JJEE

Volume 12 | Number 1 | March 2026 | Pages 189-208

Jordan Journal of Electrical Engineering

ISSN (print): 2409-9600, ISSN (online): 2409-9619

Homepage: jjee.ttu.edu.jo

Efficient Vertebrae Segmentation in MRI Using YOLOv8 and U-Net

Kheira Laazab[1*], Nadjia Benblidia[2], Ali Baaloul[3]

[1, 2, 3] Universite Saad Dahlab Blida, ouled Yaich, blida09000, Algeria

E-mail: laazab_kheira@univ-blida.dz

Received: Oct 17, 2025 *Revised: Jan 02, 2026* *Accepted: Jan 19, 2026* *Available online: Mar 19, 2026*

***Abstract*—** Manual analysis of spinal MRI scans is time-consuming and prone to human errors due to subtle anatomical variations. Recent advancements in Artificial Intelligence (AI) provide tools for automating spinal segmentation, potentially improving diagnostic efficiency. In this study, we propose a hybrid framework that integrates YOLOv8 for vertebrae localization with a U-Net architecture and the watershed algorithm for segmentation. YOLOv8 is used for its high-speed detection and localization capabilities, while the combination of U-Net and watershed facilitates precise delineation of spinal structures. Experimental results indicate a mean Average Precision (mAP) of 99.1% for localization and a segmentation accuracy of 90% with minimal loss. The integrated method may contribute to improving diagnosis and treatment planning of spinal deformities. Moreover, the proposed approach is scalable and could be adapted to other medical imaging segmentation tasks.

***Keywords*—** Vertebrae segmentation; Lumbar spine localization; YOLOv8; U-net; MRI imaging; Watershed algorithm.

1. INTRODUCTION

In 2013, the World Health Organization reported that between 250,000 and 500,000 people suffer from spinal injuries or deformities each year [1]. The spine is the main structural support of the human body and is composed of 26 vertebrae, separated by intervertebral discs. These vertebrae are organized as follows: seven cervical vertebrae, twelve thoracic vertebrae, five lumbar vertebrae, and two vertebrae forming the sacrum and coccyx [2, 3] (Fig. 1 shows the distribution of the vertebrae in the spine).

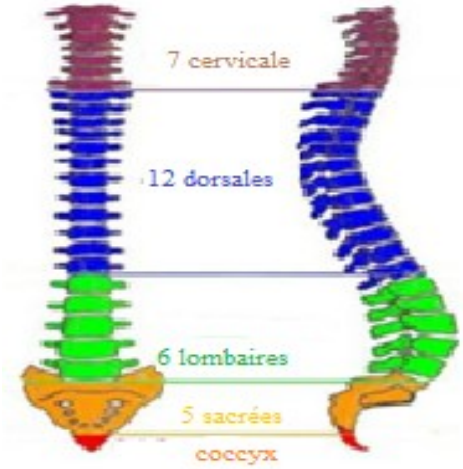


Fig. 1. Model of the human spine.

* Corresponding author

DOI: https://doi.org/10.5455/jjee.204-1760706104

Accurate analysis and segmentation of the spine in medical imaging are challenging due to the high variability in vertebral morphology and position across different patients. Deep learning approaches have demonstrated significant potential in addressing these challenges, achieving high performance in tasks such as vertebrae localization, spine segmentation, and anomaly detection [4].

helping doctors to analyze and detect the disease early quickly and accurately. In this research, we present a model using deep learning algorithms to segment vertebrae of the lumbar spine from MRI images, accurately and efficiently, to assist in early and accurate diagnosis of spinal diseases. It is guided by the main research question: RQ: How can deep learning techniques identify and segment vertebrae of the lumbar spine from MRI images, in particular the YOLOv8 and U-net models, and thus provide rapid and early detection of spinal diseases?

Our contributions include:

- Preparing the dataset by changing the size of the images to 256 x 256, and using LabelMe and Roboflow, we placed annotations on the vertebrae of the lumbar spine.
- Training the YOLOv8 model on the dataset to localize the vertebrae of the lumbar spine.
- Cropped the vertebrae of the lumbar spine by choosing a confidence level of 0.8, then group each vertebra with its mask according to the name of the vertebrae, and get rid of every file that does not contain a mask or vertebra at the same time.
- Using a set of geometry methods, such as rotating to different angles and flipping the images upside down, on the data set of the lumbar spine vertebra
- Developing a U-net model to segment the spinal vertebrae from MRI images by changing the size of the filters and increasing the number of convolution and deconvolution layers to extract the largest amount of information from the images.
- Integrating the U-net model with the watershed method to segment vertebrates more efficiently and accurately.

Our research was organized according to the following approach: Part 2 contains previous studies on the localization and segmentation of spinal vertebrae. In Part 3, we find a detailed explanation of our proposed method and a mention of the Materials used in our research. The results obtained and their interpretation are given in Part 4. We conclude our research with a conclusion in Part 5.

2. RELATED WORK

Medical image segmentation is pivotal in diagnosing and treating spinal deformities, aiding in the accurate identification and analysis of vertebrae from MRI images. In this section, we categorize related work into two groups:" Medical Image Segmentation Techniques" and" Advanced Segmentation Approaches," reflecting the evolving landscape of segmentation methodologies in spinal deformity diagnosis.

2.1. Medical Image Segmentation Techniques

Segmentation of the spine in medical images plays a crucial role in detecting spinal deformities and guiding surgical interventions. Numerous segmentation systems have been developed by researchers to address the complexity and diversity of spinal components [1]. Recent advancements in artificial intelligence have led to significant progress in medical

imaging, with various techniques being explored to assist in disease diagnosis and vertebrae segmentation [5]. An autoencoder is used for anomaly detection. It reconstructs normal regions well and poorly reconstructs abnormal ones. This allows it to detect anomalies with an accuracy of 99.1%, showing its robustness.

Several studies have reported strong performance across multiple datasets [6]. Suzani *et al.* presented a deep learning–based framework for vertebra localization in CT images, achieving an accuracy of approximately 96% with a reduced computation time [7]. In a similar context, Chu *et al.* proposed a semi-automatic approach for vertebral body localization and segmentation, obtaining similarity coefficients ranging from 88.7% to 91% on various datasets [8]. AlArif *et al.* employed an enhanced U-Net architecture combined with shape-aware information to segment cervical vertebrae, reaching a similarity score of 94.38% after applying data augmentation strategies [9]. Payer *et al.* introduced a multi-stage vertebra segmentation pipeline integrating U-Net with a Spatial Composition Network (SC-Net), which resulted in effective vertebra localization (89.78%) and a Dice coefficient of 94% [10]. In the present study, U-Net and YOLOv7 are used to detect and localize small anatomical markers in MRI images. Experimental results indicate that both models achieve reliable performance, with U-Net showing slightly higher accuracy. Compared to conventional techniques, these methods significantly reduce localization errors. Moreover, they are robust, fully automated, and capable of adapting to different MRI datasets when sufficient training data are available [11].

2.2. Advanced Segmentation Approaches

The identification and categorization of spinal structures from medical imaging play a vital role in the timely detection and swift diagnosis of spine-related conditions. Over the past few years, technological progress has led to the emergence of numerous techniques for analyzing vertebral elements, including both semi-automated procedures and fully automated methods driven by deep learning models. Contemporary research efforts have aimed to refine these segmentation strategies to boost precision and streamline the diagnostic process, especially in cases involving spinal abnormalities. Mushtaq and al [3] employed YOLOv5, total overlapping edge detection, and U-Net for lumbar spine segmentation, achieving an impressive mean precision (mAP) of 97.5% and an accurate rate o74. 5% in a dataset of 51images. Wang and al [12], introduced an image segmentation approach based on U-Net++ and U-Net, restructuring network paths to minimize differences between encoder and decoder characteristics, resulting in enhanced segmentation accuracy. While demonstrating promising results with dice similarity coefficients of 88% and 86.2% for U-Net++ and U-Net, respectively, further validation is warranted. Zhang et al [13] , employed a deep learning-based model to segment the trunks of the brachial plexus nerves using diffusion MRI images. The results demonstrated high accuracy and efficiency in handling complex and delicate anatomical structures in medical images. This work supports the growing trend of utilizing advanced artificial intelligence models in the field of medical image analysis. Table 1 summarizes some of the approved work of the two doctors in dividing the spine.

