IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3770∼3781
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3770-3781

EEB7-UNet: a deep learning framework for automated segmentation of fractured C-spine vertebrae

Abhishek Kumar Pandey[1], Pateel G. P.[2], Kedarnath Senapati[1]
[1]Department of Mathematical and Computational Sciences, National Institute of Technology Karnataka Surathkal, Mangaluru, India
[2]Department of Computer Science and Engineering, NMAM Institute of Technology, Nitte (Deemed to be university), Udupi, India

Article Info

Article history:

Received Sep 30, 2024
Revised Jun 16, 2026
Accepted Jul 9, 2026

Keywords:

Cervical spine injury
Computed tomography images
Deep learning
Semantic segmentation
U-Net
Vertebra segmentation

ABSTRACT

Accurate identification of vertebral fracture (VF) regions in computed tomography (CT) images is crucial for surgeons prior to treatment planning, but remains challenging due to irregular vertebral boundaries, low contrast, noise, and image unevenness. Recent advancements in deep learning have shown promising results compared to conventional manual diagnosis methods in detecting anomalies and segmenting regions of interest in medical imaging. In this study, a deep learning model, enhanced EfficientNetB7 U-Net (EEB7-UNet), is proposed to segment the fractured cervical vertebrae. It includes custom data augmentation to increase the data size and a hybrid learning rate scheduler strategy technique for faster convergence, which increases the generalizability and robustness of the model. The proposed model achieved an improved dice score index of 95.53% and Jaccard coefficient index of 93.85% on the test dataset. Furthermore, the EEB7-UNet has emerged as a moderate size with 98.6 MB. The approach yields superior performance in terms of dice score index and Jaccard coefficient index compared to the other state of the art convolutional neural network (CNN) used as an encoder in the U-Net. This research also compared the performance of the proposed model with three other studies in similar contexts, reported in the literature.

Corresponding Author:

Pateel G. P.
Department of Computer Science and Engineering, Nitte (Deemed to be university)
NMAM Institute of Technology
Nitte, Karkala Taluk, Udupi – 574110, Karnataka, India
Email: pateel.gp@nitte.edu.in

1. INTRODUCTION

The spinal cord located inside the spinal canal, is a crucial part of the nervous system and plays an important role in transferring signals from the brain to the entire body. Severe injury in the vertebrae may affect the nerve cells and axons in the spinal cord, which may lead to paralysis below the level of injury. Following a spinal cord injury (SCI), some patients might face permanent loss and, others could achieve partial recovery through proper treatment and rehabilitation. Research showed that the approximate cost of care for patients with SCI in the US is from $1.1 to $4.6 million/patient over their lifespan [1].

Vertebral bodies play a vital role in ensuring the spine's stability and serving as a safeguard for the spinal cord [2]. Vertebral fractures (VFs) are among the most common types of skeletal fractures [3]. VFs may result from either traumatic or non-traumatic conditions. Some traumatic causes of VFs are road traffic accidents, falls from height, sports injuries, gunshot injuries, and sudden movements that strain the back.

The non-traumatic causes include osteoporosis, bone tumors, spinal infections, degenerative disk disease, and spinal metastasis [4]. These VFs are unstable and can be dangerous, resulting in progressive deformity or neurological compromise. In a recent survey conducted in Vietnam on men and women over 50 years of age [5], the prevalence, incidence, and associated risk factors for VF were studied. Out of 401 individuals, it was found that 49 had VF, yielding a prevalence of 12.2%. The study also found that wedge fractures were another common type of VF.

Cervical spine (C-spine) injury occurs when any of the seven cervical vertebrae are broken or damaged. The symptoms of a C-spine injury are associated with neck pain, limited mobility, respiratory problems, or loss of sensation. Late diagnosis of such cases may cause sudden and severe loss of nerve function, as C-spine injuries are potentially serious and can cause serious complications if not properly treated. The available treatment options for C-spine VF can be vertebroplasty and kyphoplasty [6].

It is easier for healthcare professionals to diagnose severe fractures on computed tomography (CT) scans. However, manually diagnosing subtle or very narrow fractures can be challenging and tedious, as these fractures are not noticeable on CT scans. Accurate segmentation of vertebral bone in CT scans is necessary for various applications, such as measuring cortical thickness, regional analysis, and surgical planning. Automatic segmentation methods using artificial intelligence (AI) are faster, and more consistent, addressing the limitations of manual segmentation. However, this task is challenging due to partial volume artifacts, and unclear boundaries between the bones in the vertebrae. This study focuses on finding an AI solution to such problem.

Bae *et al.* [7] performed the segmentation of C-spine vertebrae in the axial plane and subsequently used some post-processing for the entire vertebral segmentation. To achieve this, the two-dimensional (2D) U-Net was used for segmentation, and then the result of 2D segmentation was used to detect separation points. The complete experiment was conducted on small datasets obtained from two independent hospitals.

Zhang and Wang [8] proposed a three-stage semi-automatic method for the C-spine vertebrae segmentation. In which, first, the adaptive thresholding method was used to segment the vertebra bone tissue, followed by PointNet++ for concave segmentation. PointNet++ handled 3D medical imaging data well but struggled to find the local details and large point clouds. Research by Liu *et al.* [9], a statistical method was proposed for the segmentation of vertebrae of the C-spin. The method performed well with a small dataset; however, it took a more computational resources and a longer time to get the segmentation result of a CT scan. In a recent study [10], a hybrid convolutional neural network (CNN)–vision transformers based model was proposed for lung cancer detection using CT images, which achieved impressive accuracy on benchmark datasets while also identifying different subtypes of lung cancer.

