




# Improving Lumbar Disc Bulging Detection in MRI Spinal Imaging: A Deep Learning Approach

Mohammed A. Abed<sup>1</sup>, Z. T. Al-Qaysi<sup>1</sup>\*, M. S Suzani<sup>2</sup>

<sup>1</sup>Department of Computer Science, Computer Science, and Mathematics College, Tikrit University, Tikrit, Iraq.

<sup>2</sup>Data Intelligence and Knowledge Management Special Interest Group, University Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia.

\*Corresponding Author: Z. T. Al-Qaysi

DOI: <https://doi.org/10.55145/ajest.2025.04.01.001>

Received May 2024; Accepted July 2024; Available online August 2024

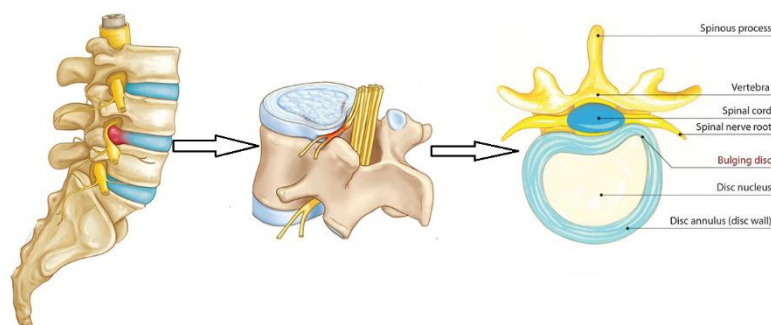
**ABSTRACT:** Lower back pain is a common ailment that affects many people resulting from various spinal diseases such as disk bulge. Disc bulging, which refers to the narrowing of the intervertebral disc within the spinal canal, can cause lower back pain and lead to lumbar spine stenosis. Deep learning algorithms based on artificial intelligence are indispensable in the field of medicine, enhancing the precision of medical image diagnosis significantly and the ability to process massive amounts of data which contributes to reducing the workload on radiologists and doctors. Therefore, Deep learning techniques have become a more helpful tool to overcome this problem. For this purpose, this study employed the YOLO-v7, YOLOv8m object detection technique to build a model to detect lumbar spine disc discords using MRI data from 291 individuals suffering from lower back pain. The image data utilized for training the model were divided into three segments: 70% for training, 20% for validation, and 10% for testing, and validate the model results according to evaluation metrics. shows the optimal model validated externally utilizing new lumbar spine MRI images and assessment with the radiologist. YOLOv8m achieved a mean average in precision (93.7%), precision (90.5%), recall (79.2%), and F1-score (88.6%), accuracy (89.8%). A deep learning model demonstrated similar agreement to subspecialist radiologists in detecting and classifying lumbar spine stenosis on lumbar spine MRI. By selecting the most suitable YOLO model, doctors can significantly enhance their ability to detect lumbar spine stenosis at an early stage and effectively mitigate potential negative consequences for their patients.

**Keywords:** Automatic Detection, Low Back Pain, Disc bulging, MRI, Deep Learning, YOLO



## 1. INTRODUCTION

A bulging disc happens when the disc's outer layer weakens or bulges out beyond its normal limits (called the annulus). Spinal stenosis, a consequence of this ailment, refers to narrowing gaps in the spinal column. This narrowing puts pressure on the spinal nerve roots that extend from the lower back to the legs [1]. Figure 1 shows the anatomical structure of the spine with a disc bulging condition.



**FIGURE 1. - Spinal anatomy with a bulging disc [2]**

Lumbar bulging disc (LBD) diagnosis based on clinical symptoms and magnetic resonance imaging (MRI) findings. MRI has gained significant popularity due to its non-invasive and painless nature. It is commonly employed to acquire comprehensive data regarding the anatomy and function of various bodily organs. [3]. MRI images can be utilized to visually examine the lumbar spine, section by section, in three different planes of view: sagittal (from the side), axial (from the top down), and coronal (from the front). However, in the case of the lumbar spine, only the sagittal and axial planes are commonly employed [4].

An adept radiologist can visually inspect MRIs without the aid of any tools and accurately categorize the indications of LBD. However, due to the demanding workload and complexity of biological images, clinicians may encounter stress and fatigue, and becoming a specialist radiologist requires many years of experience. Prominent research publications underscore the benefits of utilizing artificial neural networks (ANNs) and deep learning (DL) in computer-assisted medical diagnosis and treatment. DL in AI serves as a solution to mitigate the workload and stress faced by radiologists and clinicians [5, 6].

Recently, deep learning (DL) algorithms have shown superior performance in many engineering applications [7-10], medical applications [11], and most computer vision tasks, such as image classification [12, 13], subject identification [14, 15]. DL has significantly benefited MRI by enhancing clinical practice, reducing the workload for radiologists and MR technologists, and improving throughput [16].

DL is a method used to rapidly and consistently detect important characteristics in biological images, namely areas that are aberrant or raise suspicion during clinical diagnostics, rather than just identifying objects [17, 18].

YOLO, an abbreviation for You Only Look Once, is widely recognized as one of the most effective neural networks for object recognition in images, utilizing deep learning techniques. A single Convolutional Neural Network (CNN) is used to predict several bounding boxes and the corresponding probability for each box. During the evaluation, the single-stage YOLO detector could immediately predict class likelihoods and bounding boxes dependent on the given frame. The network partitions each training image into a uniform grid of squares, in which each square predicts bounding boxes and class probabilities for objects in its region [19].

A few of the research employed the real-time single-stage detectors, and the YOLO models to train deep learning applications for the detection and prediction of biological images. These models were developed under the technique commonly known as the anchor-based technique together with the intersection over union technique.

This study [17] A novel technique is introduced in this study for precisely localizing the position of discs between vertebrae in lumbar MRI scans. It employs YOLOv2-IVD using Enhanced Visual Geometry Group 16 as the underlying model. This research assessed a proprietary dataset of 52 patient sagittal T2-weighted lumbar spine MRIs. Results demonstrated a remarkably high accuracy of 93.59% in detecting herniations.

[5] Specializes in the employment of YOLOv3 deep learning system to enable the detection of lumbar disc herniation from MRI scans. The analysis is based on a limited number of lumbar MRI scans and supplemented by data augmentation in the form of rotating the images, changing their contrast, and adjusting the brightness level. Each of these images was then very well converted into an 8-bit grayscale JPEG, as this facilitated the correct manipulation of the brightness and darkness. According to the outcomes of the study, the YOLOv3 algorithm with data augmentation methodology sufficiently identified the Lumbar herniation. The study also sought to establish that using YOLOv3 for LDH diagnosis can be useful to radiologists in pointing out exactly where the problem is in the MRI scan. Thus, the publication discussed the crucial issue of small dataset size and further work to improve future deep learning systems for diagnosing lumbar disc herniation.

[19] Describes an implementation strategy that will facilitate the grading of intervertebral disc degeneration (IDD). Data augmentation was used in the methodology and a deep learning approach involved the use of a Convolutional Neural Network and YOLO-V5 model. The study used a dataset, which included 1000 axial T2-w Mid-sagittal lumbar spine MRIs. Pfirrmann's grading system was used to categorize these pictures. The algorithm was triangulated with the use of precision, recall, and average precision measures; which gave reasonable results. Hence, it is evident that the proposal of using the CNN model for grading IDD in sagittal T2-weighted MRI images can be of esteemed assistance to radiologists, and is also accurate and efficient as viewed from the results of the study. The major problem in the conducted study was the creation of a proper dataset for training the deep learning model, as well as other problems connected to selecting the most suitable hyper-parameters.

[20] Conducted a study to compare the effectiveness of the developed YOLOv7 algorithm on MRI scans of patients with LDH. The method used the Precision, Recall, and mAP measures of the YOLOv7 model to analyze its efficiency on a dataset of 100 lumbar spine sag-T2 weight MRI images. The precision of YOLOv7 was relatively low while detecting LDH at a level of 42.90%, and a Recall of 44.10%. And an overall performance measured by mAP of 35.00%. The study showed that in comparison to the other cases, YOLOv7 performed significantly worse specifically in LDH detection in L4-L5 and L5-S1 areas. To the authors' recommendations, further experiments using a larger number of datasets can be performed to determine the most accurate model for the detection of LDH. The main limitation faced in the course of the study was the unavailability of a large number of datasets for analysis.

[21] This study will infer and classify LDH using object detection on MRI data to make the detection of LDH more accurate. The methodology utilizes the single-stage detection approach of the YOLO versions [5, 6, 7]. The sagittal perspective obtained from both upper and lower lumbar MRI scans is utilized. The dataset, consisting of 550 photos, was

partitioned into separate training and validation sets. According to the study, the YOLOv5 algorithm demonstrated the highest performance in detecting LDH, achieving an accuracy rate of 98.22%. The challenges encompassed the restricted quantity of images in the dataset.

To our knowledge, YOLOv8 has not been assessed for its capability to recognize LDB in MRI. However, the model performs well in detecting objects. Our study aimed to evaluate the diagnostic efficacy of two convolutional neural networks (CNNs), YOLOv7 and YOLOv8, and the comparison between them. These models were trained on multiple lumbar spine (LS) MRI datasets, and the classification of lumbar spine discs was assessed as either normal or exhibiting bulging.

