume verse651 (339 slices), also restricted to the valid-voxel range. As before, row (a) shows the original CT slices, (b) the ground-truth masks, (c) the predicted masks, and (d) the segmentation maps. The same behavior is observed, reinforcing the need for models that exploit 3D spatial continuity.

4.1 Surface Reconstruction

Figure 2 presents 3D reconstructions generated from the same slices shown in Figures 3 and 1, provided here for qualitative analysis. Images (a) and (c) show reconstructions from the ground-truth annotations, while images (b) and (d) display the corresponding predictions. Overall, the predicted reconstructions approximate the anatomical structure of the cervical spine; however, some discrepancies are evident. The predicted reconstructions capture the general anatomy of the cervical spine but reveal irregularities and discontinuities at the extremities. These artifacts directly correspond to the increased FP and FN observed in the slice-level analysis, indicating the limitations of a purely 2D segmentation approach. Ground-truth reconstructions, in contrast, appear smoother and more continuous. Future work will explore fully 3D deep learning models capable of segmenting the entire volume directly, which should preserve spatial coherence across slices.

5. Conclusion

This study presented a systematic evaluation of three deep learning architectures (U-Net, FPN, and SegFormer) combined with ResNet-50 and EfficientNet-B2 encoders for the task of cervical spine segmentation in MRI volumes. The results showed that FPN with ResNet-50 achieved the best overall performance, while EfficientNet-B2 provided more stable results across different models. Data augmentation had a limited effect, improving some configurations but not consistently.

