

Leveraging DeepLabv3+ and U-Net Models for Accurate Spine Segmentation in CT Scans

Retinderdeep Singh

Department of Computer Science and Engineering,
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.
E-mail: retinderdeepsingh@gmail.com

Chander Prabha

Department of Computer Science and Engineering,
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.
E-mail: prabhanice@gmail.com

Meena Malik

Department of Computer Science and Engineering,
Chandigarh University, Punjab, India.
E-mail: meenamlk@gmail.com

Ram Kishun Lodhi

Department of Applied Science,
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
Lavale, 412115, Pune, Maharashtra, India.
Corresponding author: ramkishun.lodhi@gmail.com

(Received on February 9, 2025; Revised on April 29, 2025 & June 4, 2025; Accepted on June 18, 2025)

Abstract

Medical imaging depends heavily on spine segmentation for diagnosing spinal disorders as well as developing treatment plans. The proposed research describes a deep learning framework for spine segmentation where two state-of-the-art models known as DeepLabv3+ and U-Net are evaluated for comparison. Our team applied the developed framework using 1089 high-resolution CT images which had been annotated with segment outlines for cervical, thoracic, lumbar and sacral areas. Person working with the DeepLabv3+ model produced segmentations with 0.925 Intersection over Union (IoU) and 0.920 Dice score along with 0.920 F1 score and an accuracy rate of 98.5%. The U-Net model achieved better performance when applied with a ResNet34 backbone by reaching an IoU of 0.953 alongside Dice score 0.951 F1 score 0.951 and accuracy reaching 98.6%. U-Net outperforms existing methods of spine segmentation by achieving superior results in area detection along with boundary definition accuracy. The research provides a thorough assessment of network designs and practical utility tests to support the development of automated spine segmentation applications that can work effectively in clinical settings.

Keywords- DeepLabv3+, Deep learning, U-Net, Segmentation, Spine.

1. Introduction

Modern healthcare relies on medical imaging to identify diseases in their early stages and to deliver precise medical diagnoses which also help create effective treatment strategies. The human spine ranks as one of many essential anatomical regions because it supports body structure and enables movement and conduces nervous system signals. Spinal disorders including degenerative disc disease together with scoliosis and fractures and tumors require accurate imaging and analysis for correct diagnosis. Manual segmentation of spinal structures through radiological scan review is both labor-intensive and shows inconsistent results due to observer differences. Automated deep learning-based spine segmentation techniques have become more

significant because they both improve diagnostic workflows and increase their accuracy levels. An important component of the central nervous system is the human spine that not only maintains physical posture but also provides significant spinal cord protection. The spinal cord controls impulse movement between the brain and the rest of the body, therefore enabling sensation, motion, and other bodily functions. Comprising soft tissues, intervertebral discs, and vertebrae, the spine itself provides protection, support, and flexibility. Correct medical imaging and spine segmentation discover and treat a variety of spinal problems, including congenital defects, tumors, fractures, and degenerative diseases. As medical imaging technology advances, accurate and automated methods for spine segmentation become more important, which influences research on deep learning methods in this domain (Kim et al., 2020a).

Deep learning (DL) is a kind of artificial intelligence (AI) revolutionizing several fields, including medical diagnostics. DL methods have shown significant advances in image classification, segmentation, and detection tasks in medical imaging. Deep learning in medical diagnostics offers among its other advantages less time needed for image processing, less human error, and reliable, repeatable outcomes. In spine imaging, where exact segmentation may greatly affect patient outcomes and treatment choices, these advantages are especially important.

DeepLabv3+ and U-Net are the most suitable architectures for image segmentation among the many DL models. An expansion of the DeepLab series, DeepLabv3+ is well-known for its atrous convolution and atrous spatial pyramid pooling (ASPP) capability to gather contextual information at a multi-scale. DeepLabv3+ adds a decoder module to increase spatial resolution and adjust the segmentation boundaries, hence strengthening the DeepLab architecture. Because of its strong performance in collecting small details and managing many image scales, this model has been extensively embraced in several medical imaging uses. DeepLabv3+ may use its capabilities in the framework of spine segmentation to precisely define the complicated anatomical features of the spine, separating between many spinal segments and surrounding tissues (Liu et al., 2024).

U-Net is a well-known encoder-decoder system intended especially for segmentation of biological images. Comprising encoder and decoder, the U-Net design creates a U-shaped construction. The decoder reconstructs the image to its original resolution, conserving spatial information. The encoder compresses the input image into a lower-dimensional representation, therefore capturing fundamental properties. By letting the decoder concatenate high-resolution encoder data, U-Net's architecture improves the model's fine-grained detail capture capacity. This work uses a ResNet34 backbone (encoder) variation of U-Net pre-trained on the ImageNet dataset. This variation makes use of ResNet34's strong feature extracting power, especially suitable for grayscale medical image processing. Using ResNet34 as the U-Net architectural encoder has various benefits. Widely used CNN design ResNet34 is recognized for its residual connections, which allow the training of deeper networks and aid in reducing the vanishing gradient issue. ResNet34 pre-trained on ImageNet by use of transfer learning improves the U-Net model by use of rich feature representations obtained from a large collection of real images. Although natural and medical imaging has very diverse applications, this pretraining helps the model to more generalize to medical images. Setting channels = 1 helps the model to be more fit for grayscale images and consequently optimized for usage in medical imaging where grayscale images are often utilized (Kushnure & Talbar, 2021; Srivastava et al., 2022).

This work is relevant as it may help to develop fundamentals for many therapeutic purposes for automated spine segmentation methods. Automated segmentation enables faster and more accurate diagnosis and treatment planning, therefore helping radiologists and doctors to greatly reduce their workload. Customizing treatment plans based on appropriate segmentation also includes surgical planning, radiation

treatment, and biomechanical study. The findings of study conducted in this will provide a sensible evaluation of the advantages and drawbacks of every model, thereby directing further studies and development in this sector. The purpose of this research is to enhance automated spine segmentation in computed tomography (CT) images through the utilization and comparison between DeepLabv3+ and U-Net with a ResNet34 architecture. The main research attributes and novel aspects of this study include the following points:

- The study used a high-quality Spine Segmentation Dataset that contained 1089 CT scans with thorough segmentations for every region from cervical through thoracic to lumbar and sacral segments.
- The evaluation platform used for DeepLabv3+ and U-Net models featured consistent data preparation and standardized resizing along with resilient evaluation measurements to perform a proper comparison.
- The enhancement of U-Net involved pre-training ResNet34 on ImageNet followed by optimized processing for grayscale medical images which boosted both extraction and segmentation capabilities.
- The experimental outcomes showed that the proposed U-Net model produced better segmentations than DeepLabv3+ while reaching higher F-score and Dice Coefficient performance than SSAE and classical CNN-based models as well as other recent state-of-the-art techniques.
- The study departed from conventional research which examined solely the lumbar or segmented few spine regions by providing a full spinal (cervical, thoracic, lumbar, sacral region) segmentation for improved clinical use.
- The proposed models receive additional validation through both visual presentation and statistical analysis in addition to performance comparisons with various state-of-the-art techniques.

The broad aspects of anatomy and advanced experimental and segmentation capabilities separated the proposed work from earlier research by advancing spine imaging applications.

Section 2 presents in detail recent studies about deep learning methods applied to spine segmentation through literature review. The proposed methodology Section 3 outlines three main areas which include dataset preparation techniques and preprocessing operations with model architectural designs. A comprehensive analysis of experimental outcomes along with segmentation output visualizations and comparative assessments occurs in Section 4 of the work. The study ends with this section by presenting a summary of key discoveries coupled with research constraints and prospective improvements for spinal segmentation techniques.

2. Literature Review

Studies on the combination of AI and medical imaging are particularly active, especially in terms of enhancing the safety and accessibility of diagnostic procedures. Intending to offer a safer alternative for X-ray imaging, Ungi et al. (2020) conducted an interesting 2020 study on the application of tracked ultrasonic combined with artificial intelligence technologies for scoliosis assessment. This paper proposed an automated ultrasonic segmentation technique to solve the inherent challenges of using ultrasonic waves for spine imaging and scoliosis assessment. Researchers trained a convolutional neural network (CNN) using ultrasonic images from eight healthy adult volunteers and assessed the network on eight pediatric instances. Using patient data to lower than those of healthy volunteers, recall from 0.72 to 0.64 and accuracy from 0.31 to 0.27 were obtained utilizing the trained network. Still, binary segmentation measures showed improved performance after the thresholds for prediction maps on patient data were changed to minimize recall and accuracy fluctuations. The study efficiently reconstructed segmentation prediction maps into 3D volumes such that scoliosis could be evaluated with a maximum error of 2.2° compared to X-ray data. The researchers concluded that automated spine segmentation using tracked ultrasonic scans would make scoliosis measuring both efficient and precise, so perhaps overcoming historical limitations preventing the

adoption of ultrasonic as an acceptable alternative to X-ray in scoliosis evaluation. Chang et al. (2020) presented a 2020 upgraded method for multi-vertebrae segmentation. Sometimes traditional approaches segment certain spinal portions sequentially because of their lack of global spatial information, therefore limiting feature reuse and increasing processing work. Chang's method suggests a one-integrated solution exploiting local (label) and both global (geographical) information, therefore bypassing both these restrictions. Using a spatial graph convolutional network (GCN), the method automatically trains an adjacency matrix and creates a graph on local feature maps thereby allowing the capture of global spatial links between vertebrae. A label attention network also estimates the appearance probability of every vertebra, hence lowering uncertainty arising from different fields of vision (FOV) or similar appearances of nearby vertebrae. Trained end-to-end, the method—with an IDR of 89.28 and mIoU of 85.37—was evaluated on a dataset of 292 MRI scans with various FOVs and image properties.

