scientific reports

OPEN

An efficient deep learning based approach for automated identification of cervical vertebrae fracture as a clinical support aid

Maninder Singh[1,3]✉, **Umang Tripathi**[2,3]✉, **Kunvar Kant Patel**[3], **Kumar Mohit**[1,3] & **Shashwat Pathak**[3,4]

Cervical vertebrae fractures pose a significant risk to a patient's health. The accurate diagnosis and prompt treatment need to be provided for effective treatment. Moreover, the automated analysis of the cervical vertebrae fracture is of utmost important, as deep learning models have been widely used and play significant role in identification and classification. In this paper, we propose a novel hybrid transfer learning approach for the identification and classification of fractures in axial CT scan slices of the cervical spine. We utilize the publicly available RSNA (Radiological Society of North America) dataset of annotated cervical vertebrae fractures for our experiments. The CT scan slices undergo preprocessing and analysis to extract features, employing four distinct pre-trained transfer learning models to detect abnormalities in the cervical vertebrae. The top-performing model, Inception-ResNet-v2, is combined with the upsampling component of U-Net to form a hybrid architecture. The hybrid model demonstrates superior performance over traditional deep learning models, achieving an overall accuracy of 98.44% on 2,984 test CT scan slices, which represents a 3.62% improvement over the 95% accuracy of predictions made by radiologists. This study advances clinical decision support systems, equipping medical professionals with a powerful tool for timely intervention and accurate diagnosis of cervical vertebrae fractures, thereby enhancing patient outcomes and healthcare efficiency.

Keywords Cervical vertebrae, Traumatic injury, Axial CT-Scan, Hybrid CNN, Decision support system

The cervical spine is a versatile structure responsible for safeguarding nerve pathways to the entire body while enabling movement of the head and neck. The cervical vertebral fracture identification is very crucial in case of spine injuries[1–3]. The major causes of cervical vertebral fracture include motor vehicle accidents, falls, penetrating or blunt trauma, sports-related or driving injuries. These injuries can vary in severity, ranging from mild soft tissue injuries to more serious conditions like fractures, dislocations, or damage to the spinal cord[4–7]. The symptoms and consequences of these injuries can also vary, from neck pain and stiffness to more severe neurological issues, depending on the extent of the damage. Over the past two decades, there has been a significant rise in traumatic injuries to the cervical vertebrae globally. This rise is attributed to rapid urbanization and significant challenges including poor road infrastructure, high speeds, and inadequate workplace safety measures[8]. Furthermore, there has been a noticeable increase in cervical vertebrae fractures among the elderly due to osteoporosis and degenerative disorders, making it harder to diagnose these fractures[9–11]. Several conventional clinical support imaging techniques are used to diagnose cervical vertebrae fractures in spinal injuries such as computed tomography scan (CT-Scan), magnetic resonance imaging (MRI), and fan-beam dual-energy X-ray absorptiometry (DEXA). All these clinical imaging techniques have some limitations. They are not been able to detect very small or subtle fractures, especially if they are non-displaced or stress fractures[12]. That's why, doctors face challenges in making optimal treatment decisions. Therefore, it is necessary to improve

[1]Present address: Symbiosis Centre for Medical Image Analysis, Symbiosis International (Deemed University), Pune 412115, India. [2]Present address: Autonomy Technologies (M.Sc.), EEI Department, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) Erlangen, Erlangen 91054, Germany. [3]Electronics & Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India. [4]Atal Incubation Centre, AIC GNITS Foundation, Hyderabad 500104, India. ✉email: maninder.singh@scmia.edu.in; manibiet04@gmail.com; umang.tripathi@fau.de

patient outcomes and healthcare interventions, leveraging emerging advanced techniques based on deep neural network models can assist in accurately identifying cervical vertebrae fractures.

In cervical vertebrae fractures detection, traditional methods include adaptive thresholding, region growing, contour segmentation, and graph-based methods. Recently, the CNN based models were reported to learn the desired features and provide an outcome to diagnose fractures in cervical spine. Salehinejad et al.[13] evaluates the DCNN known as ResNet-50 along with BLSTM layer in detecting the spine fractures on CT axial images. The method determined the validation accuracy of 70.92% and 79.18% on balanced and unbalanced test datasets, respectively. J.E. Small et al.[14] employed an CT scan dataset on the CNN model to determine the cervical spine fracture. The study compared the accuracy for the CNN trained model and radiologist computed value, which found to be 92% and 93% respectively. Arunnit Boonrod et al.[15], uses the different versions of YOLO models to determine the injury in cervical spine (C-spine). The dataset includes the CT scan of lateral neck, where the YOLO v4 variant has outperformed v2 and v3 versions. It determined that the YOLO v4 model has obtained sensitivity, specificity, and accuracy of 80%, 72% and 75% respectively. Merali et al.[16] performed the experiment on the MRI images for detection of cervical spinal cord compression. It trained the MRI image dataset using the deep convolutional neural network and determined the sensitivity and specificity of 89% and 88% respectively. Others reported work includes segmentation of vertebrae using multilayer perceptron and adaptive pulse coupled neural network, with median filtering to refine the outcomes, the implementation of U-Net variants, and various sequential learning models for diagnosing fractures from axial CT images of the cervical spine. Some representative work related to variants of U-Net is highlighted in Table 1.

In this paper, the automated classification of cervical spine fractures on CT scan slices is performed, which is composed of seven vertebrae, labeled C1 to C7, starting from the top of the spine. Each vertebra is separated by intervertebral discs and connected by ligaments, and the cervical spine plays a crucial role in supporting the head's movement and protecting the spinal cord. The proposed methods involve the preprocessing of the images, followed by extracting the features using four pre-trained CNN based models, and integrating best performing model with U-net for final identification and multilabel classification of fractures across the seven vertebrae C1-C7. Unlike previous studies that primarily focus on binary classification or fracture detection at a single vertebral level, our method enables simultaneous multi-label classification across all cervical levels, enhancing diagnostic coverage and clinical relevance. Moreover, we introduce a hybrid architecture by combining Inception-ResNet-v2 as the encoder with a partial U-Net decoder, which allows the model to capture both high-level semantic features and fine-grained spatial details—a design not present in earlier works. To further improve clinical trust and model transparency, Grad-CAM visualizations are incorporated, providing interpretability that is often lacking in related studies. The performance is evaluated based on the parameters such as accuracy, precision, sensitivity, specificity, and F1 score. The primary contribution of the research work includes:

I. An enhanced novel hybrid transfer learning model is proposed for the multilabel classification of cervical vertebrae C1-C7 fractures in the CT scans slices.
II. For an effective multilabel classification, developed a learning model by integrating pre-trained Inception-ResNet-v2 with the U-Net architecture to enhance low-level feature extraction capabilities and reduce redundancy in detecting cervical spine fracture.
III. Experiment was conducted on publicly available dataset of Radiological Society of North America (RSNA). Through Grad-CAM analysis visualized the fracture by extracting features through proposed model and validate the findings for confirming the model's efficiency and effectiveness.

