MRI LUMBAR SPINE SEMANTIC SEGMENTATION USING YOLOv8

Fabio Rodrigo Fernandes de Oliveira
Control and Automation Engineering
Instituto Federal do Espírito Santo
Serra, Brazil
eng.fabiorfo@gmail.com

Luiz Alberto Pinto
Control and Automation Engineering
Instituto Federal do Espírito Santo
Serra, Brazil
luiz.pt@ifes.edu.br

Gustavo Maia de Almeida
Control and Automation Engineering
Instituto Federal do Espírito Santo
Serra, Brazil
gmaia@ifes.edu.br

Daniel Cruz Cavalieri
Control and Automation Engineering
Instituto Federal do Espírito Santo
Serra, Brazil
daniel.cavalieri@ifes.edu.br

***Abstract*—The use of Artificial Intelligence (AI) as an assistant for diagnosis in imaging exams has already proven to be effective, and is known as Computer-Aided Diagnosis (CAD). This paper evaluates the effectiveness of using a single network, YOLOv8x is the current state-of-the-art in the YOLO family, for lumbar spine detection and segmentation in Magnetic Resonance Imaging (MRI) exams. The network was used for detection, classification, and semantic segmentation, generating the masks over the vertebrae, which simplified the implementation and reduced the computational cost. Encouraging results were obtained using a dataset of 1,116 samples (images). The detection step achieved a mean average precision (mAP) of 0.989 at 50% intersection over union (IoU), mAP:50-95 of 0.886, recall of 0.98, and precision of 0.97. For bounding box marking, the following results were achieved: mAP of 0.978 at 50% IoU, mAP:50-95 of 0.882, recall of 0.971, and precision of 0.948. The semantic segmentation step achieved a mAP of 0.978 at 50% IoU, mAP:50-95 of 0.856, recall of 0.971, and precision of 0.948. These results demonstrate the effectiveness of using YOLOv8x for lumbar spine detection and segmentation in MRI exams.**

***Index Terms*—CAD, Semantic Segmentation, DeepLearning, YOLOv8x, Lumbar Spine, MRI.**

I. Introduction

Low back pain is the leading cause of disability worldwide, according to the World Health Organization (WHO) [1]. It is estimated that up to 80% of people will experience low back pain at some point in their lives. The most common of these pathologies include herniated discs, spinal stenosis, spondylolisthesis, vertebral compression, osteoarthritis, and others. These conditions can be caused by traumatic injuries, underlying medical conditions such as osteoporosis and arthritis, as well as genetic and lifestyle factors. These spine pathologies and fractures are common in individuals of all ages. Prevention and early diagnosis of these pathologies are essential to reduce the pain and disability associated with them. For this, imaging exams are commonly used, such as X-rays, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). According to the WHO, in 2010, there were 9 million osteoporotic fractures worldwide, of which 1.4 million were hip fractures [2].

Authors such as [3], [4], [5] highlight the importance of using artificial intelligence techniques for diagnosis, known as CAD - Computer-Aided Diagnosis. The main purpose of CAD is to assist the medical specialist as a second opinion, reducing false positive diagnoses. This can increase process efficiency and reduce the time to confirm a diagnosis, which can lead to less aggressive treatment and lower emotional and financial costs.

The implementation of machine learning in the medical field has significant advantages, as highlighted by [6]. These advantages include the ability to gather a large volume of information about a particular disease in a single digital tool, suppressing the human bias, and speeding up the analysis of the material to be studied. This approach can lead to safe and fast diagnoses, and can even be used as a ”second expert opinion” in more complex cases.

In the context of neural networks, the choice of using a YOLO network is justified by its simplicity of implementation, low computational cost compared to other algorithms, and ease of onboarding such a solution, as highlighted by [7]. Using YOLOv8, the current state-of-the-art in the YOLO family, for CAD in spine pathologies and fractures based on MRI is a novel approach that may offer specific advantages over previous methods. However, there are likely research gaps and challenges that lead to the use of YOLO as a solution in this context:

- MRI-based CAD for Spine Pathologies and Fractures: while CAD systems have been developed for various medical imaging tasks, the specific application of CAD to spine pathologies and fractures using MRI may not have been extensively explored or have fewer dedicated solutions. Previous research might be more focused on other imaging modalities, such as CT scans or X-rays [8].
- Object Detection in MRI: MRI images can present different challenges compared to other modalities. The appearance of anatomical structures and abnormalities in MRI can vary significantly due to differences in tissue contrast and resolution [9].