Kim et al. (2020b) investigated the use of DL for spine segmentation in computed tomography (CT) images, a vital instrument for back pain diagnosis affecting 60% to 80% of humans at some time in their lives. Kim's research covers the increasing usage of computer-aided diagnosis to support doctors in understanding medical images. Using a CNN (convolutional neural network)—more especially, the U-Net architecture produced a web-based automated spine segmentation approach producing exact segmentation findings. The work consisted on training a hierarchical data format file with 330 CT images and evaluating the technique on 14 more images. With an average Dice coefficient of 90.4%, a precision of 96.81%, and an F1-score of 91.64%, the suggested approach produced rather outstanding results. These results imply that for spine segmentation, the web-based deep learning method is both extremely accurate and useful, thus providing great possibility to improve diagnosis procedures. Crucially for applications in orthopedics, neurology, and cancer; In 2020, Buerger et al. (2020) on 3D CT images, tackled the problem of automated instance segmentation of individual vertebrae. Especially for creating mesh-based spine representations using model-based segmentation (MBS), accurate segmentation is very essential. Buerger's work suggested a deep learning-based approach to start and steer MBS, therefore avoiding mistakes in vertebra labelling brought on by the like look of neighboring vertebrae. Four phases comprised the method: a single-class U-Net for coarse spine segmentation, a multi-class U-Net for fine segmentation and individual vertebra labelling, detection and labelling of landmark coordinates, and initializing of MBS models using these landmarks. On 147 patient images, the technique was verified with a root mean squared distance (RMSDist) of 0.90 mm separating segmentation results from ground truth meshes. This method shows how well DL may improve the accuracy and robustness of medical imaging's spinal segmentation.

Li et al. (2021) examined lumbar spine segmentation in 2021 to identify lumbar spinal stenosis (LSS), the condition with increasing prevalence. Using DL methods, the research concentrated on lumbar spine MRI images aiming on exact segmentation of the spinal body, lamina, and dural sac. They included a dual-branch multi-scale attention module into the network to enhance segmentation performance. This module integrates three 3×3 convolution operators with an attention method to choose essential information at multi-scale. Over state-of-the-art approaches, a drop in average surface distance from 6.40 to 2.71 mm and an increase in average Dice similarity coefficient from 0.9008 to 0.9252 shown significant gains. These developments show how effectively advanced deep learning methods might be able to increase LSS diagnostic accuracy. Zhang et al. (2021) proposed the Sequential Conditional Reinforcement Learning network (SCRL) to address issues with vertebral body (VB) detection and segmentation from MR spine images. Traditional methods might have false positives and mistakes as they ignore local VB appearances and global spine patterns. The SCRL network creatively blends deep reinforcement learning to simultaneously detect and segment VBs using modeling of spatial correlation between VBS as sequential dynamic-interaction processes. This approach enables SCRL to focus globally on every VB and collect comprehensive local appearance features. The network consists of three fully-connected residual neural

networks for global context and accurate VB bounding box detection, an anatomic-modeling reinforcement learning network for dynamic interaction, and a Y-shaped network for precise segmentation. With an average detection IoU of 92.3%, segmentation Dice of 92.6%, and classification accuracy of 96.4%, tested on 240 people SCRL shown promise as a rapid diagnostics tool for spinal diseases. Banerjee et al. (2022) presented research on the difficulties in scoliosis diagnosis in 2022, a three-dimensional spinal distortion marked by lateral curvature in the coronal plane. Radiation exposure is a concern with conventional radiative imaging techniques used for periodic scoliosis detection. Although technology is often hampered by noise that hides important image information, 3D ultrasonic imaging is under investigation for its safety and economy. To improve bone feature recognition in ultrasonic images, Banerjee presented a new hybridized CNN architecture dubbed the multi-scale feature fusion Skip-Inception U-Net (SIU-Net), thereby supporting the safe and automated evaluation of scoliosis degree. SIU-Net significantly improves upon the fundamental U-Net, UNet++, and MultiResUNet models by combining an improvised Inception block with dense skip paths at the decoder side. Tested on 109 spine ultrasonic datasets, SIU-Net showed better segmentation ability, especially in thoracic and lumbar areas, obtaining the lowest histogram Euclidean distance of 0.011, Dice score of 0.883, and the highest average Jaccard score of 0.781. This design shows potential to further completely automated scoliosis detection methods. Pang et al. (2022) developed a novel method for spine segmentation in magnetic resonance (MR) images in 2022; this is a crucial operation for the diagnosis and treatment of many spinal illnesses. Conventional segmentation techniques suffer with inter-class similarity, that is, when adjacent spinal structures have similar forms and appearances. Pang tackles this problem with the Detection-Guided Mixed-Supervised Segmentation Network (DGMSNet). Two components define DGMSNet mostly: a segmentation path for spine segmentation predictions and a detection path (regression network) for generating keypoint heatmaps. Special detection-guided learner produced dynamic parameters provide a semantic feature map for the route of adaptive convolution segmentation. Built on datasets annotated with pixel-level and keypoint-detection data, the network combines segmentation loss with detection loss under mixed-supervised loss. Approaching limited annotated datasets, this approach improves generalization. Modern performance of T2-weighted MR images was demonstrated surpassing previous approaches on both in-house and public datasets using mean Dice similarity values of 94.39% for spinal bodies and 87.21% for intervertebral discs.

Graf et al. (2023) examined fundamental automated segmentation of spinal magnetic resonance imaging (MRI) in 2023 stressing the difficulties of recognizing posterior spine components. Work using 263 paired CT/MRI series converted T1-weighted and T2-weighted MRI data into computed tomography (CT) images. Using "peak signal-to-noise ratio", many image-to-image translating techniques including paired (Pix2Pix, denoising diffusion implicit models (DDIM)) and unpaired (SynDiff, contrastive unpaired translation)-were assessed. Dice similarity coefficients (DSC) allow one to evaluate the segmentation performance of produced CT images. Although DDIM image mode was obtaining the highest image quality, the investigation revealed that 2D paired approaches and SynDiff had identical translation performance. Especially for minor structures like the spinous process, 3D translation methods generated anatomistically accurate segmentations and a higher DSC (0.80) over the 2D approach. The study underlined the need of employing at least two landmarks per vertebra for adequate registration, therefore significantly increasing the accuracy of MRI-to-CT translation and segmentation. The summary of literature review is shown in **Table 1**.

The investigated studies show significant developments in spine imaging, diagnostic artificial intelligence and deep learning application. Although segmentation accuracy between healthy individuals and pediatric patients presents significant challenges, artificial intelligence-enhanced ultrasonic imaging was shown to be a safer alternative for X-ray scoliosis assessment. Although a comprehensive solution integrating spatial and label information substantially enhances segmentation in multi-vertebrae scenarios, a web-based spine

segmentation system using U-Net shows remarkable accuracy and promise for improved diagnostic processes in CT imaging. We emphasized the need of initialization in model-based segmentation by means of accurate vertebrae segmentation attained utilizing deep learning. Emphasizing lumbar spine segmentation and Sequential Conditional Reinforcement Learning (SCRL) network introduction, new methods shown improved diagnosis accuracy. Improved feature detection and segmentation performance for scoliosis diagnosis and spine segmentation in MR images is obtained by means of creative designs comprising the multi-scale feature fusion Skip-Inception U-Net (SIU-Net) and Detection-Guessed Mixed-Supervised Segmentation Network (DGMSNet). In the conclusion, it was reiterated the importance of suitable posterior spine segmentation and landmark utilization in MRI-to-CT translation, thereby providing the route for more accurate and reliable diagnostic tools. These studies taken together illustrate how much AI and DL increase the accuracy, efficiency, and safety of spine MRI diagnosis.

Table 1. Summary of literature review.

References	Name of journal / Conference	Technique
Ungi et al. (2020)	IEEE transactions on bio-medical engineering	Tracked ultrasonic with CNN for scoliosis assessment
Chang et al. (2020)	MICCAI 2020	Spatial Graph Convolutional Network (GCN) for multi-vertebrae segmentation
Kim et al. (2020b)	Healthcare informatics research	Web-based U-Net CNN for spine CT segmentation
Buerger et al. (2020)	Medical imaging 2020 (SPIE)	Deep learning initialization for Model-Based Segmentation (MBS) using U-Net
Li et al. (2021)	Neural computing & applications	Dual-branch multi-scale attention CNN for lumbar spine segmentation
Zhang et al. (2021)	Medical image analysis	Sequential Conditional Reinforcement Learning (SCRL) for vertebral detection and segmentation
Banerjee et al. (2022)	Biocybernetics and Biomedical Engineering	Skip-Inception U-Net (SIU-Net) for scoliosis diagnosis from 3D ultrasound
Pang et al. (2022)	Medical image analysis	Detection-Guided Mixed-Supervised Segmentation Network (DGMSNet)
Graf et al. (2023)	European radiology experimental	MRI to CT translation using paired/unpaired GANs for posterior spine segmentation

3. Proposed Methodology

The segmentation of the spine from computed tomography (CT) images is addressed in this study using a comparison of two advanced DL models, DeepLabv3+ and U-Net (**Figure 1**). Many spinal diseases need proper spine segmentation for diagnosis and treatment; thus, the use of DL techniques might significantly increase the accuracy and efficiency of this operation. This section details the employed dataset, the accompanying preprocessing steps, and the proposed model architectures used for segmentation problems.