## 2. METHOD

The research methodology framework of this study consists of four phases that were carried out sequentially for the detection of LDB utilizing the deep transfer learning technique as follows:

### Phase One: Data Pre-Processing:

- Collecting the LS MRI dataset of patients suffering from lumbar spine discords.
- Filtering the dataset based on lumbar spine disc conditions (e.g. normal, bulge, herniation).
- Performing image preprocessing techniques (e.g., resizing, cropping).

### Phase Two: LSS Model Architecture Selection:

- Choosing appropriate deep learning models (e.g., YOLO V7/V8) for the LDB detection.

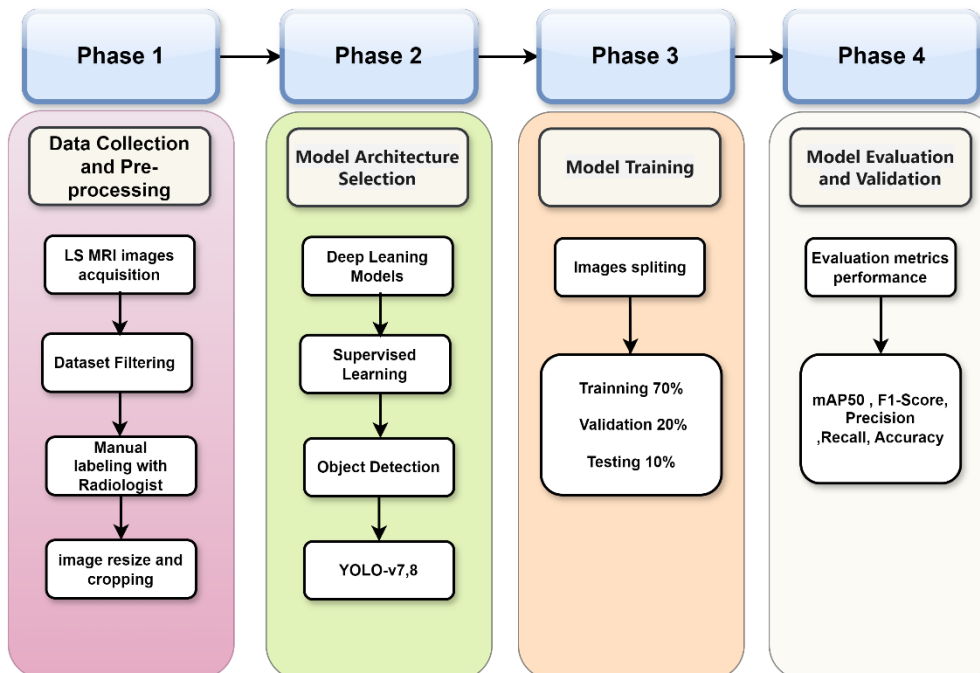
### Phase Three: Model Training:

- Datasets training set splitting (training 70%, validation 20%, testing 10%)
- Utilizing transfer learning to fine-tune the selected model on the lumbar spine dataset.

### Phase Four: LSS Model Evaluation:

- Calculating evaluation metrics (e.g., F1-score, precision, recall, accuracy) to assess model performance.
- Assessing the model's robustness by validating its performance using test images.

Figure 2. depicts all the phases of the research methodology procedure.



**FIGURE 2. - Integration Methodology Phases of LS disc Detection**

### 2.1 Phase One: Data Collection and Pre-Processing

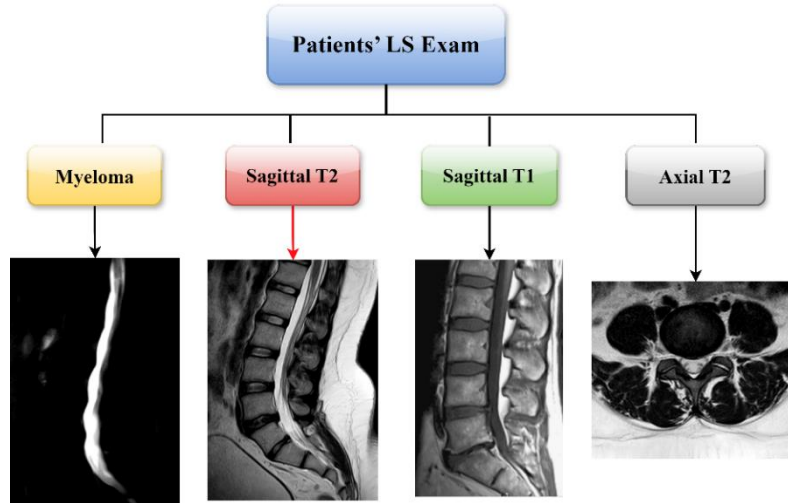
Phase one focuses on the collection and pre-processing of the LS MRI dataset of patients.

### 2.2 Data collection

LS MRI datasets were prospectively collected between September 2023 and January 2042 at Tikrit Teaching Hospital and Salah AL-Din General Hospital in Iraq after obtaining all necessary and ethical approvals.

### 2.3 MRI Modality Selection

The LS MRI examination protocol comprises several sequences, as illustrated in Figure 3.



**FIGURE 3. - LS MRI Modality**

Additionally, radiography may introduce an additional sequence; however, the default sequences for LS exams, as depicted in Figure 3, remain consistent. This study selected the Sagittal T2-weighted (Sag-T2) MRI modality due to its characteristics for displaying anatomical details of the lumbar spine.

### 2.4 Dataset Description

The dataset includes LS MRI scans, encompassing both normal and abnormal cases. The dataset comprises a comprehensive collection of MRI images from 291 patients, obtained from two institutions. Each image has dimensions of 448×448 pixels. The dataset comprises patients of all ages and both genders (male and female). Additionally, MRI protocol contains many sequence pulses and that sequence has various parameters that vary from one MRI device to another, and these variations directly impact image quality [11]. Table 1 presents details of the MRI pulse sequence parameters of acquisition images.

**Table 1. - LS MRI Sag\_T2 pulse sequence parameters**

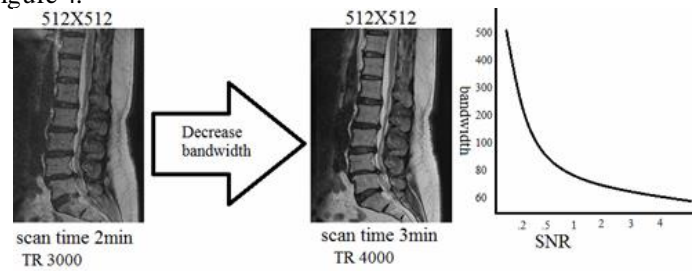
	MRI 1	MRI 2
Source	Al-Tawfiq Hospital	Salah al-Din General Hospital
Manufacture	Hitachi Airis II	Siemens
Magnet field (tesla)	0.3 T	1.5 T
Magnet field (type)	Open MRI	Closed MRI
Sequence pulse Type	Sag_T2_FSE	Sag_T2_TSE
Repetition Time (msec)	3000	3500
Echo Time (msec)	120	91

Slice Thickness (mm)	6.0	4.0
Spacing Between Slices (mm)	6.5	4.0
Field of View (mm)	350	260
Matrix (Freq. x Phase)	256 X 168	256 X 192
Bandwidth (Hz/pixel)	34.1 Hz	197 Hz
Number of slices	9	15

Here, we shall briefly discuss these sequences as follows:

**TR (Repetition Time):**

TR refers to the interval of time between consecutive pulse sequences or repetitions. The unit of measurement is milliseconds (ms). A longer TR can increase the signal intensity and is often used in sequences where T2 weighting is desired [22] as shown in Figure 4.



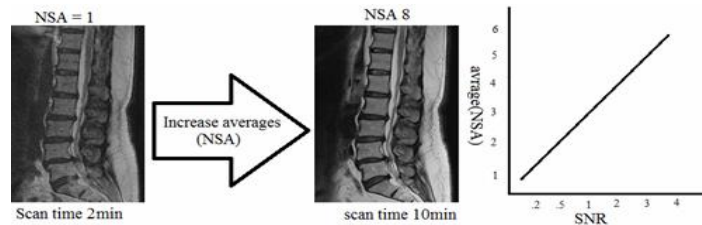
**FIGURE 4. - Effect of increased TR weight on image quality [23]**

**TE (Echo Time):**

TE refers to the duration from when the radiofrequency pulse is applied to when the signal echo reaches its highest point. The unit of measurement is milliseconds (ms). Short TE values are associated with T1-weighted images, while longer TE values are related to T2-weighted images [24].

**NEX (Number of Excitations):**

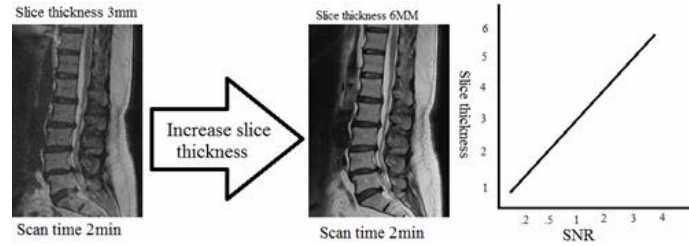
NEX represents the number of excitations (NEX) or signal averages/acquisitions (NSA). Increasing NEX improves the signal-to-noise ratio (SNR) but prolongs scan time [25] as shown in Figure 5.



