

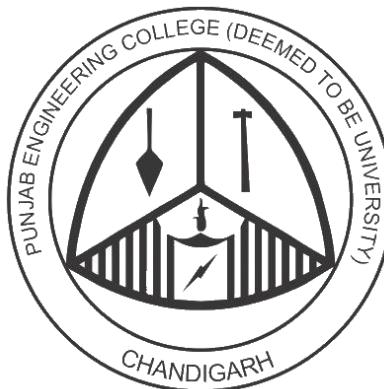
**PUNJAB ENGINEERING COLLEGE  
(DEEMED TO BE UNIVERSITY)**

Department of Computer Science and Engineering

**MINOR PROJECT  
REPORT**

**Lumbar Spine MRI Segmentation: A Deep Learning Approach  
for identifying Spine Regions**

Under the Guidance of **Dr. Alka Jindal** and **Dr. Sudesh Rani**  
Punjab Engineering College (Deemed to be University)



Submitted By:-  
Akanksha Narula(22103041)  
Chhavi Kansal (22103024)  
Mansi Kalra(22103045)  
Harshita Goyal(22103033)  
Srijan Patel(22103115)

## DECLARATION

We hereby declare that the project work entitled "**Lumbar Spine MRI Segmentation: A Deep Learning Approach for identifying Spine Regions**" is an authentic record of our own work carried out at Punjab Engineering College (Deemed to be University), Chandigarh as per the requirements of "Minor Project" for the award of the degree of B.Tech. Computer Science and Engineering, under the guidance and supervision of **Dr Alka Jindal** and **Dr Sudesh Rani**.

We further declare that the information has been collected from genuine & authentic sources and we have not submitted this project report to this or any other university for the award of diploma or degree or certificate examination.

Certified that the above statement made by the students is correct to the best of my knowledge and belief.

**Dr. Alka Jindal** (Assistant Professor, Punjab Engineering College)

**Dr. Sudesh Rani** (Assistant Professor, Punjab Engineering College)

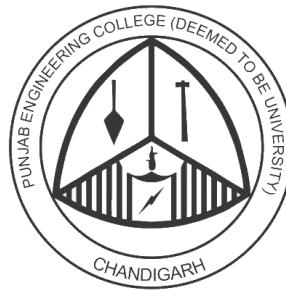
Date: 04/12/2024

Place: Punjab Engineering College, Chandigarh

Akanksha Narula(22103041)  
Chhavi Kansal (22103024)  
Mansi Kalra(22103045)  
Harshita(22103033)  
Srijan Patel(22103115)

# **PUNJAB ENGINEERING COLLEGE (DEEMED TO BE UNIVERSITY)**

Department of Computer Science and Engineering



## **CERTIFICATE**

It is Certified that the Project work entitled **Lumbar Spine MRI Segmentation: A Deep Learning Approach for identifying Spine Regions** submitted by **Akanksha Narula, Chhavi Kansal, Mansi, Harshita Goyal and Srijan Patel** for the fulfillment of Minor Project offered by Punjab Engineering College (Deemed to be University) during the academic year **2024-25** is an original work carried out by the students under my supervision and this work has not framed any basis for the award of and Degree, Diploma or such other titles. All the work related to the project is done by these candidates themselves. The approach towards the subject has been sincere and scientific.

Dr. Alka Jindal  
CSE Department, (Date 4/12/2024)  
Punjab Engineering College (Deemed to be University)

Dr. Sudesh Rani  
CSE Department, (Date 4/12/2024)  
Punjab Engineering College (Deemed to be University)

## ACKNOWLEDGEMENT

We would like to take this opportunity to thank our college Punjab Engineering College (Deemed to be University), Chandigarh and Department of Computer Science and Engineering for giving us an opportunity to work on this project.

We are immensely grateful to our project mentor **Dr Sudesh Rani** (Assistant Professor, Department of Computer Science and Engineering) and **Dr. Alka Jindal**(Assistant Professor, Department of Computer Science and Engineering) whose continuous guidance, technical support and moral support at times of difficulty helped us to achieve milestones in the given time. She has been a great source of knowledge.

We are also thankful to the Committee for evaluation who gave their valuable feedback and guidelines for the betterment and completion of this project.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Science and Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Akanksha Narula(22103041)  
Chhavi Kansal (22103024)  
Mansi Kalra(22103045)  
Harshita(22103033)  
Srijan Patel(22103115)

## ABSTRACT

The application of deep learning in the medical field has been on a constant rise over the years and many medical complexities and challenges can be modeled as standard deep learning problems. Lumbar spine MRI segmentation and classification of spine disorders is such a field which can be framed as a computer vision problem. The first step towards automatic assessment of lumbar spine MRI is segmentation of relevant anatomical structures, such as the vertebrae, intervertebral discs (IVDs) and the spinal canal. The advancement in Computer Vision models over the last few years have greatly improved the accuracy and prediction time and thus are vastly being used in real-life applications.

The following study aims at developing a mechanism to assist in assessment of lumbar spine disorders, thus reducing human effort and skill required for the task, and also solving the problem of its detection and diagnosis. On our input dataset SPIDER, we primarily apply data augmentation using the Albumentations library to broaden our training data. Normalization has then been applied to the MRI images to give better contrast and resolution to the MRI's. Then we train our data of over 3535 total unique images to identify regions. Mask labels 1-9 are L5 and cephalic(towards the head), 10 is spinal canal, 11-19 are L5/S1 disc space and cephalic. After that we constructed our segmentation model via weighted ensemble technique. **MedFusionNet** outperformed individual and other ensemble models, achieving the highest mean IoU and Dice Coefficients.

Finally, a functional web application has been developed using the Python Streamlit framework to allow the users to provide input in the form of Lumbar Spine MRI and receive the results of segmentation.

# TABLE OF CONTENTS

DECLARATION	II
CERTIFICATE	III
ACKNOWLEDGEMENT	IV
ABSTRACT	V
TABLE OF CONTENTS	VI
LIST OF FIGURES	VIII
LIST OF TABLES	IX
LIST OF ABBREVIATIONS	X
CHAPTER 1: INTRODUCTION	1
1.1 Motivation	1
1.2 Problem Statement	1
1.3 Methodology	2
CHAPTER 2: BACKGROUND	4
2.1 Anatomy and Significance of Lumbar Spine	4
2.2 Role of MRI in Lumbar Spine Diagnosis	5
2.3 Deep Learning in Image Segmentation	6
2.4 Literature Review	7
2.5 Research Gaps	10
CHAPTER 3: PROPOSED WORK	11
3.1 Image Segmentation in Deep Learning	11
3.2 Data Augmentation	11
3.3 Data Splitting with K-Fold	12
3.4 Batch Processing with Pytorch - Data Loader	12
3.5 Model Architectures	13

3.6 Ensemble Model	14
3.7 Streamlit	14
CHAPTER 4: IMPLEMENTATION DETAILS	
4.1 Data Collection	15
4.2 Data Pre-Processing	15
4.3 Tools and Technologies	16
4.4 Web Application	17
CHAPTER 5: RESULTS AND DISCUSSION	
5.1 Evaluation Metrics	18
5.2 Model Training	19
5.3 Model Testing	20
5.4 Performance Analysis	22
CHAPTER 6: Conclusion and Future Work	
REFERENCES	25

## LIST OF FIGURES

Figure 01 :	WorkFlow Diagram	03
Figure 02 :	Lumbar Spine	04
Figure 03 :	Lumbar Spine MRI	05
Figure 04 :	A Survey on Medical Image Segmentation Based on Deep Learning Techniques	06
Figure 05 :	Basic Architecture of Image Segmentation Model	11
Figure 06 :	U-NET Architecture	13
Figure 07 :	U-NET++ Architecture	13
Figure 08 :	ResUNet Architecture	14
Figure 09 :	DeepLabV3 Architecture	14
Figure 10 :	Initial Dataset	15
Figure 11 :	Transformed Images after Augmentation	17
Figure 12 :	Web Application	18
Figure 13 :	Training of Model	19
Figure 14 :	Algorithm used in weighted ensemble	20
Figure 15 :	Weighted Ensemble Results	21
Figure 16 :	Graphical Comparison of Testing Results	22

## LIST OF TABLES

Table 1	Literature review analysis	09
Table 2	Comparison of Testing results	22

## LIST OF ABBREVIATIONS

<b>LBP</b>	Lower Back Pain
<b>IVD</b>	Intervertebral Discs
<b>MRI</b>	Magnetic Resonance Imaging
<b>CNN</b>	Convolutional Neural Network
<b>SPIDER</b>	Spine Public Image Dataset for Evaluating Research
<b>T1</b>	T1-Weighted Imaging (MRI sequence)
<b>T2</b>	T2-Weighted Imaging (MRI sequence)
<b>U-Net</b>	Universal Network
<b>IoU</b>	Intersection over Union
<b>MedFusionNet</b>	Medical Fusion Network (proposed ensemble method)

# CHAPTER 1: INTRODUCTION

The lumbar spine, consisting of vertebrae, intervertebral discs (IVDs), and the spinal canal, plays an essential role in supporting, flexing, and protecting the spinal cord and nerve roots. Disorders like low back pain (LBP) are among the most prevalent causes of disability globally, emphasizing the importance of accurate diagnostics. Lumbar MRI is a crucial tool in diagnosing such conditions due to its ability to provide detailed views of spinal structures. However, only a small fraction of lumbar MRI scans, approximately 13%, contribute effectively to clinical decision-making, highlighting inefficiencies in current diagnostic practices. Manual segmentation of anatomical structures is labor-intensive, prone to inconsistencies, and difficult to scale for large datasets, necessitating automated and reliable solutions for precise segmentation.