- Multi-class Object Detection: YOLO is capable of multi-class object detection, allowing it to identify and localize various abnormalities and structures in MRI images relevant to spine pathologies and fractures. This is important as spine pathologies can involve different types of abnormalities and fractures in various regions.
- Spatial Context: YOLO considers the entire image at once, allowing it to capture spatial context and dependencies between different regions of the spine. This is beneficial for accurate detection and localization of abnormalities, especially in the context of spine pathologies.
- Real-time Performance: YOLO's architecture is designed for real-time object detection tasks, and this capability can be advantageous for providing timely second opinions to medical specialists during clinical assessments.
- Transfer Learning: YOLO can leverage transfer learning, which means it can be pre-trained on large datasets from other domains (e.g., natural images) and then fine-tuned on a smaller MRI dataset for object detection. This can be valuable when medical datasets are limited, as is often the case due to privacy concerns and data availability.
- Generalization: YOLOv8 has improved generalization capabilities, meaning it can perform well on unseen data and potentially adapt to different MRI protocols and acquisition settings.
- Automation: By integrating YOLO into a CAD system, the detection process can be automated, reducing the burden on medical specialists and potentially increasing efficiency in diagnosing spine pathologies and fractures.

Several techniques are currently used in CAD, either isolated or combined, with the purpose of making all the processes as automated and efficient as possible, as listed in Section 3: Related Work.

The application of deep learning in lumbar spine MRI images is proposed by this paper. Its main objective is to evaluate the YOLOv8 network and its performance on the same dataset for detection and semantic segmentation, generating masks for each vertebra to be used in a subsequent step for image detection, which will be addressed in future work.

This paper is divided as follows: Section 2 describes some basic concepts necessary for understanding the related work, which is covered in Section 3. The methodology used in the development of this paper is presented in Section 4, and its results are discussed in Section 5. Suggestions for future work are presented, and Section 6 presents the conclusions reached by this paper.

II. Concepts

This section introduces the key concepts used in this paper. First, a brief overview of the spine is given, outlining the main imaging studies used. Finally, an overview of YOLO networks is given, with a focus on YOLOv8.

A. Spine, Pathologies and Examinations

Figure 1 presents 5 spine regions, with its labels [10]:

- Cervical Spine: 7 vertebrae in the neck region, from C1-C7;
- Thoracic Spine: 12 vertebrae in the thorax region, from T1 to T12;
- Lumbar spine: 5 vertebrae in the final back portion, from L1 to L5;
- Sacrum: 5 vertebrae in the hip region, from S1 to S5;
- Coccix: End portion of the spine;

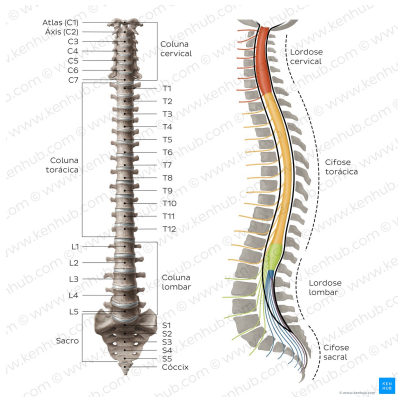

Fig. 1. Basic spine regions [10]

Several pathologies can affect the spine in all its extension, such as herniated discs, scoliosis, stenosis, tumors, infections and inflammations, and vertebral fractures [11]. For diagnosis, the imaging exams that provide the most detail are CT and MRI, as shown in figure 2. CT uses X-rays to create detailed images of the bones and tissues in the body. MRI does not use X-rays, but instead uses a strong magnetic field and radio waves to produce images of the soft tissues in the body. The main difference between CT and MRI is that CT is better at visualizing bones and other calcified tissues, while MRI is better at visualizing soft tissues, such as the spinal cord and nerves, according to [12].

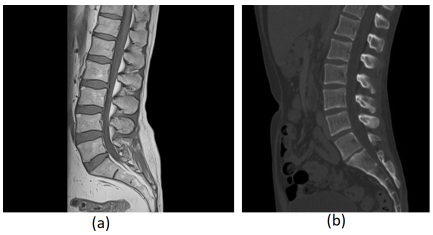

Fig. 2. Examination images: a) MRI b) CT

MRI can also detect other fractures, such as spondylolysis and spondylolisthesis. Spondylolysis is a fracture at the back of the vertebra, while spondylolisthesis is a condition in which a vertebra slides forward of the underlying vertebra. Both conditions can cause back pain, loss of balance, and muscle weakness.

B. YOLO network

Convolutional Neural Networks (CNNs) are widely used in computer vision applications, including object detection in images. One of the most popular architectures for object detection is known as You Only Look Once (YOLO). YOLO

networks are known for their ability to perform real-time detection and provide accurate results for different types of objects, even on images with many objects in the scene. Since its first release in 2015, YOLO has continuously evolved, with versions YOLOv2 and YOLOv3 providing featuring significant improvements in accuracy and speed [13].

YOLOv8 is the current state-of-art for object detection and image segmentation, created by Ultralytics. Like YOLOv5, it is divided into 3 components: backbone, neck and head. There are 5 versions: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x, highlighting the difference in the amount of trained parameters [14]. A structure diagram in YOLOv8 and a better explanation can be found in [14].