**FIGURE 5. - Effect of increased NEX average on image quality [23]**

**Slice Thickness:**

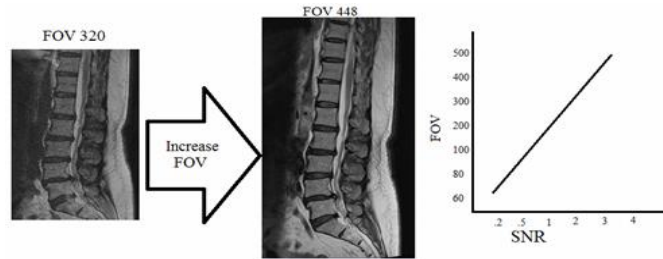
Slice thickness refers to the thickness of each imaging slice and is typically measured in millimeters (mm). Thinner slices provide higher resolution but may increase scan time [26] as shown in Figure 6.



**FIGURE 6. - Effect of increasedlice thickness on image quality [23]**

**FOV (Field of View):**

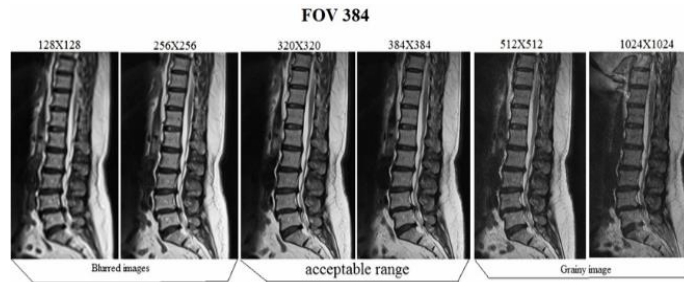
FOV refers to the physical dimensions of the imaging area and is typically quantified in millimeters. It determines how much anatomy is covered in each image. A larger FOV captures more anatomy but may result in a lower resolution [27] as shown in Figure 7.



**FIGURE 7. - Effect of IncreasedFOV on image quality [23]**

**Matrix:**

The matrix represents the total number of pixels in both the rows and columns of the image. Larger matrix sizes result in higher spatial resolution but may increase data acquisition time and storage requirements [28] as shown in Figure 8.



**FIGURE 8. - Effect of IncreasedMatrix size on image quality [23]**

The MRI lumbar spine scan includes many sequences such as (SAG\_T2, SAG\_T1, MYLO, Ax\_T1, and Ax\_T2), in our study, we chose T2 weight. The image contrast in T2-weighted imaging (T2WI) is manipulated by modifying the repetition time (TR) and echo time (TE). T2-weighted images utilize a lengthy repetition time (TR) ranging from 3000 to 6000 milliseconds and a prolonged echo time (TE) between 90 and 110 milliseconds. On T2-weighted images (T2WI), fluids have a bright appearance, often known as hyperintensity. Consequently, regions that contain cerebrospinal fluid (CSF), such as the ventricles of the brain or the canal of the spinal cord, will exhibit a bright appearance in T2-weighted images [29]. Figure 9 shows the brightness level of the lumbar spine side view anatomy. The color gradients and variations between image pixels are crucial for deep learning algorithms. That's why we utilized T2WI to get the most accurate diagnosis for the condition of the intervertebral discs.



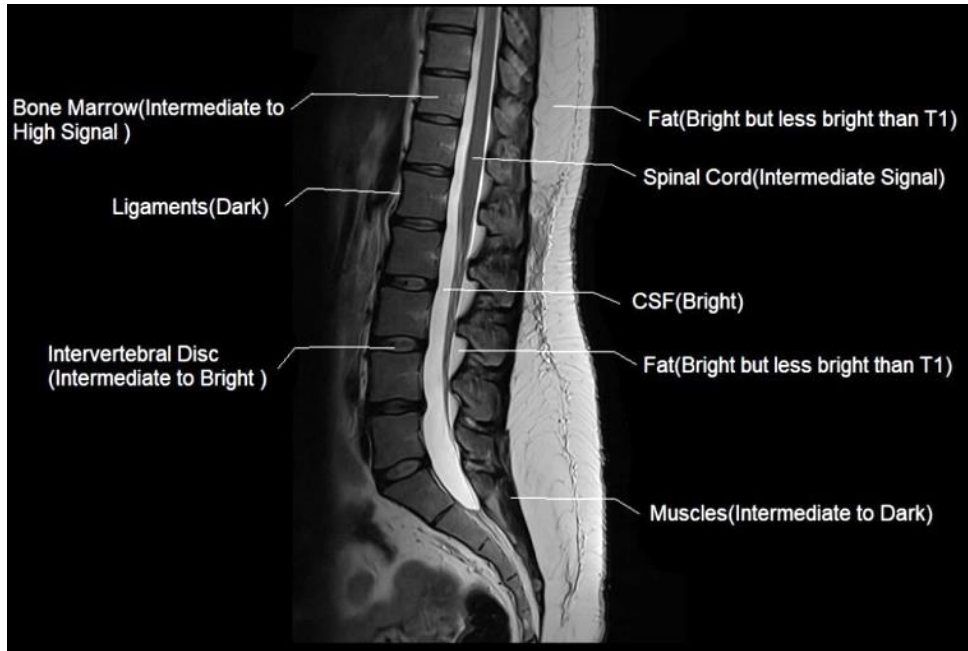


FIGURE 9. - T2-WI MRI Image of the lumbar spine [30]

### 2.5 Dataset filtering

Fundamentally, with the assistance of a radiologist with 10 years of expertise in diagnosing lumbar spine diseases, the dataset was filtered and classified according five most lumbar spine degenerative changes. Noise and artifact images were excluded. Our work is divided into two classes: normal and bulging discs and excludes other lumbar spine disease cases as shown in Figure 10.

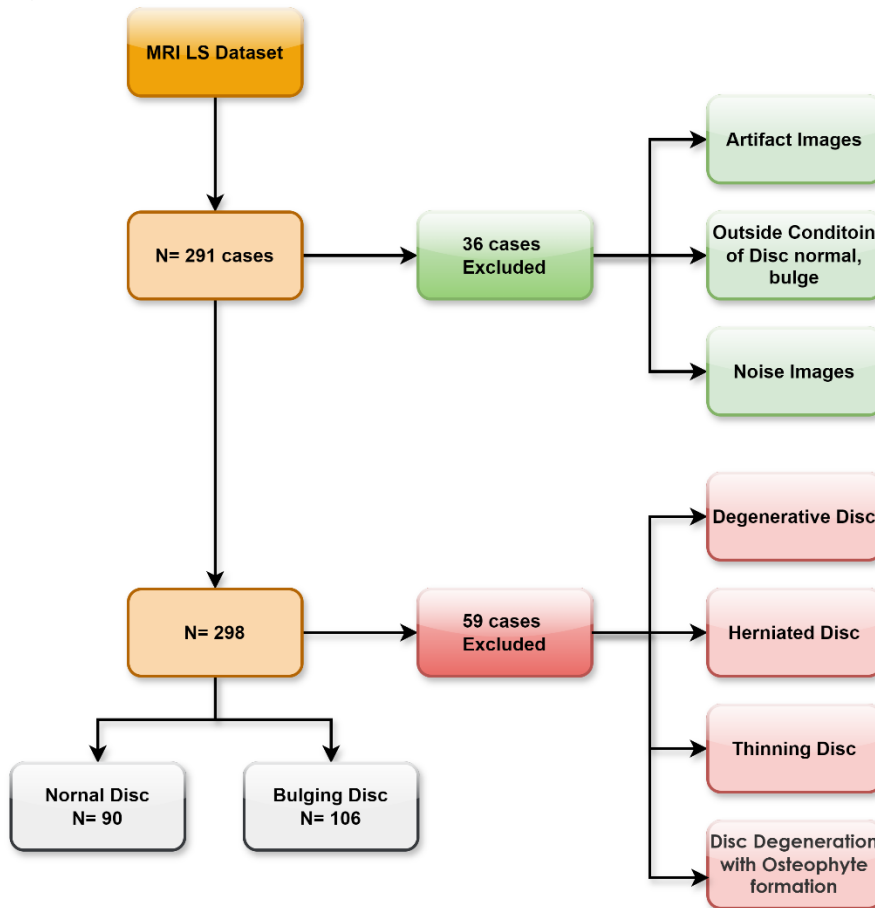


FIGURE 10. - MRI dataset filtering and classification

After the filtering process, we must balance images between normal and bulging cases by choosing three slices (between 6 to 9) that show the best view of the sagittal view of the lumbar spine as shown in Figure 11.