3.1 Dataset

This study takes use of a Spine Segmentation Dataset based on high-resolution CT scans (Wasserthal, 2024). The collection consists of ten thousand nine images, and each one of them has precisely delineated spinal parts. This study uses 1089 high-resolution CT images from the dataset along with skillfully segmented NIfTI files. Targeting certain spinal vertebrae, these segmentations divide them into four primary groups: lumbar, thoracic, cervical, sacral, vertebrae. This wide segmentation not only enables a thorough examination of the architecture of the spine but also enables deep learning models to regularly differentiate among the many spinal types. Every scan provides a three-dimensional image of the spine that allows one to document the complex anatomical elements needed for appropriate segmentation. The dataset was chosen for its complete representation of spinal anatomy as well as for its high-quality annotations—which are absolutely required for training and validation of DL models.

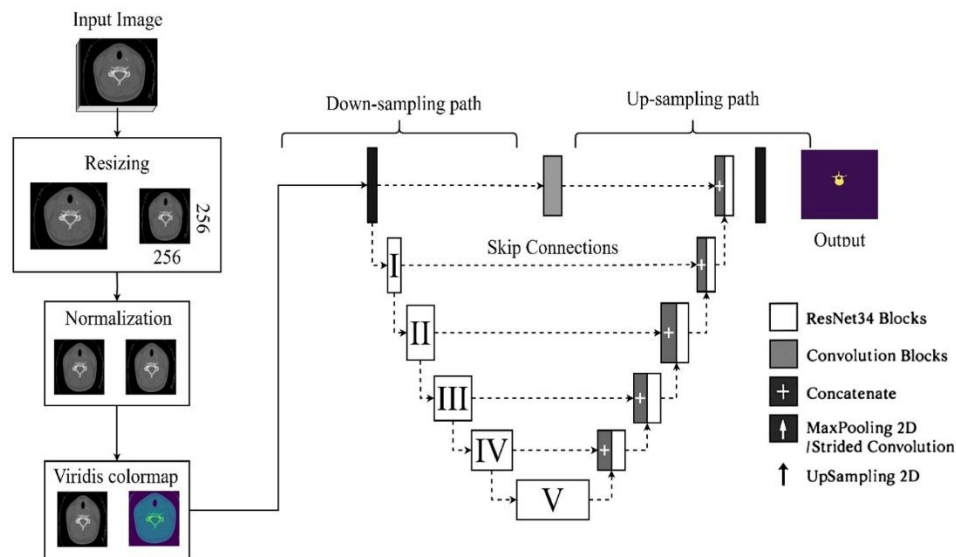


Figure 1. Proposed spine segmentation model.

3.2 Preprocessing

Pre-processing is essential for the method as it provides consistent and fit data for training the deep learning systems. Along with corresponding segmentation masks, high-resolution CT images saved in NIfTI form constitute the dataset (**Figure 2**). The process of preparation comprises many crucial stages:

- a) **Loading and normalization:** From the dataset, segmentation masks and CT images are loaded. To give consistent input values throughout many scans and hence stabilize the training process, pixel intensity values for the images are normalized to a standard range, generally between 0 and 1. Reaching normality requires one to divide every pixel value by the greatest pixel intensity in the image. Any negative value is zero in order to guarantee consistency.
- b) **Resampling and shrinking:** The CT images and segmentation masks are resampled and shrunk to a constant size assures homogeneity in image dimensions and aids to permit efficient training. This work uses images downsized to 256×256 pixels. This scaling helps to equalize the input data, thereby allowing the models to learn relevant elements free from impact from changes in the original image sizes.
- c) **Processing the segmentation masks:** The segmentation masks are preprocessed ensures their binary nature, wherein every voxel is found to be either background or a vertebra. By thresholding the mask values, this practically turns the mask into a binary form; any value greater than 1 is given 1. This phase simplifies the segmentation choreography as it corresponds with the masks with the desired output format for both the DeepLabv3+ and U-Net models.
- d) **Splitting:** The dataset is splitted into training, validation, and test sets helps to accurately assess the model's performance. The dataset consists of 10% for testing, 10% for validation, and 80% for training. This division assures that the model receives training on a diversified set of data while keeping another part for final evaluation and another for validation to change hyperparameters and prevent overfitting.

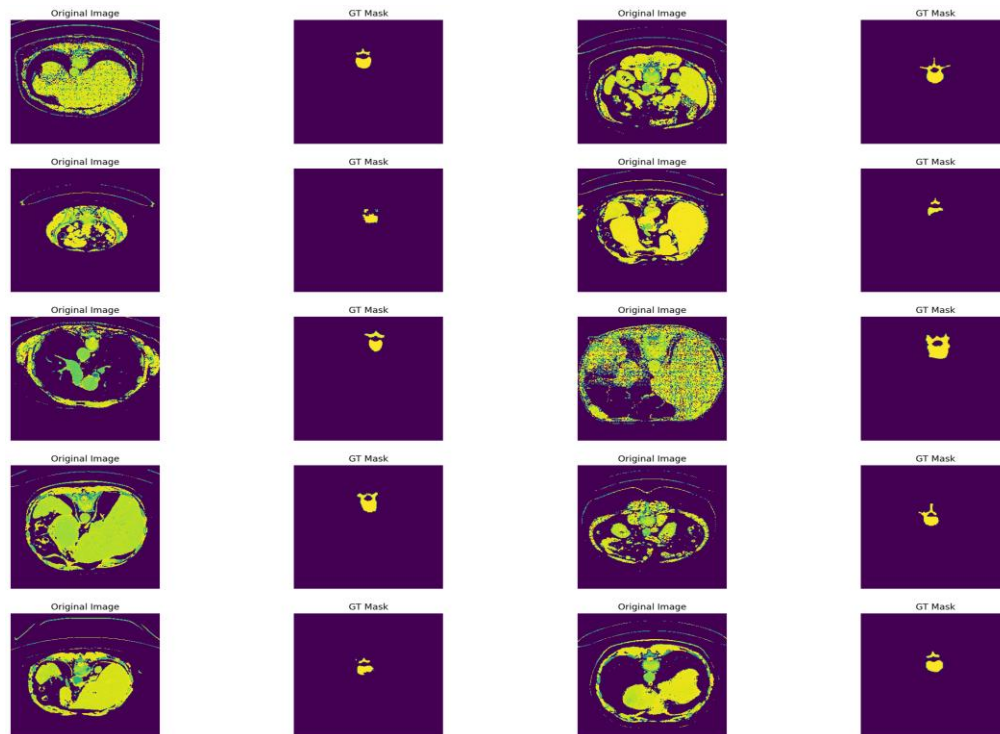


Figure 2. Dataset.

3.3 DeepLabV3+

The study used DeepLabv3+ model to automate spine segmentation tasks from CT images. The deep convolutional neural network DeepLabv3+ excels at semantic segmentation duties because it maintains both detailed anatomical structures and contextual understanding for complex anatomical structures. The network achieves enhanced spatial accuracy because it contains both an encoder-decoder framework with atrous convolution elements that expand receptive field without compromising feature map resolution. The main characteristic of DeepLabv3+ involves its Atrous Spatial Pyramid Pooling (ASPP) module that executes parallel convolutional operations with altered dilation rates. The model acquires multiple scale contextual data because this feature is vital when distinguishing different spine structure dimensions and shapes throughout different medical regions. Boundary segmentation refinement occurs through the decoder module by implementing an upsample process that adds lower-level features from upcoming network stages (**Figure 3**). The upsampling process recovers information that gets lost when the input is downscaled. The softmax activation function produces pixel-wise probability information which divides the image into vertebral areas and background sections. The system achieves effective multi-scale context aggregation and boundary refinement through its design principles which render it useful for medical image segmentation of spinal regions.

The last layer generates a probability map $P(y|x)$ for every voxel by means of softmax activation function $\sigma(z)$, therefore showing the likelihood P of every voxel either from a spine or the background (Prakash et al., 2020). The model is set in this work with multiple output classes $n_{cls}=2$, corresponding to the binary classification of vertebrae against background, hence indicating that the input pictures be grayscale with $in_channels=1$. The Cross Entropy Loss function is denoted as $L_{ce}(y, \hat{y})$, is applied during the training

phase to maximize the model's capacity to recognize spinal structures in the CT images from background regions. Here the real label is y and $\hat{y}$ is the expected probability (Qadri et al., 2023). One might create the loss function mathematically as follows shown in Equation (1).

$$L_{ce}(y, \hat{y}) = - \sum_{c=1}^{n_{cls}} y_c \log(\hat{y}_c) \quad (1)$$

where, y_c and $\hat{y}_c$ are the true and predicted probabilities for class c , respectively. This loss function is particularly suitable for classification tasks where each input belongs to one of several classes (Wang et al., 2024).