Author	Methodology	Advantages	Limitations	Key Results
Sha et al. (2021)[17]	Introduced a modified U-Net with dilated convolution and attention module to enhance lesion segmentation in spinal CT images.	Improved accuracy in lesion segmentation, increased receptive field, reduced clinical diagnosis time.	Limited focus on specific regions and potential overfitting due to attention mechanisms.	Achieved better segmentation compared to baseline U-Net, reducing manual diagnosis errors.
Shim et al. (2022)[18]	Tested four U-Net variations (U-Net, Attention U-Net, Residual U-Net and Attention Recurrent Residual U-Net) on X-ray images for cervical spine segmentation, particularly for traumatic atlanto-occipital dislocation (TAOD) diagnosis.	Achieved accuracy of 99% with minimal manual intervention and rapid segmentation.	Struggled with false-positive results in more complex bone regions.	Attention U-Net showed highest sensitivity and dice coefficient among models tested.
Sha et al. (2020)[19]	Applied U-Net to spinal fracture lesion segmentation with preprocessing (Gauss, Laplace) and data augmentation techniques for enhancing training.	Simplifies segmentation for real-time clinical use with reasonable accuracy (88%).	False positive and negative rates were notable, needing further refinement.	The model achieved 87.9% accuracy, with potential for clinical diagnostic support.
Bae et al. (2020)[20]	Fully automated 2D U-Net for 3D segmentation of cervical vertebrae in CT images.	Achieved high segmentation accuracy comparable to manual expert work, reduced diagnosis time.	Validation needed across more diverse datasets to enhance generalization.	Dice coefficient of 96.23%, demonstrating high efficiency in 3D cervical vertebrae segmentation.
Paul et al. (2023)[21]	Modified transfer learning with MobileNetV2, Inception v3, and ResNet50V2 for cervical spine fracture detection.	High classification accuracy (99.75%) in detecting fractures using data augmentation techniques.	Limited to CT images, lacking evaluation on other modalities such as MRI, X-ray, and no web-based deployment.	MobileNetV2 achieved highest accuracy, demonstrating robustness for clinical diagnosis of spine fractures.
Xu et al. (2023)[22]	Combined Residual U-Net and Transformer (RUnT) for spinal CT image segmentation.	Improved segmentation boundaries and global/local feature fusion with Transformer, reduced error rates.	High computational complexity and longer inference times due to Transformer structure; limited performance on smaller datasets like VerSe 20.	Achieved state-of-the-art performance with DSC of 88.4% on CTSpine1K, improving boundary precision and robustness compared to prior models.

Table 1. Summary of reported work related to variants of U-Net.

Methodology

This section presents the detailed description of the proposed framework for the identification of cervical vertebrae fractures. The methodology involves preprocessing of the cervical images, training of the images using proposed hybrid transfer learning based on Inception-ResNet-v2 U-Net architecture and performance is evaluated by comparing with different existing pre-trained CNN models, followed by Grad-CAM analysis for validating the detected cervical vertebrae fracture. The detailed discussion of each step is explained as:

Preprocessing of the dataset

The preprocessing of CT scan slices involves several essential steps. The proposed approach for the preprocessing of DICOM (Digital Imaging and Communications in Medicine) images, adopts a comprehensive approach to ensure optimal data fidelity and compatibility for subsequent analysis. Initially, we resize the input image to a standardized 128 × 128 resolution using nearest-neighbor interpolation to preserve pixel integrity. Due to the limited number of samples, we chose to resize images to 128 × 128 to reduce model complexity and mitigate overfitting.

Since the unit of measurement in CT scans is the Hounsfield Unit (HU), which represents radiodensity, it is crucial to note that these scanners are precisely calibrated to ensure accurate measurements. However, the values provided by default may not initially be in Hounsfield Units, requiring further processing to convert them into this standard measurement for analysis. To mitigate inconsistencies in image intensity caused by non-tissue regions, we identify and zero out pixels corresponding to values below a threshold indicative of non-scan areas, typically − 1000 Hounsfield Units (HU). Following this, we apply calibration to convert pixel values to Hounsfield Units (HU) as given in Eq. 1, accounting for inherent variations in image acquisition. This involves scaling the pixel intensities by the slope factor and adding the intercept term obtained from DICOM metadata. In our implementation, the HU values are stored as signed 32-bit integers (int32) to preserve the full range of intensity values including negative numbers.

$$HU\ value\ =\ pixel\ value\ *\ Slope\ +\ Intercept \tag{1}$$

Furthermore, to standardize intensity ranges across images, we normalize pixel values. Finally, we ensure color consistency by converting the grayscale image to a 3-channel RGB representation using the YBR_FULL photometric interpretation. In this step, grayscale pixel values are stored as unsigned 8-bit integers (uint8), which is standard for YBR_FULL format and compatible with most visualization and processing libraries. This conversion from 32-bit Hounsfield values to 8-bit may result in minor precision loss due to dynamic range compression; however, it is a common practice to ensure compatibility with deep learning frameworks and visualization tools. Figure 1 represents the preprocessed images along with their pixel density distribution. It can be observed that after conversion to Hounsfield scale the pixel density distribution is normalized around 0 and − 1000.

Architecture of the proposed approach

The study adopted two distinct modeling methodologies. In the initial approach, image training is conducted on pre-trained models through the application of transfer learning principles. The second approach introduces a novel hybrid model structured upon the encoder-decoder block architecture. The encoder component in this methodology is instantiated with the optimal performing transfer learning model identified in the first approach. For the decoder, we leverage the up-sampling module of the U-Net architecture to perform low-level feature extraction, subsequently integrating fully connected layers to enable image classification. Later, we fine-tuned

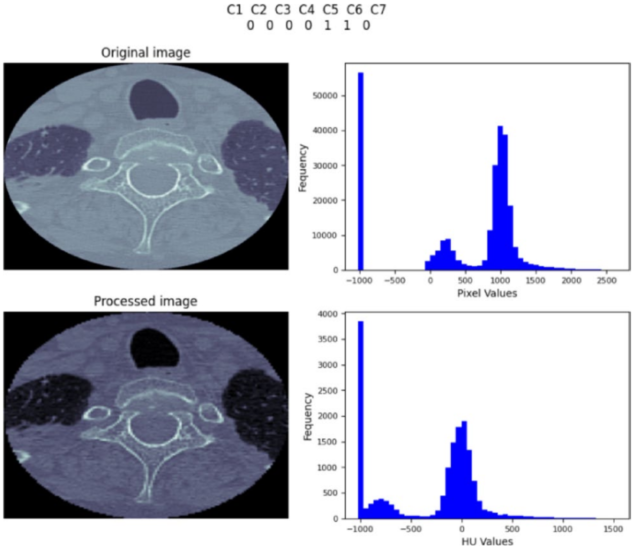


Fig. 1. Original and processed CT scan slice for patient id ‘1.2.826.0.1.3680043.1868’[23].

the final model to further enhance its performance. The three adopted pre-trained deep CNN architectures, namely DenseNet121, Inception v3, and Inception-ResNet-v2, were selected based on their suitability for transfer learning in limited-data settings, making them appropriate and reliable baselines for comparison. These pre-trained CNN architectures as feature extractors for the multilabel image classification task are as follows:

DenseNet-121

DenseNet121[24] offers an improvement over the traditional CNN models by alleviating the vanishing-gradient problem. Layers in a DenseNet121 architecture are linked by dense blocks, which means that each layer builds a feature map that feeds data to all subsequent layers by utilizing inputs from all preceding layers. Our study uses it as a baseline model to evaluate classification performance on cervical spine fracture images.

Inception v3

Inception v3[25], an extension of the renowned GoogLeNet[26], has demonstrated strong classification performance across various biomedical applications. Using transfer learning, we employ it as a baseline for comparison in our multilabel cervical spine fracture classification task.

Inception-ResNet-v2

The Inception structure and the Residual connection are combined to form the basis of Inception-ResNet-v2[27], enabling efficient training of deep networks while mitigating degradation in performance. The study uses it both as a standalone baseline model and as the encoder backbone in our proposed hybrid architecture. Figure 2 illustrates the fundamental network architecture of Inception-ResNet-v2.