A web based spine segmentation model was introduced by Kim *et al.* [11], where U-Net was utilized for segmentation. However, the model lacked generalization capability as it utilized very small dataset for training. A hybrid method combining CNN and fully convolutional network (FCN) was used for automatic spine segmentation, which reported 94% dice similarity index (DSI) and 93% Jaccard coefficient index (JCI) [12]. The motivation arises from the importance of locating the accurate fractured area before surgery to prevent any damage to nearby healthy tissue. Enhancement of outline structures of fractures or tumors is really useful in surgical planning strategies. By observing the success of U-Net in many medical imaging applications, this work aims to utilize U-Net as baseline for the segmentation of fractured C-spine vertebrae using axial CT data. Also, in a recent survey, it was found that very little research work on AI assisted VFs detection were done till 2022, indicating that this field is still in its budding stage [13].

This research makes several key contributions. First, an enhanced EfficientNetB7 U-Net (EEB7-UNet) is introduced for accurate fractured C-spine vertebrae segmentation, leveraging fine-tuned EfficientNetB7 [14] as an encoder for precise feature extraction. Second, adaptive learning rate schedulers, along with a specific configuration of hyperparameters, including the use of the AdamW optimizer, are utilized in the proposed method for model optimization. Finally, an extensive comparative analysis of EEB7-UNet against various CNN encoders in the U-Net is conducted to validate the results obtained from the approach.

This work is an extension of the previous study [15]. The present manuscript specifically utilizes only fractured vertebra dataset to assist medical professionals in accurately delineating fractures prior to treatment planning. Furthermore, this work performs model interpretability analysis using gradient-weighted class activation mapping (Grad-CAM) [16] and presents a detailed error analysis.

The rest of the paper is organized as follows. The data preprocessing, and proposed methodology are presented in section 2. The performance of the proposed model is evaluated using different encoder networks, analyzed, and presented in section 3. A comparison between the proposed model and state-of-the-art (SOTA) methods is also given in section 3. The conclusion and future research directions are presented in section 4.

2. METHOD

The CT scan is often the first-line choice of imaging modality in cases of acute trauma or suspected fractures in the spine. It provides a detailed visualization of the bony structures and may identify fractures, dislocations, or other injuries [17], [18]. The dataset used in this study is originally CT data in DICOM format, which is obtained from the open source data repository [19]. Some sample images of C-spine with fracture, are shown in Figure 1, where the red arrow indicates fracture location.

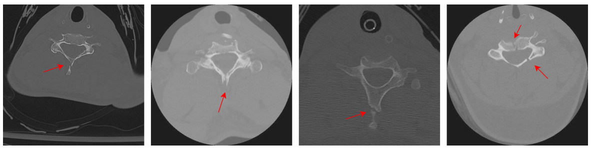


Figure 1. Sample fractured CT images of C-spine vertebrae in the axial plane

2.1. Data pre-processing

To address the difficulties involved in training neural network using DICOM images, the CT data are converted to JPG format in the data pre-processing step. Then the JPG images, along with the corresponding masks, are utilized for training. The masks are nothing but the ground truth segmentation performed by the medical experts and provided with the dataset. In the given dataset, the masks are in the sagittal orientation, whereas the training data are in the axial orientation, therefore, to ensure proper alignment, the transpose operation on each mask is performed. The overview of data pre-processing is displayed in Figure 2. The 2D axial CT slices is regularly being used with routine clinical workflows, where fracture assessment is commonly performed slice wise. Compared to 3D volumetric approaches, 2D models offer reduced model computational complexity and less inference time with test data, which makes them more practical for large-scale datasets and potential real-time deployment on cloud-based services.

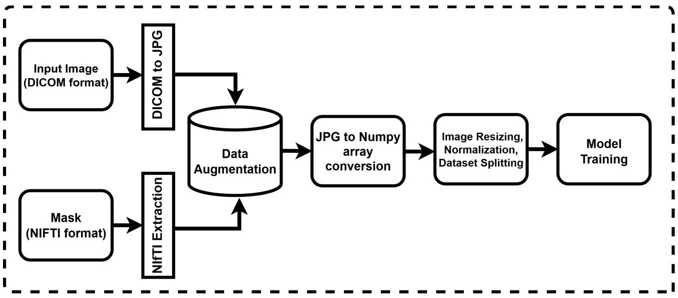


Figure 2. Workflow of data pre-processing

2.2. Custom data augmentation

To improve the model robustness and generalization, data augmentation operations on training data, including resizing, horizontal and vertical flips, shifting, scaling, rotation, grid distortion, optical distortion, elastic transformation, and coarse dropout, are considered. Coarse dropout is used to apply in coarse blocks of pixels as it avoids overfitting by randomly removing rectangular regions from the image [20]. These augmentation techniques are applied with different probabilities p to each image during the training process. For instance, $p = 0.65$ for horizontal flip indicates, 65% probability that the horizontal flip will be applied to a given image during the data augmentation. This p values adjustment helps to control the intensity and frequency of specific augmentation. Higher the p value for any operation leads to a higher chance of being applied that particular augmentation operation, while a lower value decreases the likelihood. This allows us to experiment with different augmentation strategies to get a good balance between diversity in training data, avoiding excessive perturbations. The data, along with augmented images are then fed into network for training.

2.3. Introduction of novel hyperparameters and their optimization

The learning rate is an important hyperparameter, and it controls the convergence and overall performance of the model. This experiment utilizes a combination of three learning rate scheduler strategies, as discussed as follows, to adaptively adjust the learning rate [21]. Warm-up scheduler gradually increases the learning rate at the beginning of training, which allows the model to converge faster. It is also proven that the deployment of a warm restart with the AdamW optimizer accelerates the training process by a factor of 10 [22]. The ReduceLROnPlateau scheduler (RLROnP) dynamically adjusts the learning rate to improve convergence and avoid local minima. The cosine annealing scheduler (CosAnneal) is used to vary the learning rate cyclically between a minimum and maximum value that exhibits the cosine function. In CosAnneal, the learning rate initially starts at a maximum value, gradually decreases in a smooth cyclical pattern, and then returns to the maximum value. By deploying the hybrid learning rate schedulers strategy, the model is able to explore different regions of the loss landscape, which potentially improves both the convergence and generalization [23]. In this study, grid search method is utilized to obtain the optimal hyperparameter settings for the U-Net with different CNNs, as illustrated in Table 1.