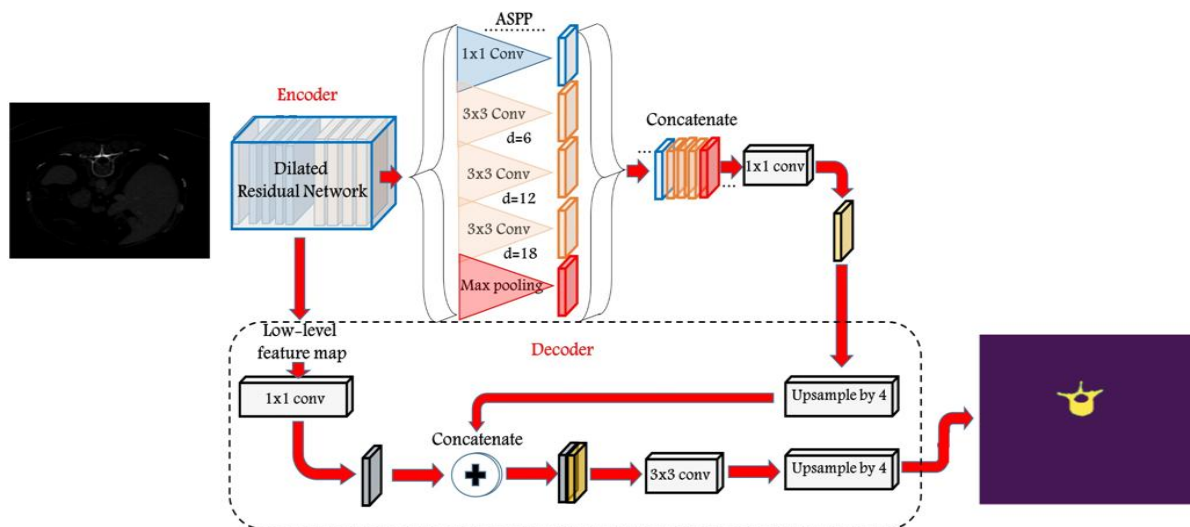


Figure 3. DeepLabv3+ model architecture.

3.4 Proposed U-Net with ResNet-34 Backbone Model

A U-Net model was designed as a comparative segmentation tool and made better by including a ResNet34-based encoder, so features could be picked out and gradient flow could improve. Many researchers use U-Net for biomedical image segmentation because its symmetric coding-decoding layer designs allows it to look at both small and big features at once.

Creating the encoder parts (Adding ResNet34)

In our model, the encoder path employs ResNet34 which has learned from the ImageNet dataset ahead of training. Because of this connection, it is possible to train the model with huge quantities of natural images to use their learned features for examining grayscale CT images in medical fields. Thanks to ResNet34's residual blocks that use identity maps and skip connections, vanishing gradient problems are less likely and training in deeper networks is possible. Because of these residual connections, the training process is more stable and the network can detect the structures of vertebrae.

Skip connections and the decoder

To keep useful details from being missing, in the decoder, U-Net connects the outputs of each encoder layer directly with its corresponding decoder layer. Because of this, clear and precise details about spinal structures can be seen in the fused features of the encoder and decoder.

The decoder is made up of transposed convolutions that do the upsampling and progressively make the maps restore their original size. The upscaled features are connected to the encoder's outputs via skip connections and then go through convolutional layers to improve the segmentation output.

Mathematically, combining the encoder feature map with the upsampled decoder version U_i gives the resultant feature map C_i .

$$C_i = F_i + U_i \quad (2)$$

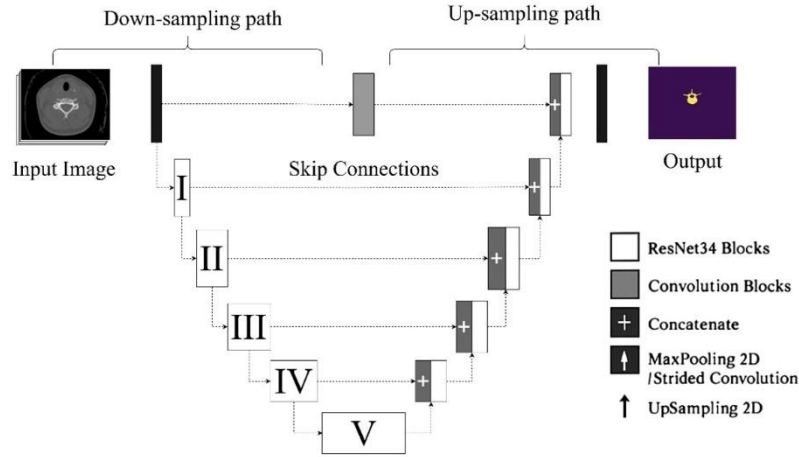


Figure 4. U-Net model architecture.

This combination preserves significant geographical data, hence improving segmentation results. The decoder enhances segmentation across the skip connections by up-sampling the features to match the original image resolution. This method ensures the precise acquisition of minor characteristics about the vertebrae, thereby directing correct boundary delineation from an up-sampling operation U followed by a convolution operation $Conv$. This may be stated mathematically as Equation (3).

$$\hat{F}_i = Conv(U(C_i)) \quad (3)$$

where, F_i is the refined feature map in the decoder.

The final output of the U-Net is generated using a softmax activation function $\sigma(z)$, which produces a voxel-wise probability map $P(y|x)$. This probability map indicates the likelihood P of each voxel belonging to a particular class, facilitating accurate segmentation. The softmax function is mathematically defined as Equation (4).

$$\sigma(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^{n_{cls}} \exp(z_j)} \quad (4)$$

Considering n_{cls} as the total count of classes and z_i as the input for the softmax function for class i . This framework allows U-Net to effectively differentiate complex anatomical components including spinal CT scan vertebrae in biomedical images (Saravagi et al., 2023; Tanwar et al., 2023). This architectural design allows the U-Net with ResNet34 encoder to capture both high-level contextual and fine-grained spatial information—making it highly suitable for accurate segmentation of complex spinal CT images. **Figure 4** has been updated to illustrate this architecture, showing residual blocks in the encoder, skip connections, and decoder upsampling paths.

4. Results

This section discusses the training and validation results of both models. Visualizations of the segmentation of the test dataset is also provided along with the comparative analysis of both the models and the state-of-the-art model. The hyperparameters used in the study for the deepLabv3+ and U-Net models are shown in **Table 2**.

Table 2. Hyperparameters of the study.

Hyperparameter	DeepLabv3+	U-Net
Model backbone	Xception	ResNet34
Input image size	256 × 256	256 × 256
Batch size	16	16
Learning rate	0.001	0.001
Optimizer	Adam	Adam
Loss function	Cross-entropy loss	Cross-entropy loss
Epochs	7	8
Activation function	Softmax	Softmax
Pretrained weights	Yes (ImageNet)	Yes (ImageNet)
Data augmentation	Random flip, rotation	Random flip, rotation
Normalization	Pixel values scaled to [0,1]	Pixel values scaled to [0,1]

4.1 Model Performance

The performance of the deepLabv3+ and U-Net models are shown in this sub-section. This sub-section discusses the model's performance at every crucial point. Graphs of accuracy, loss, F1 score, Mean IoU and dice score are also shown.

4.1.1 DeepLabv3+

The DeepLabv3+ model was trained over seven epochs with constant improvement in both training and validation criteria. With an Intersection over Union (IoU) of 0.653 and a Pixel Accuracy (PA) of 0.968, the model shown in the first epoch showed a training loss of 0.122. With both the Dice score of 0.468 and F1 score of 0.468, the start in exactly segmenting the spine structures was really bad. The validation metrics were somewhat better: 0.029, PA of 0.984, IoU of 0.832, Dice and F1 scores of 0.799.

By the second epoch, the DeepLabv3+ model improved. The training loss was reduced to 0.222 and the training PA came to 0.983. Each rose earns a 0.828 IoU and a 0.794 Dice score respectively. The validation loss dropped to 0.016 and the PA topped 0.985. The validation IoU declined to 0.864 as dice and F1 scores rose to 0.844. Following this trend in the future epochs, the training loss reduced even further to 0.013 by the third epoch while the IoU and Dice scores reached 0.880 and 0.865 respectively. The validation metrics also improved with a loss of 0.012, PA of 0.984, IoU of 0.890, Dice score of 0.878, and F1 score of 0.878.

By the fourth epoch, the model achieved a training loss of 0.010 with an IoU of 0.894 and a Dice score of 0.883. With an IoU of 0.909 and a Dice score of 0.901, the validation loss was reduced by another 0.009. The last training results for the seventh epoch were a training loss of 0.006, PA of 0.982, IoU of 0.921, and Dice and F1 scores of 0.915. The validation results showed some variation with the last loss at 0.006, PA at 0.985, IoU at 0.925, and Dice and F1 scores at 0.920 (as shown in **Figures 5 to 9**).