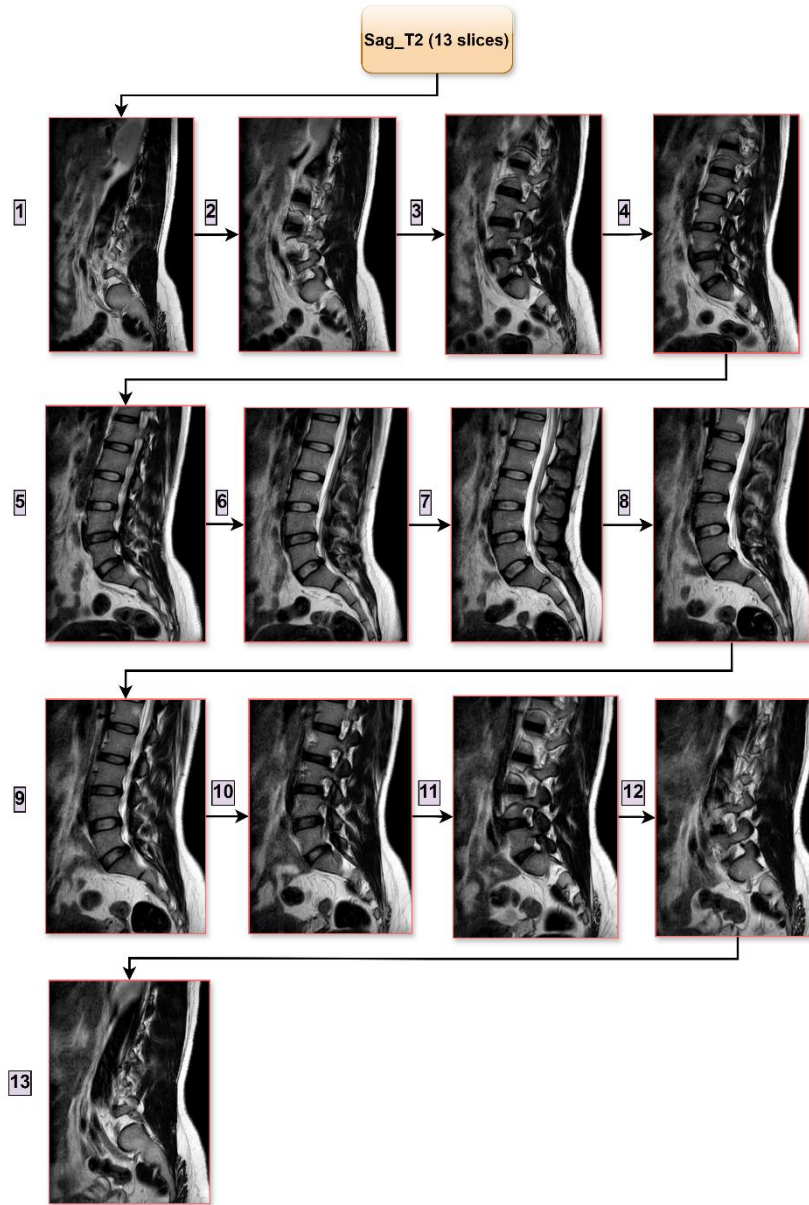


FIGURE 11. - MRI Sag\_T2 images slices

## 2.6 Dataset pre-processing

The dataset pre-processing includes the following steps:

1. Use the Radiant Dicom viewer software to convert DICOM file types to jpeg images, and each image has a resolution of 448 x 448 pixels.
2. Using the polygon annotation technique (by Robflow tool), it has manually labeled the mid-sagittal lumbar spine disc selected at the anatomical level of the intervertebral disc between L4 and L5 (the vertebra most prone to spinal degenerative changes due to its specific morphology and function) as shown in Figure 12.
3. Stretch the image size to  $640 \times 640$  to prepare it for training using the YOLOv7,8 models.
4. Color System: Gray Scale
5. Horizontal crop 15 %.



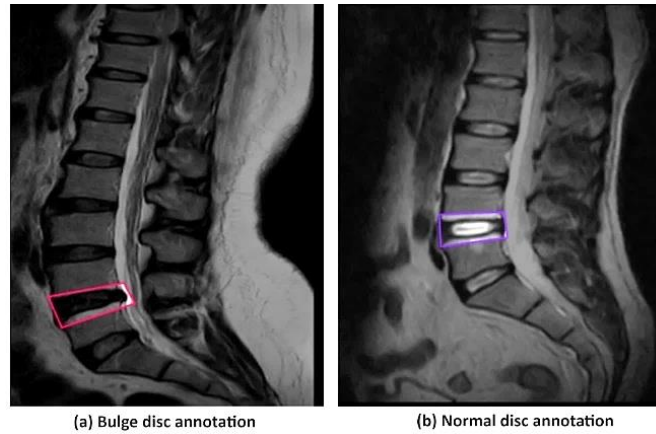


FIGURE 12. - MRI dataset annotation

### 2.7 Phase Two: Model Architecture Selection

Phase two focuses on choosing the deep transfer learning object detection YOLO model which is a commonly used single-stage target detection algorithm in real-time systems because of its high running speed and accuracy in detecting small objects in images [31].

### 2.8 YOLOv7 Object Detection Architecture

The publication of YOLOv7 [32] Occurred in July 2022 on ArXiv, authored by the same individuals who created YOLOv4 and YOLOR. At that moment, it outperformed all existing object detectors in terms of both speed and accuracy, operating within a range of 5 frames per second (FPS) to 160 FPS. Similar to YOLOv4, it underwent training solely using the MS COCO dataset, without utilizing pre-trained backbones. YOLOv7 introduced several architectural modifications and a range of bag-of-freebies techniques, resulting in improved accuracy without any impact on inference speed, solely impacting the training duration. Figure 13 shows the detailed architecture of YOLOv7.

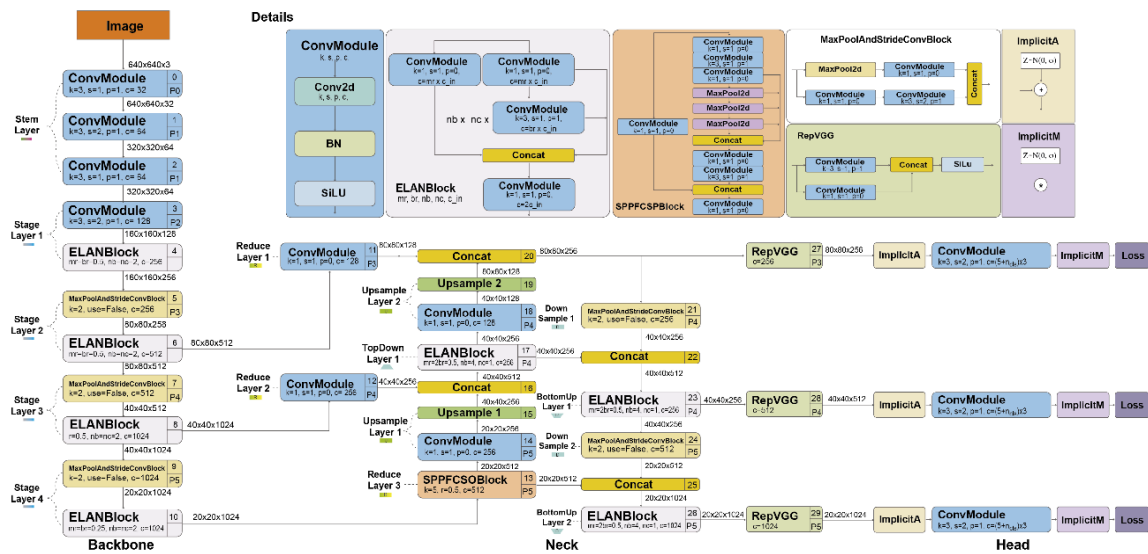


FIGURE 13. - Architecture detail of YOLOv7 [33]

The YOLOv7 employed the Extended efficient layer aggregation network (E-ELAN) backbone, model scaling, and model re-parameterization to achieve a compromise between detection efficiency and precision [33].

### 2.9 YOLOv8 Object Detection Architecture

Currently, YOLO is considered to be one of the most rapidly expanding and superior algorithms. The latest version, YOLOv8, was launched in 2023. By implementing enhancements to the detection features in YOLOv8, we anticipate that the accuracy of LSS detection can be improved. YOLOv8 is a highly recommended option for various applications involving object identification, instance segmentation, and image classification because of its fast speed, precise accuracy, and user-friendly interface [34].

Ultralytics, the company responsible for the development of YOLOv5, published YOLOv8 [35] In January 2023. YOLOv8 has five different scaled versions: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large). YOLOv8 is capable of performing many vision tasks including object identification, segmentation, pose estimation, tracking, and classification.

Figure 14 shows the detailed architecture of YOLOv8. YOLOv8 utilizes a comparable underlying structure as YOLOv5 but with modifications to the CSPLayer, which is now referred to as the C2f module. The C2f module, which incorporates two convolutions and a cross-stage partial bottleneck, enhances detection accuracy by integrating high-level characteristics with contextual information.