## 1.1 Motivation

The motivation behind this research is to enhance diagnostic efficiency and accuracy in lumbar spine MRI analysis. Traditional manual segmentation methods face key limitations like labor-intensive and delays timely diagnosis, variability between operators affects accuracy, not suitable for handling large datasets and poor image quality and unclear boundaries introduce errors.

Deep learning, particularly CNNs, can address these challenges by automatically learning relevant features from MRI images, providing high-quality, consistent segmentation with minimal human intervention. This research aims to leverage advanced deep learning models and ensemble techniques to support radiologists and orthopedic specialists in improving diagnostic accuracy.

## 1.2 Problem Statement

This research seeks to develop an automated system that accurately segments lumbar spine regions from MRI images. The goal is to design a model capable of detecting and classifying spinal conditions at varying severity levels, thereby reducing human error and enhancing diagnostic capabilities in clinical settings.

**Problem Statement: Automatic Segmentation of Lumbar Spine from MRI Images Using Deep Learning Techniques.**

## 1.3 Methodology

To develop an efficient and accurate deep learning-based system, the following methodology is proposed. The methodology employed in this research is visually represented in Figure 1.

**1. Review of Existing Literature:** A thorough review of current state-of-the-art techniques is conducted to identify suitable models for lumbar spine segmentation, focusing on U-Net, DeepLabV3, and ensemble learning strategies.

**2. Dataset Collection:** Utilize the SPIDER dataset, a publicly available multi-center collection of sagittal T1 and T2 MRI scans, which provides diverse data representative of different spinal conditions.

### 3. Preprocessing of Dataset:

Image Resizing: All images are resized to 128x128 pixels for uniform input.

Normalization: Pixel intensities are normalized to [0, 1].

Data Augmentation: Techniques like rotation, flipping, and translation are applied to enhance model generalization and reduce overfitting.

**4. Model Selection and Training:** Train individual models, including U-Net, U-Net++, ResUNet, and DeepLabV3 on the preprocessed dataset.

**5. Ensemble Approach:** Develop an ensemble model, using weighted average method taking pairwise each of the above models.

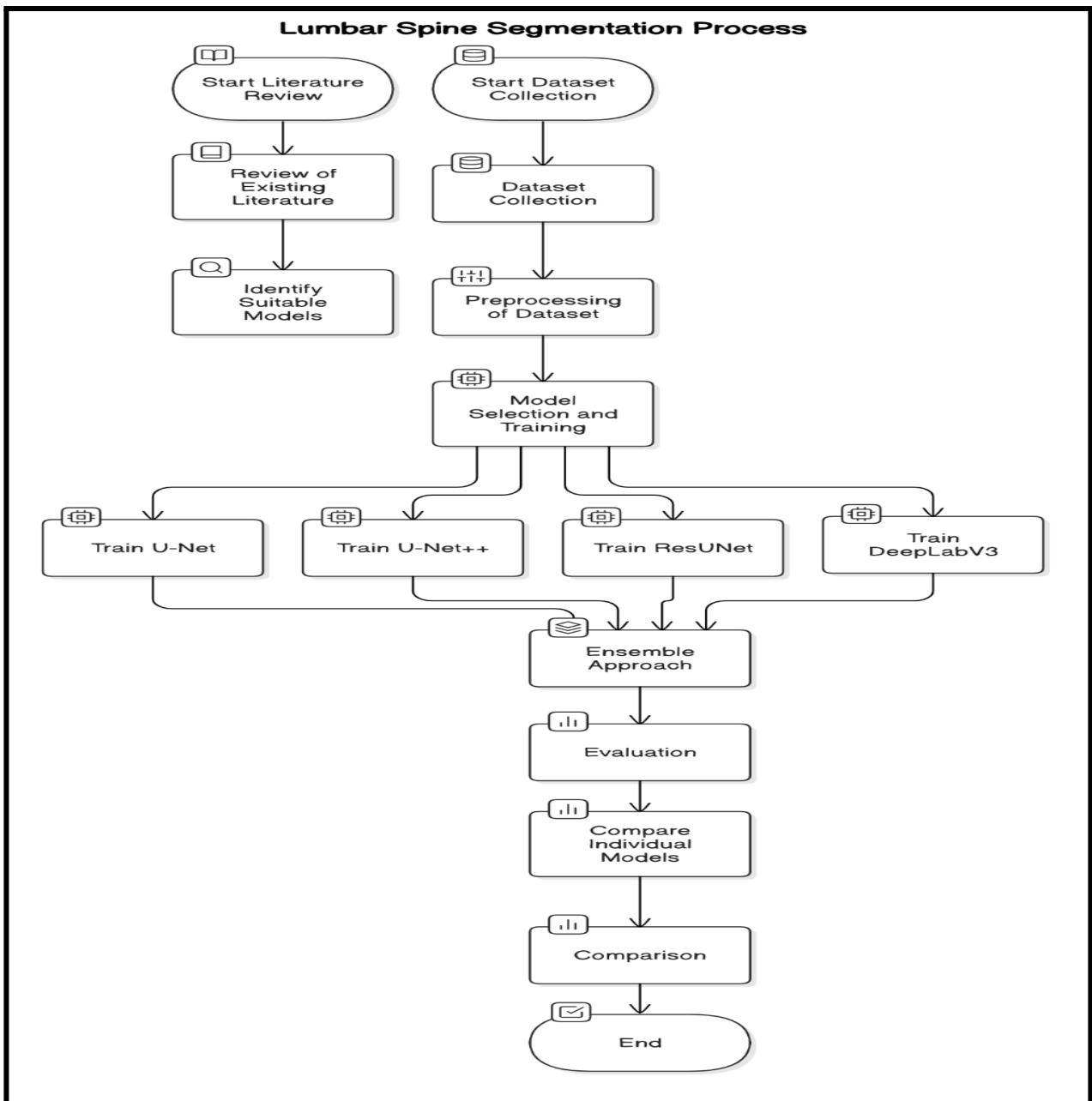
### 6. Evaluation:

Performance is assessed using:

Dice Score: Measures overlap between predicted and ground truth masks.

Intersection over Union (IoU): Evaluates the ratio of overlap to the union of predicted and ground truth areas.

**7. Comparison:** Results from individual models and the ensemble approach are compared. The ensemble method is expected to outperform individual models, offering higher accuracy and robustness by integrating diverse feature representations and named MedFusionNet.



**Figure 1:** Workflow Diagram

## CHAPTER 2: BACKGROUND

### 2.1 Anatomy and Significance of the Lumbar Spine

The lumbar spine, located in the lower back, consists of five vertebrae (L1-L5) and plays a critical role in supporting the body's weight, enabling flexibility, and protecting the spinal cord and associated nerve roots. Structurally, the lumbar spine includes intervertebral discs, facet joints, and the spinal canal, each contributing to spinal stability and mobility.

The lumbar region is particularly susceptible to degenerative disorders due to the significant mechanical stress it endures. Conditions such as disc herniation, degenerative disc disease, and spinal stenosis frequently affect this area, leading to chronic back pain, reduced mobility, and in severe cases, neurological deficits. According to the **Global Burden of Disease (GBD) 2019** study, low back pain is the leading cause of disability worldwide, with approximately **540 million people** affected at any given time. In the United States alone, low back pain is responsible for annual healthcare expenditures exceeding **\$100 billion**, highlighting the socioeconomic impact of lumbar spine disorders.