YOLOv8 uses the EfficientNet backbone for efficient and accurate object detection. The backbone consists of 10 convolutional layers, followed by a pooling layer and an object detection layer. The pooling layer reduces the size of the feature map, making it easier for the object detection layer to find objects in the image. The object detection layer detects objects and provides location and size information [15].

III. Related works

The field of diagnostic medicine is a constant target of studies for deep learning applications, especially for imaging exams such as MRI and CT. Accurate segmentation of vertebrae in these images is a fundamental step for diagnosis, especially as an automated solution. The large variation in spinal anatomy between patients and the scarcity of available samples for many of them make this task challenging.

In their paper, Kuang et al. [16] address the challenge of classifying spinal diseases with a small training dataset. They propose a supervised hybrid generation model called SpineGEM, which is divided into two steps: first, generating initial models based on a large-scale trained CNN; and then, fitting these initial models based on the limited training data available. The resulting models are then combined into an ensemble to improve classification accuracy.

The results show that SpineGEM outperforms other spine classification methods in terms of classification accuracy, even with a very small training dataset. An accuracy of 87.6% was achieved on the spinal pathology classification task using only 58 training images. These results suggest that the SpineGEM approach can be an effective solution for spinal disease classification with a limited training dataset. This may be especially relevant in clinical settings where image availability is limited.

Candemir et al. [17] address strategies for model training in scenarios with limited data, insufficiently labeled data, and/or limited expert resources. They discuss strategies for expanding the data sample, reducing the time required for of manual supervised labeling, adjusting the architecture of the neural network to improve model performance, applying semi-supervised approaches, and leveraging the efficiency of pre-trained models. They argue that patient privacy, tedious annotation processes, and the limited availability of radiologists pose challenges to building such datasets.

A two-stage deep learning approach was proposed by [18], for the automatic localization and segmentation of vertebrae in CT images of the spine. The first stage used a 2D-Dense-U-Net network to localize the vertebrae and detect their centroids. The second stage used a 3D-Dense-U-Net network to segment each specific vertebra into a region of interest, based on the centroid detected in the first stage. The approach achieved high accuracy, with a 100% detection rate and a Dice coefficient of 0.877 ± 0.035.

A CNN based on the VGG19 architecture was used by [19] to detect lumbar disc herniations on MRI images. The results were promising, an were successfully tested on more than 200 patients, achieving 100% accuracy and respective intervertebral disc (IVD) identification accuracies for L1-L2, L2-L3, L3-L4, L4-L5, and L5-S1 IVDs of 92.7%, 84.4%, 92.1%, 90.4%, and 84.2%, respectively.

Wang et al. [20] proposed an improved Attention U-Net algorithm for lumbar spine image segmentation. The algorithm uses attention-based fusion and residual modules to improve performance. It was tested on 1000 MRI images and achieved better performance than other algorithms in terms of accuracy, recall, and Dice similarity coefficient.

A Fully Convolutional Network (FCN) was proposed by [21] to segment and label vertebrae. The FCN was combined with a memory component to store information about previously segmented vertebrae. The network then searched for nearby vertebrae and predicted whether they were visible enough to be segmented. The method was evaluated on five different datasets and achieved an average Dice score of 94.9%, an anatomical identification accuracy of 93%, and a visibility identification accuracy of 97%. The method is fast, flexible, generalizable, and compares favorably with state-of-the-art methods.

Yolov5 as an object detector and the HED U-NET as segmentation model to segment vertebrae was used by Mushtaq et. al [22],deep learning models to localize and segment lumbar spine vertebrae in MRI images. The results showed that the proposed method was able to accurately localize and segment the vertebrae, with a mean average precision (mAP) of 0.975 for localization and a Dice coefficient of 0.94 for segmentation. The proposed method was also able to diagnose lumbar deformities with an accuracy of 74.5%.

Mercaldo et al. [23] applied a YOLOv8 model to automatically detect and localize brain cancer in MRI images. The model achieved a precision of 0.943 and a recall of 0.923 for brain cancer detection, and an mAP0.5 of 0.941 for brain cancer localization. These results demonstrate the effectiveness of the YOLOv8 model for brain cancer detection and localization.

At the time of writing this paper, no studies have been found on the segmentation of the lumbar spine using YOLOv8 applied to MRI images.

IV. Materials and Methods

The work described in this paper was performed on the Google Colaboratory (Colab) platform, using the paid environment to have access to more GPU availability and high memory availability. The code was developed in Python, and the pipeline is illustrated in Figure 3.