Table 1. Chosen hyperparameter values for different encoder architectures in U-Net

Hyperparameter	Experimented	Chosen value			
		VGG16	ResNet34	MobileNetV2	EfficientNetB7
Initial LR	$[0.0001 - 0.01]$	0.0001	0.005	0.001	0.001
Training batch size	$[16 - 512]$	32	32	32	32
Learning optimizer	–	Adam	SGD	Adam	AdamW
Dropout rate	$[0 - 0.5]$	0.2	0.4	0.4	0.5
Weight decay	$[10^{-6} - 10^{-2}]$	10^{-4}	10^{-4}	10^{-3}	10^{-4}
Width scaling	$[1.0 - 2.0]$	–	–	–	1.4
Depth scaling	$[1.0 - 3.0]$	–	–	–	2.6
LR scheduler	–	StepLR	RLROnP	StepLR	RLROnP+CosAnneal
Warmup steps	$[0 - 500]$	–	200	–	200
Warmup epochs	$[0 - 10]$	–	5	–	5
Mode	[min, max]	–	min	–	min
Factor	$[0.1 - 0.9]$	–	0.2	–	0.1

Although grid search is used for hyperparameter tuning, the motivation behind choosing specific hyperparameters is based on the architecture of the chosen encoder. EfficientNetB7 contains a large number of trainable parameters and has deep feature hierarchies, that makes its training unstable. If the large value of learning rates is taken initially, it increases the risk of overfitting. To address this issue, a lower learning rate is employed along with dropout and weight decay regularization, while the AdamW optimizer is used to ensure stable convergence [22]. The initial learning rate and weight decay are set to a value of 10^{-3} and 10^{-4}, respectively. Then, the warm-up scheduler slowly increases the learning rate and helps the model to gradually start learning meaningful patterns. The parameters 'warm-up step' and 'warm-up epoch' in this scheduler are chosen to be 200 and 5, respectively. In the ReduceLROnPlateau scheduler, the mode is chosen as 'min' mode after observing no improvement in validation loss. The 'Factor' parameter reduces the learning rate, which further helps to decrease validation loss. The value for the parameter 'Factor' is chosen to be 0.1. The value of training batch size is chosen as 32, based on the system configuration and graphics processing unit (GPU) memory. The width and depth scaling coefficients are key parameters of EfficientNetB7, which balance the compound scaling. The value of the width and depth scaling of EfficientNetB7 are chosen as 1.4 and 2.6, respectively.

2.4. Proposed model

U-Net based model has been widely used for various image segmentation tasks, especially in the biomedical domain by leveraging skip connections between sub-blocks of encoders and decoders. The performance of U-Net based model strongly depends on the encoder block of the model [24]–[26]. This study believe that utilizing EfficientNetB7 [14] as an encoder in the U-Net along with custom data augmentation techniques and various learning rate scheduler could be a promising model for segmenting fine structure in medical imaging. In this study, a U-Net based model is proposed for accurate segmentation of the fractured vertebrae, which is referred as EEB7-UNet. To validate the proposed model, several experiments are conducted by with the various SOTA CNNs, such as visual geometry group-16 (VGG16) [27], residual network with 34 layers (ResNet34) [28], and MobileNetv2 [29] as encoders in U-Net. The architecture of the proposed model is displayed in Figure 3, where the blue lines represent the skip connection, which obtain context from different blocks of the encoder.

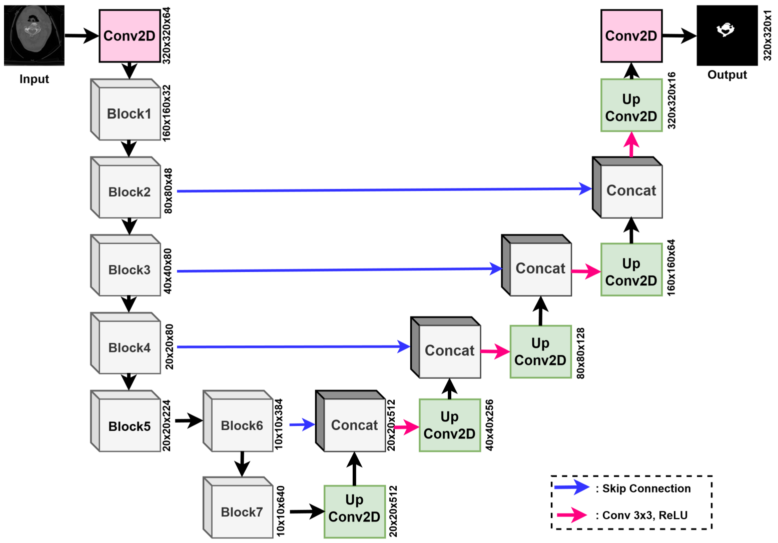


Figure 3. Architecture of the EEB7-UNet model

2.4.1. EfficientNet as encoder

In this study, several experiments are carried out with different CNNs as encoders of U-Net, and found that EfficientNetB7 resulted in better performance. EfficientNetB7 is extensively validated in medical image segmentation tasks and provides a well-established hierarchical feature extraction mechanism, which is particularly suitable for subtle anatomical structure delineation. It is structured multi-stage feature maps that integrate seamlessly with U-Net skip connections, which facilitates effective multi-scale feature fusion. Moreover, EfficientNetB7 benefits from widely available and stable pretrained weights, enabling reliable transfer learning in medical imaging applications. Therefore, considering architectural compatibility, segmentation robustness, and practical deployment stability, EfficientNetB7 is adopted as the encoder backbone in the proposed framework. The EfficientNetB7 incorporates depth wise separable convolution, which reduces the number of parameters, and the floating point operations per second, while preserving image resolution. EfficientNet is one of the SOTA CNN models that exhibits superior performance through adjustments in network width, depth, and resolution scaling. The balance between these three scaling factors is achieved through the compound scaling technique [14]. The architectural diagram of EfficientNetB7 is provided in Figure 4, which contains seven different subblocks with the basic building block is the mobile inverted bottleneck (MB) convolution block.