Proposed encoder-decoder architecture

In this section, we present our model based on an encoder-decoder architecture. At its core, the encoder-decoder architecture operates on principles of feature extraction and data transformation. This hybrid model combines Inception modules[26], which are well-suited for capturing multi-scale contextual information, with residual connections[29] that enhance gradient flow and training stability in deep networks. While originally adapted from a segmentation framework used in prior work by Aghayari et al.[31], the architecture here is repurposed for classification by modifying the output layers. The design retains a symmetric encoder-decoder structure, similar in spirit to U-Net[27], but is adapted for classification rather than pixel-wise prediction. In the following sections, the U-Net and Inception-ResNet-v2 U-Net architectures are explained.

a) U-Net

U-Net[28] is a fully connected convolution network proposed by Ronneberger et al. in 2015 which mainly finds application in image segmentation tasks. U-Net integrates an encoding path, or contracting path, with a decoding path referred to as the expanding path. The U-shaped appearance of the architecture as represented in Fig. 3 is the source of its name. This design enables the network to capture both local features and global context, facilitating precise segmentation outcomes. The contracting, or encoder, path of the U-Net architecture employs successive convolutions with a 3 × 3 kernel size, same padding, and stride one, followed by Rectified Linear Unit (ReLU) activation, batch normalization, and max pooling operations. This process enhances the depth of feature layers while simultaneously reducing the spatial dimensions of the input image. Notably, the absence of fully connected layers distinguishes this model, allowing for interchangeability with pre-trained architectures. Each

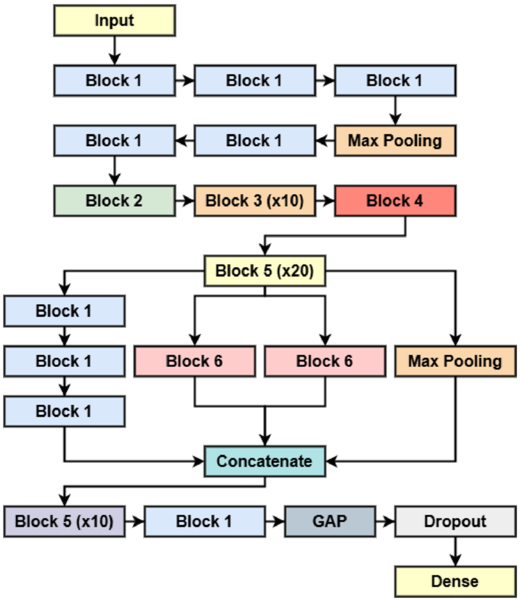


Fig. 2. Inception-ResNet-v2[27] architecture presented in terms of functional blocks.

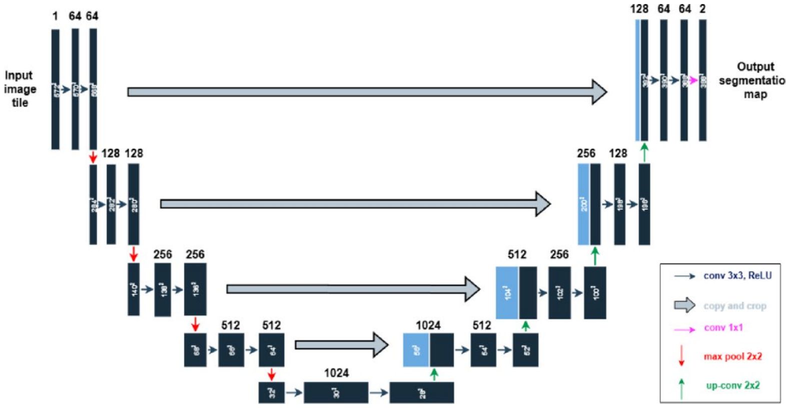


Fig. 3. U-Net architecture[28].

down-sampling step doubles the number of feature maps. Conversely, the expansive, or decoder, path utilizes up-convolution operations to restore the original spatial dimensions of the image while reducing the number of feature maps. Each up-convolutional step results in a halving of the feature count. To mitigate information loss, feature concatenation from the contracting path is integrated during up-sampling. In this phase, standard convolutional layers are applied to the concatenated features to ensure an effective synthesis of both local and global features during the up-sampling process. Figure 3 illustrates the U-Net architecture as presented in the proposed paper. The encoder is depicted on the left, and the decoder on the right. The input dimensions of the image in our model are set at 128 × 128 pixels, with a corresponding output image size of 128 × 128 pixels within the model's framework. In our work, the U-Net architecture is adapted not for segmentation, but to support slice-level multi-label classification. The decoder path is retained to enable reconstruction of semantically rich, spatially aware features before applying global average pooling for classification, leveraging the encoder-decoder together to enhance the model's representational capacity.

b) Inception-ResNet-v2 U-Net architecture

Encoder The encoder is based on the pre-trained Inception-ResNet-v2 network and comprises 37 functional blocks, totaling 164 layers, with its architectural design derived from Aghayari et al.[31]. It extracts progressively abstract representations of input CT slices while reducing spatial dimensions. As shown in Fig. 4, Block 3 is repeated 10 times, yielding a feature map of size 13 × 13 × 320, and Block 5 is repeated 20 times to produce a 6 × 6 × 1088 feature map. These blocks (see in Fig. 5) incorporate parallel convolutional paths and residual connections, improving representational capacity and convergence. The residual connections in Block 3 and Block 5 are implemented using a Lambda layer, which performs residual addition between the block input and the convolution output. To leverage transfer learning, we fine-tuned the model by freezing Blocks 1 through 3, which contain lower-level feature extractors, and allowed the remaining layers to be trainable. This strategy preserves the general feature extraction capabilities of the encoder. The final model contains 36.79 million parameters, of which 32.42 million are trainable, and 4.37 million are non-trainable.

Decoder Although segmentation is not the goal, we retain a decoder path inspired by U-Net[27] to upsample and recover spatial detail necessary for detecting localized fractures. This is because the classification task is performed at the slice level, and fracture cues often appear as small, localized anomalies within the anatomical context of the cervical spine. Simply using the encoder output followed by Global Average Pooling (GAP) could result in a loss of spatial detail crucial for accurate vertebra-specific prediction. The decoder consists of six up-sampling blocks and integrates skip connections from the encoder, enabling the network to combine high-level semantic features with lower-level structural cues. This design mirrors the U-Net structure and facilitates better localization of discriminative features (see in Fig. 4). At the final stage, the output is upsampled to 128 × 128 × 64, matching the input resolution. A Global Average Pooling layer is then applied, followed by Dropout and a Dense output layer with seven sigmoid-activated nodes, each indicating the presence or absence of a fracture in vertebrae C1 to C7. Applying GAP at full spatial resolution helps aggregate fine-grained and spatially distributed features, which is critical in this medical imaging task where co-occurring and vertebra-specific fractures are common. Thus, the decoder is not used for generating pixel-wise segmentations but to enhance the spatial richness of the feature maps prior to classification. This approach is consistent with prior findings that spatial context plays a crucial role in classification tasks involving localized pathologies[30].