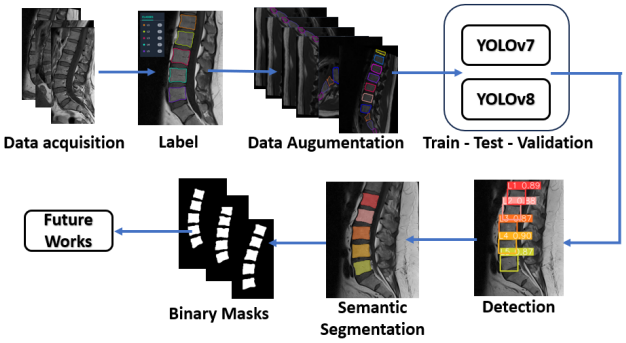

Fig. 3. Pipeline of the steps performed

A. Dataset

The choice of data is an important step, as the quality of the samples directly afects the results. For this paper, we used the Lumbar Spine Composite dataset, available on Mendeley Data [24]. This dataset contains 514 individual sagittal plane MRI images, which were derived from the MRI dataset [25]. The original dataset was obtained from anonymized clinical MRI studies of 515 patients with symptomatic back pain, with each study containing axial or sagittal slices of the three lower vertebrae and the intervertebral discs. The dataset contains a total of 48,345 MRI slices with an accuracy of 12 bit per pixel and a resolution 320x320 pixels for most slices. The dataset generated by [24] is already available for use in .png format for use. From this total, 3 samples were excluded due to noise, and the final dataset used has a total of 512 central sagittal plane images.

B. Label Annotation

Currently, there are several resources available for annotating labels, such as LabelME, LabelIMG, V7, Roboflow, and others. As recommended by Ultralytics due to the use of YOLOv8, Roboflow was used, as shown in Figure 4. Only vertebrae related to the lumbar spine, from L1 to L5, which are the object of this study, were annotated.

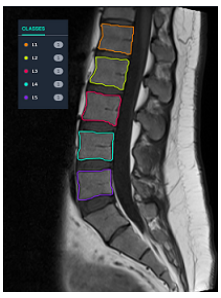

Fig. 4. Annotation process example

From the total set of 512 images, 50 were separated for testing, and only 462 images were annotated. In the first stage, without using data augmentation, 375 images were used for training (70%) and 137 images were used for validation (30%).

C. Data Augmentation

Data augmentation is a solution applied to improve algorithm training and reduce overfitting [25]. It is used when there is no access to large databases, such as in medical image analysis [17]. Data augmentation comprises a set of techniques that increase the size and quality of training datasets in order to build better deep learning models [25].

This process was performed directly in Roboflow, and the ratios were changed automatically through the following steps: the images were inverted horizontally, rotated clockwise, counterclockwise, and upside down, clipped to a minimum zoom of 0º and a maximum zoom of 20º, rotated between -23° and +23°, tilted up to 15° horizontally and vertically.

The training dataset has been expanded from 462 images to 1,116, and is automatically adjusted to 981 training images (87,9%), 103 for validation (9,2%) and 32 (2,87%)for test, the maximum allowed for the free version of Roboflow.

D. Training, Detection and Segmentation

The decision to use YOLOv8x, which produced the best results, was made after the training and validation process described below and shown in figure 5:

- 1st: Images were reset to 320x320 and trained on the YOLOv7 network;
- 2nd: Images reset to 640x640 and trained on the YOLOv7 network
- 3rd: Images from previous step, applied Data Augmentation as described below and trained on YOLOv7;
- 4th: Upon reaching a satisfactory result, the previous dataset was trained on YOLO8n;
- 5th: Detection training performed on YOLOv8x;
- 6th: Detection validation of the 50 images with the previously trained weights;
- 7th: Segmentation training on YOLOv8x;
- 8th: Validation of segmentation of the 50 images with the previously trained weights;

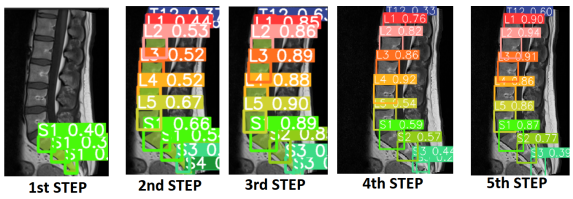

Fig. 5. Networking training steps

Due to the small number of samples, the model was pretrained on a general COCO 2017 dataset. This allowed the model to learn general object features without losing the generalization of the pretrained model. The hyperparameters were optimized using an iterative trial-and-error process.

To achieve the best results, several training runs were performed for both detection and segmentation. Adjustments were made, such as varying the batch size from 5 to 30 and the number of epochs from 50 to 300. The best results were obtained with a batch size of 20 and 300 epochs. However, the training was terminated with 169 epochs for detection

training and 108 epochs for segmentation due to early stopping, when no further improvements were observed.

The hyperparameters shown in table I were used for the task of detecting and segmentating vertebrae in lumbar spine MRI images. The use of a large batch size (20) and a large number of workers (8) helps to improve the training efficiency of the model. The use of a constant learning rate (0.01) throughout the training process helps to prevent the model from overfitting the training data. Using an intersection over union (IOU) threshold of 0.7 for evaluation is a reasonable choice, as it balances the precision and recall of the model. The use of a random seed helps to ensure that the reproducibility of the training results.

Overall, the hyperparameters in the table I are a good starting point for training a YOLOv8 model for vertebrae detection and segmentation in lumbar spine MRI images. However, it is important to note that these hyperparameters may need to be adjusted depending on the specific dataset and application.