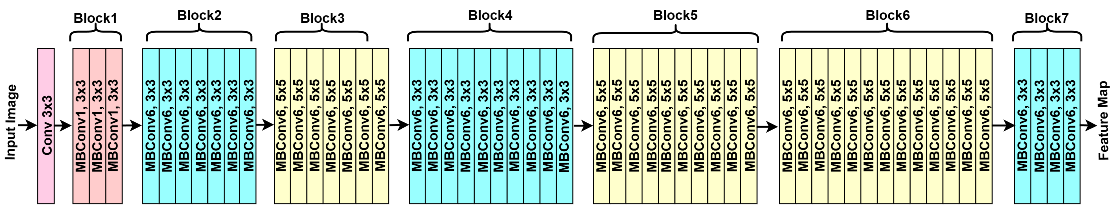


Figure 4. Block diagram of EfficientNetB7 with basic building blocks as MBConv

Figure 5 represents the architecture of the MB convolution block, which integrates the squeeze and excitation (SE) block with the inverted residual network [30]. In the context of feature extraction from an image, certain channels in the image may carry more relevant information than others. Selecting such channels for convolution operation, enhances the network's focus on critical features, potentially leading to improvements in both representational efficiency and overall segmentation performance. Furthermore, integration of the SE block within the MBConv facilitates channel wise feature recalibration by explicitly assigning more weight to the relevant channels. This block identifies and points out the more informative channels during training.

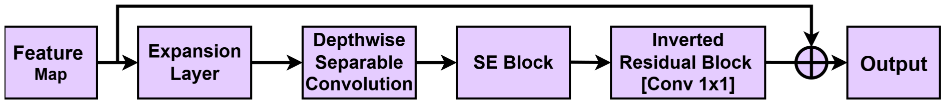


Figure 5. Block diagram of MBConv block

2.4.2. Decoder

In the decoder part of EEB7-UNet, up-sampling of the extracted feature maps is performed to obtain a high resolution segmentation mask using transpose convolution. In the encoder block of EEB7-UNet, the feature maps from the input image are being extracted, but the spatial resolution of the original image is downsampled as it goes through several blocks. The decoder block consists of two major blocks, the Up Conv2D and the Concat. The Up Conv2D blocks upsample the feature using transpose convolution, and the Concat blocks concatenate the feature maps received from encoder block through skip connections and from the previous Up Conv2D blocks. These two operations also minimize the vanishing gradient problem. By combining low level and high level features through skip connections, the model learns to generate more accurate segmentation masks.

2.5. Experimental setup and model training

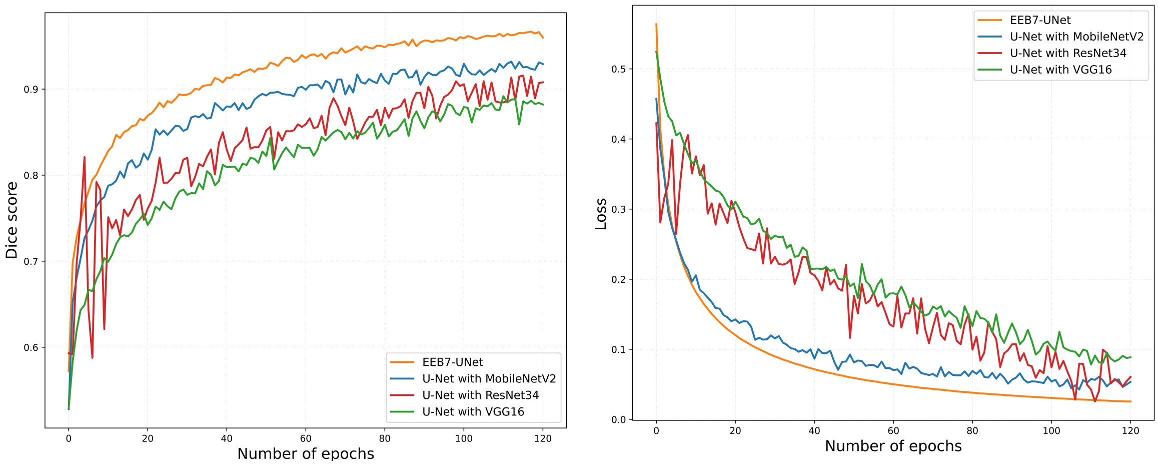
The proposed model is implemented using PyTorch 1.9.0 with CUDA 11.1 support and Python 3.8.3, while maintaining a random seed value of 42. It took approximately 30 hours (1.25 days) to train the EEB7-UNet model. This study utilized 29,658 images from 724 patients. The entire dataset is divided patient wise into training, validation, and testing sets in a ratio of 70%, 15%, and 15%, respectively. To prevent data leakage, the dataset is divided on a patient wise basis. Strict separation between the training, validation, and testing sets is maintained to ensure that no information from the test set is used during training. The training set, comprising the CT images and their corresponding masks, are used to train the model. The binary cross entropy with logits (BCEWithLogitsLoss) is chosen as loss function, which is expressed as (1).

$$L(x_n, y_n) = -\frac{1}{N}\sum_{i=1}^{N}(y_n \log(\sigma(x_n)) + (1-y_n)\log(1-\sigma(x_n))) \tag{1}$$

Where $\sigma(x_n) = \frac{1}{1+e^{-x_n}}$ is the sigmoid function, x_n is a input to the model, and y_n is true label for the n^{th} sample which is either 0 or 1.