Results and discussion

Experimental dataset

In this study, we have utilized the dataset from the RSNA 2022 Cervical Spine Fracture Detection competition (https://www.rsna.org/rsnai/ai-image-challenge/cervical-spine-fractures-ai-detection-challenge-2022). It consists of CT-scans from a total of 2019 distinct patients[23]. However, the segmentation data which contains the annotated labels for the CT scan slices is just available for 87 patients. Consequently, our analysis and model development will be confined to the subset of axial view of CT-scans for which this annotated data is accessible,

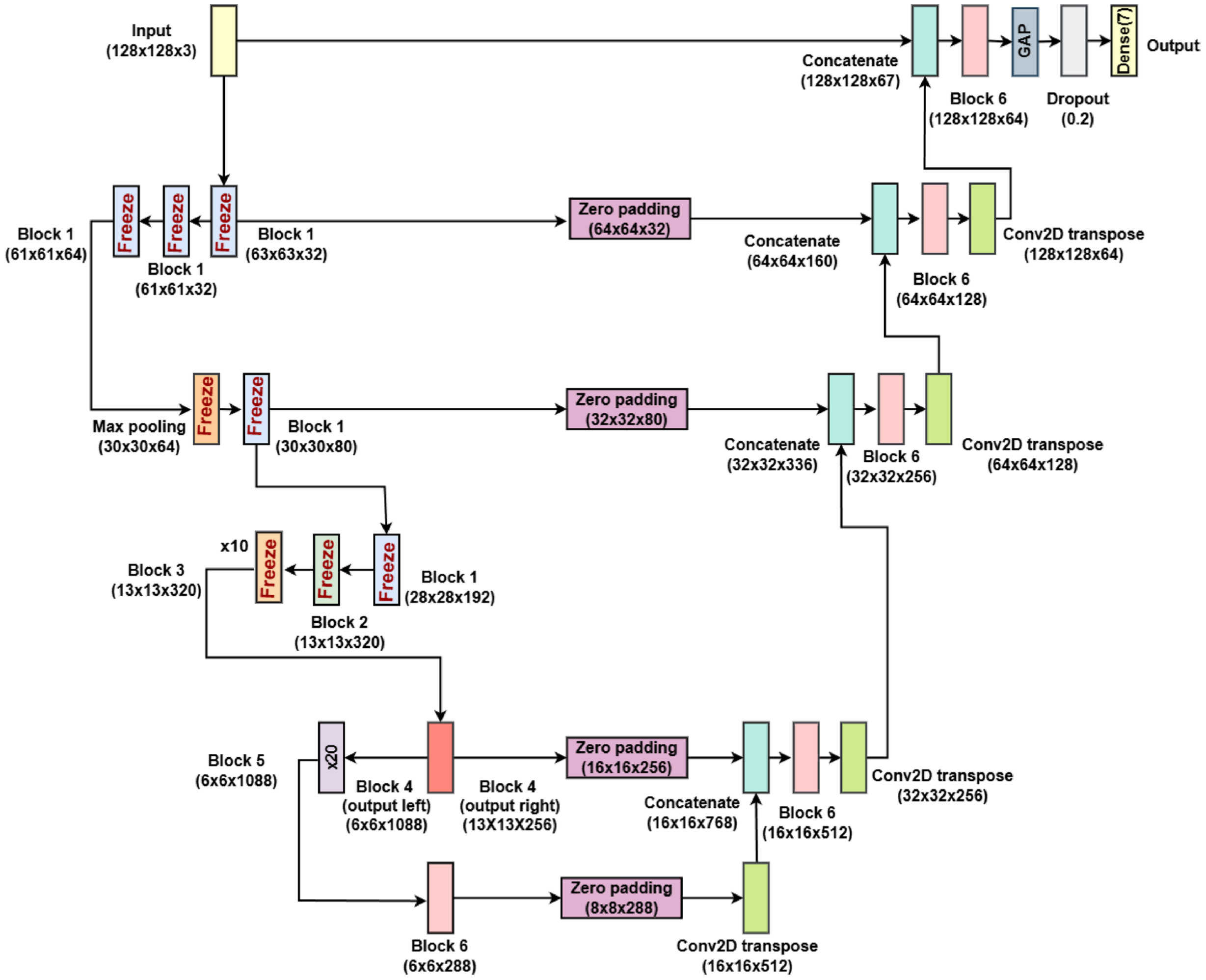


Fig. 4. Proposed Inception-ResNet-v2 U-Net architecture for classification of cervical spine fracture.

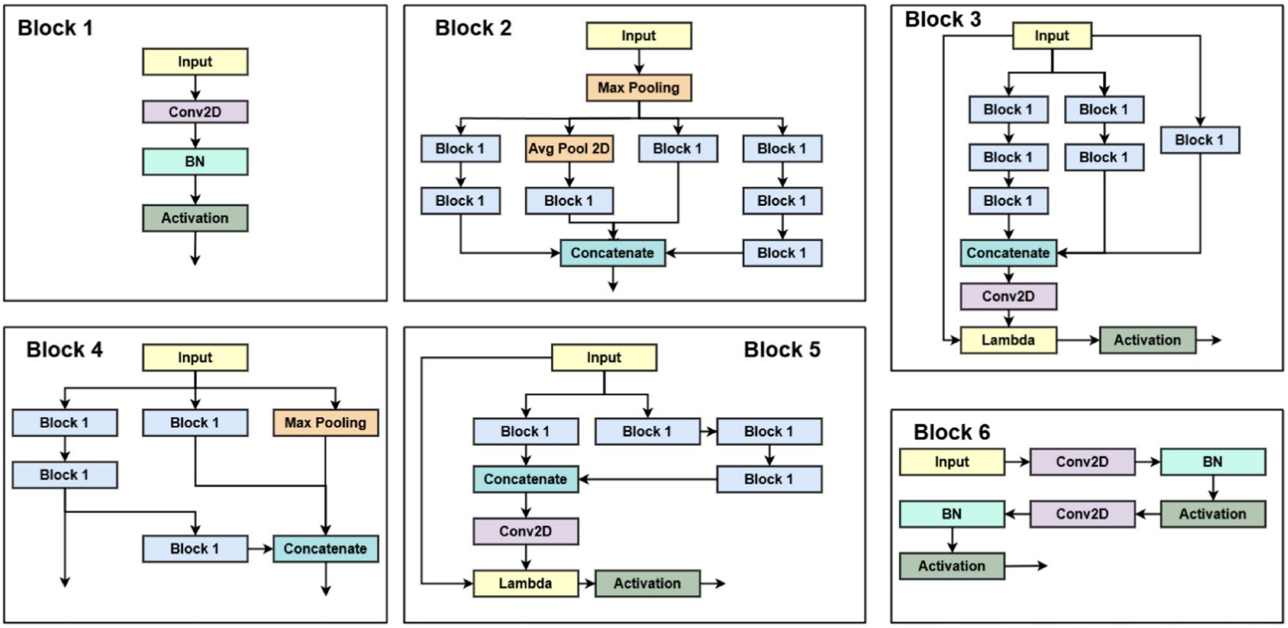


Fig. 5. Details of the blocks used in Inception-ResNet-v2 U-Net architecture in Fig. 4.

focusing on the 29,832 CT-scan slices from 87 patients. The CT-scan are available in the DICOM file format. The DICOM image files have a bone kernel, a slice thickness and axial orientation of less than 1 mm. Figure 6 (a) represents the number of fractures in each vertebra of the cervical spine. It can be observed that the number of fractures in C6 and C7 are more evident as compared from C1 to C5. Several patients have more than one fracture. If multiple fractures occur on a single patient, they tend to occur in vertebrae close together, e.g. C5 and C6 as opposed to C1 and C7 which is evident from the correlation matrix in Fig. 6 (b). Further, the dataset is split into 26,848 training and 2,984 test scan slices, with 10% of the training data further reserved for validation.

Potential dataset biases

Although the RSNA 2022 dataset is widely used and appropriate for cervical spine fracture classification, it has inherent limitations that may introduce bias. The subset of 87 patients utilized in this study does not fully capture the demographic and anatomical diversity of the larger cohort. Additionally, the dataset lacks comprehensive demographic metadata (e.g., age, sex, trauma type), limiting the ability to assess population-level biases. Finally, there is an uneven distribution of fractures across cervical vertebrae, with certain levels being overrepresented, which could affect model training.