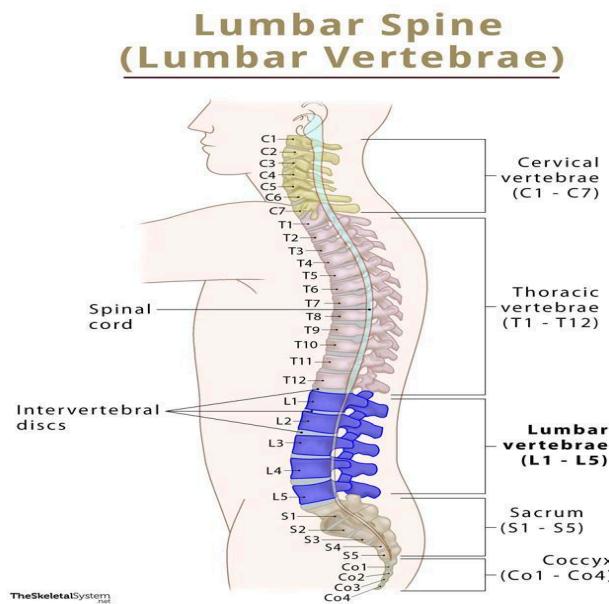
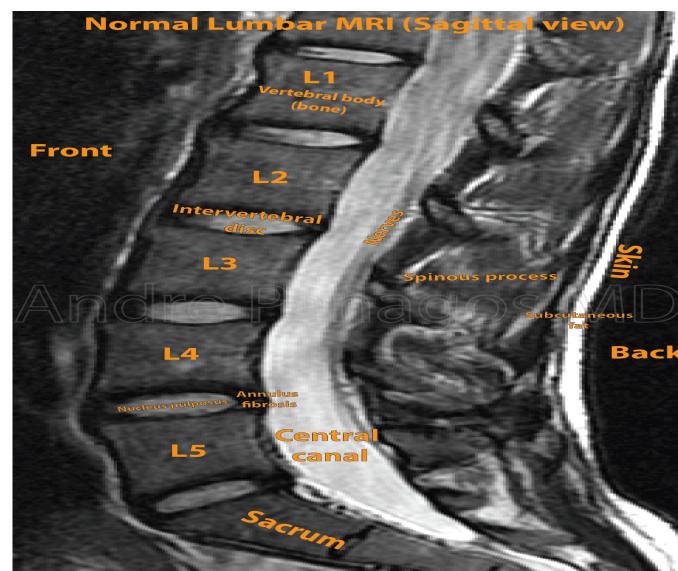


Figure 2 : Lumbar Spine

## 2.2 Role of MRI in Lumbar Spine Diagnosis

Magnetic Resonance Imaging (MRI) has emerged as the gold standard for diagnosing lumbar spine disorders due to its ability to provide high-resolution, three-dimensional images of soft tissues without the use of ionizing radiation. Unlike other imaging modalities such as X-ray and CT scans, MRI can accurately visualize the intervertebral discs, spinal cord, nerve roots, and surrounding soft tissues, making it invaluable in diagnosing conditions like herniated discs, spinal stenosis, and tumors.

Despite its advantages, the interpretation of lumbar spine MRI scans remains a time-consuming and labor-intensive process. Radiologists must manually segment and analyze the images, which can lead to variability in diagnosis due to subjective interpretation. Studies indicate that only **13% of lumbar spine MRI scans** significantly influence clinical decision-making, underscoring the need for automated segmentation tools to enhance diagnostic accuracy and reduce the workload on healthcare professionals.



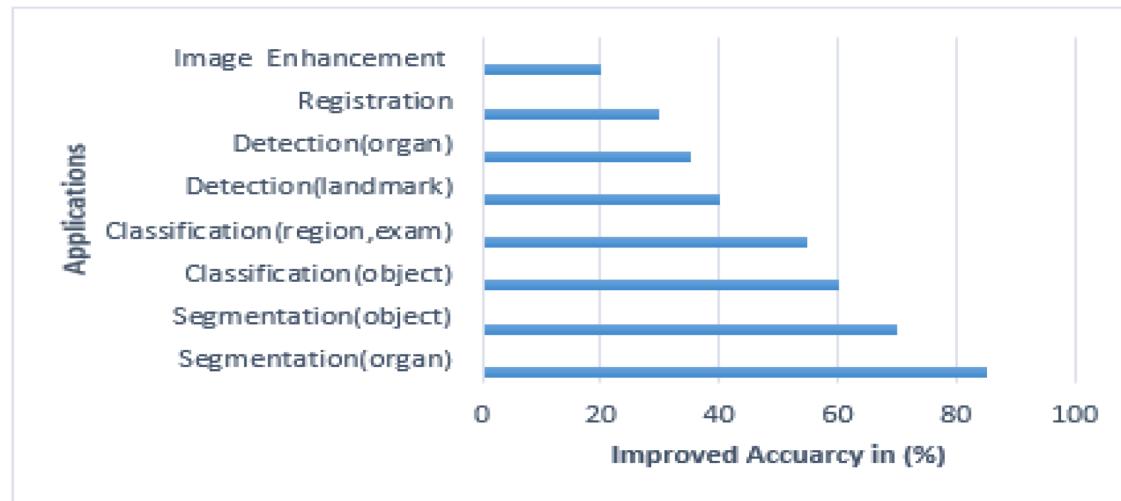
**Figure 3 :Lumbar Spine MRI**

## 2.3 Deep Learning in Medical Image Segmentation

Deep learning (DL), a subset of artificial intelligence, has revolutionized medical image analysis by enabling computers to automatically learn and extract relevant features from complex datasets. Convolutional Neural Networks (CNNs) are particularly well-suited for image segmentation tasks due to their ability to detect spatial hierarchies and patterns in images.

In medical image segmentation, CNN-based architectures such as U-Net, ResUNet, and DeepLabV3 have demonstrated remarkable performance. U-Net, a fully convolutional network, has become a benchmark for biomedical image segmentation, achieving Dice scores exceeding **85%** in various applications. ResUNet, which integrates residual connections, and DeepLabV3, which employs atrous convolutions for multi-scale context, have further improved segmentation accuracy, particularly in complex anatomical structures like the lumbar spine.

These architectures have been applied to a wide range of medical imaging tasks, from brain tumor segmentation to lung nodule detection. However, their application to lumbar spine MRI segmentation remains an emerging field, with significant potential to improve diagnostic workflows and patient outcomes.



**Figure 4 :** A Survey on Medical Image Segmentation Based on Deep Learning Techniques

## 2.4 Literature Review

In the field of lumbar spine MRI segmentation, deep learning techniques have shown great promise in automating diagnostic processes, particularly in identifying abnormalities and guiding treatment decisions. A variety of model architectures have been explored to address the unique challenges of lumbar spine imaging, such as class imbalance, complex anatomical features, and limited data availability. Models such as U-Net, SegNet, and Attention U-Net have demonstrated excellent performance in segmenting key structures like vertebrae, spinal canal, and intervertebral discs (IVDs). Advanced variations, including U-Net++ and nn-UNet, have further refined segmentation capabilities, offering improved accuracy and boundary delineation. Additionally, hybrid models incorporating residual and attention mechanisms, such as RAU-Net, have been proposed to enhance feature extraction and classification. The combination of custom loss functions like IoU and Dice Loss has proven effective in mitigating class imbalance, ensuring precise segmentation. These innovations, along with robust preprocessing techniques, have significantly advanced the state-of-the-art in lumbar spine MRI segmentation, informing the approach adopted in this study.

Table 1: Literature Review Analysis

Sr No.	Title	Publication and Year	CNN Model	Results
1	Lumbar Spine Segmentation in MR Images: A Dataset and a Public Benchmark	Scientific Data(2024)	nnU-Net, Iterative Instance Segmentation (IIS)	nnU-Net achieved a mean Dice coefficient of 0.92 for spinal canal, 0.93 for vertebrae; ASD for vertebrae was 0.48 mm

2	Deep Learning Based Vertebral Body Segmentation with Extraction of Spinal Measurements and Disorder Disease Classification	Biomedical Signal Processing and Control(2022)	ResNet-UNet (achieved DSC: 0.97, IoU: 0.86), UNet, VGG-UNet, MobileNets-UNet, PSPNet, MobileNets-SegNet, ResNet-PSPNet, ResNet-SegNet, VGG-SegNet
3	Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks	IEEE Access (2019)	Achieved excellent segmentation performance for lumbar spinal stenosis detection with high inter-rater agreement scores and Cohen's kappa > 0.80. SegNet-TL80 identified as the best-performing .
4	Automatic Segmentation of Lumbar Spine MRI Images Based on Improved Attention U-Net	Computational Intelligence and Neuroscience (2022)	Achieved accuracy of 95.50%, recall of 94.53%, and Dice similarity coefficient of 95.01%, outperforming SVM, FCN, R-CNN, U-Net, and Attention U-Net.
5	Pioneering Precision in Lumbar Spine MRI Segmentation with Advanced Deep Learning and Data Enhancement	Preprint (2024)	Achieved superior segmentation accuracy for vertebrae, spinal canal, and IVDs with a mean Dice coefficient > 96%. Incorporated leaky ReLU and Glorot initializer to address class imbalance and improve training stability..