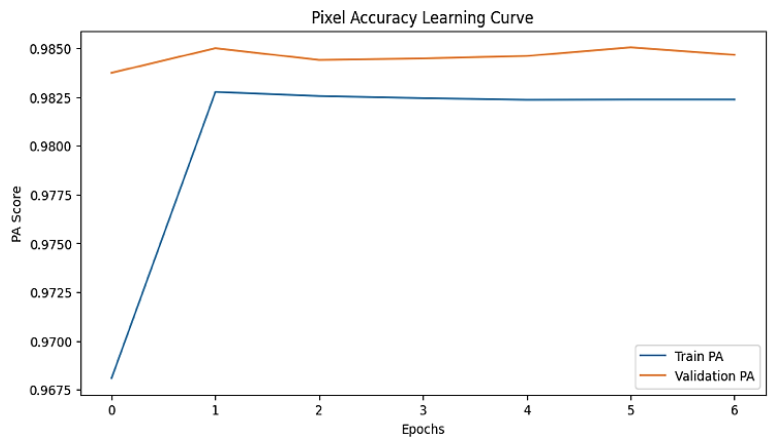


Figure 5. Training and validation accuracy graph (DeepLabv3+).

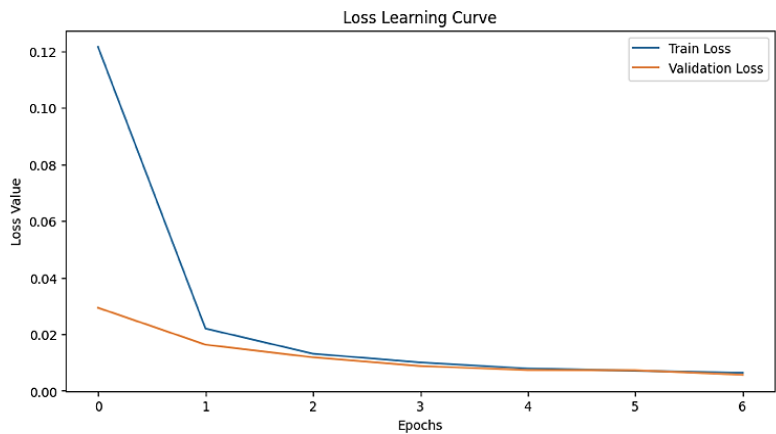


Figure 6. Training and validation loss graph (DeepLabv3+).

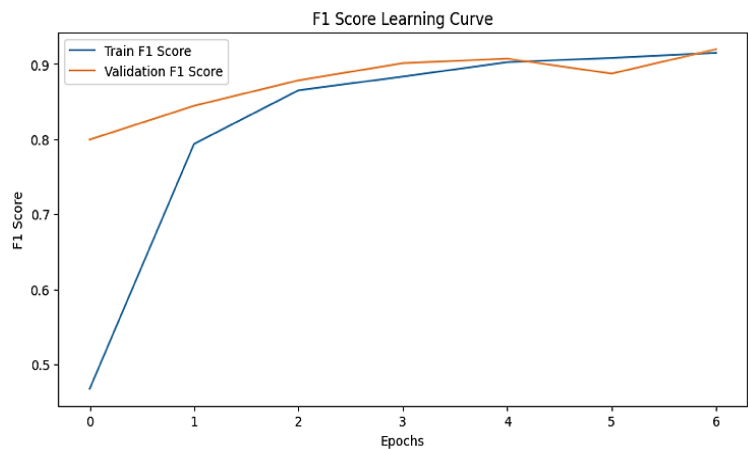


Figure 7. Training and validation F1 score (DeepLabv3+).

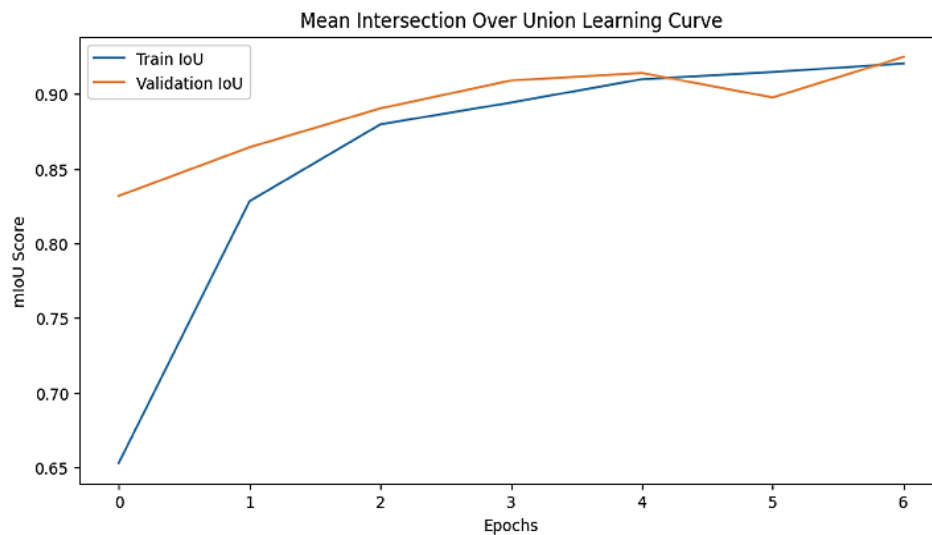


Figure 8. Training and validation mean IoU (DeepLabv3+).

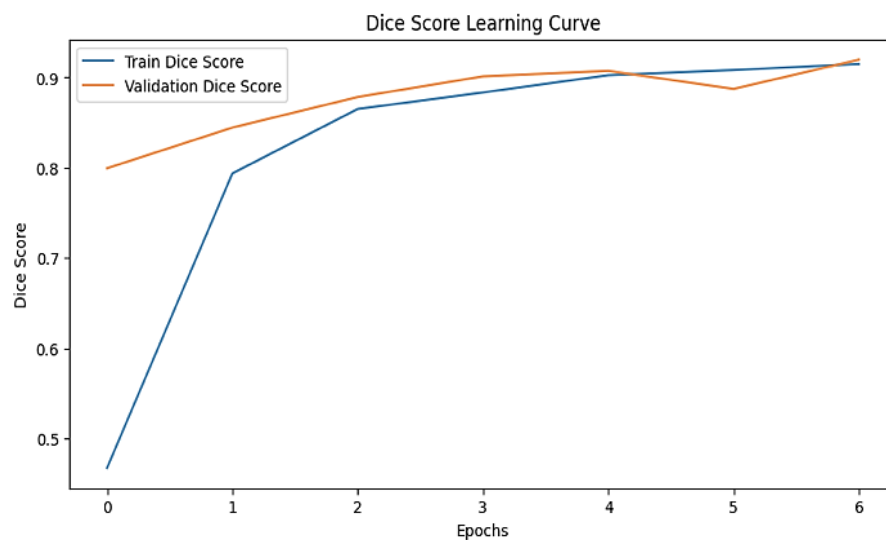


Figure 9. Training and validation dice score (DeepLabv3+).

4.1.2 UNet

The U-Net model trained throughout eight epochs as well. First epoch's training loss was 0.105; IoU was 0.799 and PA was 0.975. The F1 and Dice scores came back at 0.726 each. The valid set yielded Dice and F1 scores of 0.882, IoU of 0.893, PA of 0.985, and a loss of 0.282. U-Net improved substantially throughout the second quarter. Training loss dropped drastically to 0.019 using a PA of 0.982 and IoU of 0.907. The F1 and Dice scores came out at 0.898. Validations losses down to 0.014, PA to 0.686, IoU to 0.905, Dice and F1 scores to 0.897. The third epoch followed similar path with a training loss down to 0.011 and PA stabilizing at 0.982. Turning up respectively as 0.919 and 0.925 were the IoU and Dice scores. The validation criteria reported a loss of 0.009, PA of 0.985, IoU of 0.935, and Dice and F1 scores of 0.931.

With an IoU of 0.938 and a Dice score of 0.935, the training loss dropped yet further to 0.008 in the fourth epoch. The validation statistics revealed a loss of 0.006, PA of 0.985, IoU of 0.945, and Dice and F1 scores of 0.943. The U-Net model had a training loss of 0.004 at the ninth epoch, an IoU of 0.511 and Dice score of 0.499. The most recent validation findings showed a little fluctuation in the loss at 0.005, PA at 0.985, IoU at 0.938, Dice and F1 scores at 0.935 as shown in **Figures 10 to 14**.

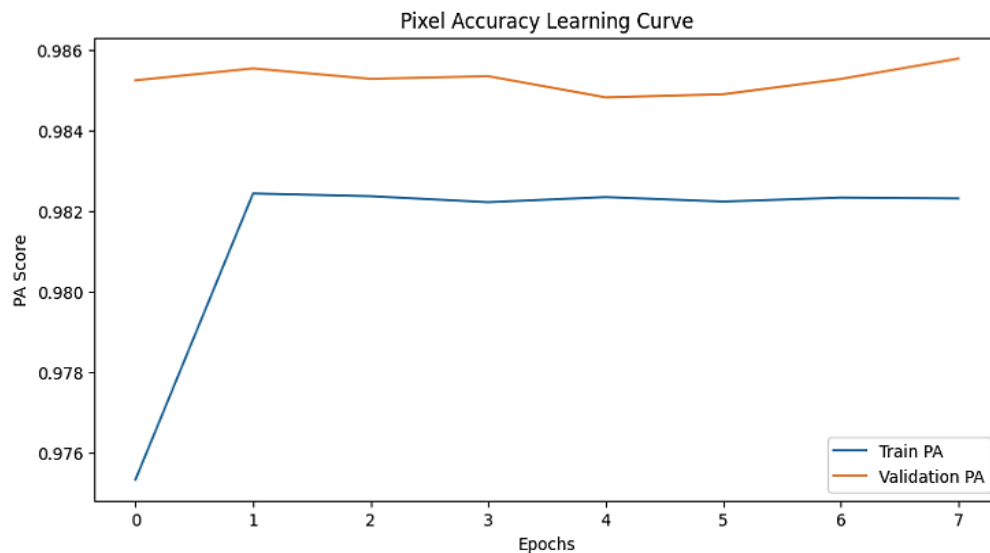


Figure 10. Training and validation accuracy graph (U-Net).