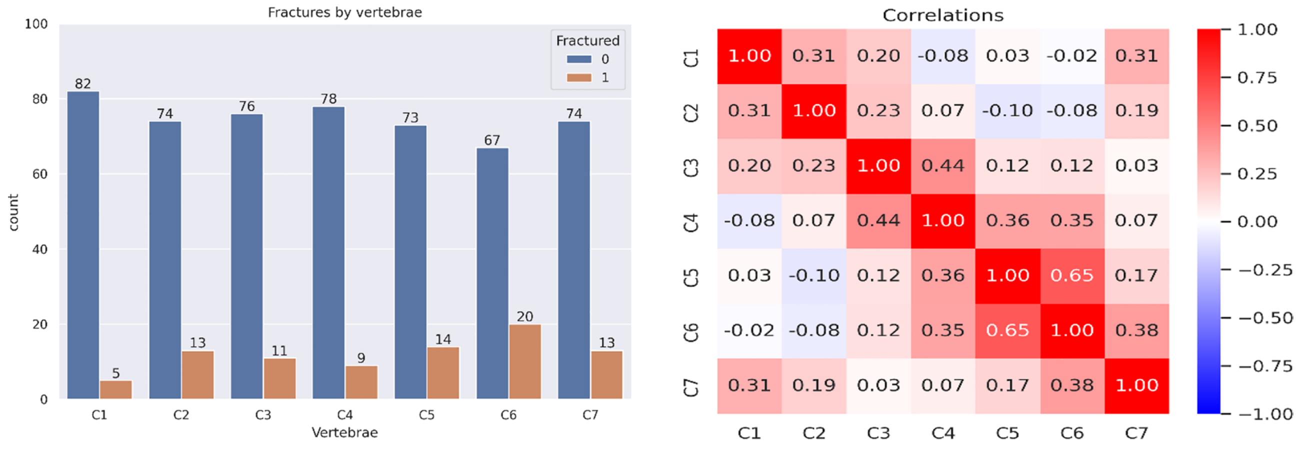
Experimental settings

Experimental tests are conducted using TensorFlow[32], with the dataset randomly divided into 90% for training and 10% for testing. From the training portion, 10% is further set aside as a validation set to monitor. Given the nature of the problem as multilabel classification, binary cross-entropy loss function is employed. Stochastic Gradient Descent (SGD)[33] optimizer with Nesterov momentum[34] is utilized, initialized with an initial learning rate of 0.01, momentum of 0.9, and a weight decay of 1e-6. SGD with Nesterov momentum enhances convergence speed by anticipating future gradients, allowing for more stable and faster learning compared to traditional SGD. Compared to optimizers like Adam, SGD with momentum excels in minimizing overfitting, as Adam can over-adapt to noise, potentially harming performance on unseen medical images. Additionally, the ReduceLROnPlateau callback is implemented, adjusting the learning rate dynamically based on validation accuracy, with reduction factor and minimum learning rate set to 0.5 and 1e-5 respectively. To mitigate overfitting, early stopping callback with a patience of 9 epochs is incorporated, alongside model checkpoint callback to preserve the best-performing model. Training occurs with a batch size of 32, optimizing resource utilization across the 100-epoch training phase. Both baseline and proposed models initialize every weight randomly. Evaluation metrics encompass sensitivity, specificity, and F1-score.

Performance evaluation and discussion

To gauge the overall effectiveness of our proposed model, we conducted a comparative analysis against four benchmark models. Using sensitivity and specificity and F1-score as assessment metrics, Table 2 shows the performance evaluation of the proposed model (both with and without fine-tuning) and other baseline models on the dataset. DenseNet121 exhibits a decline in performance metrics, particularly in later layers, despite initially high Sensitivity and Specificity. Inception v3 presents competitive performance across all metrics, maintaining a balanced Sensitivity and Specificity throughout its layers. Inception-ResNet-v2 displays consistent performance with high Sensitivity and Specificity, especially in early layers, and sustains a strong F-1 Score across most layers.

The Inception-ResNet-v2 U-Net (Proposed) model introduces notable enhancements, surpassing previous architectures by demonstrating improvements in Sensitivity (up to 2.0% increase), Specificity (up to 3.0% increase), and F-1 Score (up to 2.0% increase) across various layers. However, the Inception-ResNet-v2 U-Net (fine-tuned) model exhibits further enhancements, boasting increased Sensitivity (up to 1.0% increase), Specificity (up to 1.5% increase), and F-1 Score (up to 1.0% increase) across diverse layers, notably in deeper layers. There is a balance trade-off between correctly identifying positive and negative instances as the sensitivity

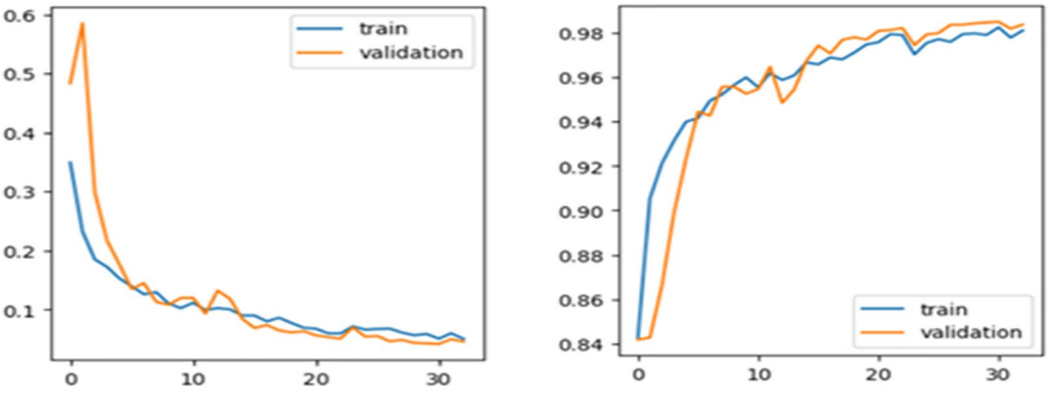


(a) (b)

Fig. 6. (**a**) distribution of fracture by vertebrae, (**b**) correlation matrix for C1-C7.

CNN Model	Parameters	C1	C2	C3	C4	C5	C6	C7
DenseNet121	Sensitivity	0.917	0.895	0.968	0.983	0.915	0.852	0.963
	Specificity	0.968	0.986	0.896	0.780	0.953	0.989	0.946
	F-1 Score	0.942	0.938	0.931	0.807	0.934	0.916	0.954
	Precision	0.968	0.985	0.896	0.684	0.953	0.990	0.945
Inception v3	Sensitivity	0.900	0.903	0.968	0.969	0.912	0.872	0.935
	Specificity	0.930	0.970	0.854	0.776	0.957	0.978	0.903
	F-1 Score	0.914	0.936	0.908	0.861	0.934	0.922	0.920
	Precision	0.928	0.971	0.855	0.774	0.957	0.978	0.905
Inception-ResNet-v2	Sensitivity	0.932	0.935	0.976	0.940	0.905	0.950	0.947
	Specificity	0.944	0.965	0.862	0.855	0.933	0.935	0.913
	F-1 Score	0.938	0.950	0.914	0.896	0.918	0.942	0.930
	Precision	0.944	0.965	0.859	0.855	0.931	0.934	0.913
Inception-ResNet-v2 U-Net (Proposed)	Sensitivity	0.922	0.953	0.950	0.976	0.957	0.878	0.973
	Specificity	0.955	0.973	0.942	0.907	0.9305	0.990	0.954
	F-1 Score	0.940	0.963	0.945	0.940	0.943	0.931	0.963
	Precision	0.958	0.973	0.940	0.906	0.929	0.990	0.953
Inception-ResNet-v2 U-Net (fine-tuned) (Proposed)	Sensitivity	0.940	0.955	0.964	0.966	0.919	0.934	0.973
	Specificity	0.951	0.976	0.930	0.906	0.962	0.988	0.940
	F-1 Score	0.945	0.966	0.947	0.935	0.940	0.960	0.956
	Precision	0.950	0.977	0.930	0.905	0.962	0.987	0.939

Table 2. Comparison of the various CNN based model in classifying the spine fracture vertebrae C1-C7.