6	Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning	arXiv (2018)	Multi-input ResNeXt-50 and U-Net architecture	Achieved: Vertebral segmentation success rate of <b>94%</b> , stenosis grading accuracy: <b>80.4% (spinal canal)</b> and <b>78.1% (foraminal stenosis)</b> when using combined sagittal and axial inputs. Also achieved AUC: <b>0.983 (spinal canal)</b> and <b>0.961 (foraminal stenosis)</b> for binary classification
7	Variability of Manual Lumbar Spine Segmentation	International Journal of Spine Surgery (2012)	None (Manual Segmentation)	Achieved root-mean-square error: <b>0.39 mm (accuracy)</b> , <b>0.35 mm (inter-user precision)</b> , and <b>0.33 mm (intra-user precision)</b> .
8	Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models	Sensors (2022)	YOLOv5s and HED U-Net	Achieved mean average precision (mAP): <b>0.975</b> for YOLOv5. Mean errors: <b>LLA: 0.29°</b> , <b>LSA: 0.38°</b> . Classified lumbar lordosis (hypo, normal, hyper) with <b>100% accuracy</b> using edge-based segmentation.
9	Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach	European Spine Journal (Springer-Verlag GmbH Germany, part of Springer Nature) Published: 12 March 2019	A fully convolutional neural network	Addresses conditions including: Sagittal deformities, Coronal deformities, Degenerative phenomena A fully convolutional neural network was trained to predict the location of 78 landmarks and use data from 493 spine reconstructions

10	Using a deep learning network to recognise low back pain in static standing	Human Factors and Ergonomics in Manufacturing & Service Industries Published: 3 July 2018	Deep Learning networks	High Accuracy: Deep learning networks achieved up to 97.2% precision and recall in distinguishing chronic low back pain patients from healthy individuals

## 2.5 Research Gaps

Although extensive research has been conducted in this field, still some major research gaps are left unexplored. These gaps include:

**Limited Dataset Diversity:** Models trained on homogeneous datasets fail to generalize across diverse populations and imaging conditions.

**Imbalanced and Inconsistent Data:** Annotated datasets often have class imbalances and variability in ground truth due to subjective labeling.

**Lack of High-Quality Datasets:** Creating annotated datasets is labor-intensive, limiting their availability.

**Underexplored Techniques:** Advanced models like U-Net variants and lightweight architectures such as MobileNet remain underutilized.

**High Model Variance:** Many models overfit to training data, reducing accuracy on unseen cases.

**Limited Integration of Hybrid Approaches:** Combining traditional methods with deep learning is rarely explored.

**Minimal Multi-task Learning:** Simultaneous segmentation, classification, and detection tasks are underdeveloped.

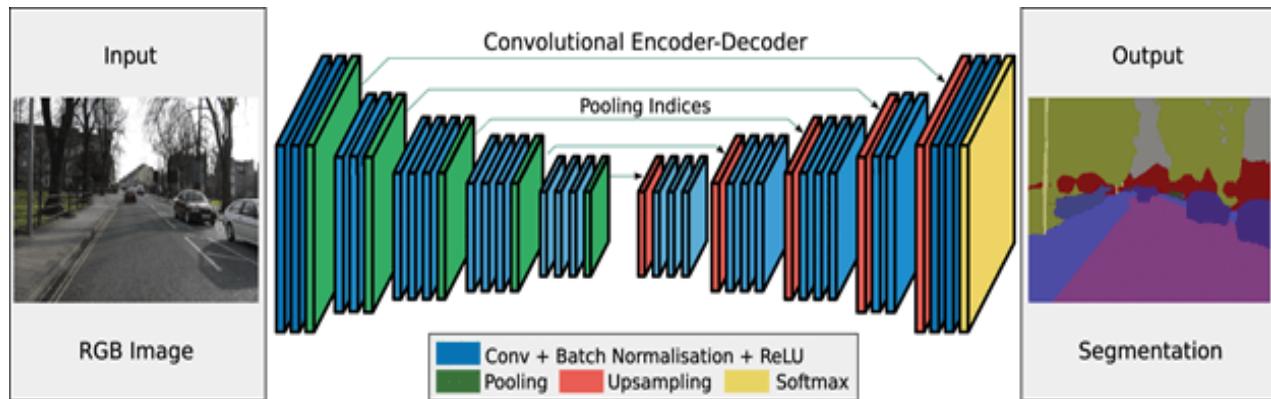
Efforts have been made to reduce the above mentioned research gaps to an appreciable extent through this project.

## CHAPTER 3: PROPOSED WORK

Our work followed an organized approach to implement image classification using deep learning. We utilized Albumentations libraries for data augmentation to reduce the effect of the Spine MRI dataset being imbalanced. This was followed by image quality improvement using RGB conversions and normalization. After that k-cross validation and batch processing techniques were used for efficient training of various models including ensembles.

### 3.1 Image Segmentation in Deep Learning

Image segmentation is a crucial task in computer vision, aimed at dividing an image into distinct regions or objects. Unlike image classification, which assigns a single label to an entire image, segmentation provides pixel-level predictions, making it a fundamental approach in applications requiring detailed scene understanding. In medical imaging, segmentation highlights regions of interest, such as tissues or abnormalities, enabling detailed analysis and supporting clinical decision-making. Our segmentation task has 20 classes. Mask labels 1-9 are L5 and cephalic(towards the head), 10 is spinal canal, 11-19 are L5/S1 disc space and cephalic. The entire procedure for segmentation is shown in the figure below:



**Figure 5:** Basic Architecture of Image Segmentation Model

### 3.2 Data Augmentation

Usually datasets are imbalanced in distribution with a few particular classes forming up the major portion of the dataset and the others being in a marginal proportion. Hence, there is the need to get

more images of the classes having less number of samples. For this data augmentation techniques are used. Data augmentation is a technique of artificially increasing the training set by creating modified copies of a dataset using existing data. It includes making minor changes to the dataset or using deep learning to generate new data points. A widely used technique for this is the Albumentations library. Albumentations efficiently implements a rich variety of image transformation operations that are optimized for performance, and does so while providing a concise, yet powerful image augmentation interface for different computer vision tasks, including object classification, segmentation, and detection.

### 3.3 Dataset Splitting with K-Fold Cross-Validation

K-Fold cross-validation is a robust technique in machine learning that involves partitioning the dataset into  $k$  equal-sized folds to ensure thorough evaluation and training. Each fold is used as a validation set exactly once, while the remaining  $k-1$  folds are used for training. This process is repeated  $k$  times, and the model's performance is averaged across all iterations, providing a comprehensive assessment of its generalization ability. In our work, this method was crucial for addressing potential data imbalances and ensuring that the model is exposed to diverse data distributions during training. By employing K-Fold cross-validation, we reduced the risk of overfitting, as the model was tested on unseen data in every fold. In our study, the dataset was split into 70% for training, 25% for testing, and 5% for validation, ensuring a balanced approach for model training, evaluation, and fine-tuning.

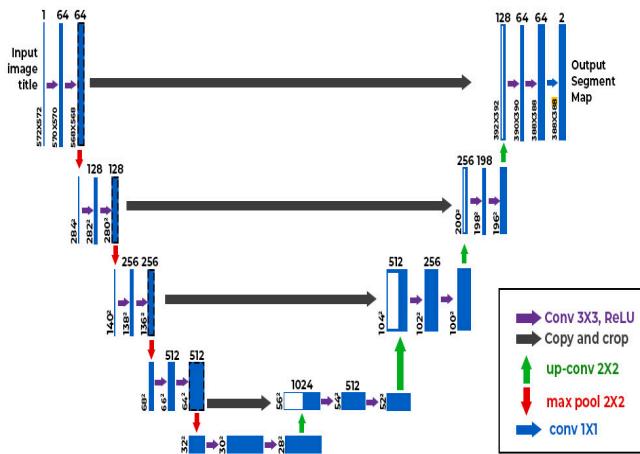
### 3.4 Batch Processing with PyTorch DataLoader

Batch processing is a core technique in deep learning that divides the dataset into smaller subsets, or mini-batches, for model training. This approach improves training efficiency by allowing the model to process multiple samples simultaneously, leveraging GPU parallelism. Using PyTorch's DataLoader, data is loaded and preprocessed on-the-fly, ensuring efficient memory management and reducing input-output bottlenecks. The DataLoader also facilitates shuffling and augmentation, ensuring diversity in batches and preventing the model from learning spurious patterns. Batch processing stabilizes gradient updates by averaging over the mini-batch, leading to smoother convergence during optimization. Furthermore, by dividing the dataset into mini-batches, the likelihood of overfitting is reduced, as the model generalizes better across smaller, diverse subsets of data.