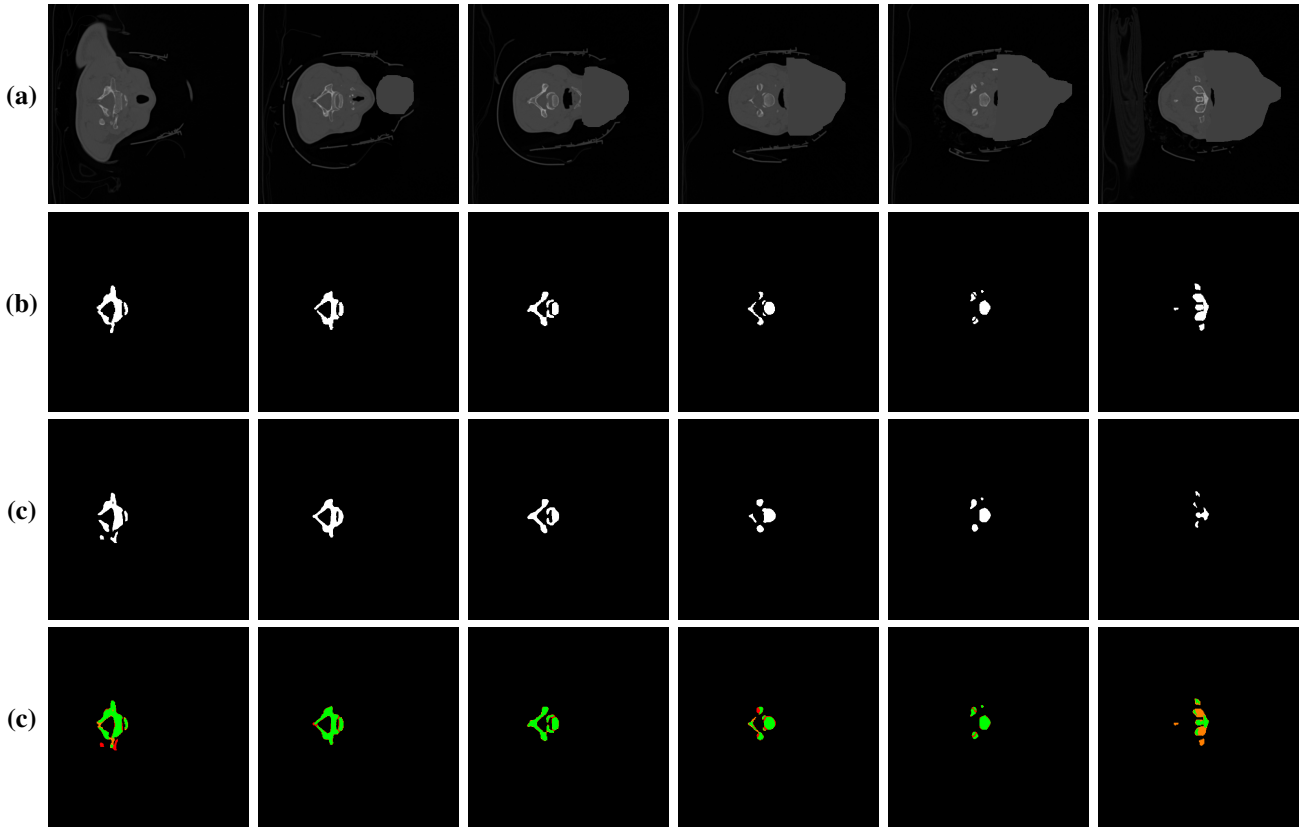
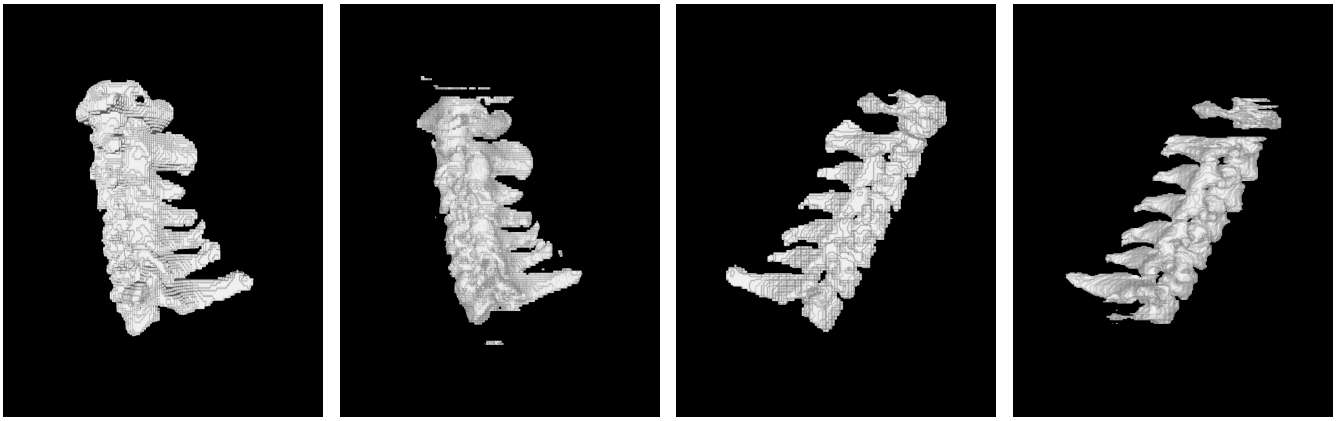


Figure 1. Six equally spaced slices extracted from the informative range of the CT volume `verse651` (339 slices), corresponding to slice indices 96, 124, 152, 181, 209, and 237. Row (a) shows the original CT slices, (b) the corresponding ground-truth segmentation masks, (c) the predicted segmentation masks, and (d) the segmentation map.



(a) GT (gl108) **(b)** Pred. (gl108) **(c)** GT (verse651) **(d)** Pred (verse651)

Figure 2. 3D surface reconstructions generated from the segmented slices of volumes gl108 and verse651. Panels (a) and (c) show the ground-truth segmentations, while panels (b) and (d) display the corresponding predicted segmentations.