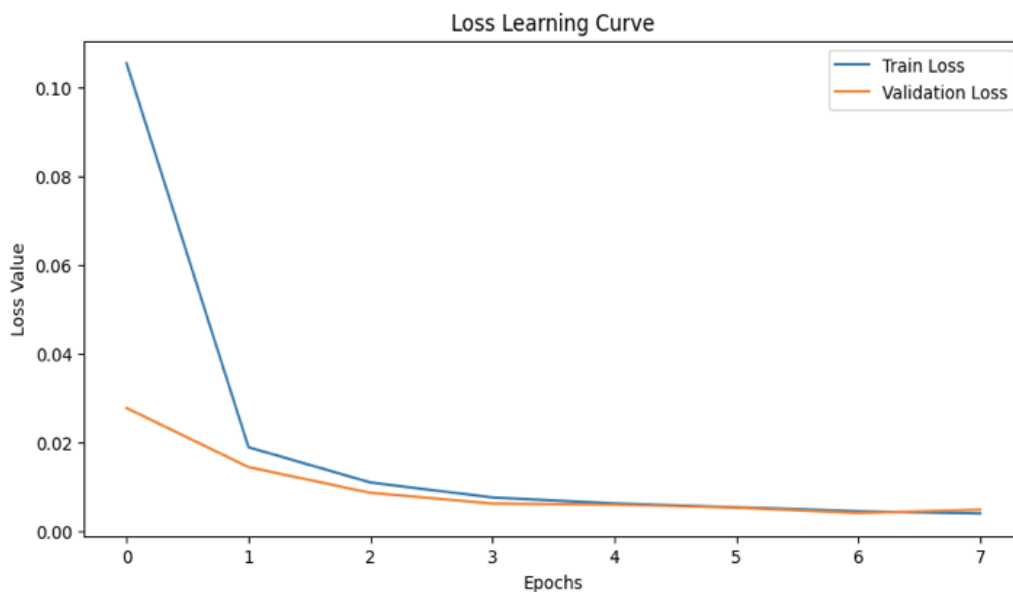


Figure 11. Training and validation loss graph (U-Net).

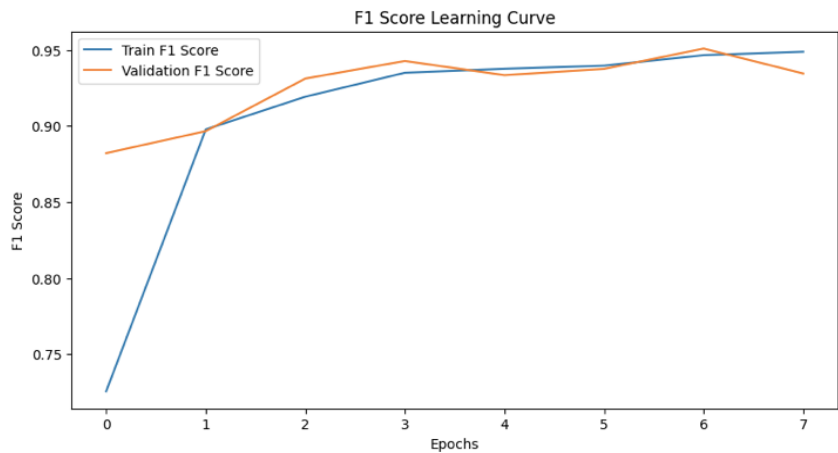


Figure 12. Training and validation F1 score (U-Net).

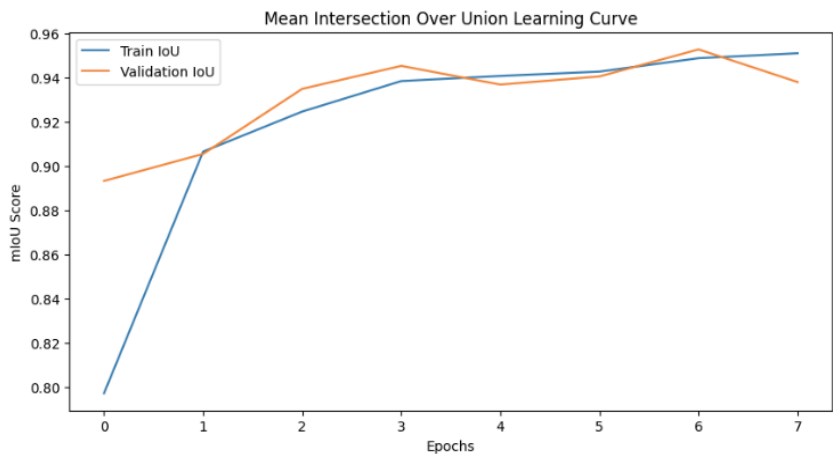


Figure 13. Training and validation mean IoU (U-Net).

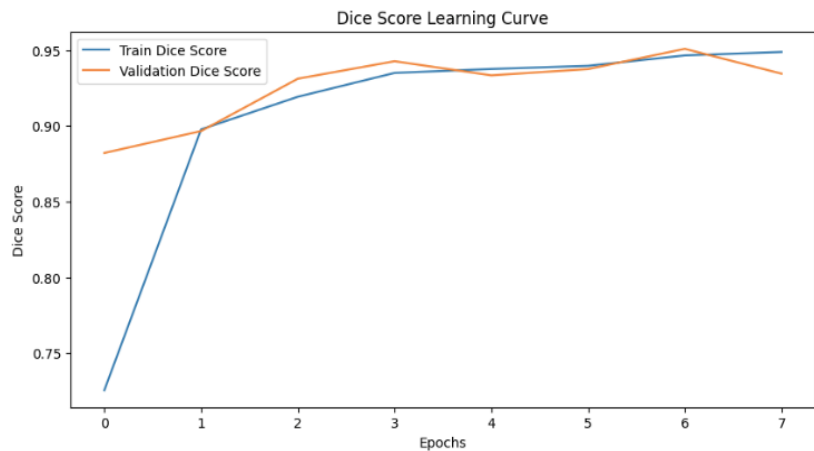


Figure 14. Training and validation dice score (U-Net).

4.2 Visualization of Segmentation

On a test set, the performance of both models was assessed; results from visible segmentation helped to gauge their accuracy in spinal structure characterization. The lumbar, sacral, thoracic, and cervical spinal bone segmentation of the models was vividly shown by the visual aids. When segmenting the vertebrae with somewhat varying limits between various spinal components, the DeepLabv3+ model showed really good performance. Especially good in describing the cervical and lumbar areas—which are sometimes difficult because to their anatomical complexity and the presence of thick tissues—the model accurately mirrored the overall form and alignment of the spine (**Figure 15**). Differentiating between neighboring vertebrae in places like the thoracic region where the components are tightly packed also offered some difficulties. Sometimes little segmentation mask errors resulted from the model blurring of the limits.

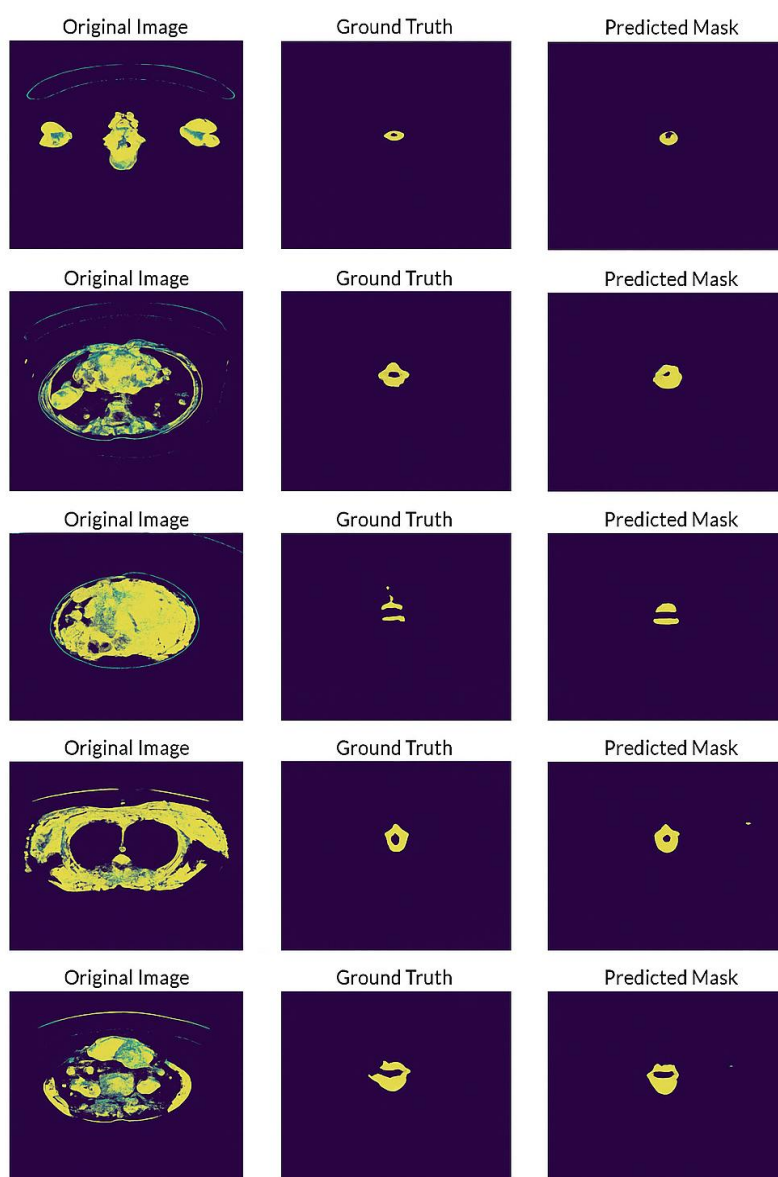


Figure 15. Visualization of segmentation of DeepLabv3+.