TABLE I
YOLOv8 HYPERPARAMETERS

Hyperparameter	Train	Segmentation
Epoch	300	300
Patience	50	50
Batch size	20	20
Image Size	640	640
Workers	8	8
Optimizer	SGD	SGD
Seed	6	6
IOU	0,7	0,7
LR0	0,01	0,01
LRf	0,01	0,01

The loss function was VarifocalLoss, a robust variant of binary cross entropy loss for small objects and objects with low probability.

The results of this study demonstrate that the YOLOv8 network is a promising approach for vertebral detection and segmentation in lumbar spine MRI images. The model achieved a mAP of 0.975 for detection and a mean IOU of 0.900 for segmentation. These results are comparable to or better than those reported in other studies using similar approaches.

The YOLOv8x detrending step predicted the vertebrae, from L1 to L5, with bounding box marking, highlighting its probability for each vertebra.

Semantic segmentation was used to predict and mark the masks on the vertebrae individually on each pixel, using different colors for each vertebra. Marking individual masks is important to support other diagnostic steps, such as deviations and fractures. When necessary, this process can also be used to segment other structures. This process has been performed by other authors using other available networks, each one with its own purpose.

The last step was the binary mask extraction of each image, saving it in a specific .png file and its corresponding .json file so that it can be used in other applications, if necessary. Figure 6 shows an example of each step described.

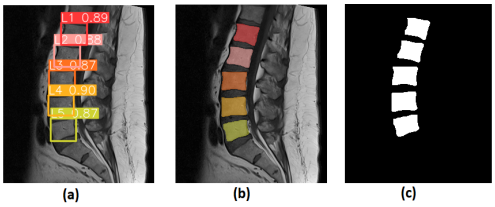

Fig. 6. Validation steps: (a) Detection with bounding box; (b) Semantic Segmentation; (c) Binary Masks

E. Evaluation Methods

Evaluating the results is important to understand the behavior of the network and makes it possible to make various adjustments, such as to its hyperparameters or even to the model itself, to improve its performance or even to replace it.

The main criteria for evaluating a neural network model are Precision, Recall and Mean Average Precision mAP, which are presented below:

- Precision: Equation 1 shows the calculation as the ratio of the total number of true positive (TP) samples to the total number of samples classified as positive (correct TP or incorrect FP). It measures the accuracy of the network in classifying a sample as positive, which reflects how reliable the model is in classifying samples as positive:

$$Precision = \frac{TP}{(TP + FP)} \tag{1}$$

- Recall: Equation 2 calculates this as the ratio of the number of true positive (TP) samples correctly classified as positive to the total number of positive samples (correctly TP or incorrect FN) and measures the ability of the model to detect positive samples. The higher the recall, the more positive samples are detected:

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

- mAP (Mean Average Precision): mAP is a metric used to evaluate the precision of an object detection model at different detection thresholds. It is usually calculated using the average mean average precision (AP) for different object categories and is the average of the precision at different recall values. The general formula for calculating AP is AP is calculated for each object category individually and is the average of the Precision (P) at different Recall (R) value:

$$AP = \int_0^1 P(R)\, dr \tag{3}$$

The mAP represents the average of these APs for all object categories considered in object detection using YOLOv8. The formula to calculate the mAP is as follows

$$mAP = \frac{1}{n} \sum_{i=1}^{N} APi \tag{4}$$

The most commonly used metric, as listed above, is mAP. It is also satisfactory, since AP represents the average Precision for each class (in this case, 5) and mAP represents the average for all classes. mAP0.5 represents a value of mAP between a fixed intersection over union of 0.5. mAP 0.5:0.95 represents the IOU values from 0.5 to 0.95. All results graphs indicate that the network performed well for training on the dataset used, together with its hyperparameters.

V. Results and Discussion

This paper explores the potential of using a single network, specifically YOLOv8x, for vertebral detection and semantic segmentation in central sagittal MRI images. Despite the use of a limited training dataset, the results obtained in this work were very promising and successfully achieved the main objective of the study. In particular, the semantic segmentation results were extremely satisfactory, indicating the possibility of using a single network to perform both steps.

The analysis of the results of the detection and semantic segmentation process will be discussed as follows.

A. Detection Results

Table II shows the results of the detection test s using 103 separated images for vertebrae from L1 to L5, including results for all classes. The training process ran for 169 epochs. It was stopped due to no improvement in the results was observed after epoch 119. The table shows the results for Bounding Box (BOX), Precision (P), Recall (R), mAP:50, and mAP:50-95.

TABLE II
YOLOv8 Detect Training Results

Metric	All Vertebrae	L1	L2	L3	L4	L5
Precision	0.97	0.946	0.971	0.981	0.971	0.98
Recall	0.98	0.971	0.987	0.958	0.985	1
mAP:50	0.989	0.981	0.99	0.983	0.995	0.994
mAP:50-95	0.886	0.877	0.909	0.898	0.883	0.865

The overall precision of 0.97 means that, on average, 97% of the vertebra instances were correctly classified. The overall recall of 0.98 means that, on average, 98% of actual vertebra instances were detected. The overall mAP of 0.989 means that, on average, vertebra instances were detected with a precision of at least 98.9%.