2.3. Research Gap and Our Position

Previous studies on vertebrae localization and segmentation have notable limitations. Many of these works treat localization and segmentation as separate stages, without an

integrated pipeline linking precise vertebrae detection to fine-grained segmentation, which may affect overall performance and consistency across tasks. Early approaches often relied on traditional CNN or patch-based methods, which struggle to capture global context and anatomical variability in spinal images [23]. While U-Net and its variants remain widely used for segmentation due to their contextual feature extraction capabilities, these methods alone may not address the variability and structural complexity inherent in MRI spine data without an explicit localization stage [23]. Other methods have proposed two-stage architectures for vertebrae detection and segmentation, such as Dense-U-Net based models for CT scans that localize vertebrae centroids before segmentation [24], illustrating the benefit of hierarchical processing. Additionally, recent systems combine object detection models like YOLOv5 with segmentation networks for spinal localization and edge-based segmentation [11], but these rely on older YOLO versions and may not fully benefit from recent advances in anchor-free detection and feature fusion. Evaluations across different imaging modalities and datasets further complicate direct comparison and benchmarking, limiting generalizability and making consistent performance assessment difficult.

In this context, our approach addresses these gaps by proposing a unified and sequential framework that integrates YOLOv8 for vertebrae localization with Watershed-based pre-segmentation and U-Net for fine-grained segmentation. By leveraging YOLOv8's anchor-free detection architecture and improved feature fusion capabilities, the localization stage provides accurate ROI extraction that simplifies and enhances the segmentation task. This integrated pipeline not only improves segmentation accuracy but also streamlines the analysis process, contributing to more efficient and effective diagnostic workflows for spinal deformity assessment in MRI. Unlike previous works, our study explicitly defines the interaction between localization and segmentation stages and evaluates the entire framework specifically on lumbar spine MRI, ensuring methodological consistency and demonstrating clear advancement over existing approaches.

Table 1. Summary of various approaches in vertebrae segmentation.

Mode	Year	Author	Modality	Categories	Database	Dice Score
Semi-automatic	2012	Benayed and al [14]	IRM	Graph cut	15 mid-sagittal images	85%
	2014	Zukic and al [15]	IRM	Inflation-based method	234 vertebrae	79.3%
	2016	Hille and al [16]	IRM	Hybrid level sets	34 vertebrae	84.8%
	2021	Altini and al [17]	CT scans	KNN & CNN	VerSe 2020 dataset	90.09%
Automatic	2018	Janssens and Zheng [18]	CT scans	FCN	xVertSeg of MICCAI'16	95.77%
	2019	Lessmann and al [5]	CT scans	FCN	50 MRI volumes "PROMISE2012"	93%
	2021	Altini and al [17]	CT scans	CNN	214 CT scans "VerSe '20"	89.17%
	2023	Klein and al [19]	CT scans	GCN	26 vertebral "VerSe dataset"	88.3%
	2023	Wang, Xiao, and Tan [20]	MRI	U-Net++	195 MRI images	88%
	2024	Moller and al [21]	MRI	nnUNet & Sliding window	SPIDER dataset	90%
	2024	Kawathekar and Aparna[22]	CT scans	Unbalanced-Unet, 3D SCN & U-Net	VERSE'19 database	93.07%

3. METHODS AND MATERIAL

In the pursuit of accurate spinal deformity diagnosis and treatment planning, the following section delineates the materials and methodologies employed. Leveraging advanced

image processing techniques and state-of-the-art neural network architectures, our approach aims to enhance the identification and analysis of spinal structures in MRI images. The methodology begins with the precise localization of vertebrae using YOLOv8, followed by segmentation utilizing the watershed transformation method combined with the U-Net architecture. This sequential approach ensures the integrity and quality of the data used for analysis, facilitating fine-grained segmentation of individual vertebrae. Additionally, robust dataset annotation and preprocessing steps further enhance the accuracy and efficacy of the analysis. This section provides a comprehensive overview of the materials and methodologies utilized in our study, elucidating the intricate processes involved in spinal structure localization and segmentation. Moreover, Fig. 2, depicting the overview of the approach, is included to offer a visual representation of the methodology's workflow.

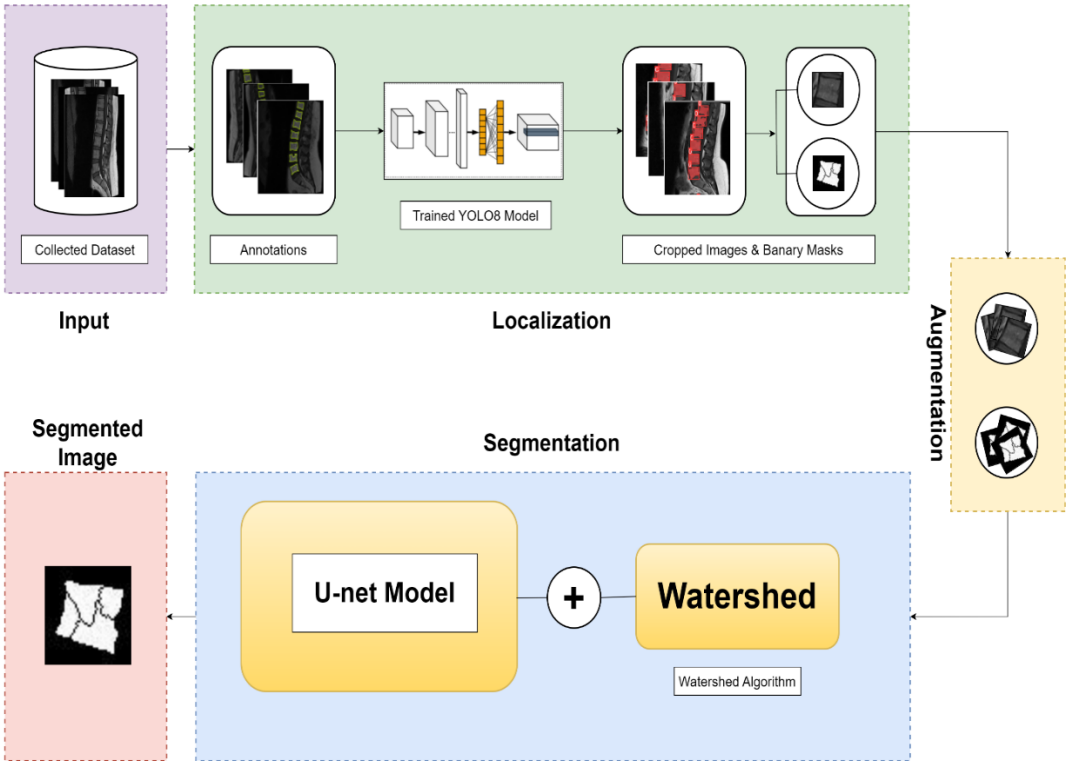


Fig. 2. Overview of the methodology: sequential localization and segmentation of spinal structures in MRI images.

The overall workflow, from image acquisition to final segmentation, is summarized in Algorithm 1, providing a step-by-step illustration of the hybrid approach employed in this study.

3.1. Materials

To facilitate the diagnosis and treatment of various spinal diseases, a comprehensive dataset was curated from the Jordanian Specialized Hospital, spanning from September 2015 to July 2016 [17]. This dataset comprises MRI scans obtained from 575 subjects with spinal ailments, from which 60 scans were excluded due to distortion and noise, resulting in a curated collection of 514 high-quality images for analysis [23]. The dataset focuses on complex lumbar spine cases and is sourced from Mendeley Data [25]. Each image is accompanied by segmentation-ready false-color labels and ground-truth annotations, facilitating precise analysis and diagnosis [25]. The annotations specify the locations of six classes: the lumbar

vertebrae (L1–L5) and the sacrum (S), in pixel coordinates, aiding surgeons in treatment planning and patient monitoring. For each class, the number of annotated instances is provided to clearly describe class distribution and coverage across the dataset. All images are standardized to 320 × 320 pixels. Figure 3 showcases a representative image from the dataset, illustrating the sagittal view of the spine along with corresponding ground-truth annotations in false colors. Figure 3 showcases a representative image from the dataset, illustrating the sagittal view of the spine along with corresponding ground truth annotations in false colors.