The U-Net model produced extremely finely detailed segmentation findings with remarkable boundary delineation across all spinal areas. U-Net allowed the model to preserve high-density characteristics by use of skip connections, therefore enabling accurate segmentation of spinal bodies and intervertebral discs. The U-Net model demonstrates great power when segmenting the sacral vertebrae, a region that commonly provides challenges due to their complicated character and closeness to other vertebrae structures (**Figure 16**). U-Net's segmentation masks were more correct and consistent than DeepLabv3+ especially in areas with minor anatomical traits.

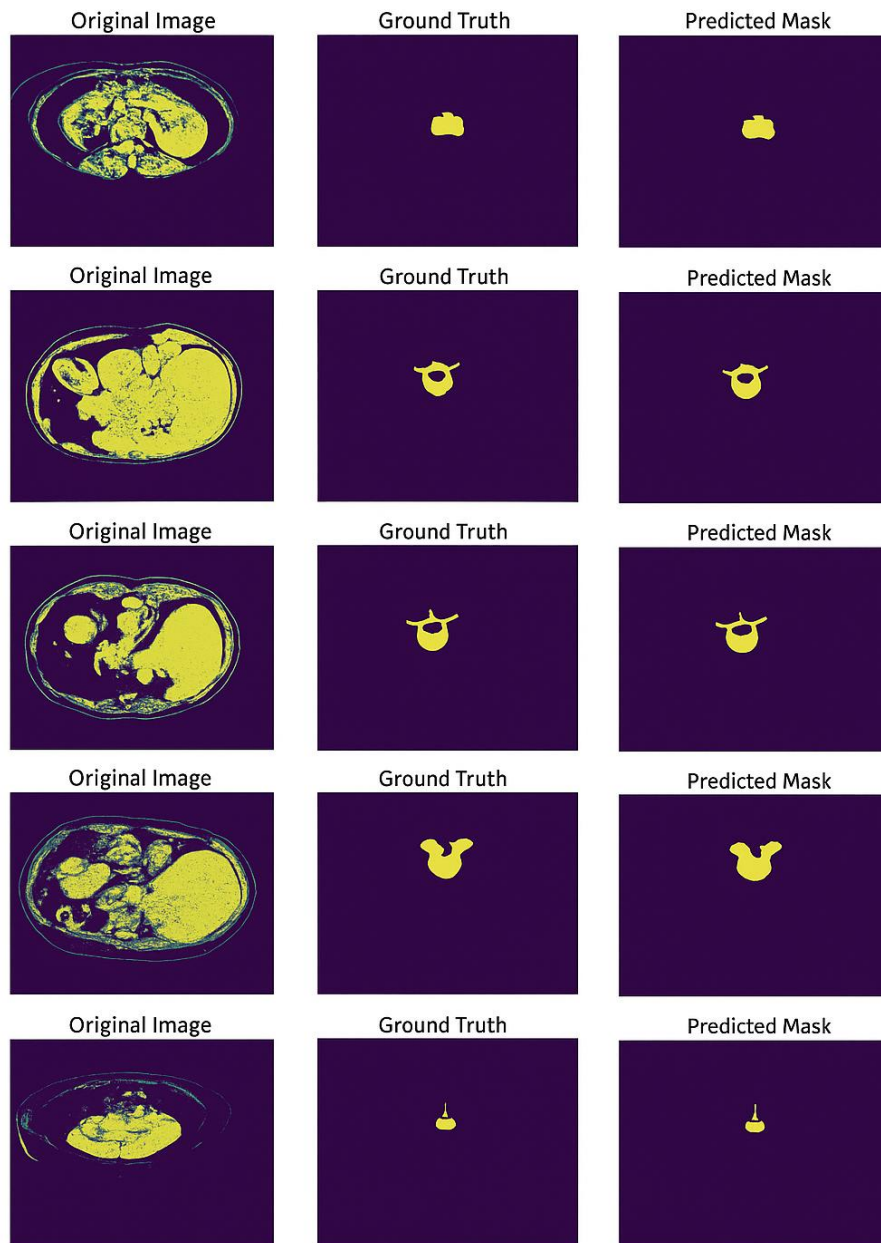


Figure 16. Visualization of segmentation of U-Net.

4.3 Comparative Analysis

4.3.1 DeepLabv3+ vs U-Net

Both models demonstrated strong performance, but U-Net consistently outperformed DeepLabv3+ in key metrics such as IoU, dice score, and F1 score, indicating better segmentation accuracy and boundary delineation as shown in **Table 1**.

4.3.2 Proposed Model vs State-of-the-art

Our DeepLabv3+ model alongside the U-Net model received comparison testing against multiple state-of-the-art approaches which were previously used for vertebrae segmentation purposes. The stacked sparse autoencoder (SSAE) architecture provided promising findings in vertebrae segmentation by reaching precision at 89.9% combined with recall at 90.2% and accuracy at 98.8% and F-score at 90.4% as well as Intersection over Union (IoU) at 82.6% and Dice Coefficient (DC) at 90.2%. The SSEA approach achieved better results than the conventional three-layered TSAE model by scoring higher precision (87.3%), recall (86.9%), accuracy (95.7%), F-score (82.8%), and IoU (77.8%) and DC (85.2%) values in evaluation metrics. The vertebrae segmentation analysis included evaluation of deep learning techniques U-Net and SpineParseNet and also involved the study of PaDBNs as well as TSAE and butterfly FCN and OP-convNet models and Mask R-CNN. **Table 3** demonstrates the SSAE-based method generated better results than conventional approaches in all evaluation parameters by reaching superior F-scores and IoU values and dice coefficients. U-Net surpassed DeepLabv3+ by achieving superior test set results including an F-score of 93.5% and IoU of 93.8% and dice coefficient of 93.5%. DeepLabv3+ obtained F-score 92.0%, IoU 92.5% and dice coefficient of 92.0% on the test data. The assessment indicates our U-Net model produces comparable results which exceed SSAE-based method performance through superior measurements in both Dice score and IoU.

Table 3. State-of-the-art comparison.

Methods	F-score (%)	IoU (%)	DC (%)
Classical U-Net (Ronneberger et al., 2015)	81.4	71.9	83.7
SpineParseNet (Pang et al., 2021)	87.6	77.5	87.3
PaDBN Model (Qadri et al., 2019)	84.9	75.6	86.1
TSAE (Xu et al., 2016)	82.8	73.9	85.2
Butterfly FCN (Sekuboyina et al., 2018)	86.4	76.9	87.0
OP-convNet (Qadri et al., 2021)	90.2	82.3	89.9
Mask R-CNN (Wang et al., 2021)	70.1	53.1	69.2
SSAE (L2 regularized) (Qadri et al., 2023)	90.4	82.6	90.2
Proposed DeepLabv3+	92.0	92.5	92.0
Proposed U-Net	93.5	93.8	93.5

5. Conclusion

This study analyzes the DeepLabv3+ and U-Net models for the segmentation of spinal vertebrae from CT images using a dataset of finely segmented NIfTI files. The results showed that both models can achieve exceptional accuracy in spine segmentation; U-Net somewhat outperforms DeepLabv3+ across several measures, including the dice coefficient and intersect over Union (IoU). Particularly about skip connections, U-Net's design allows more exact border delineation—needed in medical imaging applications where suitable segmentation may significantly affect diagnosis and therapy planning. The paper emphasizes the relevance of deep learning methods in automating difficult segmentation tasks, therefore providing potential benefits in patient outcomes and healthcare systems. Moreover, the results underline the need of constant innovation in model design and training techniques to address issues like anatomical complexity and different image quality. All things considered, this study offers insightful analysis of the use of deep learning in medical picture segmentation, thus laying basis for next projects improving the accuracy and efficiency of automated diagnostic tools.