The results for each individual vertebra class are also good. All classes have a precision of at least 0.94 and a recall of at least 0.95. L1 has the lowest precision and recall, but it is still very good. L2, L3, and L4 have similar results. L5 has the highest precision and recall.

Figure 7 shows the YOLOv8 confusion matrix, with the accuracy results for each grid, all of which are above 0.95. The background column indicates that these instances should have been classified but were not. This allows us to better analyze the model built, and the errors found.

Figure 8 shows the F1 scores for different confidence levels. The best F1 score was 0.98, which is very high. This means that the classifier achieved 98% accuracy and 98% recall. In other words, the classifier correctly classified 98% of all instances, and it correctly classified 98% of the true positives.

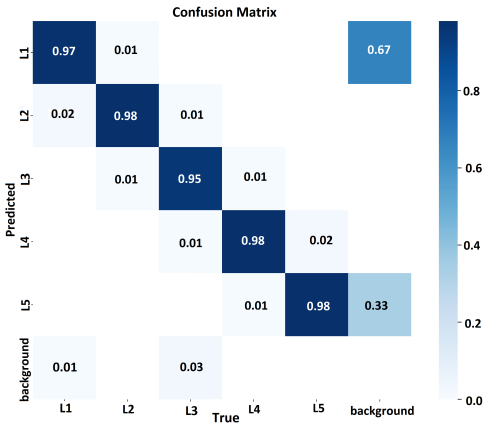

Fig. 7. Confusion matrix of the detection training model

The confidence level of 0.781 indicates that the classifier is 78.1% confident in its predictions. This is a good confidence level, but it is important to note that no classifier is perfect. There will always be some instances where the classifier misclassifies.

The results of this study suggest that the classifier is a very effective tool for detecting objects. The high F1 score and confidence level indicate that the classifier is accurate and reliable.

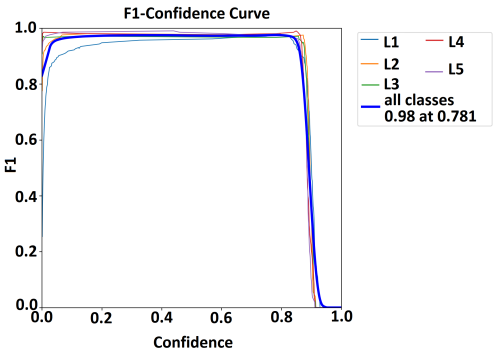

Fig. 8. F1 scores curve of the detection training model

Figure 9 shows the accuracy vs. confidence curve for all class values (L1 through L5). The curve shows that the accuracy is very high for all class values, with all values above 0.926. This is a very satisfying result, as it indicates that the classifier is very accurate in its predictions.

Figure 9 shows that the classifier achieves high accuracy and confidence for all class values. This indicates that the classifier can accurately and reliably detect objects of all types.

Overall, the results are very satisfactory, demonstrating that the classifier is well-suited for object detection tasks.

B. Semantic Segmentation Results

Table III shows the results of the semantic segmentation test for vertebrae from L1 to L5, using 103 validation images, including results for all classes. The training process ran for 108 epochs. It was stopped due to no improvement

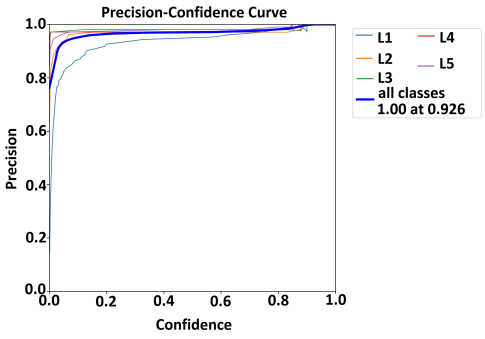

Fig. 9. Accuracy vs. Confidence curve, detection training

in the results observed after epoch 58. The best results were obtained at epoch 58. The table shows the results for Bounding Box (BOX) and Masks (MASK), Precision, Recall, mAP50 and mAP50-95.

The overall results are very good, with a precision of 94,8%, a recall of 0.971, and a mAP of 97,8% for the bounding box and 0.948, a recall of 0.971, and a mAP of 97,8% for the mask. The results for each individual vertebra class are also good, with all classes having a precision of at least 0.94 and a recall of at least 95%.