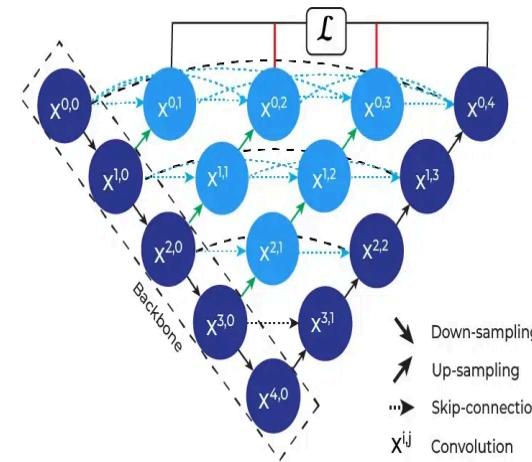
### 3.5 Model Architectures

The segmentation process employed advanced deep learning models, each tailored to enhance the performance of medical image segmentation:

1. U-Net: U-Net is a Convolutional Neural Network (CNN) architecture specifically designed for biomedical image segmentation. Its encoder-decoder structure captures both local and global features effectively. The encoder extracts feature representations, while the decoder reconstructs the segmented output. Skip connections bridge the encoder and decoder, ensuring the preservation of spatial information and facilitating the propagation of gradients, making U-Net highly effective in handling complex medical images.
2. U-Net++: An improved version of U-Net, U-Net++ introduces nested and dense skip pathways, allowing for more detailed feature extraction and refinement. These nested skip connections mitigate the semantic gap between encoder and decoder features, leading to improved segmentation accuracy. U-Net++ is particularly beneficial for applications requiring precise boundary delineation.

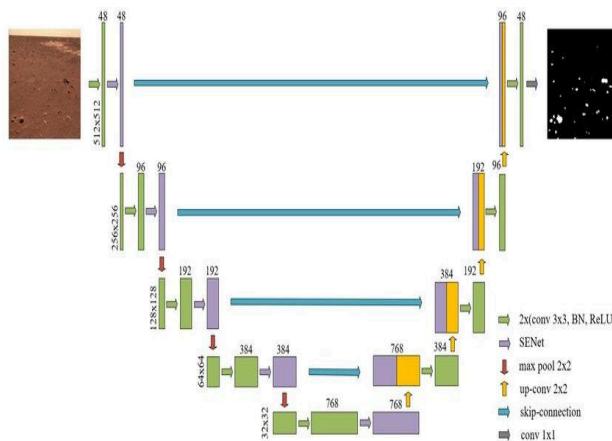


**Figure 6 :** U-Net Architecture

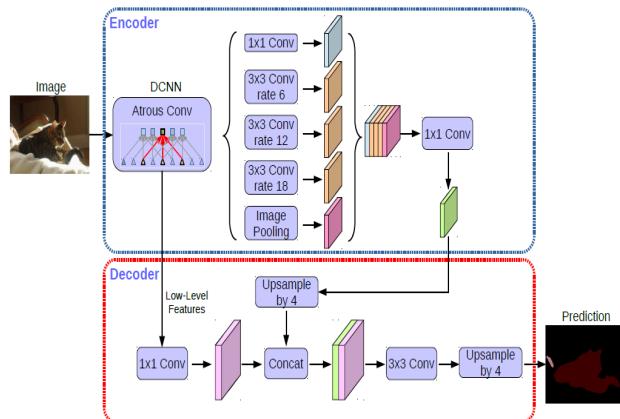


**Figure 7 :** U-Net++ Architecture

3. DeepLabV3: DeepLabV3 employs atrous spatial pyramid pooling (ASPP) to capture multi-scale contextual information. Atrous convolutions allow for a larger field of view without increasing computational complexity, enabling the model to detect features at varying resolutions. This makes DeepLabV3 suitable for segmenting regions of interest with complex structures, such as spinal anatomy in MRI scans.
4. ResUNet: ResUNet combines the strengths of U-Net with residual connections, which help mitigate the vanishing gradient problem and facilitate the training of deeper networks. Residual connections enable the model to reuse features from earlier layers, enhancing performance by ensuring efficient gradient flow and improving the representation of complex patterns in medical images.



**Figure 8 :** ResUNet Architecture



**Figure 9 :** DeepLabV3 Architecture

### 3.6 Ensemble Model

In the ensemble strategy, the predictions from each model (e.g., U-Net and U-Net++) were weighted and combined to form a final output. The weights for each model were optimized through an iterative process to maximize the overall IoU score. This process involved testing various weight combinations to identify the configuration that achieved the best segmentation results. By adjusting the influence of each model, the ensemble was able to leverage the complementary strengths of different architectures, leading to improved performance and reduced susceptibility to individual model error.

### 3.7 Streamlit

Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. Web apps can be instantly developed and deployed easily using Streamlit. Streamlit allows us to write an app the same way we write python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app. Hence, it is widely used to build web applications for machine learning models.

## CHAPTER 4: IMPLEMENTATION DETAILS

### 4.1 Data Collection

A large publicly available multi-center lumbar spine magnetic resonance imaging (MRI) dataset SPIDER was used, with reference segmentations of vertebrae, intervertebral discs (IVDs), and spinal canal. The dataset includes 447 sagittal T1 and T2 MRI series from 218 studies of 218 patients with a history of low back pain. The data was collected from four different hospitals in .mha format.

Hospital	Studies	T1	T2	T2 SPACE	Voxel size range (min - max)(mm)	Sex (% female)
UMC	41	39	39	41	(3.24 × 0.27 × 0.47) - (3.34 × 0.59 × 0.85)*	55%
RH1	43	43	37	0	(0.46 × 0.46 × 4.20) - (9.63 × 1.06 × 1.06)	58%
RH2	44	24	44	0	(0.46 × 0.46 × 4.20) - (5.17 × 1.00 × 1.23)	59%
OH	90	90	90	0	(3.15 × 0.24 × 0.24) - (3.39 × 0.83 × 1.02)	68%
Total	218	196	210	41	(3.15 × 0.24 × 0.24) - (9.63 × 1.06 × 1.23)*	63%

**Figure 10 :** Initial Dataset

### 4.2 Data Preprocessing

#### 1. Data Loading and Preprocessing with SimpleITK

The MRI scans were initially in MetaImage (.mha) format. SimpleITK was used to read and process these images, converting them into a manipulatable format. The scans were reoriented to a standard RPI (Right, Posterior, Inferior) coordinate system using the DICOMOrientImageFilter() function. The reoriented pixel data was extracted as NumPy arrays using sitk.GetArrayFromImage().

#### 2. Conversion of Images

The MRI slices were converted to PNG format to ensure compatibility with deep learning frameworks. Each image was resized to a fixed dimension of 128×128 pixels using the PIL library. This standardization ensured that the input size matched the requirements of the segmentation models.

#### 3. RGB Conversion

Even though the MRI scans are grayscale, they were converted to RGB format using image.convert('RGB'). This step was necessary as many deep learning models expect three-channel (R, G, B) input data.

#### 4. Normalization

The pixel intensity values of the images were normalized to the range [0, 1] by dividing by 255.0. Normalization speeds up training by ensuring consistent input scales for the model.

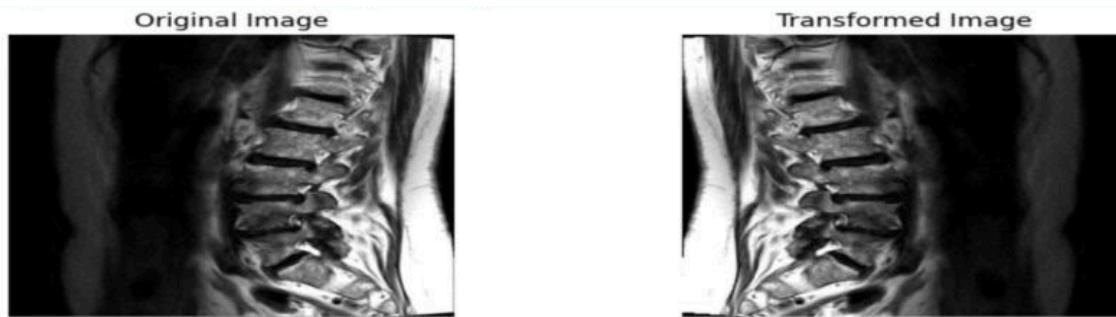
#### 5. Mask Processing

Segmentation masks were preprocessed to match the number of output classes required by the model. Masks were restricted to valid class ranges (20 in this case) and one-hot encoded using PyTorch's `torch.nn.functional.one_hot` function.

### 4.3 Tools and Technologies

#### For Data Pre-Processing:

- Albumentations library was used for data augmentation to generate a larger number of images.\



**Figure 11 :** Transformed images after augmentation

#### For Model Building, Integration and Testing:

- Visual Studio Code and Kaggle were the platforms where all the code was written.
- Pytorch was used as the primary library for importing the pretrained architectures and implementation of our segmentation model.

Pytorch provides a collection of workflows with intuitive, high-level APIs for both beginners and experts to create machine learning models in numerous languages. Developers have the option to deploy models on a number of platforms such as on servers, in the cloud, on mobile and edge devices,

in browsers, and on many other JavaScript platforms. This enables developers to go from model building and training to deploy much more easily.

### **Data Used:**

Total of 447 Spine X-MRI images were present in the original dataset. These were augmented to 3535 images and their corresponding masks, and used to train the model.