Algorithm 1: Hybrid Framework for Vertebrae Localization and Segmentation

```
Begin
Input: Raw MRI image database D
Output: Pixel-wise segmentation masks M
1. Data Preprocessing and Augmentation
For each MRI image in D
   Resize image to 256 x 256
   Apply rotations and flips for data augmentation
   Normalize pixel intensities
End For
2. Automatic Vertebrae Localization (YOLOv8)
Train YOLOv8m on annotated images
For each image
   Predict bounding boxes for L1-L5
   Select ROIs with confidence >= 0.8
End For
3. Pre-segmentation (Watershed)
For each ROI
   Apply Watershed to get initial contours
   Separate adjacent or overlapping vertebrae
   Remove irrelevant regions to reduce noise
End For
4. Fine Segmentation (U-Net)
Feed pre-segmented ROIs into the modified U-Net
For each ROI
   Encoder: Extract deep semantic features
   Decoder: Reconstruct spatial resolution using skip connections
   Generate final mask M with target accuracy ~90%
End For
End
```

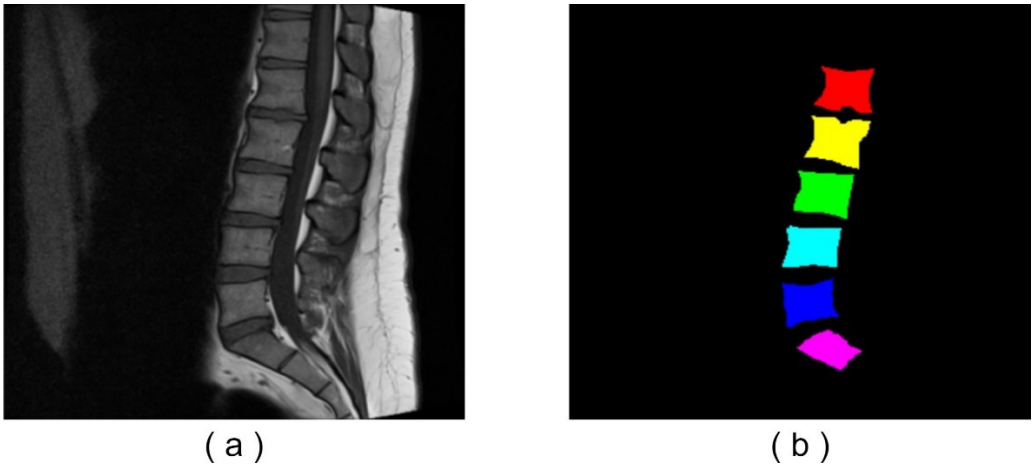


Fig. 3. a) Sagittal image of the spine; b) Ground truth annotations with false-color labels.

3.2. Methods

In our study, we employ advanced localization and segmentation techniques to facilitate the detection and analysis of spinal deformities. Our methodology focuses on automating the identification and analysis of lumbar spine deformities through a series of precise steps. Initially, we utilize YOLOv8 for automated vertebrae localization, training the model on annotated images and corresponding masks. Subsequently, we employ the U-Net architecture in conjunction with the watershed method [19] for fine-grained segmentation of the vertebrae from the spinal region.

3.1.1. Vertebrae Localization

To localize objects in images, bounding boxes are commonly used, delineating the regions of interest [20]. In the context of vertebrae localization, bounding boxes are drawn around each vertebra, facilitating their precise extraction from the images. This process is achieved through the utilization of YOLOv8 applied to the dataset, enabling efficient and accurate localization of vertebrae regions.

The YOLOv8 model was selected due to its anchor-free detection strategy, improved feature fusion through the C2f module, and decoupled prediction head, which collectively enhance localization accuracy and stability. These properties make YOLOv8 particularly suitable for precise vertebrae localization in MRI images, especially when compared to earlier YOLO versions such as YOLOv5 and YOLOv7.

Prepossessing

serves as a pivotal initial step in most segmentation workflows, laying the foundation for subsequent model training and analysis. Our preprocessing pipeline begins by resizing the images to a standardized résolution of 256 x 256 pixels, ensuring uniformity and facilitating efficient processing. To enhance the quality and diversity of the data while addressing the limited dataset size, we employ standard data augmentation techniques. It is important to clarify that our study relies exclusively on real clinical MRI scans; no synthetic or AI-generated images were used. The augmentation process, applied only to the training set, includes random rotations (±15°), horizontal and vertical flips, and brightness adjustments. This ensures the model learns from authentic anatomical variations while maintaining a strict separation between real original images and augmented versions to prevent data leakage. These preprocessing steps are essential for mitigating data variability, improving model generalization, and ultimately optimizing segmentation performance.

Annotations of the data

In the data annotation phase, we utilized LabelMe and Roboflow Annotate [3] to annotate the dataset and correct any preexisting annotations. Vertebrae in each image were meticulously labeled, with labels indicating different vertebral units. Annotations were saved in JSON format, with each file sharing the same name as its corresponding image and containing relevant information such as vertebra coordinates, width, and height. Since YOLO requires TXT format files, Roboflow was used to convert the JSON annotations seamlessly to TXT format. In our annotation schema, the class name 'V' was used for vertebrae, en compassing a sagittal deficit and the five lumbar vertebrae, resulting in six distinct labels. Figure 4 illustrates a representative annotated spine.

To ensure annotation consistency, all labels were reviewed and corrected using LabelMe and Roboflow. Inconsistent or conflicting annotations were resolved through manual

inspection by experts. Images with unresolved ambiguity were excluded from the dataset to maintain high-quality annotations

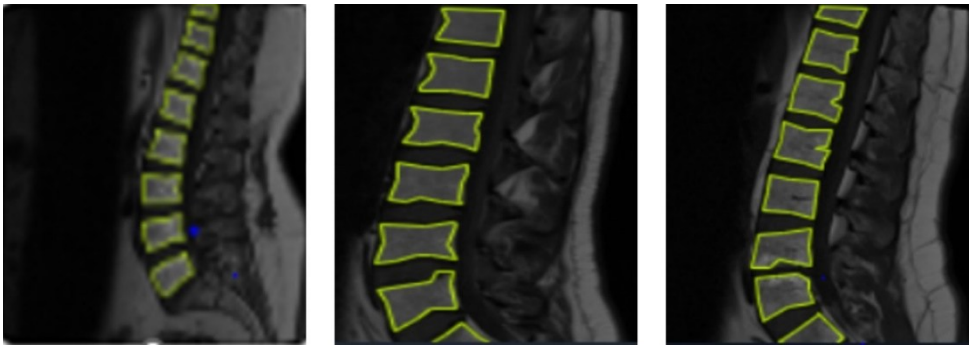


Fig. 4. Illustration of spine vertebrae annotation: Each image's corresponding text file contains essential information, including vertebrae coordinates, height, and width.

Training of YOLOv8m

In the training phase, YOLOv8m was trained on a dataset of images annotated with bounding boxes indicating vertebrae locations. YOLOv8, released by Ultralytics in January 2023, represents a recent advancement in object detection Methods [21]. Since the introduction of the YOLO family in 2015, thèse model have been widely adopted for their fast and accurate detection, enabling efficient prédiction of object boundaries using bounding boxes. Successive versions, culminating in YOLOv8, have progressively improved detection performance [9, 21]. Figure 5 illustrates the évolution of YOLO algorithms over time.

In this work, YOLOv8m was trained on images annotated with bounding boxes indicating vertebrae locations. The dataset was split into 85% training, 10% validation, and 5% testing sets to ensure reliable evaluation.

A systematic hyperparameter tuning strategy was employé, in which the Learning rate, batch size, and number of epochs were varied. The configuration that achieved the best validation performance—Learning rate = 0.001, batch size = 32, and epochs = 100—was selected to ensure stable convergence and robust detection performance.

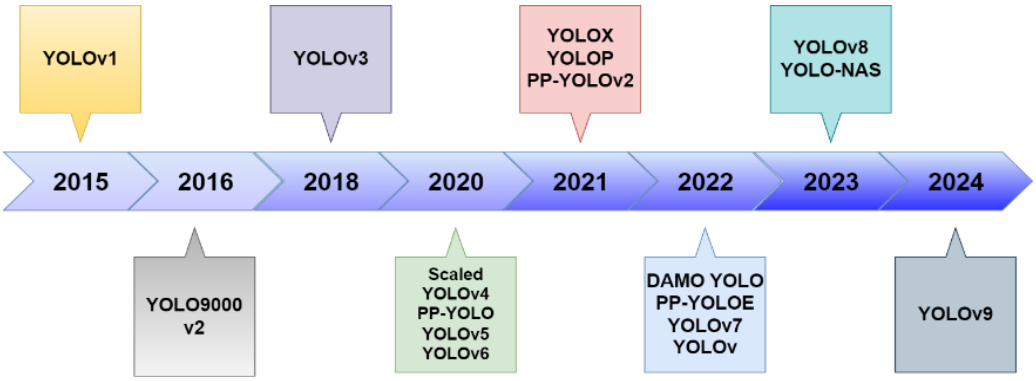


Fig. 5. A timeline of YOLO versions.

YOLOv1 has a fast detection speed but Is less efficient when it comes to smaller targets and positioned objects. While YOLOv2 improved detection accuracy, especially for Small objects, by extracting features that can adapt to images. To enhance detection capabilities across different scales and semantic details, YOLOv3 presented a model that combines Spatial Pyramid Pooling (SPP) and Feature Pyramid Network (FPN) modules. To increase accuracy, Mish was activated in YOLOv4. YOLOv5 enhances feature disruption detection with SPPF and C3 modules. To identify and represent semantic features and increase the accuracy of object

detection, effective transitional units, E-ELAN, and reparameterization structures were combined in YOLOv7[22]. YOLO was then developed into YOLOv8 through the following stages:

a) **Backbone Network:**

• Based on the CSP concept, use CBS, C2f, and SPPF modules for feature extraction (inherited from prior YOLO versions).