The L1 vertebra has the lowest precision and recall for both the bounding box and the mask. This is probably due to the fact that the L1 vertebra is the smallest vertebra in the lumbar spine and is therefore more difficult to detect and segment. The L2 vertebra has the highest precision and recall for the bounding box, but the lowest precision and recall for the mask. This is probably due to the fact that the L2 vertebra is located in the middle of the lumbar spine and is therefore more difficult to segment than the L1 or L5 vertebrae. The L5 vertebra has the highest precision and recall for the mask, but the lowest precision and recall for the bounding box. This is probably due to the fact that the L5 vertebra is the largest vertebra in the lumbar spine and is therefore easier to segment than the L1 or L2 vertebrae.

Figure 10 presents the confusion matrix for YOLOv8x.

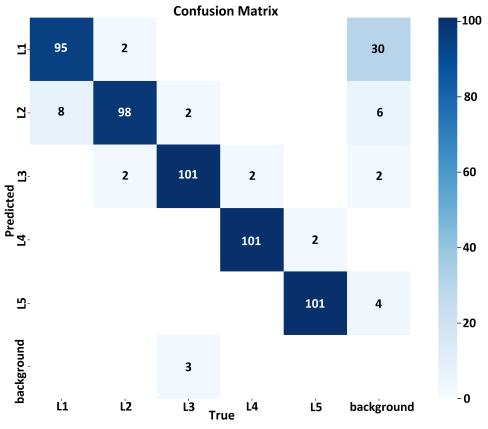

Fig. 10. Confusion matrix of the Semantic segmentation training model

Figure 11 shows the graph F1 scores for the confidence values.

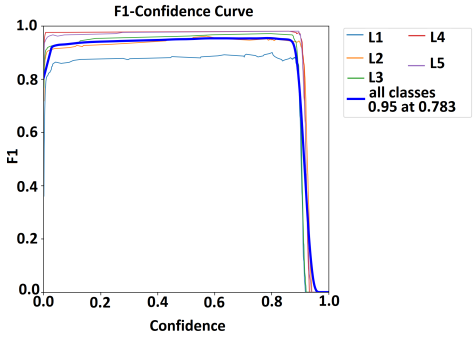

Fig. 11. F1 scores curve of the semantic segmentation training model

Figure 12 shows the curve accuracy vs confidence curve for all class values (L1 to L5).

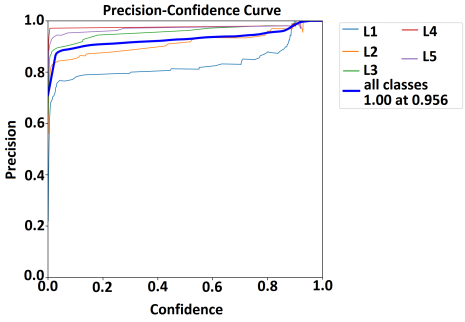

Fig. 12. Accuracy vs. Confidence curve, detection training

C. Discussion

Despite the positive results, further scrutiny of the final data is required. Some images showed mislabeling, possibly influenced by anatomical variations between patients, warranting individual analysis of the detected and segmented images.

The limitations of the study stem from a dataset with a small sample size and unbalanced data, which requires evaluation by experienced radiologists to validate the effectiveness of the process.

The study presents an innovative technical approach by integrating classification, semantic segmentation, and binary mask extraction into a single deep learning network, YOLOv8x, selected for its exceptional performance.

In order to increase the efficiency o the model, the incorporation of an image enhancement preprocessing step and an increase in the number of points in the annotation process are recommended strategies. These improvements can further increase the overall performance of the system.

VI. Conclusion

This paper presents the application of the YOLOv8 network for the detection, semantic segmentation, and binary mask extraction of lumbar spine vertebrae from L1 to L5 in MRI scans. The proposed methodology used a single, lighter, more reliable, and easier to implement network to perform both tasks. The results showed that the methodology

TABLE III
YOLOv8 SEMANTIC SEGMENTATION TRAINING RESULTS

Metric	All Vertebrae	L1	L2	L3	L4	L5
Precision (BOX)	0.948	0.875	0.942	0.971	0.969	0.981
Recall (BOX)	0.971	0.951	1	0.959	0.971	0.971
mAP:50 (BOX)	0.978	0.969	0.987	0.965	0.98	0.991
mAP:50-95 (BOX)	0.882	0.866	0.904	0.902	0.88	0.856
Precision (MASK)	0.948	0.875	0.942	0.971	0.969	0.981
Recall (MASK)	0.971	0.951	1	0.959	0.971	0.971
mAP:50 (MASK)	0.978	0.969	0.987	0.965	0.98	0.991
mAP:50-95 (MASK)	0.856	0.831	0.879	0.864	0.858	0.847

was able to achieve satisfactory performance for semantic segmentation, without the need for additional refinements. This methodology has the potential to perform the entire pathology detection process with high accuracy, which can help radiologists to make more accurate diagnoses.

As a suggestion for future work, the authors plan to include digital image processing in the initial stage of the methodology. This would include grouping the samples according to the pathologies to be treated, as well as including edge detection steps to calculate the dimensions of each vertebra. The authors also plan to apply the methodology to image classification tasks.