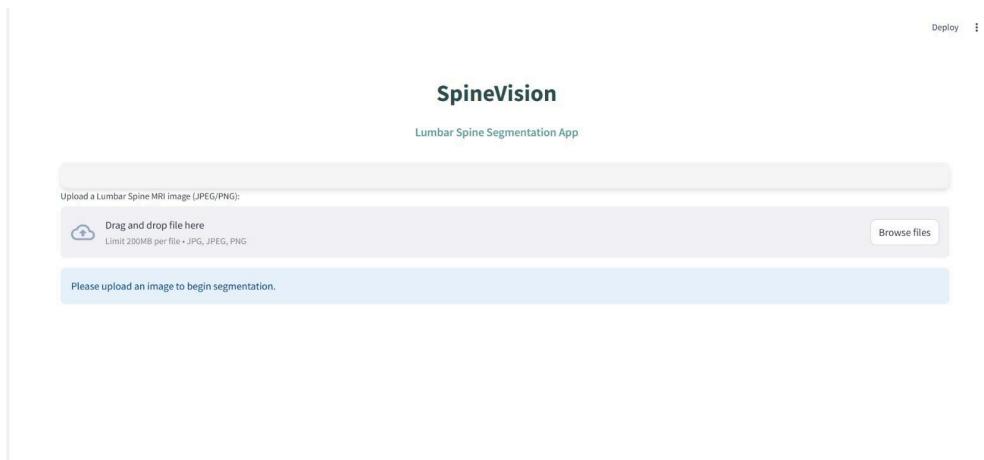
### **Labeling of Images:**

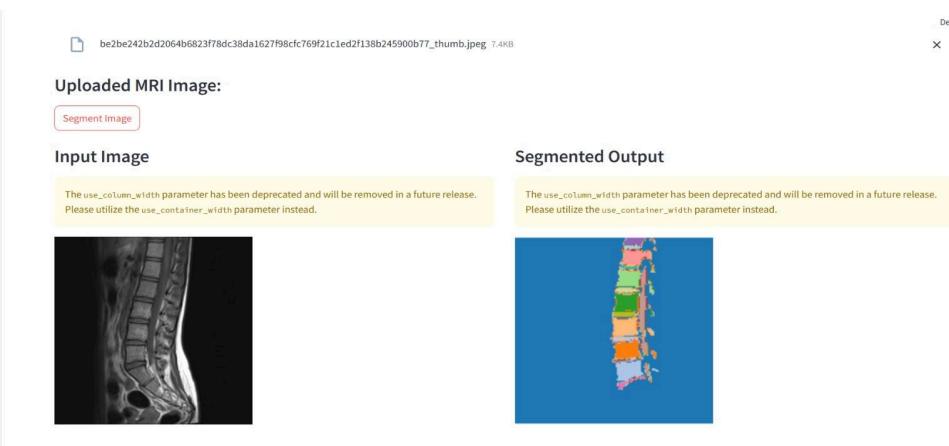
The .png images were segmented according to the labels:

1: L5, 2: L4, 3: L3, 4: L2, 5: L1, 6: T12, 7: unknown, 8: unknown, 9: unknown, 10: spinal canal, 11: L5-S1, 12: L4-L5, 13: L3-L4, 14: L2-L3, 15: L1-L2, 16: T12-L1, 17: unknown, 18: unknown, 19: unknown

## **4.4 Web Application**

The trained model was then saved in an .hdf5 file format. This model file was then integrated in a Python application using the Streamlit framework. A user interface was built where users can provide Spine MRI images as input and the application will run the model and predict the segmented mask as output. The screenshots of the functional web application have been provided below.





**Figure 12 :** Web Application

## CHAPTER 5: RESULTS AND DISCUSSION

### 5.1 Evaluation Metrics

#### Dice Score (Dice Coefficient):

- Measures the overlap between the predicted segmentation and the ground truth.
- Calculated as twice the area of overlap between the predicted and ground truth, divided by the total number of elements in both samples.
- Ranges from 0 (no overlap) to 1 (perfect overlap).
- Higher values indicate better segmentation performance.

#### Intersection over Union (IoU):

- Measures the ratio of the area of overlap between the predicted and ground truth masks to the area of their union.
- Ranges from 0 to 1, where a higher IoU indicates better performance.
- Higher values reflect a closer match between the predicted mask and the ground truth.

### 5.2 Model Training

- **Training Setup:** The models (U-Net, U-Net++, ResUNet, DeepLabV3, and their pairwise ensemble) were trained on the SPIDER dataset, with each image resized to 128x128 pixels.
- **Optimization:** The Adam optimizer was used with a learning rate of 0.001, and the models were trained for 50 epochs with a batch size of 64.
- **Train-Test Split:** The dataset was divided into 70% for training, 5% for validation and 25% for testing.
- **Data Augmentation:** Techniques such as rotation, flipping, and brightness adjustment were applied to improve generalization.



Figure 13(a): UNet

Epoch: 37/50 | Train Loss: 0.5215 | Val Loss: 0.5812 | Dice Score: 0.9861  
Epoch: 38/50 | Train Loss: 0.5214 | Val Loss: 0.5806 | Dice Score: 0.9861  
Epoch: 39/50 | Train Loss: 0.5215 | Val Loss: 0.5820 | Dice Score: 0.9861  
Epoch: 40/50 | Train Loss: 0.5219 | Val Loss: 0.5826 | Dice Score: 0.9860  
Epoch 00041: reducing learning rate of group 0 to 1.0000e-08.  
Epoch: 41/50 | Train Loss: 0.5215 | Val Loss: 0.5799 | Dice Score: 0.9861  
Epoch: 42/50 | Train Loss: 0.5217 | Val Loss: 0.5798 | Dice Score: 0.9862  
Epoch: 43/50 | Train Loss: 0.5216 | Val Loss: 0.5804 | Dice Score: 0.9861  
Epoch: 44/50 | Train Loss: 0.5218 | Val Loss: 0.5788 | Dice Score: 0.9862  
Epoch: 45/50 | Train Loss: 0.5213 | Val Loss: 0.5809 | Dice Score: 0.9861  
Epoch: 46/50 | Train Loss: 0.5215 | Val Loss: 0.5797 | Dice Score: 0.9862  
Epoch: 47/50 | Train Loss: 0.5215 | Val Loss: 0.5798 | Dice Score: 0.9862  
Epoch: 48/50 | Train Loss: 0.5214 | Val Loss: 0.5802 | Dice Score: 0.9861  
Epoch: 49/50 | Train Loss: 0.5217 | Val Loss: 0.5795 | Dice Score: 0.9862  
Epoch: 50/50 | Train Loss: 0.5214 | Val Loss: 0.5806 | Dice Score: 0.9861

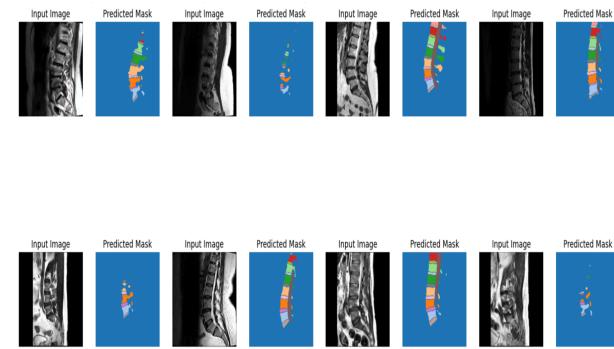


Figure 13(b): UNet++

Epoch [31/50], Train Loss: 0.4060, Validation Loss: 0.3234, Dice Coefficient: 0.8170  
Epoch [32/50], Train Loss: 0.3307, Validation Loss: 0.3083, Dice Coefficient: 0.8593  
Epoch [33/50], Train Loss: 0.2774, Validation Loss: 0.4338, Dice Coefficient: 0.8584  
Epoch [34/50], Train Loss: 0.2096, Validation Loss: 0.4395, Dice Coefficient: 0.8633  
Epoch [35/50], Train Loss: 0.2661, Validation Loss: 0.3092, Dice Coefficient: 0.8742  
Epoch [36/50], Train Loss: 0.2637, Validation Loss: 0.3997, Dice Coefficient: 0.8900  
Epoch [37/50], Train Loss: 0.1776, Validation Loss: 0.3208, Dice Coefficient: 0.8900  
Epoch [38/50], Train Loss: 0.2019, Validation Loss: 0.4300, Dice Coefficient: 0.8900  
Epoch [39/50], Train Loss: 0.1883, Validation Loss: 0.3566, Dice Coefficient: 0.8900  
Epoch [40/50], Train Loss: 0.2215, Validation Loss: 0.2935, Dice Coefficient: 0.8900  
Epoch [41/50], Train Loss: 0.2508, Validation Loss: 0.4285, Dice Coefficient: 0.8900  
Epoch [42/50], Train Loss: 0.2178, Validation Loss: 0.2558, Dice Coefficient: 0.8900  
Epoch [43/50], Train Loss: 0.1433, Validation Loss: 0.3439, Dice Coefficient: 0.8900  
Epoch [44/50], Train Loss: 0.2312, Validation Loss: 0.3387, Dice Coefficient: 0.8900  
Epoch [45/50], Train Loss: 0.2755, Validation Loss: 0.3949, Dice Coefficient: 0.8900  
Epoch [46/50], Train Loss: 0.1731, Validation Loss: 0.3737, Dice Coefficient: 0.8900  
Epoch [47/50], Train Loss: 0.2897, Validation Loss: 0.3717, Dice Coefficient: 0.8900  
Epoch [48/50], Train Loss: 0.2488, Validation Loss: 0.3154, Dice Coefficient: 0.8900  
Epoch [49/50], Train Loss: 0.2287, Validation Loss: 0.3442, Dice Coefficient: 0.8900  
Epoch [50/50], Train Loss: 0.1996, Validation Loss: 0.1985, Dice Coefficient: 0.8900