The BCEWithLogitsLoss handles class imbalance that is common in segmentation tasks, as background class is often dominant. The BCEWithLogitsLoss, combined with a sigmoid activation function squashes the model's predictions into the range [0, 1], which represents the probabilities of belonging to the foreground class, making it well-suited for a binary segmentation task. The choice of AdamW as the optimizer, incorporates the weight decay and turns out to be effective for regularization. The synergy between AdamW and different learning rate schedulers enhances the overall performance of our model. Early stopping technique, based on the validation score is also employed in the model, to ensure that it halts the training process at the appropriate juncture.

The performance curves of the proposed model, in terms of dice score, along with those of U-Net models using VGG16, ResNet34, and MobileNetV2 encoders, are presented in Figure 6. High variations and oscillations in the dice score in Figure 6(a) and loss curves in Figure 6(b) can be observed for the U-Net model with VGG16, ResNet34, and MobileNetV2 as encoders separately. In contrast, the proposed model demonstrates reduced variation as well as the best dice score and lowest loss evolution among all models.



(a) (b)

Figure 6. Performance curves of the proposed model during training of (a) dice score variation over epochs and (b) loss variation over epochs

3. RESULTS AND DISCUSSION

In this section present a detailed analysis of the experimental outcomes of the work and discuss their implications.

3.1. Evaluation metrics

The regions of interest in the CT image correspond to fractures in the vertebrae, which are located in some tiny parts of the image. Therefore, the class imbalance is inevitable as the background class dominates. DSI and JCI are less affected by class imbalance, as they penalize both false positives and false negatives. Therefore, these metrics are appropriate as performance measures for our model. Both metrics quantify the degree of overlap between the predicted regions and ground truth regions. The DSI and JCI are mathematically expressed as (2) and (3).

$$\text{DSI} = \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \hat{y}_i^2} \tag{2}$$

$$\text{JCI}\ (y_i, \hat{y}_i) = \frac{|y_i \cap \hat{y}_i|}{|y_i \cup \hat{y}_i|} \tag{3}$$

Where y_i and $\hat{y}_i$ represent the ground-truth and predicted masks respectively of i^{th} sample. The term $\sum_{i=1}^{N} y_i \hat{y}_i$ calculates the element-wise product (intersection) of the ground truth y and predicted mask $\hat{y}$. The values of both DSI and JCI range from 0 to 1, where 0 indicates no overlap and 1 signifies perfect overlap, between the predicted mask and ground truth masks.

3.2. Performance analysis with different encoders

In Table 2, increasingly better performance in terms of DSI and JCI are observed on the test dataset for all the encoders used in U-Net. However, the EEB7-UNet achieved superior performance, with the best DSI and JCI of 95.53% and 93.85%, respectively. Mixed results, in terms of model size, are observed in our experiment. Among all the models listed in Table 2, U-Net with MobileNetV2 as encoder turns out to be the smallest model with size 18.2 MB. However, its performance is least among all the models considered therein. Also, U-Net with VGG16 is comparatively lightweight with a size of 46.9 MB, but this comes with the trade-off of compromised performance, with lower DSI and JCI compared to EEB7-UNet. In fact, the EEB7-UNet has outperformed all the models, achieving the highest test score of DSI and JCI with 95.53% and 93.85%, respectively. However, it has emerged as a moderate-sized model at 98.6 MB, which can be practically deployed on the cloud or integrated with a portable clinical system.

Table 2. Comparison of model performance with different encoder network

Model	DSI (%)	JCI (%)	Model size (MB)
U-Net with VGG16 encoder	72.19	69.2	46.9
U-Net with ResNet34 encoder	73.36	70.26	138.4
U-Net with MobileNetV2 encoder	72.13	68.49	18.2
Proposed model (EEB7-UNet)	95.53	93.85	98.6

3.3. Analysis of results

Figure 7 illustrates the segmentation results obtained from the U-Net with different encoders and also highlights the effectiveness of the proposed model by zooming into the fractured area. The first column in Figure 7(a) indicates the original image, while the subsequent columns represent the result obtained from the U-Net with VGG16, ResNet34, MobileNetV2, and the proposed model, respectively. Comparing the original images with the corresponding predicted masks obtained from different models, it is evident from Figure 7(a) that the proposed model (EEB7-UNet) is able to predict the accurate profile, in terms of shapes and edges, of the vertebrae.

Figure 7(b) demonstrates the results of the EEB7-UNet on some sample test data. The first column in Figure 7(b) shows a few random sample images from the test set, with the injury area marked in a red rectangular box. The second and third columns represent the ground truth and predicted masks, respectively. Then, the predicted masks obtained from the proposed model are superimposed with the corresponding ground truth masks and displayed in the fourth column. It is observed that, most of the areas are perfectly overlapped. The non-overlapping regions are depicted in dark red color. A part of a non-overlapping region is zoomed in and displayed in a yellow rectangular box in top right corner of the respective sub-figures. The last column illustrates the overlap of the original image with the predicted mask, in which the overlapped regions in the superimposed images are highlighted in red. It can be observed that, the bony structure and fractures of the vertebrae are not clearly visible in the original CT images in the first column, but in the last column, one can identify them clearly. It can also be observed that the predicted mask obtained from the EEB7-UNet is very close to the ground truth, which demonstrates the effectiveness of the proposed model.