The authors believe that the proposed methodology has the potential to be a valuable tool for radiologists and other healthcare professionals. It can be used to automate the detection and segmentation of vertebrae in MRI scans, which can freeing radiologists' time to focus on other tasks. The methodology can also be used to improve the accuracy of diagnoses, which can lead to better patient outcomes.

REFERENCES

[1] W. H. Organization *et al.*, "Musculoskeletal conditions. 2019," Disponível em https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions Acessado em: 05/05/2023.

[2] ——, "Spinal cord injury," Disponível em https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury Acessado em: 05/05/2023.

[3] M. K. Santos, J. R. Ferreira Júnior, D. T. Wada, A. P. M. Tenório, M. H. Nogueira-Barbosa, and P. M. d. A. Marques, "Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo à medicina de precisão," *Radiologia brasileira*, vol. 52, pp. 387–396, 2019.

[4] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, "Deep learning-based cad schemes for the detection and classification of lung nodules from ct images: A survey," *Journal of X-ray Science and Technology*, vol. 28, no. 4, pp. 591–617, 2020.

[5] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, no. 1, pp. 1–13, 2016.

[6] H. Lu, M. Li, K. Yu, Y. Zhang, and L. Yu, "Lumbar spine segmentation method based on deep learning," *Journal of Applied Clinical Medical Physics*, p. e13996, 2023.

[7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[8] D. Baur, K. Kroboth, C.-E. Heyde, and A. Voelker, "Convolutional neural networks in spinal magnetic resonance imaging: A systematic review," *World Neurosurgery*, vol. 166, pp. 60–70, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1878875022009895

[9] S. Kato, T. Hozumi, K. Yamakawa, M. Saito, T. Goto, and T. Kondo, "Meta: an mri-based scoring system differentiating metastatic from osteoporotic vertebral fractures," *The Spine Journal*, vol. 15, no. 7, pp. 1563–1570, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1529943015002235

[10] *Disponível em: https://www.kenhub.com/pt/library/anatomia/coluna-vertebral-espinha*, Acessado em: 05 de maio de 2023.

[11] *Disponível em: https://www.colunar.com.br/problemas/*, Acessado em: 05 de maio de 2023.

[12] R. S. I. Carter, "Mri vs. ct scan; diagnosing spine neck injuries degenerative diseases," Disponível em https://josephspine.com/mri-vs-ct-scan-diagnosing-spine-neck-injuries-degenerative-diseases/ Acessado em: 05/05/2023.

[13] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 to yolov8 and beyond," 2023.

[14] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[15] B. A. Thompson, P. M. Frinchaboy, T. Spoo, and J. Donor, "The binary information from open clusters using seds (binocs) project: Reliable photometric mass determinations of binary star systems in clusters," *The Astronomical Journal*, vol. 161, no. 4, p. 160, 2021.

[16] X. Kuang, J. P. Y. Cheung, X. Ding, and T. Zhang, "Spinegem: A hybrid-supervised model generation strategy enabling accurate spine disease classification with a small training dataset," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*. Springer, 2021, pp. 145–154.

[17] S. Candemir, X. V. Nguyen, L. R. Folio, and L. M. Prevedello, "Training strategies for radiology deep learning models in data-limited scenarios," *Radiology: Artificial Intelligence*, vol. 3, no. 6, p. e210014, 2021.

[18] P.-C. Wu, T.-Y. Huang, and C.-J. Juan, "Automatic spine vertebra segmentation in ct images using deep learning," in *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2019, pp. 1–2.

[19] W. Mbarki, M. Bouchouicha, S. Frizzi, F. Tshibasu, L. B. Farhat, and M. Sayadi, "Lumbar spine discs classification based on deep convolutional neural networks using axial view mri," *Interdisciplinary Neurosurgery*, vol. 22, p. 100837, 2020.

[20] S. Wang, Z. Jiang, H. Yang, X. Li, and Z. Yang, "Automatic segmentation of lumbar spine mri images based on improved attention u-net." *Computational Intelligence & Neuroscience*, 2022.

[21] N. Lessmann, B. Van Ginneken, P. A. De Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical image analysis*, vol. 53, pp. 142–155, 2019.

[22] M. Mushtaq, M. U. Akram, N. S. Alghamdi, J. Fatima, and R. F. Masood, "Localization and edge-based segmentation of lumbar spine vertebrae to identify the deformities using deep learning models," *Sensors*, vol. 22, no. 4, p. 1547, 2022.

[23] F. Mercaldo, L. Brunese, F. Martinelli, A. Santone, and M. Cesarelli, "Object detection for brain cancer detection and localization," *Applied Sciences*, vol. 13, no. 16, p. 9158, 2023.

[24] R. F. Masood, T. Hassan, H. Raja, B. Hassan, J. Dias, and N. Werghi, "A composite dataset of lumbar spine images with mid-sagittal view annotations and clinically significant spinal measurements," in *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*. IEEE, 2022, pp. 1–5.

[25] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.