(a) Training and Validation Loss Curve (b) Training and Validation Accuracy Curve

Fig. 7. (**a**) displays the training and validation curves of binary loss and, (**b**) accuracy for the fine-tuned Inception-ResNet-v2 U-Net model.

values ranging from 0.919 to 0.973 and specificity values ranging from 0.906 to 0.988. Furthermore, its F1-score, which combines precision and recall, ranges from 0.935 to 0.966, indicating its robustness in capturing both true positives and true negatives while reducing the occurrence of false positives and false negatives. These findings underscore the efficacy of fine-tuning techniques in optimizing model performance for medical image classification tasks. Despite the promising results, this study has several limitations that may affect model performance and generalizability. The relatively small subset of 87 patients used in our analysis limited the diversity of training data, potentially constraining the model's ability to generalize to the broader population. Additionally, due to lack of accessible clinical datasets with appropriate annotations, external validation could not be performed; an important future step to assess robustness.

The dataset also presented challenges such as class imbalance, with certain vertebral levels (e.g., C6, C7) being overrepresented and others (e.g., C1) underrepresented, which may bias the model toward more frequent classes. Variability and noise in the CT scans, including anatomical differences and artifacts, posed further challenges for consistent feature extraction. To address these issues in future work, we plan to incorporate advanced data augmentation techniques, explore loss functions that account for class imbalance, and validate the model on external datasets to enhance generalizability. Figure 7 displays the curves provide a visual representation of the model's performance over time, highlighting both accuracy and loss metrics throughout the process. The curves shows that the model performs better as the losses are minimized and accuracy is maximized for training and validation. Table 2 indicates the accuracies of the different CNN architectures utilized in the study. Among the different CNN models, the Encoder-Decoder (Inception-ResNet-v2 U-Net fine-tuned) model determined the highest accuracy of 98.44%. Further, the Table 3 highlights related studies comparison with the proposed model. The proposed architecture also performs better in comparison to various other existing studies.

Ablation studies

The study aims to demonstrate the effectiveness of the proposed hybrid model. The analysis was performed to evaluate the contribution of various components to the overall performance of the deep learning models. The study systematically removes or modify specific parts of a model to determine their impact on key metrics such as accuracy, sensitivity, and specificity. In the Sha et al.[17] model, the attention mechanism and dilated convolution layers were individually ablated to assess their importance in lesion segmentation. This experiment demonstrated that removing the attention module led to a decrease in the model's ability to focus on crucial lesion regions, while eliminating the dilated convolution reduced the receptive field, affecting boundary detection performance. These results confirm the critical role of both components in achieving high segmentation accuracy. Shim et al.[18] model, which incorporated attention mechanisms within a U-Net architecture, revealed that excluding the attention layer reduced sensitivity and precision, particularly in complex bone regions, underscoring the importance of attention in cervical spine segmentation. For the Paul et al.[21] transfer learning-based model, removing specific components (MobileNetV2, Inception v3, ResNet50V2) demonstrated that MobileNetV2 contributed most to achieving high accuracy and speed in detecting cervical spine fractures. Furthermore, omitting data augmentation techniques during the study resulted in decreased robustness and increased overfitting, emphasizing the role of augmentation in improving model generalization. In the case of Xu et al.[22], which combined a residual U-Net with Transformer networks, ablation of the Transformer module reduced the model's ability to capture global contextual information, while removing the residual connections led to gradient instability, negatively impacting segmentation consistency. Finally, in the Inception-ResNet-v2 U-Net model used in this study, we observed that removing the Inception-ResNet architecture decreased sensitivity and specificity, demonstrating the importance of deeper architectures in accurate fracture detection. Table 4 highlighting test accuracies observed for different CNN model, where Inception-ResNet-v2 U-Net (fine-tuned) determined the accuracy of 98.44.

Grad-CAM analysis for classification of cervical vertebrae

The visualization of cervical vertebrae fractures is achieved using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique, which highlights the regions within a CT scan slice that most influence the model's classification decisions. This approach enables a qualitative assessment of whether the model focuses on anatomically relevant areas associated with fractures across vertebrae C1 to C7. Figure 8 presents Grad-CAM outputs from the final batch_normalization_209 layer of the fine-tuned Inception-ResNet-v2 U-Net model, applied to test images. In the Fig. 8, 'A' denotes the actual class label and 'P' represents the predicted probability. The highlighted regions provide insight into the areas the model attends to when detecting

Author	No of Sample	Accuracy	Sensitivity	Specificity	Comments
Sha et al[17].	7836	98.6%	83.8%	97.9%	U-Net with dilated convolution and attention for spinal lesion segmentation in CT, improving diagnostic speed.
Shim et al[18].	707	99.13%	90.44%	99.51%	Compared U-Net variations, Attention U-Net performed best for cervical spine segmentation in X-rays.
Sha et al[19].	N/A	87.9%	N/A	N/A	U-Net with preprocessing and data augmentation for spinal fracture segmentation in clinical use.
Bae et al[20].	1684 disease slices, 3490 healthy slices	96.23%	N/A	N/A	Fully automated 2D U-Net for 3D cervical vertebra segmentation, achieved high DSC scores in validation.
Paul et al[21].	4200	99.75%	N/A	N/A	MobileNetV2 model achieved highest accuracy for cervical spine fracture classification in CT images.
Xu et al[22].	CTSpine1K, VerSe 20	88.4% (CTSpine1K)	N/A	N/A	Residual U-Net combined with Transformer for vertebral edge feature fusion, state-of-the-art segmentation.
A. Dastgir et al[35].	10,000 X-ray images (VinDr-SpineXR)	96.81%	97.95%	N/A	MAFMv3: MobileNetV3 + CBAM + ASPP + histogram-equalized views for spine lesion classification
M.U. Saeed et al[36].	80 CT test samples (VerSe19/20)	N/A	96.52% (VerSe20) 94.64% (VerSe19)	N/A	3D MFA (Multi-Feature Attention): MobileNetV3 + Reverse CBAM + FPP + ASPP; lightweight and pruned architecture.
Salehinejad et al[13].	3666	79.18%	57 — 64%	77 — 84%	Uses a bidirectional long-short term memory (BLSTM) layer on CT imaging for the cervical spine fracture detection
J.E. Small et al[14].	665	92% (CNN) 95% (radiologists)	76% (CNN) 93% (radiologists)	97% (CNN) 96% (radiologists)	CNN was designed to detect cervical spine fractures on CT and compared it to that of radiologists.
Arunnit Boonrod et al[15].	625	75%	80%	72%	Trained different variants of YOLO network (v2, v3, and v4). Among these, YOLO v4 used to detect cervical spine injury.
Presented Work	29,832 CT-scan slices	98.44%	92 — 97%	90 — 98%	Inception-ResNet-v2 U-Net (fine-tuned) used to detect the spine fracture.

Table 3. Comparison of the proposed work with the studies. N/A — Not Available

CNN Model	Test accuracy (%)
DenseNet-121	97.73
Inception v3	97.25
Inception-ResNet-v2	97.83
Inception-ResNet-v2 U-Net	**98.31**
Inception-ResNet-v2 U-Net (fine-tuned)	**98.44**

Table 4. Test accuracies for different CNN architectures used on the RSNA dataset.