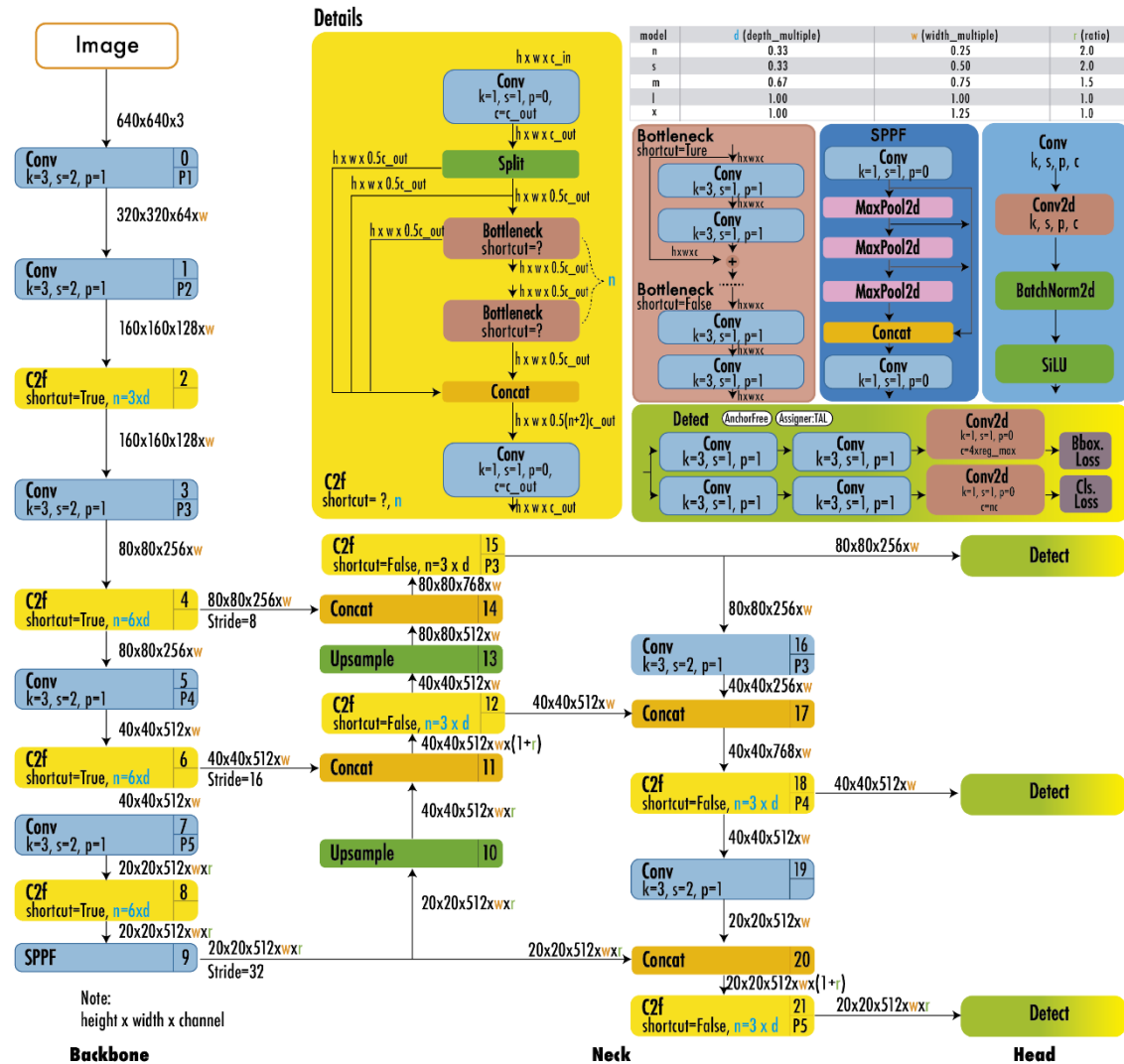


FIGURE 14. - Architecture detail of YOLOv8 [33]

The YOLOv8 model has a redesigned CSPDarknet53 backbone and introduces the C2f module as a replacement for YOLOv5's CSPLayer. The model includes a spatial pyramid pooling fast (SPPF) layer to efficiently pool features. Every convolution operation incorporates batch normalization and SiLU activation. The head is separated to allow for autonomous processing of objectness, classification, and regression tasks. Furthermore, YOLOv8 provides a semantic segmentation model called YOLOv8-Seg, which utilizes a CSPDarknet53 backbone and two segmentation heads. YOLOv8 demonstrates superior performance in object detection and semantic segmentation benchmarks, while also maintaining exceptional speed [34].

### 2.10 Phase Three: Model Training

Phase three will split filtered images into three groups (training images 70%, validation images 20%, testing images 10%). Google Colaboratory environment Python 3.7 version was used to train and validate the algorithms with batch size 16, epoch 100, and image size 640 × 640 pixels.

## 2.11 Phase Four: Model Evaluation and Validation

Phase four checks the model performance through evaluation metrics.

Evaluation metrics are essential for any project as they are used to measure the quality of the model. There exists a wide variety of evaluation metrics. We employed the mAP, Precision, Recall, and F1-score criteria to evaluate the YOLOv7 and YOLOv8 models.

The evaluation metrics are explained as follows:

**Precision:** Precision on the other hand is a metric used to determine a network accuracy in identifying targets at a given threshold that is formulaically defined as:

$$precision = TP / (TP + FP) \quad (1)$$

**Recall:** is one of the metrics, which provide information about the network's ability to find its target and equals :

$$recall = TP / (TP + FN) \quad (2)$$

Considering the specifics of this particular dataset, it is sufficient to work with a single F1-score metric as an estimate of the model's accuracy. The F1-score is often used to assess the model's accuracy since it is derived from the synthesis of the Precision and Recall indicators, which means that both the precision and recall rates are taken into account. It combines all of the model performance metrics into one formula, which gives a clear and nearly precise estimate of the general performance of the model.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

**The Average Precision (AP),** or Mean Average Precision (mAP), is one of the most common evaluation measures used when determining object detection models' efficiency.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (4)$$

The evaluation metrics are computed using the formula mentioned previously, taking into account the following conditions:

- **True Positive (TP):**
  - There are situations in which the model precisely categorizes the image as “bulge”, the actual label as well.
  - Regarding the discriminant of the study about LSSD's detection, TP is the probability of correctly classifying patients with bulging discs.
- **True Negative (TN):**
  - These are cases where the model correctly predicts “normal” (negative class) when the actual label is also “normal.”
  - TN corresponds to correctly identifying patients without bulging discs.
- **False Positive (FP):**
  - These are cases where the model predicts a “bulge” (positive class), but the actual label is “normal.”
  - FP represents false alarms—patients predicted to have bulging discs when they don't.
- **False Negative (FN):**
  - These are cases where the model predicts “normal” (negative class), but the actual label is “bulge.”
  - FN corresponds to missing cases—patients with bulging discs that the model fails to detect.

In summary:

- **TP:** Correctly identified bulging discs.
- **TN:** Correctly identified normal discs.
- **FP:** Incorrectly predicted bulging discs (false alarms).
- **FN:** Missed bulging discs (false negatives).

The confusion matrix is a valuable tool for assessing the effectiveness of a model, particularly in the context of medical diagnosis tasks. These values can be used to calculate precision, recall, F1-score, and accuracy. To ascertain the

model that would yield the maximum level of accuracy for LDB identification, we intend to compare the performance of YOLOv7 and YOLOv8 models based on these metrics.

### 3. RESULTS

The suggested approach is based on our private dataset, testing, and comparing YOLOv7, and YOLOv8 models. Every model has a different architecture, accuracy, speed, and mathematical operation; we compared the results of all the models to choose the best model.

#### 3.1 Result Obtained by YOLOv7

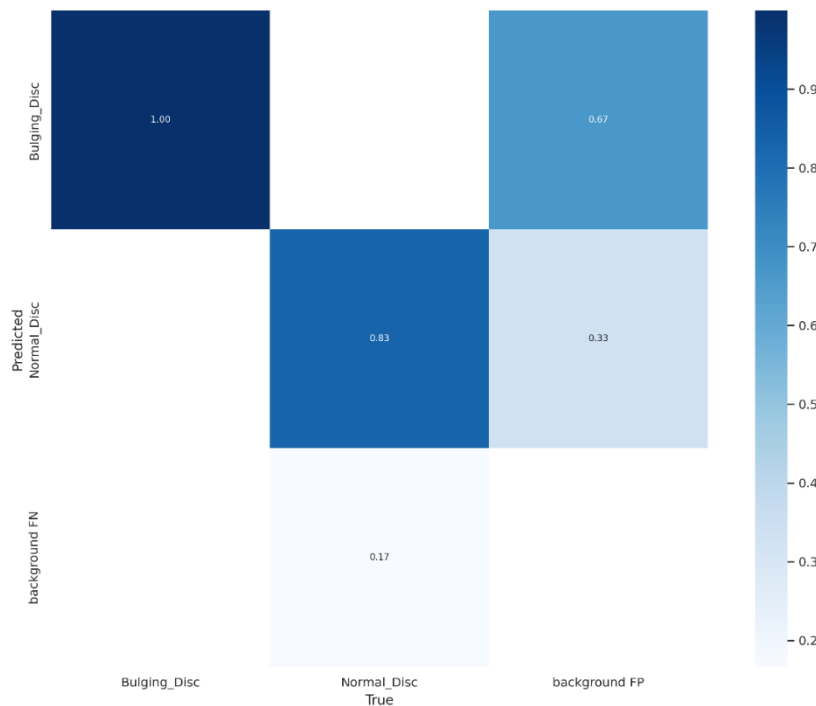
For this experiment, we employ YOLOv7 to train the gathered data. Table 2 displays the outcomes obtained by employing the YOLOv7.

**Table 2. - YOLOv7 Results**

Classes	mAP 0.5	precision	Recall	F1-Score
Normal-Disk	88.1%	89.9%	75%	81.7%
Bulging-Disc	91.9%	71.9%	95.8%	82.1%
Average	90%	80.9%	85.4%	81.9%

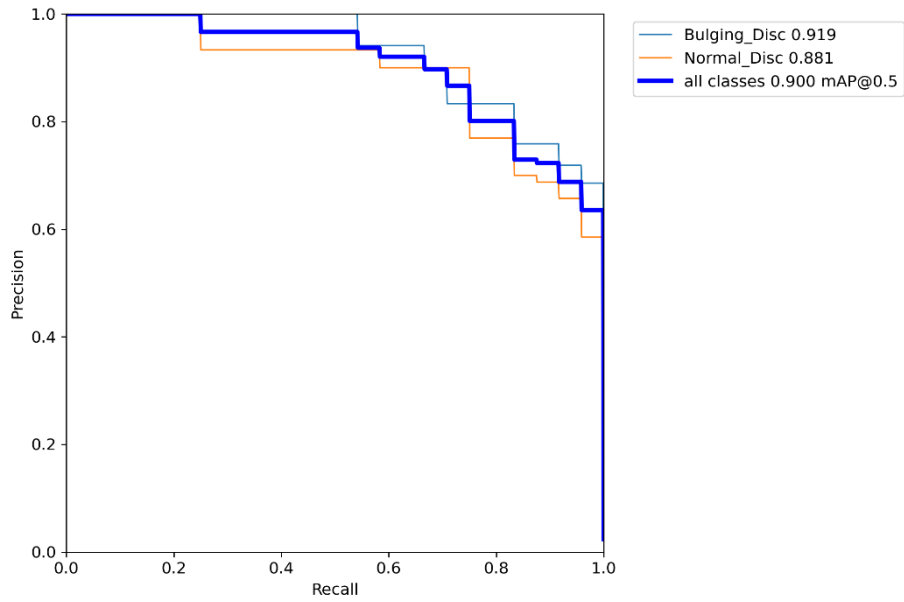
The results in Table 2 indicate that the metrics provide insights into the performance of a lumbar spine disc detection system.