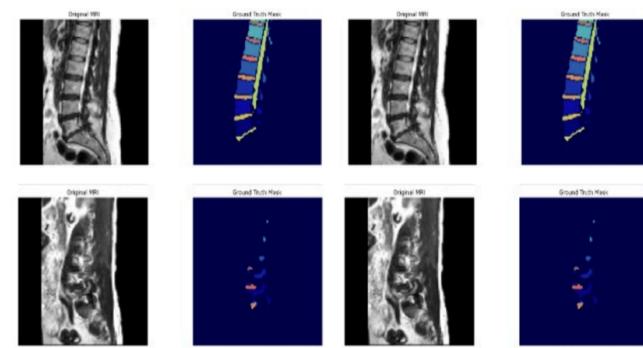
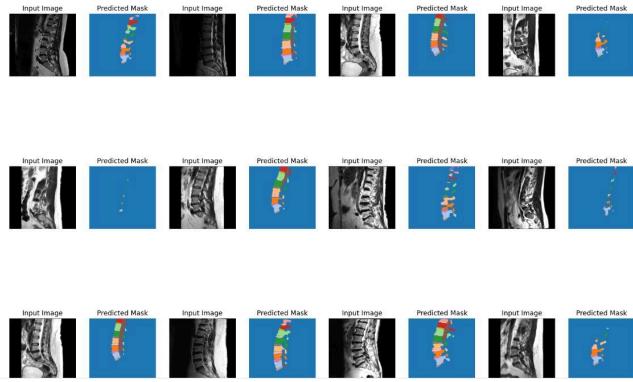


Figure 13(c): ResUnet

```

Epoch: 28/50 | Train Loss: 0.6069 | Val Loss: 0.7467 | Dice Score: 0.9704
Epoch: 29/50 | Train Loss: 0.6063 | Val Loss: 0.6934 | Dice Score: 0.9756
Epoch: 30/50 | Train Loss: 0.6060 | Val Loss: 0.7565 | Dice Score: 0.9697
Epoch: 31/50 | Train Loss: 0.6062 | Val Loss: 0.6625 | Dice Score: 0.9778
Epoch: 32/50 | Train Loss: 0.6053 | Val Loss: 0.6618 | Dice Score: 0.9778
Epoch: 33/50 | Train Loss: 0.6055 | Val Loss: 0.6633 | Dice Score: 0.9777
Epoch: 34/50 | Train Loss: 0.6056 | Val Loss: 0.6632 | Dice Score: 0.9776
Epoch: 35/50 | Train Loss: 0.6057 | Val Loss: 0.6622 | Dice Score: 0.9777
Epoch: 36/50 | Train Loss: 0.6053 | Val Loss: 0.6621 | Dice Score: 0.9778
Epoch: 37/50 | Train Loss: 0.6052 | Val Loss: 0.6619 | Dice Score: 0.9778
Epoch: 38/50 | Train Loss: 0.6060 | Val Loss: 0.6611 | Dice Score: 0.9778
Epoch: 39/50 | Train Loss: 0.6057 | Val Loss: 0.6624 | Dice Score: 0.9777
Epoch: 40/50 | Train Loss: 0.6056 | Val Loss: 0.6619 | Dice Score: 0.9778
Epoch: 41/50 | Train Loss: 0.6047 | Val Loss: 0.6638 | Dice Score: 0.9776
Epoch: 42/50 | Train Loss: 0.6057 | Val Loss: 0.6616 | Dice Score: 0.9779
Epoch: 43/50 | Train Loss: 0.6053 | Val Loss: 0.6616 | Dice Score: 0.9778
Epoch: 44/50 | Train Loss: 0.6053 | Val Loss: 0.6623 | Dice Score: 0.9777
Epoch: 45/50 | Train Loss: 0.6052 | Val Loss: 0.6624 | Dice Score: 0.9778
Epoch: 46/50 | Train Loss: 0.6053 | Val Loss: 0.6627 | Dice Score: 0.9777
Epoch: 47/50 | Train Loss: 0.6049 | Val Loss: 0.6622 | Dice Score: 0.9777
Epoch: 48/50 | Train Loss: 0.6052 | Val Loss: 0.6614 | Dice Score: 0.9778
Epoch: 49/50 | Train Loss: 0.6051 | Val Loss: 0.6634 | Dice Score: 0.9777
Epoch: 50/50 | Train Loss: 0.6048 | Val Loss: 0.6628 | Dice Score: 0.9778

```



**Figure 13(d): DeepLabV3**

**Figure 13 : Training of model**

### 5.3 Evaluating Ensemble

The provided code demonstrates a process for determining the optimal combination of weights in an ensemble model to maximize performance metrics such as the Dice Coefficient and Intersection over Union (IoU).

```

# Loop through different weight combinations (w1 + w2 = 1)
for w1 in [i * 0.05 for i in range(21)]: # Testing weights from 0 to 1 with step size 0.05
    w2 = 1 - w1 # w2 is automatically determined
    DC_ensemble = w1 * DC_unet + w2 * DC_resunet
    IoU_ensemble = w1 * IoU_unet + w2 * IoU_resunet

    weights_and_results.append((w1, w2, IoU_ensemble, DC_ensemble, diff))

weights_and_results.sort(key=lambda x: x[4])

```

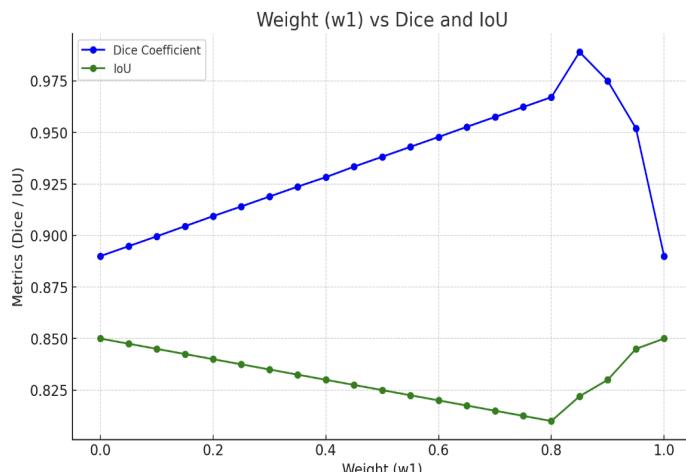
**Figure 14 : Algorithm used in weighted ensemble**

```

w1=0.80, w2=0.20, IoU=0.8100, Dice=0.9671, diff=0.0339
w1=0.75, w2=0.25, IoU=0.8125, Dice=0.9623, diff=0.0362
w1=0.70, w2=0.30, IoU=0.8150, Dice=0.9575, diff=0.0385
w1=0.65, w2=0.35, IoU=0.8175, Dice=0.9527, diff=0.0408
w1=0.60, w2=0.40, IoU=0.8200, Dice=0.9478, diff=0.0432
w1=0.55, w2=0.45, IoU=0.8225, Dice=0.9430, diff=0.0465
w1=0.50, w2=0.50, IoU=0.8250, Dice=0.9382, diff=0.0538
w1=0.45, w2=0.55, IoU=0.8275, Dice=0.9334, diff=0.0611
w1=0.40, w2=0.60, IoU=0.8300, Dice=0.9286, diff=0.0684
w1=0.35, w2=0.65, IoU=0.8325, Dice=0.9237, diff=0.0758
w1=0.30, w2=0.70, IoU=0.8350, Dice=0.9189, diff=0.0831
w1=0.25, w2=0.75, IoU=0.8375, Dice=0.9141, diff=0.0904
w1=0.20, w2=0.80, IoU=0.8400, Dice=0.9093, diff=0.0977
w1=0.15, w2=0.85, IoU=0.8425, Dice=0.9045, diff=0.1050
w1=0.10, w2=0.90, IoU=0.8450, Dice=0.8996, diff=0.1124
w1=0.05, w2=0.95, IoU=0.8475, Dice=0.8948, diff=0.1197
w1=0.00, w2=1.00, IoU=0.8500, Dice=0.8900, diff=0.1270

```

Best weights chosen:  
w1=0.85, w2=0.15, IoU=0.8220, Dice=0.9890



**Figure 15 :** Weighted ensemble results

## 5.4 Model Testing

The trained models, including U-Net, U-Net++, ResUNet, DeepLabV3, and the ensemble model MedFusionNet, were evaluated on the testing set comprising 25% of the SPIDER dataset.