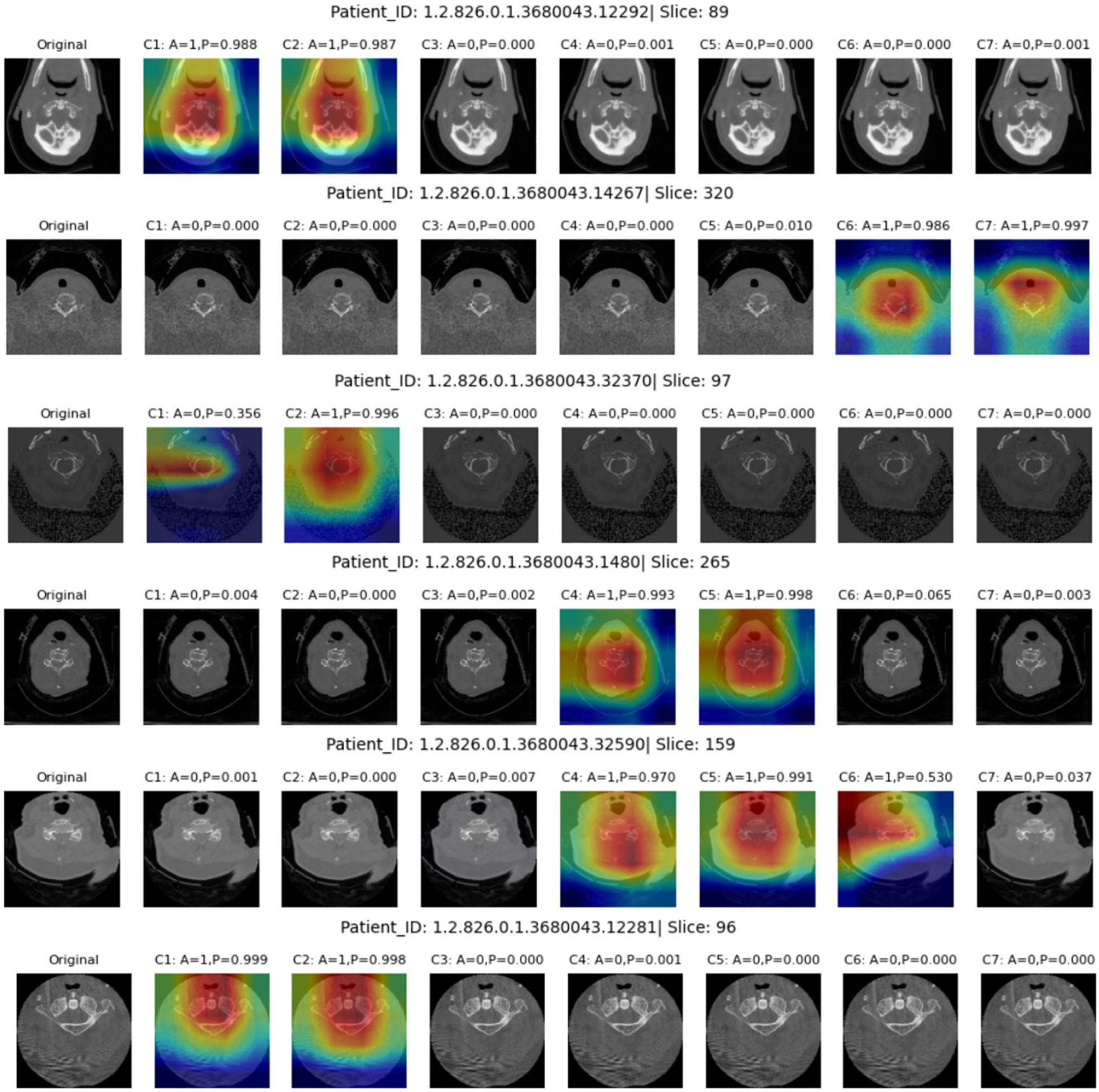


Fig. 8. Grad-CAM analysis for final `*batch_normalization_209*` layer in Inception-ResNet-v2 U-Net fine-tuned model on test images. `*A*` represents the actual label for the class and `*P*` represents predicted probability for the class.

fractures, supporting interpretability and clinical relevance. Although the dataset does not explicitly annotate the smallest or most subtle fractures, the high overall accuracy and consistent attention to relevant anatomical areas suggest the model's potential effectiveness in detecting subtle abnormalities. While the visualizations are informative, it is noted that some activation maps exhibit broad and diffuse patterns. This may be attributed to the U-Net-like architecture, where the final feature map used for classification retains the full input resolution. Such spatially extended attention allows the model to incorporate contextual anatomical cues but can reduce localization sharpness. Clinically, these broader activation regions may still be useful, as radiologists often examine surrounding tissue when assessing fractures. However, we acknowledge this as a limitation in precise interpretability and suggest that future work explore attention mechanisms or architectural modifications to improve focus localization without compromising classification performance.

Potential clinical applications

The proposed model demonstrates strong performance in classifying cervical vertebrae fractures from CT scan slices, indicating its potential utility in clinical practice. By providing rapid and automated assessments, the model could serve as a valuable clinical decision support tool, assisting radiologists in detecting fractures more efficiently and accurately. In emergency departments or trauma centers, where timely diagnosis is critical, the integration of this model could reduce diagnostic delays and workload burden. Moreover, embedding the model within existing hospital imaging platforms or radiology information systems could streamline workflow, allowing for seamless evaluation alongside routine radiological assessments. The proposed method may also be particularly beneficial in resource-limited or remote settings where access to expert radiologists is scarce. Through telemedicine applications, the model could offer preliminary evaluations, enabling earlier intervention and triage.

Furthermore, the interpretability provided by methods like Grad-CAM enhances clinician trust by highlighting relevant anatomical regions influencing model predictions. This transparency is crucial for clinical adoption, ensuring that artificial intelligence outputs complement rather than replace expert judgment. Overall, the integration of automated fracture classification models has the potential to improve diagnostic accuracy, optimize patient management, and contribute to better clinical outcomes. Future work will focus on validating this model across diverse clinical populations and integrating it into real-world healthcare workflows.

Conclusion

The proposed work presents an efficient method for automated classification of cervical vertebrae fracture C1-C7 in CT scan slices. The proposed work performs a preprocessing of the data and, using the Inception-ResNet-v2 U-Net architecture, detects the abnormality in cervical vertebrae fracture. Our experimental results have demonstrated the superior performance of our proposed model compared to baseline models. Achieving an accuracy of 98.44%, our model has exhibited high precision in identifying cervical vertebrae fractures. This exceptional accuracy underscores the efficacy and reliability of our approach, suggesting its potential to improve the diagnostic process for cervical spine injuries significantly. Integrating our automated classification method into a decision support system will hold immense promise for clinical practice. By providing timely and accurate information regarding the presence of cervical vertebrae fractures, our system can assist healthcare professionals in making informed decisions regarding patient care. It includes facilitating early diagnosis, guiding treatment planning, and improving patient outcomes.

Data availability

The data used in this study are publicly available from the RSNA 2022 Cervical Spine Fracture Detection competition, accessible at https://www.rsna.org/rsnai/ai-image-challenge/cervical-spine-fractures-ai-detection-challenge-2022.