The model YOLOv7 achieved results for the average of all classes of mAP0.5 of 90%, precision of 80.9%, recall of 85.4%, and an F1 score of 81.9%. For the Normal-Disc class, YOLOv7 achieved the following values: mAP0.5 of 88.1%, precision of 89.9%, recall of 75%, and an F1 score of 81.7%. YOLOv7 achieved the following values for the Bulging-Disc class: mAP0.5 of 91.9%, precision of 80.9%, recall of 85.4%, and an F1 score of 81.9%.



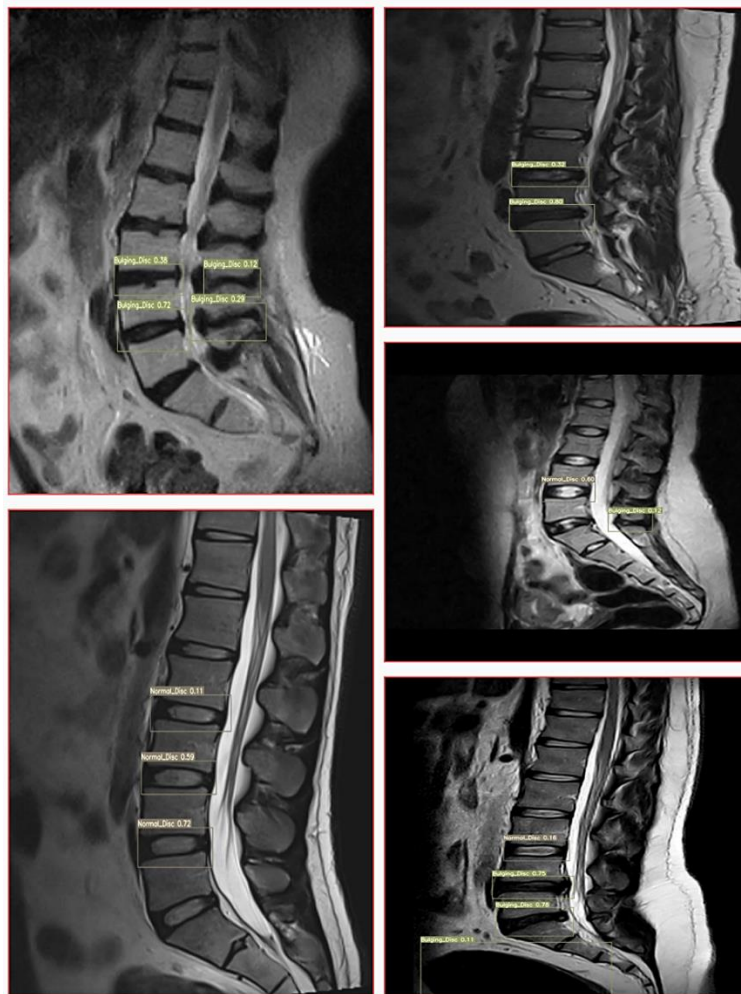
**FIGURE 15. - Confusion Matrix Using YOLOv7**

The classification accuracy values for the bulging and normal Disc, as shown in Figure 15, are 1.00 and 0.83, respectively, according to the confusion matrix values.



**FIGURE 16. - Precision-Recall Curve Using YOLOv7**

Figure 16 indicates that the lumbar spine disc bulge had a precision-recall value of 0.91, whereas the normal average had a precision-recall value of 0.88.



**FIGURE 17. - YOLOv7 Detection**

Figure 17 illustrates the detection outcomes for lumbar spine disc L4-L5 using YOLOv7. For bulge discs, the detection accuracy was between 70% to 80%, while 72% for normal discs.

### 3.2 Results Obtained by YOLOv8m

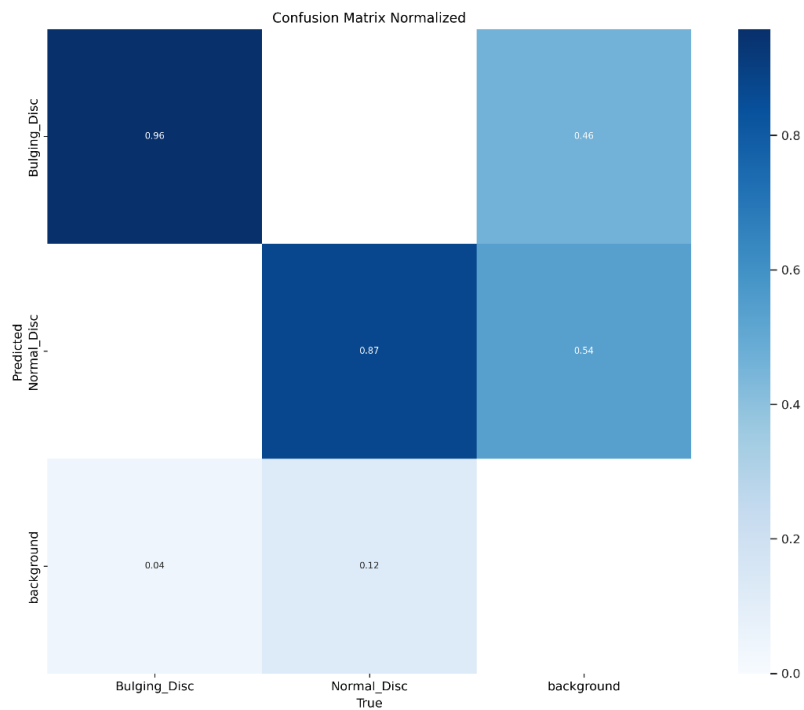
For this experiment, we employ YOLOv8m to train the gathered data. Table 3 displays the outcomes obtained by employing the YOLOv8m.

**Table 3. - YOLOv8m Results**

Classes	mAP 0.5	precision	Recall	F1-Score
Normal-Disk	89.5%	89.1%	79.2%	83.8
Bulging-Disc	98%	91.9%	94.5%	93.1
Average	93.7%	90.5%	86.8%	88.6

The results in Table 3 indicate that the metrics provide insights into the performance of a lumbar spine disc detection system. The high recall for bulging discs demonstrates that the model rarely misses them, while the precision for normal discs suggests accurate identification.

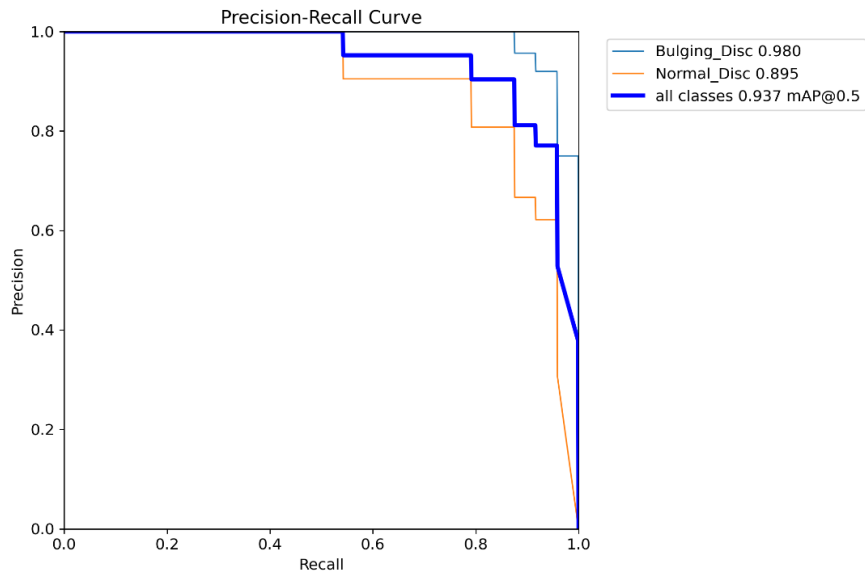
The model YOLOv8m achieved results for the average of all classes mAP0.5 of 93.7%, precision of 90.5%, recall of 86.8%, and an F1 score of 86.4%. For the Normal-Disc class, YOLOv8m achieved the following values: mAP0.5 of 89.5%, precision of 89.1%, recall of 79.2%, and an F1 score of 87.9%. YOLOv8m achieved the following values for the Bulging-Disc class: mAP0.5 of 98 %, precision of 91.9%, recall of 94.5%, and an F1 score of 85%.