Performance metrics, including Dice Score and Intersection over Union (IoU), were computed to assess the accuracy and robustness of each model.

Model	IoU Score	Dice Score	Performance
Unet	0.410	0.670	Individual
Unet ++	0.800	0.986	Individual
ResUnet	0.850	0.890	Individual
DeepLabV3	0.750	0.977	Individual
Unet + Unet++	0.773	0.881	Lower Performance
Unet + ResUnet	0.751	0.826	Good Performance
Unet + DeepLabV3	0.751	0.878	Decent Performance
Unet+++ ResUnet	0.822	0.989	Best Performance
Unet+++ DeepLabV3	0.780	0.981	Great Performance
ResUnet + DeepLabV3	0.810	0.918	High Performance

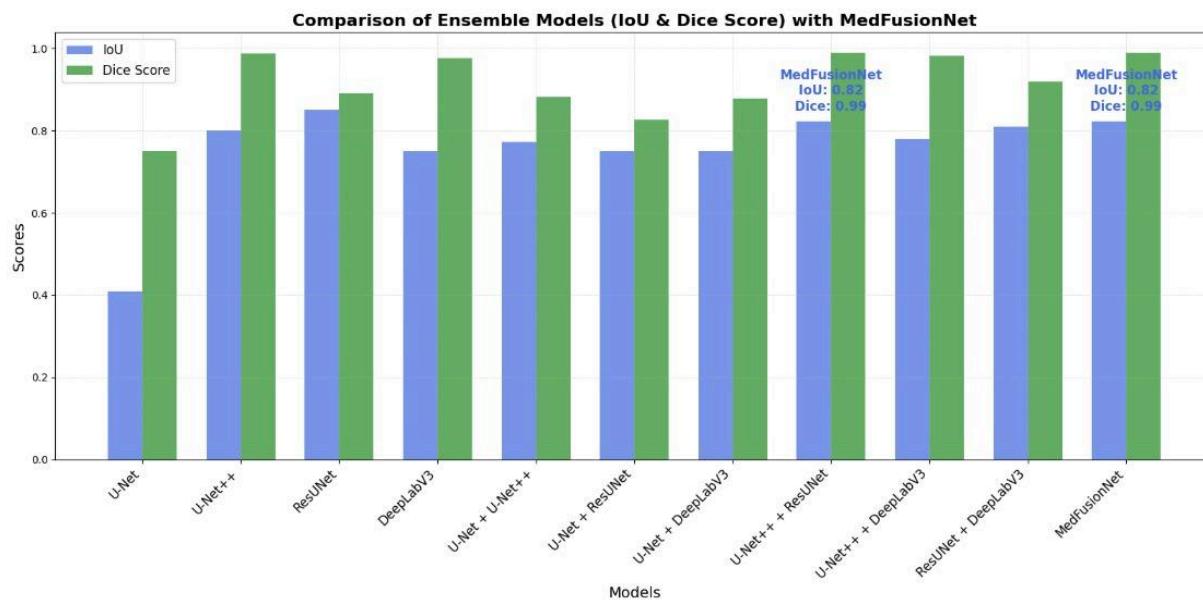
**Table 2:** Comparison of Testing Results

## 5.5 Performance Analysis and Model Selection

After evaluating the individual models (U-Net, U-Net++, ResUNet, and DeepLabV3) and comparing their performance metrics, it was observed that the combination of **ResUNet** and **U-Net++** yielded the best results in terms of Dice Score and Intersection over Union (IoU). To further enhance segmentation accuracy and leverage the strengths of both models, a **pairwise ensemble approach** was adopted.

This ensemble model, named **MedFusionNet**, combines the predictions of **ResUNet** and **U-Net++**, aiming to improve segmentation performance by mitigating the individual model's weaknesses and

capitalizing on their complementary strengths.



**Figure 16 :** Graphical Comparison of Testing Results

## CHAPTER 6: CONCLUSION AND FUTURE WORK

In this study, we addressed several critical research questions to advance the segmentation of lumbar spine MRI scans. First, we explored the most effective segmentation model architecture for the SPIDER dataset, identifying the optimal approach among U-Net, its variants, and other models. While U-Net and its derivatives are widely adopted, our study highlights the importance of dataset-specific evaluations to determine the best-performing architecture. Second, we investigated how ensemble techniques can improve segmentation accuracy, demonstrating their effectiveness in leveraging the strengths of different models such as U-Net, U-Net++, DeepLabV3, and ResUNet. This is particularly significant as ensemble methods remain underexplored for lumbar spine MRI segmentation. Third, we developed a systematic strategy for determining the correct weights for ensemble models, optimizing the contributions of each model to achieve robust performance, as reflected in improved Dice and IoU scores. Lastly, we acknowledged the limited availability of literature and datasets in lumbar spine MRI, which poses challenges for segmentation algorithm development. By utilizing the SPIDER dataset, we addressed this gap and provided a foundation for future research. Moving forward, we aim to explore more advanced ensemble techniques, such as weighted voting and stacking, while also validating our models and methods on other large publicly available datasets, such as those hosted on platforms like Mendeley, to ensure broader applicability and generalizability of our findings.

## REFERENCES

- [1] van der Graaf, J.W., van Hooff, M.L., Buckens, C.F.M. et al. Lumbar spine segmentation in MR images: a dataset and a public benchmark. *Sci Data* 11, 264 (2024). <https://doi.org/10.1038/s41597-024-03090-w>
- [2] Masood, Rao Farhat & Taj, Imtiaz & Khan, Muhammad & Qureshi, Muhammad & Hasan, Taimur. (2022). Deep Learning based Vertebral Body Segmentation with Extraction of Spinal Measurements and Disorder Disease Classification. *Biomedical Signal Processing and Control.* 71. 10.1016/j.bspc.2021.103230.
- [3] A. S. Al-Kafri et al., "Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks," in IEEE Access, vol. 7, pp. 43487-43501, 2019, doi: 10.1109/ACCESS.2019.2908002.
- [4] Wang, Shuai & Jiang, Zhengwei & Yang, Hualin & Li, Xiangrong & Yang, Zhicheng. (2022). Automatic Segmentation of Lumbar Spine MRI Images Based on Improved Attention U-Net. *Computational Intelligence and Neuroscience.* 2022. 1-10. 10.1155/2022/4259471.
- [5] Ahmed, I., Hossain, M.T., Nahid, M.Z.I., Sanjid, K.S., Junayed, M.S.S., Uddin, M.M. and Khan, M.M., 2024. Pioneering Precision in Lumbar Spine MRI Segmentation with Advanced Deep Learning and Data Enhancement. *arXiv preprint arXiv:2409.06018.*
- [6] Mohammed, Abdul Sami, et al. "Knee Osteoarthritis Detection and Severity Classification Using Residual Neural Networks on Preprocessed X-ray Images." *Diagnostics* 13.8 (2023): 1380.
- [7] Lu, Jen-Tang, et al. "Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning." *Machine Learning for Healthcare Conference.* PMLR, 2018
- [8] Cook, Daniel & Gladowski, David & Acuff, Heather & Yeager, Matthew & Cheng, Boyle. (2012). Variability of manual lumbar spine segmentation. *The International Journal of Spine Surgery.* 6.

167-173. 10.1016/j.ijsp.2012.04.002.

- [9] Mushtaq M, Akram MU, Alghamdi NS, Fatima J, Masood RF. Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models. *Sensors (Basel)*. 2022 Feb 17;22(4):1547. doi: 10.3390/s22041547. PMID: 35214448; PMCID: PMC8879729.
- [10] Galbusera, F., Niemeyer, F., Wilke, HJ. et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J* 28, 951–960 (2019). <https://doi.org/10.1007/s00586-019-05944-z>
- [11] Hu B, Kim C, Ning X, Xu X. Using a deep learning network to recognise low back pain in static standing. *Ergonomics*. 2018 Oct;61(10):1374-1381. doi: 10.1080/00140139.2018.1481230. Epub 2018 Jul 3. PMID: 29792576.
- [12] Liang YW, Fang YT, Lin TC, Yang CR, Chang CC, Chang HK, Ko CC, Tu TH, Fay LY, Wu JC, Huang WC, Hu HW, Chen YY, Kuo CH. The Quantitative Evaluation of Automatic Segmentation in Lumbar Magnetic Resonance Images. *Neurospine*. 2024 Jun;21(2):665-675. doi: 10.14245/ns.2448060.030. Epub 2024 Jun 30. PMID: 38955536; PMCID: PMC11224749.
- [13] Xiong, X.; Graves, S.A.; Gross, B.A.; Buatti, J.M.; Beichel, R.R. Lumbar and Thoracic Vertebrae Segmentation in CT Scans Using a 3D Multi-Object Localization and Segmentation CNN. *Tomography* **2024**, *10*, 738-760. <https://doi.org/10.3390/tomography10050057>