• C2f module enhances feature fusion and improves speed.

• The SPPF module increases the area of the image a neuron can "see," enabling it to combine local and global features for richer object descriptions.

b) **Head:**

• Makes final predictions.

• Separates bounding box and class prediction for improved accuracy.

• Anchor-free detection: Uses an Anchor-Free approach for potentially better flexibility in object detection.

c) **Loss Functions and Matching:**

• Utilizes different loss functions (BCE, DFL Loss + CIOU) for classification and regression. Fig. 6 [22], explains the different parts of the YOLOv8 model

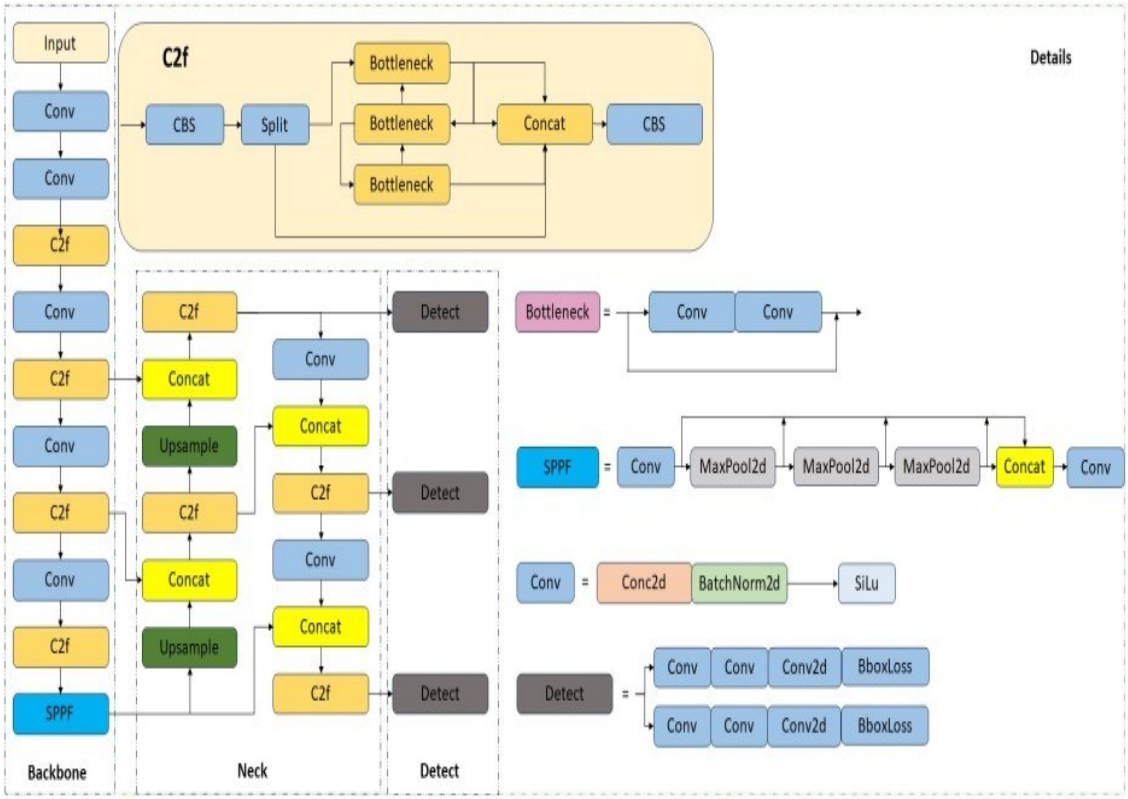


Fig. 6. Architecture of YOLOv8.

3.1.2. Segmentation

After the successful localization of vertebrae, the segmentation process proceeds to delineate individual vertebrae boundaries with precision. Watershed is applied before U-Net as a pré-segmentation step to separate overlapping vertebrae, while U-Net performs refined pixel-Wise segmentation, enabling Advanced analyses in medical imaging applications.

Watershed Transformation

The Watershed Transformation [24] method is employed to address overlapping vertebral regions and separate them into distinct segments. This technique treats the image as

a topographic surface, where pixel intensities represent elevations. Water- shed lines are constructed along the ridges of the surface to delineate boundaries between adjacent regions. By leveraging both intensity gradients and spatial information, Watershed Transformation effectively separates overlapping vertebrae regions, facilitating accurate segmentation.

U-Net Architecture

After the Watershed segmentation process, the U-Net architecture is employed for precise segmentation of the lumbar spine from MRI images. The U-Net network, based on the fully convolutional network (FCN) method of encoder-decoder type [24], is specifically chosen for its effectiveness in biomedical image segmentation tasks. With its Intricate design tailored for pixel-wise classification, the U-Net offers a robust framework for delineating individual vertebral structures.

The architecture of our proposed U-Net model is detailed in Table 2, providing insights into the network's configuration and layer specifications. By leveraging the U-Net architecture post-Watershed segmentation, our segmentation approach ensures accurate and detailed delineation of individual vertebrae boundaries, facilitating precise analysis and diagnosis in medical imaging applications.

Table 2. Illustrates the architecture of the proposed U-Net model for spinal vertebrae segmentation, delineating the encoding and decoding blocks along with their respective image sizes and filters.

Block	Path	Input Size → Output Size	Layers	Filters (Kernel)
1	Encoder	256×256→ 128×128	Conv + ReLU ×2, MaxPooling	16 (3×3), 16 (3×3)
2	Encoder	128×128 → 64×64	Conv + ReLU ×2, MaxPooling	32 (3×3), 32 (3×3)
3	Encoder	64×64 → 32×32	Conv + ReLU ×2, MaxPooling	64 (3×3), 64 (3×3)
4	Encoder	32×32 → 16×16	Conv + ReLU ×2, MaxPooling	128(3×3),128(3×3)
5	Bottleneck	16×16 → 16×16	Conv + ReLU ×2	256(3×3),256(3×3)
4	Decoder	16×16 → 32×32	Conv2DTranspose, Conv + ReLU ×2	128(2×2),128(3×3)
3	Decoder	32×32 → 64×64	Conv2DTranspose, Conv + ReLU ×2	64 (2×2), 64 (3×3)
2	Decoder	64×64 → 128×128	Conv2DTranspose, Conv + ReLU ×2	32 (2×2), 32 (3×3)
1	Decoder	128×128→ 256×256	Conv2DTranspose, Conv + ReLU ×2	16 (2×2), 16 (3×3)
Out	Output	256×256→ 256×256	Conv + Sigmoid	1 (1×1)

Contracting Path (Encoder)

The contracting path of the U-Net serves as the feature extraction stage, capturing high-level features through successive convolutional and pooling layers. This path downsamples the input image, gradually reducing spatial dimensions while increasing feature representation.

Expanding Path (Decoder)

In the expanding path, feature maps obtained from the contracting path are upsampled through transpose convolutions or upsampling layers. This path recovers spatial information lost during downsampling, enabling precise pixel-wise classification.

Skip Connections

U-Net incorporates skip connections between corresponding encoder and decoder layers to facilitate information flow across different scales. These connections enable the network to leverage both low-level and high-level features, enhancing segmentation accuracy and preserving fine details.

Pixel-wise Classification

At the final layer, pixel-wise classification assigns each pixel to its corresponding class label, delineating individual vertebrae boundaries with high precision. U-Net's architecture enables end-to-end training for fine-grained segmentation, optimizing performance for biomedical image analysis tasks.

A graphical representation of the U-Net architecture can be observed in Fig. 7, illustrating the intricate network topology and information flow.

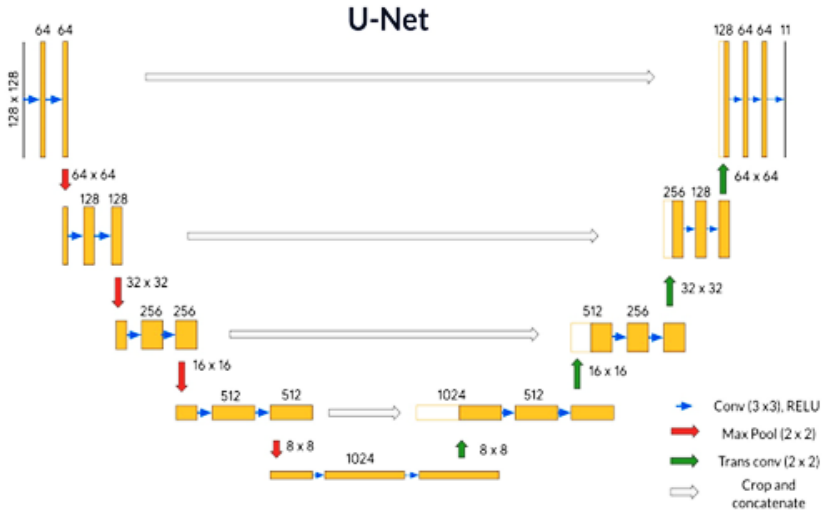


Fig. 7. Model architecture for u-net the lumbar spine segmentation.