6. Future Scope and Limitations

The promising research outcomes from DeepLabv3+ and U-Net-based framework segmentation of CT spine images create several research opportunities for future work. The main drawback of this research project stems from executing training and evaluation on a single dataset that fails to represent the complete range of clinical diversity across multiple institutions. The models have restricted applicability to multiple scanning protocols combined with various patient groups and different scanner platforms. Future investigations need to conduct model performance evaluation on external data collections as well as explore domain adaptation approaches for increased institution-wide robustness. The segmentation of tightly packed irregular vertebrae within regions with severe pathology proves challenging for the U-Net model despite its performance excellence in most evaluation metrics. The segmentation of tightly packed or irregular vertebrae can benefit from implementing shape priors and spatial attention approaches as well as anatomical restrictions for improved boundary accuracy. The research should investigate three-dimensional (3D) convolutional networks as a possible solution to fully capitalize on volumetric data for achieving better spatial consistency between slice levels. The implementation of the solution for clinical translation should emphasize real-time performance as well as workflow compatibility while exploring ways to explain the model outputs effectively to practitioners. The addition of disease classification and vertebral labeling components to the segmentation task would lead to a more complete computer-aided diagnosis solution for spinal problems.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the editor and anonymous reviewers for their comments that helped improve the quality of this work.

AI Disclosure

The author(s) declare that no assistance is taken from generative AI to write this article.

References

- Banerjee, S., Lyu, J., Huang, Z., Leung, F.H., Lee, T., Yang, D., Su, S., Zheng, Y., & Ling, S.H. (2022). Ultrasound spine image segmentation using multi-scale feature fusion skip-inception U-Net (SIU-Net). *Biocybernetics and Biomedical Engineering*, 42(1), 341-361. <https://doi.org/10.1016/j.bbe.2022.02.011>.
- Buerger, C., von Berg, J., Franz, A., Klinder, T., Lorenz, C., & Lenga, M. (2020). Combining deep learning and model-based segmentation for labeled spine CT segmentation. In *Medical Imaging 2020: Image Processing* (Vol. 11313, pp. 307-314). SPIE. Houston, Texas, United States. <https://doi.org/10.1117/12.2549566>.
- Chang, H., Zhao, S., Zheng, H., Chen, Y., & Li, S. (2020). Multi-vertebrae segmentation from arbitrary spine MR images under global view. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racocanu, D., Joskowicz, L. (eds) *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2020*. Springer Cham. pp. 702-711. https://doi.org/10.1007/978-3-030-59725-2_68.
- Graf, R., Schmitt, J., Schlaeger, S., Möller, H.K., Sideri-Lampretsa, V., Sekuboyina, A., & Kirschke, J.S. (2023). Denoising diffusion-based MRI to CT image translation enables automated spinal segmentation. *European Radiology Experimental*, 7(1), 70. <https://doi.org/10.1186/s41747-023-00385-2>.
- Kim, H.J., Yang, J.H., Chang, D.G., Suk, S.I., Suh, S.W., Song, K.S., Park, J.B., & Cho, W. (2020a). Adult spinal deformity: current concepts and decision-making strategies for management. *Asian Spine Journal*, 14(6), 886-897. <https://doi.org/10.31616/asj.2020.0568>.

- Kim, Y.J., Ganbold, B., & Kim, K.G. (2020b). Web-based spine segmentation using deep learning in computed tomography images. *Healthcare Informatics Research*, 26(1), 61-67. <https://doi.org/10.4258/hir.2020.26.1.61>.
- Kushnure, D.T., & Talbar, S.N. (2021). MS-UNet: a multi-scale UNet with feature recalibration approach for automatic liver and tumor segmentation in CT images. *Computerized Medical Imaging and Graphics*, 89, 101885. <https://doi.org/10.1016/j.compmedimag.2021.101885>.
- Li, H., Luo, H., Huan, W., Shi, Z., Yan, C., Wang, L., Mu, Y., & Liu, Y. (2021). Automatic lumbar spinal MRI image segmentation with a multi-scale attention network. *Neural Computing and Applications*, 33(18), 11589-11602. <https://doi.org/10.1007/s00521-021-05856-4>.
- Liu, Y., Bai, X., Wang, J., Li, G., Li, J., & Lv, Z. (2024). Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism. *Engineering Applications of Artificial Intelligence*, 127(part A), 107260. <https://doi.org/10.1016/j.engappai.2023.107260>.
- Pang, S., Pang, C., Su, Z., Lin, L., Zhao, L., Chen, Y., Zhou, Y., Lu, H., & Feng, Q. (2022). DGMSNet: spine segmentation for MR image by a detection-guided mixed-supervised segmentation network. *Medical Image Analysis*, 75, 102261. <https://doi.org/10.1016/j.media.2021.102261>.
- Pang, S., Pang, C., Zhao, L., Chen, Y., Su, Z., Zhou, Y., Huang, M., Yang, W., Lu, H., & Feng, Q. (2021). SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation. *IEEE Transactions on Medical Imaging*, 40(1), 262-273. <https://doi.org/10.1109/tmi.2020.3025087>.
- Prakash, N., Manconi, A., & Loew, S. (2020). Mapping landslides on EO data: performance of deep learning models vs. traditional machine learning models. *Remote Sensing*, 12(3), 346. <https://doi.org/10.3390/rs12030346>.
- Qadri, F.S., Ai, D., Hu, G., Ahmad, M., Huang, Y., Wang, Y., & Yang, J. (2019). Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images. *Applied Sciences*, 9(1), 69. <https://doi.org/10.3390/app9010069>.
- Qadri, S.F., Lin, H., Shen, L., Ahmad, M., Qadri, S., Khan, S., Khan, M., Zareen, S.S., Akbar, M.A., Heyat, M.B.B., & Qamar, S. (2023). CT-based automatic spine segmentation using patch-based deep learning. *International Journal of Intelligent Systems*, 2023(1), 1-14. <https://doi.org/10.1155/2023/2345835>.
- Qadri, S.F., Shen, L., Ahmad, M., Qadri, S., Zareen, S.S., & Khan, S. (2021). OP-convNet: a patch classification-based framework for CT vertebrae segmentation. *IEEE Access*, 9, 158227-158240. <https://doi.org/10.1109/access.2021.3131216>.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds) *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Springer Cham, pp. 234-241. ISBN: 978-3-319-24574-4. https://doi.org/10.1007/978-3-319-24574-4_28.
- Saravagi, D., Agrawal, S., Saravagi, M., Jain, S.K., Sharma, B., Mehbodniya, A., Chowdhury, S., & Webber, J.L. (2023). Predicting lumbar spondylolisthesis: a hybrid deep learning approach. *Intelligent Automation & Soft Computing*, 37(2), 2133-2151. <http://dx.doi.org/10.32604/iasc.2023.039836>.
- Sekuboyina, A., Kukačka, J., Kirschke, J.S., Menze, B.H., Valentinitzsch, A. (2018). Attention-driven deep learning for pathological spine segmentation. In: Glocker, B., Yao, J., Vrtovec, T., Frangi, A., Zheng, G. (eds) *Computational Methods and Clinical Applications in Musculoskeletal Imaging*. Springer Cham, pp. 108-119. https://doi.org/10.1007/978-3-319-74113-0_10.
- Srivastava, A., Jha, D., Chanda, S., Pal, U., Johansen, H.D., Johansen, D., Riegler, M.A., Ali, S., & Halvorsen, P. (2022). MSRF-Net: a multi-scale residual fusion network for biomedical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(5), 2252-2263. <https://doi.org/10.1109/jbhi.2021.3138024>.

- Tanwar, V., Anand, V., Chauhan, R., & Rawat, D. (2023). Improving patient care through the use of a highly accurate CNN model for the detection of cervical spine fractures. In *2023 3rd International Conference on Smart Generation Computing, Communication and Networking* (pp. 1-5). IEEE. Bangalore, India. <https://doi.org/10.1109/smartgencon60755.2023.10442236>.
- Ungi, T., Greer, H., Sunderland, K.R., Wu, V., Baum, Z.M.C., Schlenger, C., Oetgen, M., Cleary, K., Aylward, S.R. & Fichtinger, G. (2020). Automatic spine ultrasound segmentation for scoliosis visualization and measurement. *IEEE Transactions on Biomedical Engineering*, 67(11), 3234-3241. <https://doi.org/10.1109/tbme.2020.2980540>.
- Wang, K.N., Li, S.X., Bu, Z., Zhao, F.X., Zhou, G.Q., Zhou, S.J., & Chen, Y. (2024). SBCNet: scale and boundary context attention dual-branch network for liver tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, 28(5), 2854-2865. <https://doi.org/10.1109/jbhi.2024.3370864>.
- Wang, R., Voon, J.H.Y., Ma, D., Dabiri, S., Popuri, K., & Beg, M.F. (2021). Vertebra segmentation for clinical CT images using Mask R-CNN. In: Jarm, T., Cvetkoska, A., Mahnič-Kalamiza, S., Miklavcic, D. (eds) *8th European Medical and Biological Engineering Conference*. Springer International Publishing Cham, pp. 1156-1165. ISBN: 978-3-030-64610-3. https://doi.org/10.1007/978-3-030-64610-3_130.
- Wasserthal, J. (2024). Spine segmentation from CT scans [Data set]. *Kaggle*. <https://doi.org/10.34740/KAGGLE/DSV/8422484>.
- Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., & Madabhushi, A. (2016). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Transactions on Medical Imaging*, 35(1), 119-130. <https://doi.org/10.1109/tmi.2015.2458702>.
- Zhang, D., Chen, B., & Li, S. (2021). Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy. *Medical Image Analysis*, 67, 101861. <https://doi.org/10.1016/j.media.2020.101861>.



Original content of this work is copyright © Ram Arti Publishers. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.