**FIGURE 18. - Confusion Matrix Using YOLOv8m**

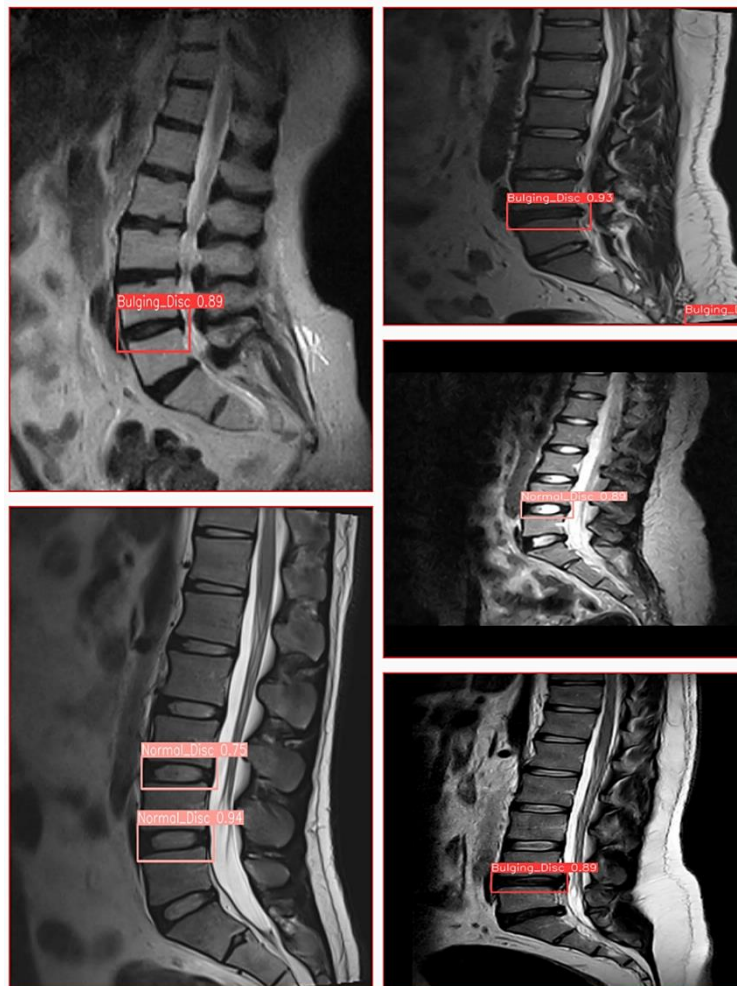
From Figure 18 for the bulging and normal Disc, the classification accuracy values are 0.96 and 0.87, respectively, based on the confusion matrix values.





**FIGURE 19. - Precision-Recall Curve Using YOLOv8m**

Figure 19 indicates that the lumbar spine disc bulge had a precision-recall value of 0.98, whereas the normal average had a precision-recall value of 0.89.



**FIGURE 20. - YOLOv8m Detection**

Figure 20 illustrates the detection outcomes for lumbar spine disc L4-L5 using YOLOv8m. For bulge discs, the detection accuracy was 93%, while 94% for normal discs.

#### 4. DISCUSSION

A bulging disc occurs when the outer layer of the spinal disc weakens or protrudes outward, while in a herniated disc, the outer covering of the disc has a hole or tear. In this research study, we investigated the feasibility of employing YOLOv7 and YOLOv8m models for the automated detection of lumbar spine disc bulges in MRI imaging. Our models were trained on diverse lumbar spine MRI datasets (Sag\_T2 weighted) sourced from various providers and MRI manufacturers.

Table 1 presents the evaluation metrics results for both normal and bulging discs using the YOLOv7 model. However, during validation with test images, the accuracy of detecting the L4-L5 lumbar spine vertebra was found to be low as shown in **Error! Reference source not found.** . In contrast, Table 2 demonstrates improved evaluation metrics for the YOLOv8m model compared to the YOLOv7 model, both trained on the same dataset. In comparison in validation test images, YOLOv8m shows the best results of detection rather than the YOLOv7 model (see **Error! Reference source not found.****Error! Reference source not found.**).

Among the YOLO models evaluated, YOLOv8m emerged as the most suitable candidate based on comprehensive evaluation metrics.

Table 4 shows a comparative analysis with related studies that used YOLO object detection techniques

**Table 4. - Comparative Analysis of Lumbar Spine Disorder Detection Studies**

Ref.	Year	Lumbar spine Condition	Method	Dataset (Public, Private)	No. of Subjects	Dataset Modality	MRI Magnetic Field		Annotation Technique
							Open	Closed	
[17]	2023	Disc Herniation	YOLO-v2	Private	52	Single	✗	✓	Rectangle
[19]	2023	Disc Degeneration	YOLO-v5	Public	200	Single	✗	✓	Polygon
[21]	2023	Disc Herniation	YOLO-v5,6,7	Public	110	Single	✗	✓	Rectangle
[5]	2021	Disc Herniation	YOLO-v3	Private	168	Single	✗	✓	Rectangle
[20]	2023	Disc Herniation	YOLO-v7	Public	20	Single	✗	✓	Rectangle
Our Study	2024	Disc Bulge	YOLO (7,8)	Private	298	Multi	✓	✓(0.3T, 1.5T)	Polygon

Valarmathi, G. et al in 2023 [17] Shows that the proposed system achieves high accuracy in both IVD localization and classification using 2600 MR images from 52 patients it achieves an mAP of 92%.

W. A.-O. Liawrungrueang et al in 2023 [19] Achieved high accuracy in detecting lumbar spine degenerative changes based on the Pfirman classification using the YOLOv5 models.

A. A. Prisilla et al in 2023 [21] T2-weighted images in the sagittal view and YOLO algorithms are utilized to detect lumbar disc herniations. The YOLOv5x non-AUG model had superior results in object detection, exhibiting the highest precision and recall scores.

Y. Tsai et al in 2021[5] It has been demonstrated that deep learning can effectively utilize a small dataset consisting of a limited number of medical images. Additionally, the use of data augmentation techniques has been found to enhance the performance of deep learning models. The YOLOv3 detection findings exhibited remarkable performance in identifying the region affected by lumbar disc herniations. The mean average precision (mAP) value exhibits an increase as the dataset undergoes various alterations, ranging from 50-aug to 550-aug, through the utilization of data augmentation techniques. Additionally, each lumbar level demonstrates precision that is correlated with the number of cases. The percentage of recall was 92% in the sample of 550 individuals in August. The mean average precision (mAP) of the 550-augmentation group achieved the maximum performance at 92.4%. Both the 500-augmentation and 350-augmentation groups exhibited the same performance level of 87%.

A. A. Prisilla et al in 2023 [20] Assessed the efficacy of YOLOv7 in detecting Lumbar Disc Herniation (LDH) in MRI images. The results indicated that YOLOv7 exhibited subpar detection ability for LDH, with a Precision of 42.90%, Recall of 44.10%, and mAP of 35.00% using the 100 datasets utilized. The YOLOv7 model demonstrated the highest

average precision (AP) in the L3-L4 region, with a value of 65.50%, while the lowest AP was seen in the L4-L5 region, with a value of 11.60%. These findings indicate that although YOLOv7 is excellent at detecting common objects, it still needs to be further studied. Additional research using larger datasets could assist in identifying the most reliable LDH detection model.

Based on the information provided, we can infer that the previous studies, as indicated in Table 4, achieved positive results for the detection of lumbar spine discords using both private and public MRI data. However, it is important to note that the variation in the quality of MRI images across different MRI devices and pulse sequence parameters used significantly affects the accuracy of the model prediction. In this study, we have made significant contributions in addressing the challenges found in previous studies, as detailed below:

**Methodological Limitations:** Although previous models, such as YOLOv2-v7 models, have shown effectiveness, they are surpassed by newer architectures in terms of both accuracy and speed. YOLOv8 introduces enhancements, including improved backbone networks and more efficient training strategies. YOLOv8 not only enhances detection accuracy but also minimizes computational overhead, rendering it better suited for real-time applications when compared to older models.

**Dataset and Diversity:** Previous studies frequently relied on smaller datasets with limited diversity, often derived from single MRI devices. This limitation can hinder the model's generalization of new, unseen data. In contrast, this study utilizes a proprietary dataset from multi-center MRI devices, including open and closed magnetic field strengths (0.3T, and 1.5T)."

**Labeling and Annotation Technique:** The choice of annotation technique plays a crucial role in preparing training images for an algorithm. Given the slightly curved angle of lumbar spine discs, particularly in vertebra L4-L5. Our approach uses the polygon technique to ensure accurate and detailed annotations.

For future work to create a robust and reliable system for clinical assistance, it is essential to train the model using a large-scale dataset of different MRI magnetic types (0.3T, 0.4T, 1.5T, and 3T).

## 5. Conclusion

The study demonstrates the effectiveness of deep learning-based object detection techniques, specifically YOLO models, in automating the identification and categorization of lumbar disc bulging from MRI data. Notably, the YOLOv8m model outperformed the YOLOv7 model in detection accuracy. While YOLOv7, like other models, relies on bounding boxes for localization, these boxes may not precisely capture small annotations, leading to reduced accuracy. This automated approach has the potential to enhance patient care, alleviate radiologists' workload, and improve diagnostic accuracy in spinal imaging. The study also underscores the challenges faced by doctors and radiologists in precise spinal disorder diagnosis, emphasizing the importance of early detection for effective therapy. Further research using larger datasets from diverse MRI sources is recommended to refine the lumbar spine disc bulge detection model. This study substantially improves previous work by addressing key limitations. Future research could explore further enhancements in model accuracy, integration with other diagnostic modalities, and validation of even larger datasets. Notably, leveraging multi-center MRI data has undeniably enhanced diagnostic accuracy, empowering clinical practitioners and radiologists to make informed decisions.