By integrating Watershed Transformation with the U-Net architecture, our segmentation approach ensures robust and accurate delineation of individual vertebrae structures, facilitating advanced analysis and diagnosis in medical imaging applications.

The best results were with partitioning (20% testing, 80% training), epoch size 20, learning rate 0.001, and batch size 16. To improve, we used Adam [25].

3.1.3. Evaluation Criteria

To evaluate the lumbar spine localization model using the YOLOv8 method, recall, precision, and mAP ratio were calculated. The lumbar spine segmentation model was evaluated by calculating the accuracy (accuracy) and mean error (ME).

Equation 1 (Recall): This shows how well your model finds the vertebrae. Recall is the result of dividing the true positive cases by the total number of cases [26] as in the following Eq. (1):

$$R = \frac{TP}{TP+FN} \quad (1)$$

where TP is True Positives and FN is False Negatives

Equation 2 (mAP): is the average accuracy and is considered a criterion for evaluating an object localization model containing different classes. The mAP value ranges from 0 to 1. The number 1 indicates that the accuracy is complete. The average precision is determined and

calculated by integrating the precision p with a function of calling r, meaning p(r) [32]. As in the following Eq. (2):

$$\text{mAP} = \int_0^1 \text{p(r)dr} \quad (2)$$

Equation 3 (Mean Error (ME)): ME represents the loss error between the expected ground and truth values. The lower the ME percentage, the more accurate the result will be. ME is calculated by the relation Eq. (3):

$$\text{ME} = \frac{1}{\text{L}}\sum_{\text{i=0}}^{\text{L}-1}\left|\widehat{\text{Y}}_{\text{i}} - \text{y}_1\right| \quad (3)$$

where L is the total number of images, Yˆi represents the expected lumbar spine segment for the i-th image, and Yi is the ground truth of the lumbar spine for the i-th image.

Equation 4 (Accuracy): Accuracy represents the percentage of correct predictions out of the total number of cases. The accuracy value is limited to [0,1]. The better the segmentation, the greater the accuracy. ME is calculated by the relation Eq. (4):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

- **TP (True Positives):** Correctly detected vertebrae.
- **TN (True Negatives):** Correctly identified background (empty space).
- **FP (False Positives):** Model says there is a vertebra where there isn't one.
- **FN (False Negatives):** Model misses a vertebra that is actually there.

3.1.4. Environment Experimental

Our model was implemented in Python, using TensorFlow, which is an image-processing library. First, in the stage of preparing the spinal data, we used Jupyter, and to divide the database, we used Spyder in Anaconda, and this was on a laptop equipped with a fast Intel(R) Core (TM) i7-6500U processor. 2.50 GHz, 2601 MHz core, 4 logical processors, and RAM. 8.00 GB installed. The U-Net was trained on a Colab platform with a type of GPU processor.

4. RESULT AND DISCUSSION

Different deep-learning methods were employed with an identical dataset; The lumbar spine was located using an object identification program. The data was then annotated using the LabelMe Python module and saved in the YOLO format. The localization of vertebrae has an extremely high degree of confidence. To remove boxes with low scores, an empirical thresholding method was used by setting the confidence score to 0.8.

4.1 YOLOv8 Localization Results

YOLOv8 achieved a mean Average Precision (mAP) of 99.1% for vertebral localization in our experiments. This metric signifies the precision and recall of the model in accurately detecting vertebral landmarks across the dataset. The high mAP underscores the proficiency of YOLOv8 in precisely localizing vertebral structures, a critical step for subsequent segmentation and diagnostic analysis.

Comparative analysis with other state-of-the-art models in vertebral localization, as depicted in Table 3, further highlighted the performance of YOLOv8. While previous models have demonstrated commendable performance, YOLOv8 surpassed them with high accuracy

and robust performance. This comparison underscores the advancements achieved with YOLOv8 and highlights its potential for real-world applications in clinical settings.

The significance of achieving a superior mAP of 99.1% with the proposed YOLOv8 model cannot be overstated. This level of accuracy holds immense promise for enhancing clinical workflows in spinal imaging, enabling clinicians to pinpoint vertebral landmarks with minimal manual intervention. By automating the localization process with such high accuracy, YOLOv8 streamlines diagnostic workflows, reduces interpretation time, and enhances diagnostic confidence (Fig. 8).

Table 3. Comparison of YOLOv8 with various models for vertebral localization in terms of mean Average Precision (mAP).

Year	Authors	Category	Approach	mAP [%]
2022	Fatima, Joddat, et al. [27]	MRI	YOLOv5	97.5
2023	Shaaf, Zakarya Farea, et al. [28]	MRI	Faster R-CNN	97.0
2023	Sezer, Aysun [29]	MRI	YOLOv8(with augmentation)	98.9
2025	Iqbal, Rashid, et al. [30]	MRI	RetinaNet	94.25
2025	Nimmagadda, Ramya, and P. [31]	MRI	YOLOv7	98.0
2025	This work (Proposed)	MRI	YOLOv8m	**99.1**

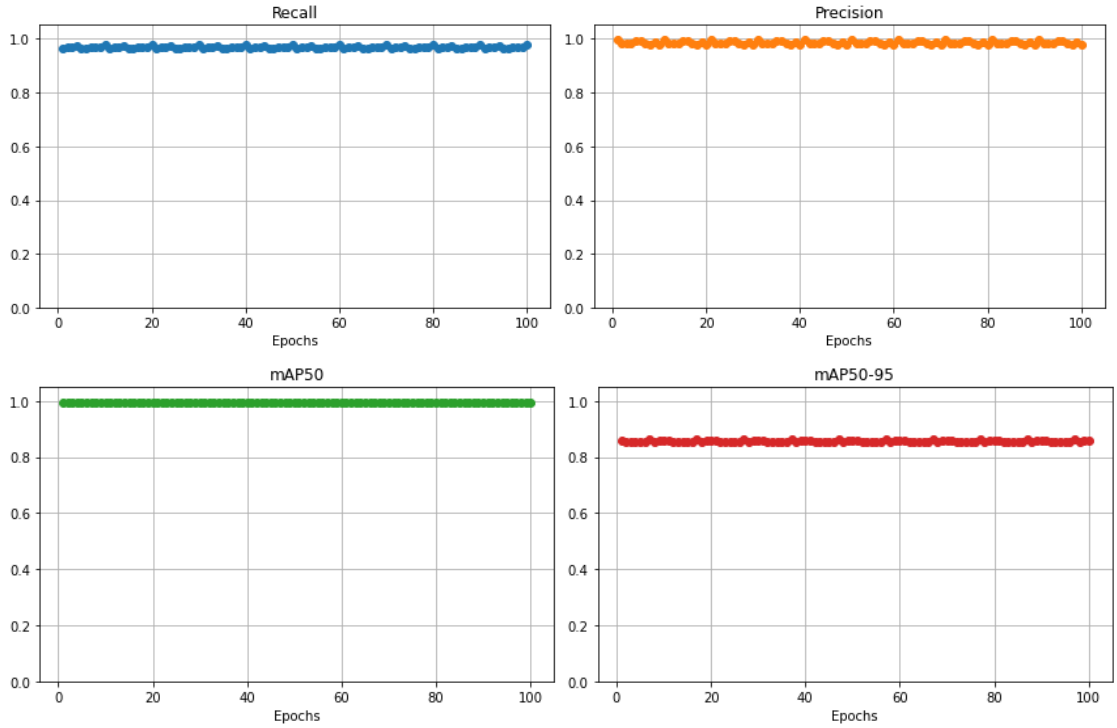


Fig. 8. YOLOv8 recall, precision, and mAP values during training, illustrating the model's learning progression and performance at different IOU thresholds.

The visual representation in Fig. 9 provides additional insights into the performance of YOLOv8 during training. It illustrates the evolution of recall, precision, and mAP values across different epochs, offering a comprehensive view of the model's learning trajectory and convergence. Additionally, it showcases the mAP values at varying Intersections over Union (IOU) thresholds, providing a detailed analysis of model performance under different evaluation criteria. Moreover, the reliability and consistency offered by YOLOv8 pave the way for improved patient outcomes and treatment planning in spinal care. Thus, the exceptional

performance of YOLOv8 in vertebral localization heralds a new era in computer-aided diagnosis and medical image analysis for spinal pathologies.