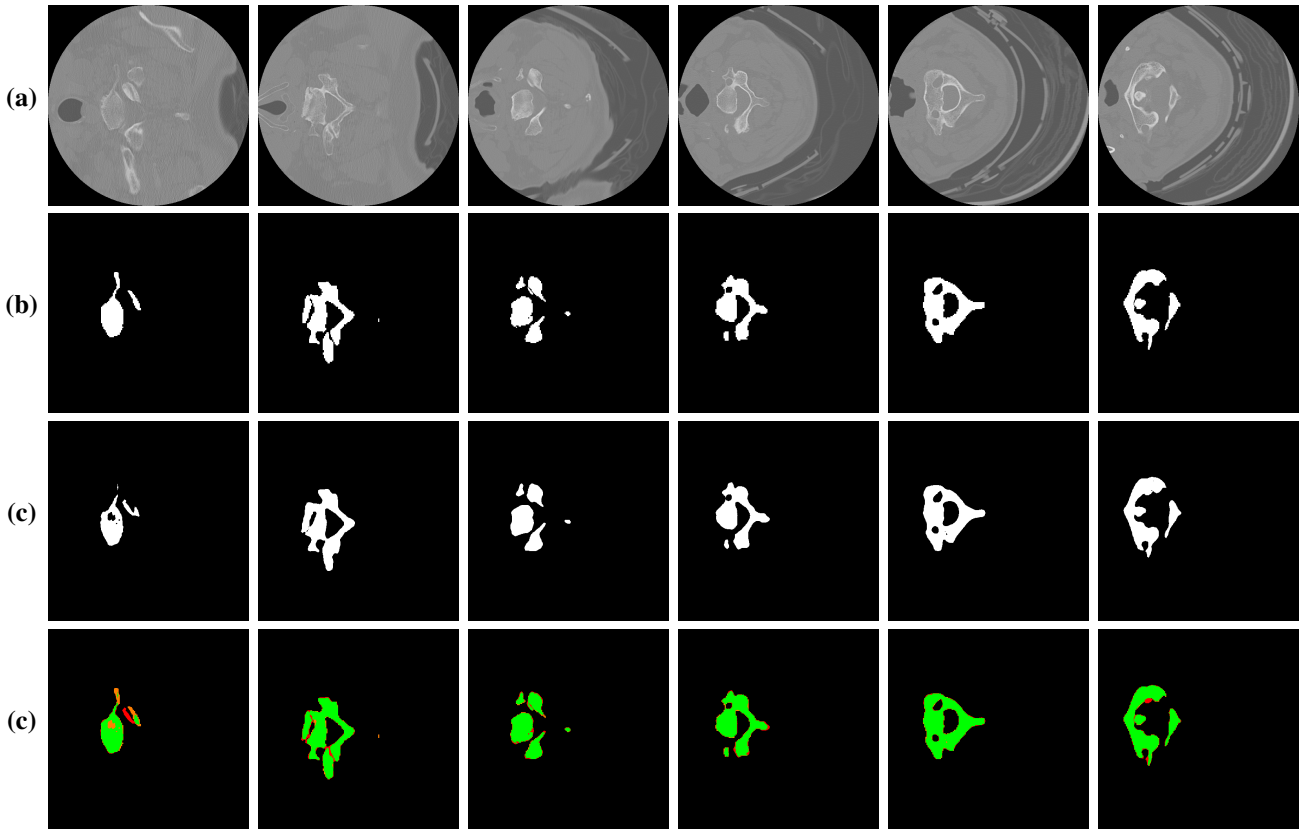


Figure 3. Six equally spaced slices extracted from the informative range of the CT volume `gl108` (175 slices), corresponding to slice indices 56, 73, 90, 106, 123, and 140. Row (a) shows the original CT slices, (b) the corresponding ground-truth segmentation masks, (c) the predicted segmentation masks, and (d) the segmentation map.

The main contributions of this work are: (i) the construction of a reproducible pipeline for spinal segmentation in MRI volumes, including dataset preparation, training, and evaluation; (ii) a comparative analysis of convolutional and transformer-based architectures in this medical imaging context; and (iii) quantitative and qualitative evidence that 2D slice-based methods face limitations, reinforcing the need to investigate 3D models that can better capture volumetric context.