## Abbreviations

CSF	Cerebrospinal Fluid
LDB	Lumbar Disc bulge
IVD	lumbar intervertebral discs
CNN	Convolutional Neural Network
ANN	Artificial Neural Networks
DL	Deep Learning
YOLO	You Only Look Once
MRI	Magnetic Resonance Imaging
TR	Repetition Time
TE	Echo Time
NEX	Number of Excitations
FOV	Field of View

## Funding

None

## ACKNOWLEDGEMENT

I want to thank radiologist Dr. Aws Qahtan Hamdi. His expertise and insights significantly enriched our understanding of lumbar spine imaging and contributed to the quality of our research. I would like to thank the engineer Joseph Varghese MRI technical consultant, as his technical expertise and support in handling MRI datasets were invaluable throughout the performance of this work.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest

## REFERENCES

- [1] A. N. Laiwalla et al., "Lumbar Spinal Canal Segmentation in Cases with Lumbar Stenosis Using Deep-U-Net Ensembles," *World Neurosurgery*, vol. 178, pp. e135-e140, Oct. 2023, doi: 10.1016/j.wneu.2023.07.009.
- [2] H. Gray and H. V. Carter, *Gray's Anatomy*. Arcturus Publishing, 2013.
- [3] X. Zhao and X.-M. Zhao, "Deep learning of brain magnetic resonance images: A brief review," *Methods*, vol. 192, pp. 131-140, Aug. 2021, doi: 10.1016/j.ymeth.2020.09.007.
- [4] A. S. Al-Kafri et al., "Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks," *IEEE Access*, vol. 7, pp. 43487-43501, 2019, doi: 10.1109/ACCESS.2019.2908002.
- [5] J. Y. Tsai et al., "Lumbar Disc Herniation Automatic Detection in Magnetic Resonance Imaging Based on Deep Learning," *Frontiers in Bioengineering and Biotechnology*, vol. 11, 2023, doi: 10.3389/fbioe.2023.1247112.
- [6] Z. T. Al-Qaysi et al., "A comprehensive review of deep learning power in steady-state visual evoked potentials," *Neural Computing and Applications*, 2024, doi: 10.1007/s00521-024-10143-z.
- [7] Z. Al-Qaysi et al., "Generalized Time Domain Prediction Model for Motor Imagery-based Wheelchair Movement Control," *Mesopotamian Journal of Big Data*, vol. 2024, pp. 68-81, 2024.
- [8] Z. Al-Qaysi et al., "Optimal Time Window Selection in the Wavelet Signal Domain for Brain-Computer Interfaces in Wheelchair Steering Control," *Applied Data Science and Analysis*, vol. 2024, pp. 69-81, 2024.
- [9] Z. Al-Qaysi et al., "A Frequency-Domain Pattern Recognition Model for Motor Imagery-Based Brain-Computer Interface," *Applied Data Science and Analysis*, vol. 2024, pp. 82-100, 2024.
- [10] Z. Al-Qaysi, A. Al-Saegh, A. F. Hussein, and M. Ahmed, "Wavelet-based Hybrid learning framework for motor imagery classification," *Iraqi Journal of Electrical and Electronic Engineering*, 2022.
- [11] Z. Al-Qaysi, A. Albahri, M. Ahmed, and S. M. Mohammed, "Development of hybrid feature learner model integrating FDOSM for golden subject identification in motor imagery," *Physical and Engineering Sciences in Medicine*, vol. 46, no. 4, pp. 1519-1534, 2023.
- [12] S. M. Samuri, T. V. Nova, B. Rahmatullah, S. L. Wang, and Z. T. Al-Qaysi, "Classification model for breast cancer mammograms," *IJUM Engineering Journal*, vol. 23, no. 1, pp. 187-199, 2022.
- [13] A. Albahri et al., "A trustworthy and explainable framework for benchmarking hybrid deep learning models based on chest X-ray analysis in CAD systems," *International Journal of Information Technology and Decision Making*, 2024.
- [14] R. A. Aljanabi, Z. Al-Qaysi, and M. Suzani, "Deep Transfer Learning Model for EEG Biometric Decoding," *Applied Data Science and Analysis*, vol. 2024, pp. 4-16, 2024.
- [15] M. Ahmed, M. D. Salman, R. Adel, Z. Alsharida, and M. Hammood, "An intelligent attendance system based on convolutional neural networks for real-time student face identifications," *Journal of Engineering Science and Technology*, vol. 17, no. 5, pp. 3326-3341, 2022.
- [16] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102-127, May 2019, doi: 10.1016/j.zemedi.2018.11.002.
- [17] G. Valarmathi and S. Nirmala Devi, "Automatic localization and classification of intervertebral disc herniation using hybrid classifier," *Biomedical Signal Processing and Control*, vol. 86, p. 105291, Sept. 2023, doi: 10.1016/j.bspc.2023.105291..

- [18] A. S. Albahri et al., "A Trustworthy and Explainable Framework for Benchmarking Hybrid Deep Learning Models Based on Chest X-Ray Analysis in CAD Systems," *International Journal of Information Technology and Decision Making*, 2024, doi: 10.1142/s0219622024500019.
- [19] W. A.-O. Liawrungrueang, P. Kim, V. A.-O. X. Kotheeranurak, K. A.-O. Jitpakdee, and P. Sarasonbath, "Automatic Detection, Classification, and Grading of Lumbar Intervertebral Disc Degeneration Using an Artificial Neural Network Model," *Diagnostics*, vol. 13, no. 4, pp. 663, 2023, doi: 10.3390/diagnostics13040663.
- [20] A. A. Prisilla et al., "Automatic Detection of Lumbar Disc Herniation Using YOLOv7," in *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, 17-19 July 2023, pp. 843-844, doi: 10.1109/ICCE-Taiwan58799.2023.10226718.
- [21] A. A. Prisilla et al., "An approach to the diagnosis of lumbar disc herniation using deep learning models," *Frontiers in Bioengineering and Biotechnology*, vol. 11, Sept. 2023, doi: 10.3389/fbioe.2023.1247112.
- [22] G. Maslebu, E. S. D. Kusrini, and A. Setiawan, "Analysis of signal to noise ratio from 1.5 tesla MRI head coil phantom image on daily quality assurance," *Journal of Physics: Conference Series*, vol. 1524, no. 1, p. 012026, Apr. 2020, doi: 10.1088/1742-6596/1524/1/012026.
- [23] Mri\_Master. "MRI Resolution and Image Quality." Available: <https://mrimaster.com/index-4/> (accessed Jul. 24, 2024).
- [24] K. D. Meadows, C. L. Johnson, J. M. Peloquin, R. G. Spencer, E. J. Vresilovic, and D. M. Elliott, "Impact of pulse sequence, analysis method, and signal to noise ratio on the accuracy of intervertebral disc T2 measurement," *JOR SPINE*, vol. 3, no. 3, p. e1102, Sept. 2020, doi: 10.1002/jsp2.1102.
- [25] V. M. Runge and J. T. Heverhagen, "Number of Averages," in *The Physics of Clinical MR Taught Through Images*, V. M. Runge and J. T. Heverhagen, Eds. Cham: Springer International Publishing, 2022, pp. 46-47.
- [26] V. M. Runge and J. T. Heverhagen, "Slice Thickness," in *The Physics of Clinical MR Taught Through Images*, V. M. Runge and J. T. Heverhagen, Eds. Cham: Springer International Publishing, 2022, pp. 48-49.
- [27] K. Lei, A. B. Syed, X. Zhu, J. M. Pauly, and S. V. Vasanawala, "Automated MRI Field of View Prescription from Region of Interest Prediction by Intra-Stack Attention Neural Network," *Bioengineering*, vol. 10, no. 1, doi: 10.3390/bioengineering10010092.
- [28] M. Strzelecki, A. Piórkowski, and R. Obuchowicz, "Effect of Matrix Size Reduction on Textural Information in Clinical Magnetic Resonance Imaging," *Journal of Clinical Medicine*, vol. 11, no. 9, doi: 10.3390/jcm11092526.
- [29] A. M. Shanechi, M. Kiczek, M. Khan, and G. Jindal, "Spine Anatomy Imaging: An Update," *Neuroimaging Clinics of North America*, vol. 29, no. 4, pp. 461-480, Nov. 2019, doi: 10.1016/j.nic.2019.08.001.
- [30] M. D. A. van Gastel et al., "T1 vs. T2 weighted magnetic resonance imaging to assess total kidney volume in patients with autosomal dominant polycystic kidney disease," *American Journal of Nephrology*, 2023, doi: 10.1159/000527874.
- [31] P. Huang, S. Wang, J. Chen, W. Li, and X. Peng, "Lightweight Model for Pavement Defect Detection Based on Improved YOLOv7," *Sensors*, vol. 23, no. 16, doi: 10.3390/s23167112.
- [32] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv e-prints*, p. arXiv:2207.02696, 2022, doi: 10.48550/arXiv.2207.02696.
- [33] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680-1716, 2023, doi: 10.3390/make5040083.
- [34] R. Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," *Scientific Reports*, 2023, doi: 10.1038/s41598-023-39482-w.
- [35] Contributors. "M. YOLOv8 by MMYOLO." Available: <https://github.com/openmmlab/mmyolo/tree/main/configs/yolov8> (accessed Jun. 25, 2024).