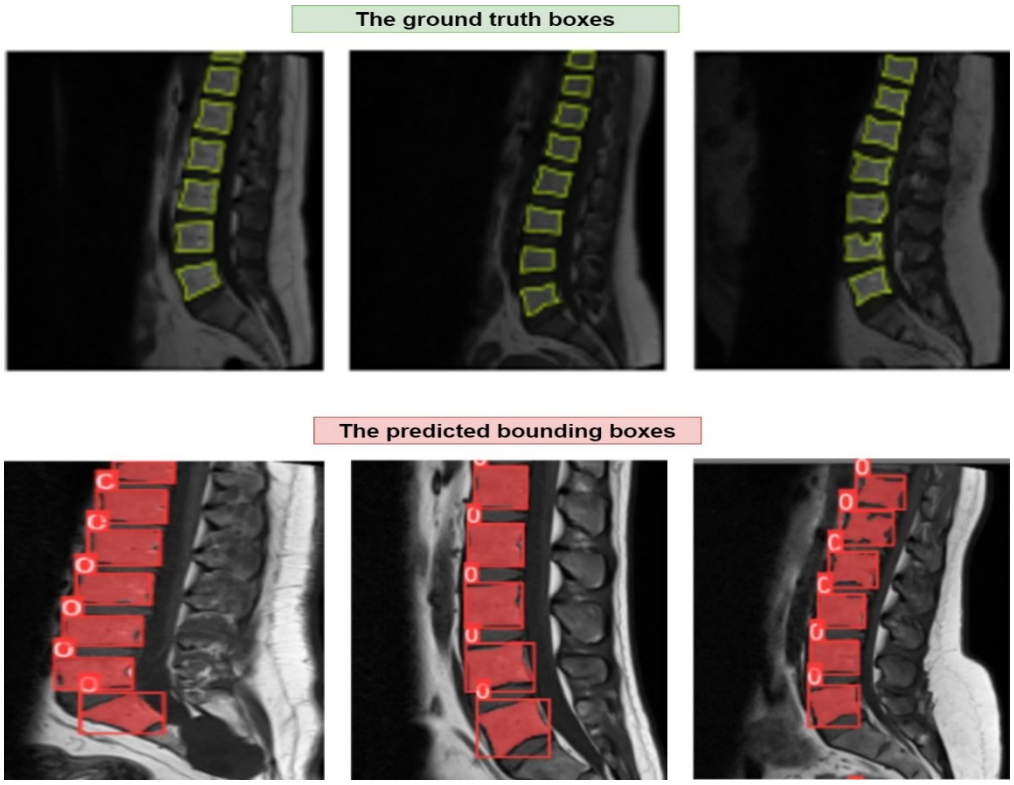


Fig. 9. (A) The ground truth boxes, represented by yellow bounding boxes, are provided by Roboflow for the annotation of spine vertebrae. (B) The predicted bounding boxes are generated using YOLOv8, depicted in red.

4.2 Data Augmentation

The augmentation of the dataset, resulting in a total of 8,900 spinal vertebrae samples, was a pivotal strategy stemming from the localization phase to enhance the dataset's diversity and prevent overfitting during the segmentation phase. Overfitting occurs when a model learns to memorize the training data rather than generalize from it, leading to poor performance on unseen data. By expanding the dataset through augmentation, we introduced variations in the data while preserving the underlying structure of the vertebrae, thereby enabling the segmentation model to learn more robust and generalizable features.

These augmentation techniques included rotations by 90° clockwise, 90° counterclockwise, and 180° upside down, along with rotation combined with translation (+45°). These transformations introduced variations in the orientation and position of the vertebrae, simulating real-world variations in patient positioning and imaging conditions. Additionally, these augmented samples help the segmentation model learn to recognize vertebrae from different perspectives, enhancing its ability to accurately segment vertebrae in unseen images.

The samples resulting from data augmentation are depicted in Fig. 10, showcasing the effectiveness of the augmentation techniques in diversifying the dataset. By augmenting the dataset in this manner, we not only expanded the dataset's size but also enriched its variability, leading to a more robust and generalizable segmentation model. This augmentation process plays a crucial role in ensuring the segmentation model's effectiveness and reliability in

clinical applications, where accurate vertebrae segmentation is essential for diagnostic and treatment purposes.

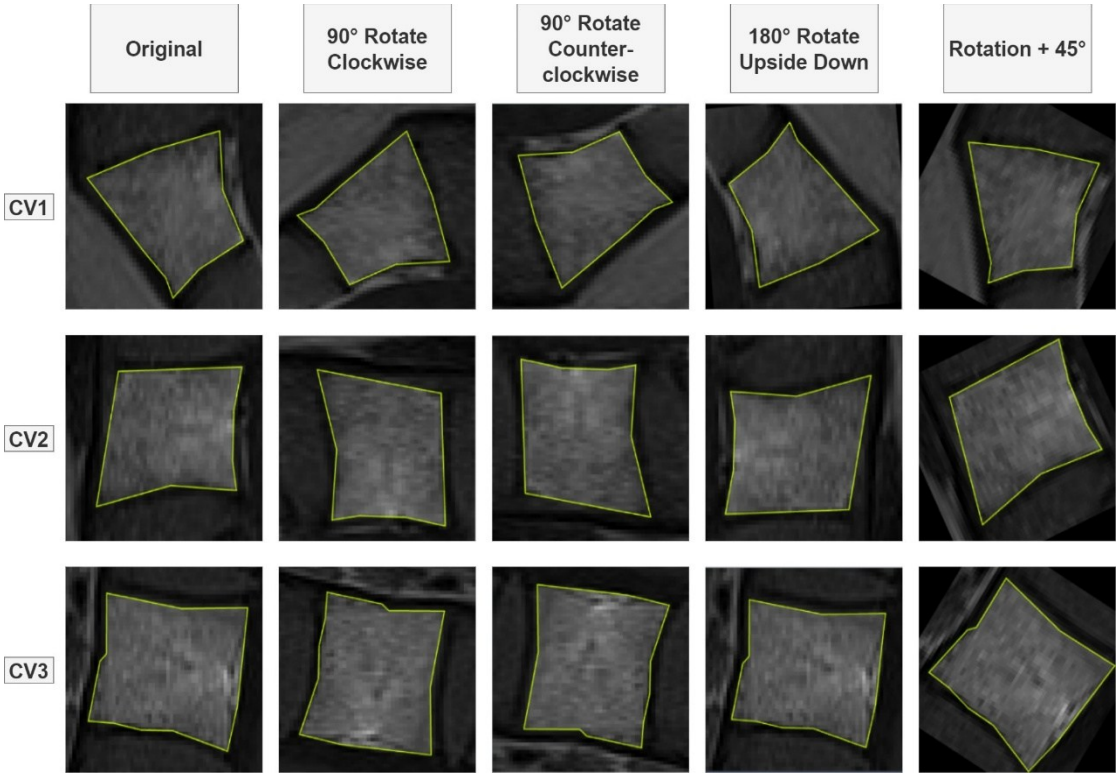
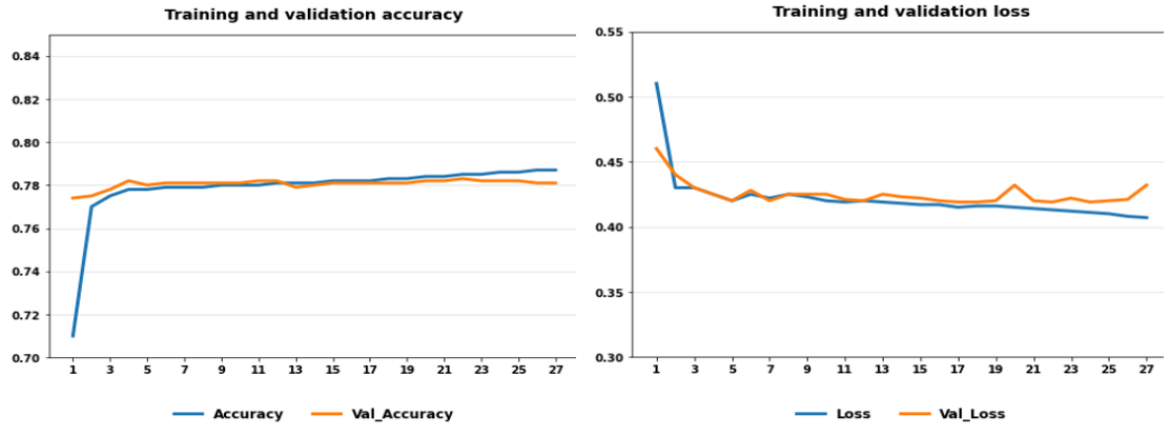


Fig. 10. Samples of spinal vertebrae after applying four data augmentation techniques: 90° clockwise, 90° counterclockwise, 180° upside down, and rotation +45°.