Future work will focus on extending these experiments to 3D approaches and exploring more advanced augmentation strategies to improve robustness.

Acknowledgements

We would like to thank CAPES, CNPq, and FAPEMIG for the financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. R. S. Carvalho received a scholarship from PIBIC/FAPEMIG.

Author Contributions

Conceptualization: R.S.C., L.H.F.P.S., and J.F.M. Methodology: R.S.C., L.H.F.P.S., and J.F.M. Code Development: R.S.C., L.H.F.P.S., and J.F.M. Experiments: R.S.C., L.H.F.P.S., and J.F.M. Discussion of Results: R.S.C., L.H.F.P.S., and J.F.M. Writing - Original Draft: R.S.C. Writing - Review & Editing: L.H.F.P.S., and J.F.M. Supervision: L.H.F.P.S., and J.F.M.

References

[1] PEREIRA, J. F. M.; MARI, J. F.; SILVA, L. H. F. P. Exploiting data augmentation strategies to improve the classification of spinal disorders in x-ray images. *Revista de Informática Teórica e Aplicada*, v. 32, n. 1, p. 257–264, 2025.

[2] BHARADWAJ, U. U.; CHIN, C. T.; MAJUMDAR, S. Practical applications of artificial intelligence in spine imaging: a review. *Radiologic Clinics*, Elsevier, v. 62, n. 2, p. 355–370, 2024.

[3] BAUR, D. et al. Convolutional neural networks in spinal magnetic resonance imaging: a systematic review. *World Neurosurgery*, Elsevier, v. 166, p. 60–70, 2022.

[4] QU, B. et al. Current development and prospects of deep learning in spine image analysis: a literature review. *Quantitative Imaging in Medicine and Surgery*, v. 12, n. 6, p. 3454, 2022.

[5] HAN, Z. et al. Spine-gan: Semantic segmentation of multiple spinal structures. *Medical image analysis*, Elsevier, v. 50, p. 23–35, 2018.

[6] MÖLLER, H. et al. Spineps—automatic whole spine segmentation of t2-weighted mr images using a two-phase approach to multi-class semantic and instance segmentation. *European Radiology*, Springer, v. 35, n. 3, p. 1178–1189, 2025.

[7] SAEED, M. U. et al. 3d mfa: An automated 3d multi-feature attention based approach for spine segmentation using a multi-stage network pruning. *Computers in Biology and Medicine*, Elsevier, v. 185, p. 109526, 2025.

[8] LÖFFLER, M. T. et al. A vertebral segmentation dataset with fracture grading. *Radiology: Artificial Intelligence*, Radiological Society of North America, v. 2, n. 4, p. e190138, 2020.

[9] SEKUBOYINA, A. et al. Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images. *Medical image analysis*, Elsevier, v. 73, p. 102166, 2021.

[10] LIEBL, H. et al. A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data. *Scientific data*, Nature Publishing Group UK London, v. 8, n. 1, p. 284, 2021.

[11] RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. [S.l.], 2015. p. 234–241.

[12] LIN, T.-Y. et al. Feature pyramid networks for object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2017. p. 2117–2125.

[13] XIE, E. et al. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, v. 34, p. 12077–12090, 2021.

[14] HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 770–778.

[15] TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2019. p. 6105–6114.

[16] IAKUBOVSKII, P. *Segmentation Models Pytorch*. [S.l.]: GitHub, 2019. <https://github.com/qubvel/segmentation_models.pytorch>.

[17] LORENSEN, W. E.; CLINE, H. E. Marching cubes: A high resolution 3d surface construction algorithm. In: *Seminal graphics: pioneering efforts that shaped the field*. [S.l.: s.n.], 1998. p. 347–353.

[18] CIGNONI, P. et al. MeshLab: an Open-Source Mesh Processing Tool. In: SCARANO, V.; CHIARA, R. D.; ERRA, U. (Ed.). *Eurographics Italian Chapter Conference*. [S.l.]: The Eurographics Association, 2008. ISBN 978-3-905673-68-5.