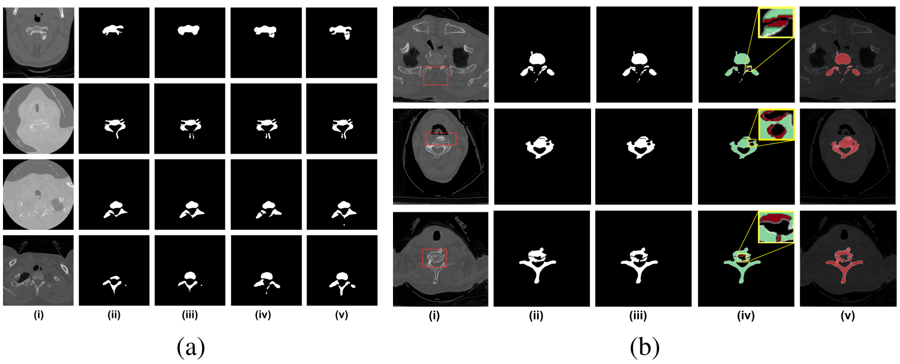


Figure 7. Segmentation results on some test samples of (a) comparison of U-Net with different encoders and the proposed EEB7-UNet and (b) zoomed-in fractured regions predicted by EEB7-UNet

3.4. Error analysis

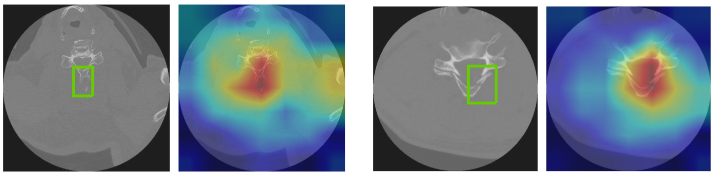
To understand where the model falls short, this study examined test samples with low DSI and JCI scores, presented in Table 3. The error analysis shown in Table 3 indicates that low-contrast CT slices lead to indistinct vertebral boundaries, which reduces the model's performance and results in under-segmentation. In cases with complex anatomical structures of vertebral bone, the model tends to produce mild over-segmentation by extending into adjacent soft tissue regions. Noisy CT images have acquisition artifacts that degrade the boundary clarity, which eventually cause irregular and inaccurate predictions. Additionally, when the area of interest occupies a very small region in the image, the model exhibits under-segmentation due to limited foreground pixels. Overall, this analysis shows that the sources of errors mainly stem from low contrast, complex anatomical structure, noisier image, and unbalanced classes or a combination of these factors. To reduce these errors, some of image preprocessing techniques, such as CT image contrast normalization [31], noise robust feature learning through denoising filters [32], and improved handling of small regions of interest using multi-scale preprocessing, can be utilized in the future.

Table 3. Quantitative error analysis of the proposed model using some test samples

Patient ID	Image characteristics	Segmentation error type	DSI (%)	JCI (%)	Description/observed cause
1.2.826.0.1.3680043.10016	Very low contrast slice	Under segmentation	62.04	58.03	Boundary ambiguity
1.2.826.0.1.3680043.10443	Anatomically complex region	Over segmentation	65.62	62.04	Irregular anatomy
1.2.826.0.1.3680043.10515	Noise or artifacts present	Boundary inaccuracy	68.38	65.27	Noise artifacts
1.2.826.0.1.3680043.10579	Small vertebral structure	Under segmentation	60.09	55.95	Small targets (class imbalance)

3.5. Interpretability analysis of EEB7-UNet using Grad-CAM

The proposed model's interpretability analysis is performed using the Grad-CAM method, and the results are shown in Figure 8. To perform this analysis, the last convolutional layer of the encoder block of EEB7-UNet is utilized. Two sample test images chosen from C1 and C4, along with their corresponding predicted class activation maps, are presented in Figures 8(a) and 8(b), respectively, which demonstrates that the proposed model effectively focuses on the fractured regions, highlighted by the green rectangular boxes in the chosen test samples. The heatmaps show that the model focuses on the true anatomical regions associated with fractures.



(a) (b)

Figure 8. Class activation map visualization on sample test images of (a) from C1 and (b) from C4

3.6. Comparison with the other existing methods

The results of the proposed model are compared with those of the closed problems found in the literature, which are given in Table 4. In this table, N and n represent the number of patients and the number of slices. All the existing works listed in Table 4 utilized CT images, but varied in dataset size due to differences in voxel spacing and slice thickness. Most of the previous studies concentrated on segmenting the vertebrae of different levels of spine, on the other hand, our work focused particularly on fractured C-spine vertebra segmentation. A few works were reported on C-spine vertebra segmentation which used the fractured cervical vertebra images for training, validation, and testing. The objective of the present work is to segment the fractured vertebra accurately so that the fracture area and its span can be clearly visible.

Table 4. Performance evaluation of the proposed model against existing approaches in the literature

Work	Method	Target	N Modality (n)	DSI (%)	JCI (%)	Segmentation time/image
Hyun *et al.* [7]	U-Net	C-spine	17 CT (1,684)	93.15 ± 3.09	87.54 ± 5.11	–
Xinxin *et al.* [9]	Hybrid (AI + SiFC)	C-spine	20 CT (60)	95.47	–	48 min
Taeyong *et al.* [33]	CNN	T-spine/L-spine	158 CT (341)	94.00	–	1.18 s
Proposed model	EEB7-UNet	C-spine	724 CT (29,658)	95.53	93.85	0.27 s

Automated segmentation is considered more effective than manual and semi-automated segmentation for spinal fracture diagnosis. Several AI-based models were found to be highly accurate in segmenting vertebrae on CT images. Bae *et al.* [7], an algorithm was proposed for vertebra segmentation, which achieved a mean DSI and JCI of 93.15% and 87.54% respectively, on test data. Liu *et al.* [9] proposed and implemented a semi-automated method based on the combination of AI and sparse intervertebral fence composition (SiFC) for cervical vertebra segmentation, which worked well with a small dataset, but it took approximately 48 minutes to get the complete segmentation result of a single CT image. Recently, a CNN algorithm trained on CT data of lumbar and thoracic spine of 158 patients to perform automated segmentation of fractured vertebral bodies, obtained DSI 94% on test data, and it took approximately 1.18 seconds to segment a vertebra [33]. The fully automated proposed model reported impressive DSI and JCI of 95.53% and 93.85%, respectively. The time taken to segment a single CT scan is 0.27 seconds, which is the fastest in terms of computational time and the best in terms of DSI and JCI compared to all other methods listed in Table 4.