Received: 1 April 2025; Accepted: 3 July 2025
Published online: 15 July 2025

References

1. Parizel, P. M. et al. Trauma of the spine and spinal cord: imaging strategies. *Eur. Spine J.* **19**, 8–17. https://doi.org/10.1007/s00586-009-1123-5 (Sep. 2009).
2. Okereke, I., Mmerem, K. & Balasubramanian, D. The management of cervical spine Injuries – A literature review. *Orthop. Res. Reviews.* **13**, 151–162. https://doi.org/10.2147/orr.s324622 (Sep. 2021).
3. Simon, L. V., Lopez, R. A. & King, K. C. (2017). Blunt force trauma.
4. Eli, I., Lerner, D. P. & Ghogawala, Z. Acute traumatic spinal cord injury. *Neurol. Clin.* **39** (2), 471–488. https://doi.org/10.1016/j.ncl.2021.02.004 (2021). Epub 2021 Mar 31. PMID: 33896529.
5. Bedbrook, G. M. *The Care and Management of Spinal Cord Injuries* (Springer Science & Business Media, 2013).
6. Myers, B. S. & Winkelstein, B. A. Epidemiology, Classification, Mechanism, and Tolerance of Human Cervical Spine Injuries, vol. 23, no. 5–6, pp. 307–409, Jan. (1995). https://doi.org/10.1615/critrevbiomedeng.v23.i5-6.10
7. Imhof, H. & Fuchsjäger, M. Traumatic injuries: imaging of spinal injuries. *Eur. Radiol.* **12** (6), 1262–1272. https://doi.org/10.1007/s00330-002-1448-5 (Apr. 2002).
8. Gururaj, G. *Injuries in India: A National Perspective. Background Papers: Burden of Disease in India Equitable Development-Healthy Future*325–347 (National Commission on Macroeconomics and Health, Ministry of Health & Family Welfare, Government of India, 2005).
9. Ferrar, L., Jiang, G., Adams, J. & Eastell, R. Identification of vertebral fractures: an update. *Osteoporos. Int.*, **16**, 7, pp. 717–728, May 2005, doi: https://doi.org/10.1007/s00198-005-1880-x
10. Sunder, A., Chhabra, H. S. & Aryal, A. Geriatric spine fractures – Demography, changing trends, challenges and special considerations: A narrative review. *J. Clin. Orthop. Trauma.* **43**, 102190. https://doi.org/10.1016/j.jcot.2023.102190 (Aug. 2023).

11. Zeytinoglu, M., Jain, R. K. & Vokes, T. J. Vertebral fracture assessment: enhancing the diagnosis, prevention, and treatment of osteoporosis. *Bone* **104**, 54–65. https://doi.org/10.1016/j.bone.2017.03.004 (Nov. 2017).
12. Sutherland, M., Bourne, M., McKenney, M. & Elkbuli, A. Utilization of computerized tomography and magnetic resonance imaging for diagnosis of traumatic C-Spine injuries at a level 1 trauma center: A retrospective cohort analysis. *Annals Med. Surg.* **68**, 102566. https://doi.org/10.1016/j.amsu.2021.102566 (Aug. 2021).
13. Salehinejad, H. et al. Deep Sequential Learning For Cervical Spine Fracture Detection On Computed Tomography Imaging, IEEE Xplore, Apr. 01, (2021). https://ieeexplore.ieee.org/document/9434126 (accessed Mar. 31, 2023).
14. Small, J. E., Osler, P., Paul, A. B. & Kunst, M. CT cervical spine fracture detection using a convolutional neural network. *Am. J. Neuroradiol.* **42** (7), 1341–1347. https://doi.org/10.3174/ajnr.a7094 (Apr. 2021).
15. Boonrod, A., Boonrod, A., Meethawolgul, A. & Twinprai, P. Diagnostic accuracy of deep learning for evaluation of C-spine injury from lateral neck radiographs. *Heliyon* **8**, e. https://doi.org/10.1016/j.heliyon.2022.e10372 (Aug. 2022). no. 8.
16. Merali, Z. et al. A deep learning model for detection of cervical spinal cord compression in MRI scans. *Sci. Rep.* **11** https://doi.org/10.1038/s41598-021-89848-3 (May 2021).
17. Sha, G., Wu, J. & Yu, B. A robust segmentation method based on improved U-Net. *Neural Process. Lett.* **53** (4), 2947–2965. https://doi.org/10.1007/s11063-021-10531-9 (May 2021).
18. Shim, J. H. et al. Evaluation of U-Net models in automated cervical spine and cranial bone segmentation using X-ray images for traumatic atlanto-occipital dislocation diagnosis. *Sci. Rep.* **12** (1). https://doi.org/10.1038/s41598-022-23863-w (Dec. 2022).
19. Sha, G., Wu, J. & Yu, B. Spinal fracture lesions segmentation based on U-net, 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Jun. (2020). https://doi.org/10.1109/icaica50127.2020.9182574
20. Bae, H. J. et al. Namkug kim, 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, **184**,2020,105119, https://doi.org/10.1016/j.cmpb.2019.105119
21. Showmick, G., Paul, A., Saha, M. & Assaduzzaman A real-time deep learning approach for classifying cervical spine fractures. *Healthc. Analytics.* **4**, 100265–100265. https://doi.org/10.1016/j.health.2023.100265 (Dec. 2023).
22. Xu, H. et al. RUnT: A network combining residual U-Net and transformer for vertebral edge feature fusion constrained spine CT image segmentation, in IEEE access, **11**, pp. 55692–55705, (2023). https://doi.org/10.1109/ACCESS.2023.3281468
23. Lin, H. et al. The RSNA cervical spine fracture CT dataset. *Radiology: Artif. Intell.* **5** (5), e230034 (2023).
24. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. & Recognition, P. IEEE Conference on Computer Vision and Densely Connected Convolutional Networks, (CVPR), Honolulu, HI, USA, 2017, pp. 2261–2269, (2017). https://doi.org/10.1109/CVPR.2017.243
25. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826, (2016). https://doi.org/10.1109/CVPR.2016.308
26. Szegedy, C. et al. IEEE Conference on Computer Vision and Going deeper with convolutions, (CVPR), Boston, MA, USA, 2015, pp. 1–9, (2015). https://doi.org/10.1109/CVPR.2015.7298594
27. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, Feb. (2017). https://doi.org/10.1609/aaai.v31i1.11231
28. Ronneberger, O., Fischer, P. & Brox, T. UNet: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241. (2015).
29. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778, (2016). https://doi.org/10.1109/CVPR.2016.90
30. Diakogiannis, F. I., Waldner, F., Caccetta, P. & Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, ISPRS Journal of Photogrammetry and Remote Sensing, vol. 162, pp. 94–114, Apr. (2020). https://doi.org/10.1016/j.isprsjprs.2020.01.013
31. Aghayari, S. et al. Jan., ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. X-4/W12022, pp. 9–17, (2023). https://doi.org/10.5194/isprs-annals-x-4-w1-2022-9-2023
32. Anwarul, S. & Joshi, D. Deep Learning with TensorFlow, Advances in Computer and Electrical Engineering, pp. 96–120, (2020). https://doi.org/10.4018/978-1-7998-3095-5.ch004
33. Herbert Robbins and Sutton Monro. A stochastic approximation method. The annals of mathematical statistics, pages400–407, (1951).
34. Ghadimi, E., Feyzmahdavian, H. R. & Johansson, M. Global convergence of the Heavy-ball method for convex optimization, IEEE Xplore, Jul. 01, (2015). https://ieeexplore.ieee.org/abstract/document/7330562 (accessed Sep. 12, 2022).
35. Dastgir, A., Bin, W., Saeed, M. U., Sheng, J. & Saleem, S. MAFMv3: an automated Multi-Scale Attention-Based feature fusion MobileNetv3 for spine lesion classification. *Image Vis. Comput.* **155**(C), 105440. https://doi.org/10.1016/j.imavis.2025.105440 (2025).
36. Saeed, M., Usman, W., Bin, J., Sheng, S. & Saleem 3D MFA: an automated 3D Multi-Feature attention based approach for spine segmentation using a multi-stage network pruning. *Comput. Biol. Med.* **185**, 109526 (2025).

Author contributions

MS and UT were involved in the conception of the study, methodology, data analysis, writing the original draft and revision of the manuscript preparation. MS, KKP, KM, and SP were involved in the supervision and revision of manuscript preparation. Finally, all the authors reviewed and approved the manuscript.

Funding

Open access funding provided by Symbiosis International (Deemed University).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.S. or U.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025