4.3 Segmentation Experiments

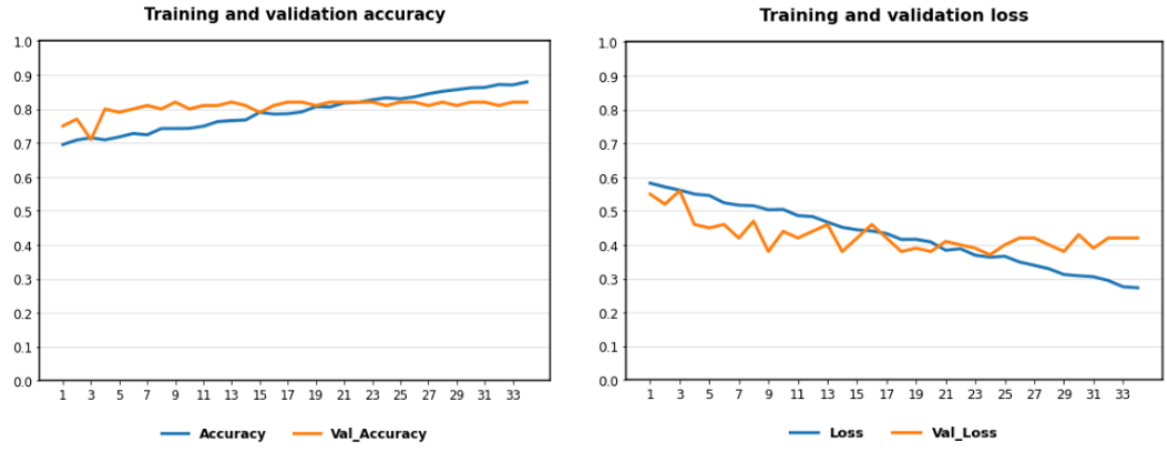
The segmentation experiments aimed to refine the localization results obtained earlier by accurately delineating spinal vertebrae. Initially, the U-Net model was trained on the augmented dataset, resulting in an accuracy of 79% (Fig. 11a) and a loss of 0.41% (Fig. 11b). Despite achieving satisfactory results, the U-Net model's performance was hindered by the loss of structural information during convolution and undersampling, leading to suboptimal segmentation outcomes.



(a) (b)

Fig. 11. a) Precision and validation accuracy of the model utilizing the u-net method; b) training and validation loss of the model.

To address this limitation and improve segmentation accuracy, the watershed method was integrated into the workflow. This addition yielded promising results, with the combined U-Net and watershed approach achieving an accuracy of 90% (Fig. 12a) and a loss of 0.35% (Fig. 12b). Notably, the watershed method facilitated the removal of extraneous regions, enhancing the precision of vertebrae segmentation.



(a) (b)

Fig. 12. a) Precision of the Model During Training and Validation accuracy by way of U-Net + watershed methods; b) Loss of the model during training and validation.

The comparative analysis presented in Table 4 highlights the performance of various segmentation methods, including the proposed modified U-Net approach and the watershed + U-Net approach. These methods were evaluated based on their accuracy in segmenting spinal vertebrae, with the proposed watershed + U-Net method outperforming other techniques with an accuracy of 90%.

Table 4. Comparison of results of different segmentation methods.

Year	Category	Approach	Accuracy
2021	CT scans	Convolutional Neural Network (CNN) [17]	89.17%
2022	MR	YOLOV5-HED-UNET [1]	74.5%
2023	CT scans	(CNN) type and the Graph Convolutional Network (GCN) [19]	88.3%
2023	MR	U-Net++ [12]	88%
Proposed	MR	U-Net Modified	79%
Proposed	MR	Watershed + U-net Method	90%

Furthermore, Fig. 13 provides visual examples of spinal vertebrae and corresponding masks obtained using the watershed method. These illustrations demonstrate the effectiveness of the segmentation approach in accurately delineating individual vertebrae from complex spinal images. The training and validation performance of the U-Net and combined watershed + U-Net models are depicted in Fig. 12, showcasing the precision, accuracy, and loss trends throughout the training process. These visualizations offer insights into the model's learning dynamics and highlight the convergence towards optimal segmentation outcomes.

Overall, the segmentation experiments demonstrate the efficacy of the proposed methodologies in accurately segmenting spinal vertebrae from medical imaging data. By leveraging advanced deep learning techniques and integrating innovative approaches such as

the watershed method, significant advancements have been made in automating and enhancing spinal image analysis for diagnostic and therapeutic purposes.

4.4. Discussion

The findings from the experimentation conducted in this study hold significant implications for the field of spinal image analysis... In the context of vertebral localization, the utilization of YOLOv8 achieved a mean Average Precision (mAP) of 99.1%, which was higher than the values obtained by the other evaluated models. This high level of accuracy underscores the potential of YOLOv8 to streamline diagnostic workflows, reduce interpretation time, and enhance diagnostic confidence in clinical settings. Additionally, comparative analysis with other contemporary methods highlights the superiority of YOLOv8, further emphasizing its significance for real-world applications in spinal imaging. This exceptionally high mAP is attributed to the highly predictable and sequential anatomical structure of the lumbar spine in sagittal views, which facilitates bounding box detection by the YOLOv8 model.

The segmentation experiments, employing U-Net and the watershed method, yielded promising results in accurately delineating spinal vertebrae from medical images. While U-Net initially demonstrated satisfactory performance, the integration of the watershed method led to significant improvements in segmentation accuracy. Visual examples and performance metrics offer insights into the effectiveness of the segmentation approach and its potential for facilitating detailed anatomical analysis and treatment planning. It is important to note that while localization is near-perfect, the actual complexity of delineating precise vertebral boundaries is more accurately reflected by our 90% segmentation accuracy, which provides a more realistic measure of the framework's performance on detailed structures.

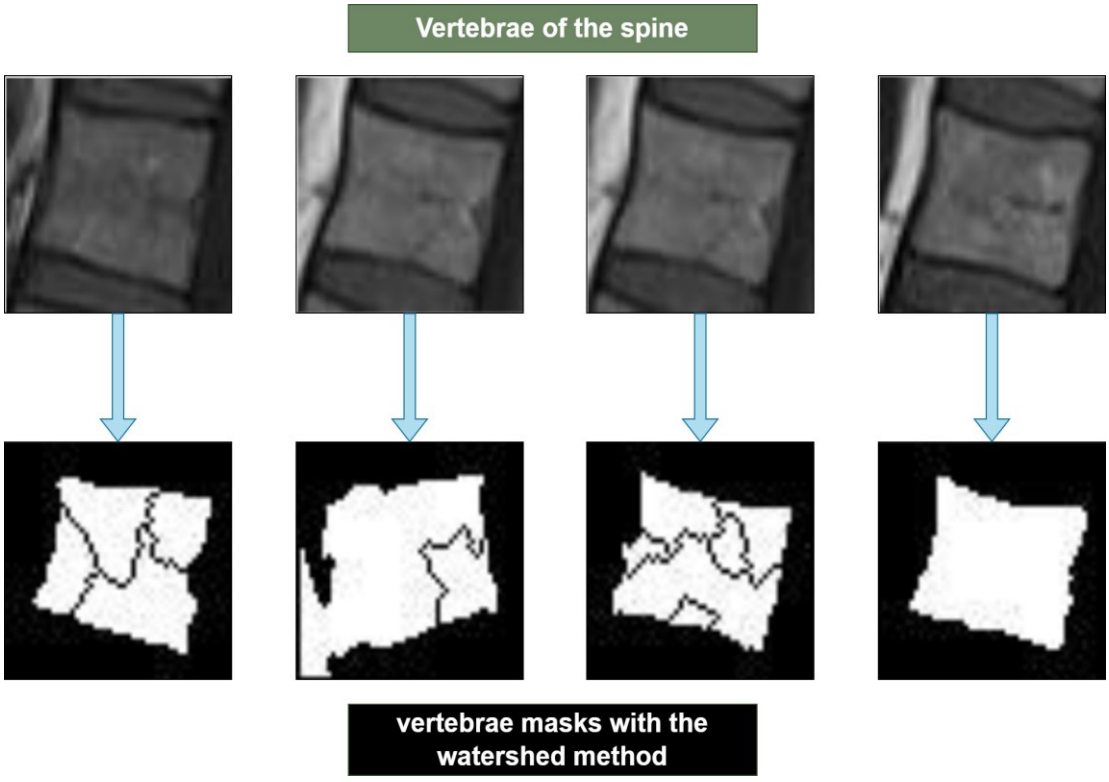


Fig. 13. Some examples of spinal vertebrae and masks obtained by the watershed method, a) vertebrae of the spine and; b) vertebrae masks with the watershed method.