Performance improvements using deep encoder–decoder architectures have also been reported in prior studies, where backbone enhancements and multi scale feature extraction strategies contributed to improved overall segmentation accuracy [34], [35]. These findings align with the results and further support effectiveness of the proposed EEB7-UNet model. Prior studies have also emphasized the importance of accurate vertebral segmentation for clinical assessment [36] and the difficulty of segmenting small vertebral structures [37].

4. CONCLUSION

This work focused on segmentation of fractured vertebrae in the C-spine, intended to assist healthcare professionals in further investigation and treatment planning. In this study, the encoder block of the proposed EEB7-UNet is optimized by utilizing a hybrid learning rate scheduler technique, custom data augmentation, and hyperparameter optimization. The overall modification in the EEB7-UNet significantly improved the model's performance and showed outstanding segmentation results, such as DSI and JCI of 95.53% and 93.85%, respectively. The time required to get the segmentation result of a single CT image from the proposed model is only (0.27) seconds, further demonstrating the efficiency of the proposed model when used in real-time applications. The proposed model may help surgeons by providing a precise delineation of the fractured vertebrae, allowing a better assessment of the location of the fracture, the extent of injury and the involvement of adjacent anatomical regions. This information is valuable during preoperative planning to determine the surgical approach. In addition, the proposed model reduces the time required for manual annotation, supports consistent interpretation across clinicians, and helps in rapid triage in emergencies. However, despite strong quantitative performance, expert clinical validation remains essential before clinical deployment, which is one of the potential limitations of this study. In future work, incorporating 3D CNNs will leverage volumetric information, which may enhance feature representation and enable segmentation of more complex anatomical regions. However, it might require volumetric segmentation for better visualization of the actual pathology. The future work could be to address these limitations by using SOTA methods, such as methods based on 3D U-Net with a suitable encoder. Furthermore, the study does not explicitly segment other anatomical structures of the spine, such as cortical and cancellous bone, and vertebral arch, due to the lack of ground truth data. Moreover, the present model is trained only on CT images, however combining image with different modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) may make the model more generalized and robust. This study emphasize the importance of model size, as a smaller model size enables faster and cost-effective, real-time deployment on the cloud to ensure accessibility for healthcare practitioners.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abhishek Kumar Pandey	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Pateel G. P.	✓		✓	✓	✓				✓		✓		✓	
Kedarnath Senapati		✓	✓			✓	✓			✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection, reference [19].

REFERENCES

[1] C. S. Ahuja *et al.*, "Traumatic spinal cord injury," *Nature Reviews Disease Primers*, vol. 3, no. 1, pp. 1-21, 2017, doi: 10.1038/nrdp.2017.18.

[2] A. Wallace *et al.*, "Percutaneous spinal ablation in a sheep model: protective capacity of an intact cortex, correlation of ablation parameters with ablation zone size, and correlation of post-ablation MRI and pathologic findings," *American Journal of Neuroradiology*, vol. 38, no. 8, pp. 1653-1659, 2017, doi: 10.3174/ajnr.A5228.

[3] J. T. Schousboe, "Epidemiology of vertebral fractures," *Journal of Clinical Densitometry*, vol. 19, no. 1, pp. 8-22, 2016, doi: 10.1016/j.jocd.2015.08.004.

[4] R. Hamid *et al.*, "Epidemiology and pathophysiology of neurogenic bladder after spinal cord injury," *World Journal of Urology*, vol. 36, pp. 1517-1527, 2018, doi: 10.1007/s00345-018-2301-z.

[5] H. T. Nguyen *et al.*, "Prevalence, incidence of and risk factors for vertebral fracture in the community: the Vietnam osteoporosis study," *Scientific Reports*, vol. 14, no. 1, p. 32, 2024, doi: 10.1038/s41598-023-50145-w.

[6] S. Prost, S. Pesenti, S. Fuentes, P. Tropiano, and B. Blondel, "Treatment of osteoporotic vertebral fractures," *Orthopaedics & Traumatology: Surgery & Research*, vol. 107, no. 1, 2021, doi: 10.1016/j.otsr.2020.102779.

[7] H.-J. Bae *et al.*, "Fully automated 3D segmentation and separation of multiple cervical vertebrae in CT images using a 2D convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 184, 2020, doi: 10.1016/j.cmpb.2019.105119.

[8] L. Zhang and H. Wang, "A novel segmentation method for cervical vertebrae based on PointNet++ and converge segmentation," *Computer Methods and Programs in Biomedicine*, vol. 200, 2021, doi: 10.1016/j.cmpb.2020.105798.

[9] X. Liu *et al.*, "Sparse intervertebral fence composition for 3D cervical vertebra segmentation," *Physics in Medicine & Biology*, vol. 63, no. 11, 2018, doi: 10.1088/1361-6560/aac226.

[10] B. Ozdemir, E. Aslan, and I. Pacal, "Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection," *IEEE Access*, vol. 13, pp. 27050-27069, 2025, doi: 10.1109/ACCESS.2025.3539122.

[11] Y. J. Kim, B. Ganbold, and K. G. Kim, "Web-based spine segmentation using deep learning in computed tomography images," *Healthcare Informatics Research*, vol. 26, no. 1, 2020, doi: 10.4258/hir.2020.26.1.61.

[12] M. Vania, D. Mureja, and D. Lee, "Automatic spine segmentation from CT images using convolutional neural network via redundant generation of class labels," *Journal of Computational Design and Engineering*, vol. 6, no. 2, pp. 224-232, 2019, doi: 10.1016/j.jcde.2018.05.002.

[13] L.-W. Cheng *et al.*, "Automated detection of vertebral fractures from x-ray images: a novel machine learning model and survey of the field," *Neurocomputing*, vol. 566, 2024, doi: 10.1016/j.neucom.2023.126946.

[14] M. Tan and Q. Le, "Efficientnet: rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105-6114.

[15] A. K. Pandey, K. Senapati, and G. Pateel, "CAEB7-UNet: an attention-based deep learning framework for automated segmentation of C-spine vertebrae in CT images," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3546324.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[17] T. O'Keefe, H. Ramirez, M. J. Huggins, W. F. Bennett, and B. Blumenkopf, "Computed tomography detection of cervical spinal cord hemangioblastoma: a case report," *Journal of Computed Tomography*, vol. 9, no. 3, pp. 249-252, 1985, doi: 10.1016/0149-936x(85)90070-0.

[18] T. Ulicny, K. Dastur, and G. J. Gray, "CT demonstration of cervical spinal cord compression-cause of reversible itching and burning in the lower extremities," *Journal of Computed Tomography*, vol. 6, no. 1, pp. 57-60, 1982, doi: 10.1016/0149-936x(82)90013-3.

[19] H. M. Lin *et al.*, "The RSNA cervical spine fracture CT dataset," *Radiology: Artificial Intelligence*, vol. 5, no. 5, 2023, doi: 10.1148/ryai.230034.

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1-48, 2019, doi: 10.1186/s40537-019-0197-0.

[21] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.

[23] L. N. Smith, "Cyclical learning rates for training neural networks," in *2017 IEEE Winter Conference on Applications of Computer Vision*, 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.

[24] M. U. Saeed *et al.*, "RMU-Net: a novel residual mobile U-Net model for brain tumor segmentation from MR images," *Electronics*, vol. 10, no. 16, 2021, doi: 10.3390/electronics10161962.

[25] E. Aslan and O. Yildirim, "Detection of road extraction from satellite images with deep learning method," *Cluster Computing*, vol. 28, no. 1, 2025, doi: 10.1007/s10586-024-04880-y.

[26] E. Aslan and Y. Ozupak, "Comparison of machine learning algorithms for automatic prediction of alzheimer disease," *Journal of the Chinese Medical Association*, vol. 88, no. 2, pp. 98-107, 2025, doi: 10.1097/JCMA.0000000000001188.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.
[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.
[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.
[31] Y. Ohno *et al.*, "Dynamic contrast-enhanced CT and MRI for pulmonary nodule assessment," *American Journal of Roentgenology*, vol. 202, no. 3, pp. 515-529, 2014, doi: 10.2214/AJR.13.11888.
[32] A. D. Cheveigné and J. Z. Simon, "Denoising based on spatial filtering," *Journal of Neuroscience Methods*, vol. 171, no. 2, pp. 331-339, 2008, doi: 10.1016/j.jneumeth.2008.03.015.
[33] T. Park *et al.*, "Automated segmentation of the fractured vertebrae on CT and its applicability in a radiomics model to predict fracture malignancy," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-10807-7.
[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision*, 2016, pp. 565-571, doi: 10.1109/3DV.2016.79.
[35] Ö.Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424-432, doi: 10.1007/978-3-319-46723-8_49.
[36] A. Sekuboyina *et al.*, "Verse: a vertebrae labelling and segmentation benchmark for multi-detector CT images," *Medical Image Analysis*, vol. 73, 2021, doi: 10.1016/j.media.2021.102166.
[37] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. M.-Hein, "NNU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203-211, 2021, doi: 10.1038/s41592-020-01008-z.

BIOGRAPHIES OF AUTHORS

Abhishek Kumar Pandey received the B.Tech. degree in Electronics and Instrumentation Engineering from Maulana Abul Kalam Azad University of Technology, Kolkata, and the M.Tech. degree in Engineering Statistics from Cochin University of Science and Technology, Kerala, India. He is currently pursuing a Ph.D. degree with the Department of Mathematical and Computational Sciences at the National Institute of Technology Karnataka, Surathkal, India. He aims to develop data-driven machine learning solutions to assist healthcare professionals and automate decision making, potentially creating a sustainable impact on society or organizations. He can be contacted at email: akumarpanndey686@gmail.com.

Pateel G. P. received the B.E. degree in Electronics and Communication Engineering and the M.Tech. degree in Digital Communication and Networking Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, in 2018. He has completed a Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India. Currently, he is working as an assistant professor in the Department of Computer Science and Engineering at NMAM Institute of Technology, Nitte, Karnataka. He is a lifetime associate member of the Institution of Electronics and Telecommunication Engineers (IETE). His research interests include image processing, machine learning, and deep learning in medical applications. He can be contacted at email: pateel.gp@nitte.edu.in.

Kedarnath Senapati received Ph.D. in Mathematics from the Institute of Mathematics and Applications, Bhubaneswar, in 2013, and a post-doctoral from IIT Bhubaneswar, in 2016. He is currently an associate professor with the Department of Mathematical and Computational Science, National Institute of Technology Karnataka, Surathkal, India. His research interests include the application of deep learning and machine learning in medical image analysis, wavelet transform, signal processing, time series analysis, and spectral analysis. He can be contacted at email: kedar@nitk.edu.in.