5. Conclusion

Accurate segmentation of spinal vertebrae is paramount for facilitating the swift and precise detection of various spinal pathologies, including scoliosis and fractures. In this research, we have introduced a comprehensive methodology aimed at localizing and segmenting spinal vertebrae. Firstly, by harnessing the power of the YOLOv8 method, we achieved accurate vertebral localization in our experiments, laying a strong foundation for subsequent segmentation tasks. Secondly, through the strategic utilization of the watershed method in conjunction with U-Net segmentation, we have successfully obtained accurate masks for vertebrae segmentation, indicating the performance of our approach. Our study demonstrates promising performance in both localization and segmentation tasks, underscoring its potential to enhance surgical treatment outcomes and facilitate early disease detection and classification. While the localization results are exceptionally high, we acknowledge that the limited dataset size may influence these scores; therefore, future work on larger, multicenter databases will be necessary to confirm the model's robustness and generalizability in clinical settings.

However, it is imperative to acknowledge the limitations inherent in our study, including the necessity for validation on larger and more diverse datasets to ensure generalizability and robustness. Moving forward, future research should focus on: **(i)** evaluating the proposed framework on multi-center and multi-vendor MRI datasets to assess robustness across different imaging protocols; **(ii)** extending the method to full 3D volumetric segmentation to exploit spatial context better and improve anatomical consistency; **(iii)** investigating more advanced architectures, such as attention-based networks, transformer models, or nnU-Net, as well as end-to-end joint localization–segmentation pipelines; **(iv)** exploring semi-supervised or self-supervised learning strategies to reduce reliance on extensive manual annotations; and **(v)** conducting clinical validation studies, including comparisons with expert radiologists and assessment of computational efficiency, to facilitate integration into real clinical workflows.

These directions will contribute to enhancing the clinical reliability, scalability, and practical applicability of automated spinal vertebrae segmentation systems in future computer-aided diagnosis tools.

Data Availability: The dataset [32] is openly available online and can be accessed through the link: https://data.mendeley.com/datasets/k3b363f3vz/2.

Declarations: The authors declare that they have no conflict of interest.

REFERENCES

[1] M. Mushtaq *et al.*, "Localization and edge-based segmentation of lumbar spine vertebrae to identify deformities using deep learning models," *Sensors*, vol. 22, no. 4, p. 1547, 2022, doi: 10.3390/s22041547.

[2] N. Lessmann, B. Ginneken, I. Isgum, "Iterative convolutional neural networks for automatic vertebra identification and segmentation in CT images," SPIE Medical Imaging 2018: Image Processing, 2018, doi: 10.1117/12.2293416.

[3] R. Bolton *et al.*, "Effects of mattress support on sleeping position and low-back pain," *Sleep Science and Practice*, vol. 6, no. 1, 2022, doi: 10.1186/s41606-022-00073-x.

[4] M. Saeed *et al*., "An automated deep learning approach for spine segmentation and vertebrae recognition using CT images," *Diagnostics*, vol. 13, no. 16, p. 2658, 2023, doi: 10.3390/diagnostics13162658.

[5] S. Chouhan, M. Bhattacharya, G. Sharma, "HD-CAE: hybrid encoding and deformable decoding autoencoder with cascade attention for visual anomaly detection," *Signal, Image and Video Processing*, vol. 19, 2025, doi: 10.1007/s11760-024-03456-y.

[6] A. Suzani *et al*., "Fast automatic vertebrae detection and localization in pathological CT scans: a deep learning approach," MICCAI Workshop, 2015, doi: 10.1007/978-3-319-24553-9_82.

[7] C. Chu *et al*., "Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method," *PLOS ONE*, vol. 10, no. 11, e0143327, 2015, doi: 10.1371/journal.pone.0143327.

[8] S. Arif, K. Knapp, G. Slabaugh, "Shape-aware deep convolutional neural network for vertebrae segmentation," MICCAI Workshop, 2018, doi: 10.1007/978-3-030-11166-3_2.

[9] C. Payer *et al*., "Coarse-to-fine vertebrae localization and segmentation with spatial configuration-net and U-Net," VISIGRAPP, 2020, doi: 10.5220/0008975201240133.

[10] C. Fiedler, S. Köhlbig, "Neural network-based localization of spherical MRI markers," *Signal, Image and Video Processing*, 2025, doi: 10.1007/s11760-024-03445-1.

[11] Z. Wang, P. Xiao, H. Tan, "Spinal MRI segmentation based on U-Net," *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 3, p. 100627, 2023, doi: 10.1016/j.jrras.2023.100627.

[12] Y. Zhang *et al*., "Brachial plexus nerve trunk segmentation using deep learning and diffusion MRI," *Neurocomputing*, vol. 548, p. 127413, 2024, doi: 10.1016/j.neucom.2024.01.127.

[13] I. Ayed *et al*., "Vertebral body segmentation in MRI via convex relaxation and distribution matching," MICCAI, 2012, doi: 10.1007/978-3-642-33418-4_64.

[14] D. Zukić *et al*., "Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images," *Computer Graphics Forum*, vol. 33, 2014, doi: 10.1111/cgf.12376.

[15] G. Hille, S. Glaßer, K. Tönnies, "Hybrid level-sets for vertebral body segmentation in clinical spine MRI," *Procedia Computer Science*, vol. 90, 2016, doi: 10.1016/j.procs.2016.07.005.

[16] N. Altini *et al*., "Segmentation and identification of vertebrae in CT scan using CNN, k-means clustering, and k-NN," *Informatics*, vol. 8, no. 2, p. 30, 2021, doi: 10.3390/informatics8020030.

[17] R. Janssens *et al*., "Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks," ISBI, 2018, doi: 10.1109/ISBI.2018.8363714.

[18] G. Klein *et al*., "VertDetect: Fully end-to-end 3D vertebral instance segmentation model," *arXiv preprint arXiv:2311.09958*, 2023, doi: 10.48550/arXiv.2311.09958.

[19] S. Li *et al*., "Identification of natural compounds with antiviral activities against SARS-associated coronavirus," *Antiviral Research*, vol. 67, no. 1, pp. 18–23, 2005, doi: 10.1016/j.antiviral.2005.02.007.

[20] H. Möller *et al*., "SpinePS—automatic whole spine segmentation of T2-weighted MR images," *arXiv preprint arXiv:2402.16368*, 2024, doi: 10.48550/arXiv.2402.16368.

[21] I. D. Kawathekar, A. S. Areeckal, and V. Aparna, "A novel deep learning pipeline for vertebra labeling and segmentation of spinal CT images," *Biomedical Signal Processing and Control Journal*, vol. 92, p. 106037, 2024, doi: 10.1016/j.bspc.2024.106037.

[22] B. Qu *et al*., "Current development and prospects of deep learning in spine image analysis: a literature review," *Quantitative Imaging in Medicine and Surgery*, vol. 12, no. 6, 2022, doi: 10.21037/qims-21-1244.

[23] P. Cheng *et al*., "Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net," *Science Reports*, vol. 11, p. 22156, 2021, doi: 10.1038/s41598-021-01635-x.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014, doi: 10.48550/arXiv.1412.6980.

[25] K. Boyd, K. Eng, C. Page, "Area under the precision-recall curve: point estimates and confidence intervals," ECML PKDD, 2013, doi: 10.1007/978-3-642-40994-3_29.

[26] J. Fatima *et al*., "Automatic localization and segmentation of vertebrae for Cobb estimation," *Intelligent Automation and Soft Computing*, vol. 34, no. 3, 2022, doi: 10.32604/iasc.2022.025816.

[27] Z. Shaaf *et al*., "Detection of left ventricular cavity from cardiac MRI images using Faster R-CNN," *Computers, Materials and Continua*, vol. 75, no. 1, 2023, doi: 10.32604/cmc.2023.033621.

[28] A. Sezer, "Anatomic region detection in MRI using CycleGAN-augmented dataset with YOLOv8," *Eurasian Journal of Medical Investigation*, vol. 8, no. 4, 2024, doi: 10.14744/ejmi.2024.18432.

[29] R. Iqbal *et al*., "Brain-RetinaNet: Detection of brain tumor using improved RetinaNet in MRI," *CAAI Transactions on Intelligence Technology*., 2025, doi: 10.1049/cit2.12402.

[30] R. Nimmagadda, P. Devi, "A deep learning approach for brain tumor classification using YOLOv7," *Frontiers in Oncology*, vol. 15, p. 1508326, 2025, doi: 10.3389/fonc.2025.1508326.

[31] R. F. Masood *et al*., "Deep learning-based vertebral body segmentation and spinal disease classification," *Biomedical Signal Processing Control*, vol. 71, p. 103230, 2022, doi: 10.1016/j.bspc.2021.103230.