Cervical Spine Fracture Detection

Abdulrahman Emad
Systems and Biomedical Engineering
Cairo University
Cairo, Egypt
abdulrahman.masoud02@eng-st.cu.edu.eg

Mourad Magdy
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
murad.ibrahim02@eng-st.cu.edu.eg

Youssef Ashraf Mohammed
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
youssef.aziz02@eng-st.cu.edu.eg

Mariam Ahmed Said
Systems & Biomedical Engineering
Cairo University
Cairo, Egypt
maryam.abdulmajeed01@eng-st.cu.edu.eg

***Abstract*—Cervical spine fractures, accounting for a significant proportion of over 1.5 million spinal fractures annually in the United States, pose a substantial challenge in medical imaging diagnostics. The rapid detection and localization of these fractures are critical for preventing neurologic deterioration and paralysis, particularly in elderly populations where degenerative disease and osteoporosis often obscure fracture visibility in computed tomography (CT) scans.**

This paper explores a machine learning-based approach to cervical spine fracture detection, leveraging advanced segmentation and classification techniques. We utilized a dataset curated by the Radiological Society of North America (RSNA), comprising approximately 3,000 CT studies annotated by expert spine radiologists. Our two-stage solution integrates segmentation with a 2.5D convolutional neural network (CNN) and U-Net architecture in the first stage to isolate cervical vertebrae, followed by a CNN combined with bi-directional gated recurrent units (BiGRU) and attention mechanisms in the second stage to classify fractures at vertebral levels.

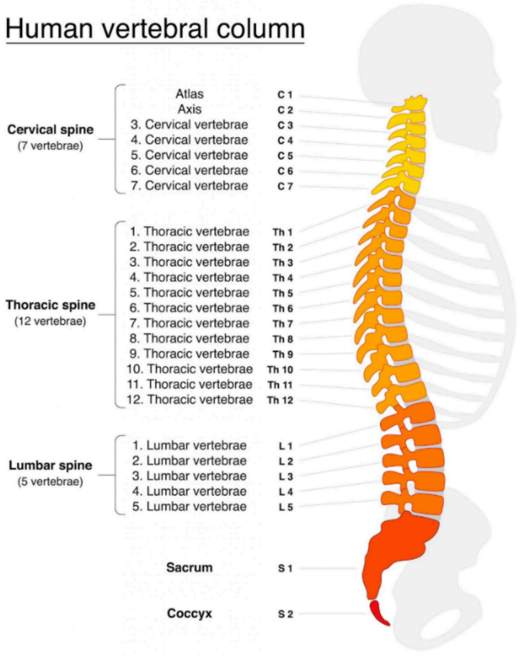
The segmentation stage employed an efficientnet-b0 backbone and data augmentation strategies to enhance generalization. In the classification stage, cropped vertebral segments were processed using efficientnet and resnest50d backbones alongside BiGRU and attention layers to capture spatial and sequential features.

***Keywords*—Cervical spine fractures, Healthcare, 2.5D convolutional neural network (CNN), Machine Learning, Medical imaging, Segmentation, Attention mechanisms**

I. Introduction

The human spine is a long column of bones extending from the neck to the lower back, consisting of vertebrae, facet joints, intervertebral disks, the spinal cord, nerves, and soft tissues. Structurally, the spine is segmented into five regions: the cervical spine, thoracic spine, lumbar spine, sacrum, and coccyx (tailbone). Among these, the cervical spine, located in the neck region, plays a vital role in supporting the head, facilitating movement, and protecting the spinal cord.

Since delayed diagnosis can result in serious sequelae, such as neurological degeneration and paralysis after trauma, radiologists must be vigilant in detecting cervical spine fractures. High-energy trauma, including car crashes or falls, frequently causes cervical spine fractures, which need to be treated right away. To reduce the chance of irreversible injury, these fractures must be detected accurately and promptly. The best diagnostic method for evaluating spinal injuries is computed tomography (CT) imaging because of its high resolution and capacity to view bony structures. However, interpreting CT scans takes a lot of time, requires radiological knowledge, and is subject to observer variation. In order to overcome these obstacles, medical imaging diagnostic workflows can be improved and automated with artificial intelligence (AI). The goal of the Kaggle-hosted RSNA Cervical Spine Fracture Detection competition is to accelerate the creation of AI models that can accurately and efficiently identify cervical spine fractures. This contest provides a platform to investigate the possibilities of deep learning in addressing actual radiological problems by offering a carefully selected dataset of 3D CT pictures and annotations.



II. Related Work

A. Introduction

Deep learning has proven to be a transformative technology for cervical spine fracture detection, leveraging advanced architectures to enhance diagnostic accuracy and efficiency. In this section we review key methodologies with their strengths and weaknesses.

B. CNN-Based Feature Extraction for Fracture Detection

Convolutional Neural Networks (CNNs) have been widely adopted for analyzing cervical spine X-rays and CT images. A notable study, *Deep Learning for Cervical Spine Fracture Detection: A Comparison of Architectures*, utilized ResNet and DenseNet, employing transfer learning to exploit the advantages of pre-trained models. These architectures efficiently extracted features, even with smaller datasets, and demonstrated high sensitivity, reducing the likelihood of missed fractures—an essential factor in clinical applications. However, these methods suffered from high false-positive rates, which could lead to over-diagnosis and unnecessary follow-ups. Additionally, the lack of explainability in predictions posed challenges for clinical trust, while the limited demographic diversity of the training data restricted the generalizability of the models.

C. Attention Mechanisms for Fracture Localization

The incorporation of attention mechanisms has significantly enhanced the localization of cervical spine fractures. In the study *Attention-Augmented Convolutional Networks for Cervical Fracture Detection*, attention layers focused on critical regions within medical images, improving fracture detection accuracy and reducing false positives by isolating areas of interest. This advancement is particularly beneficial for assisting radiologists in identifying fractures more efficiently. However, the addition of attention layers increased the computational complexity, making these models resource-intensive. Furthermore, their performance declined when applied to low-quality images or subtle fractures, and they lacked robustness across datasets with varying imaging protocols.

D. 3D CNNs for Volumetric CT Scans

Three-dimensional CNNs have been employed to process volumetric CT scans, capturing spatial relationships across multiple slices to improve diagnostic accuracy. The study *Automated Cervical Spine Fracture Detection Using 3D Convolutional Neural Networks* demonstrated the effectiveness of this approach in identifying complex fracture patterns. These models performed exceptionally well on large, annotated datasets and leveraged the spatial context to achieve superior diagnostic outcomes. However, the high computational and memory demands of 3D CNNs limited their accessibility in resource-constrained settings. Additionally, these models struggled with smaller datasets, which increased the risk of overfitting, and their time-intensive processing reduced their utility in clinical scenarios requiring rapid diagnosis.

III. Dataset and Features

A. Dataset Description

The dataset used in this study was created through a collaboration between the Radiological Society of North America (RSNA), the American Society of Neuroradiology (ASNR), and the American Society of Spine Radiology (ASSR). It includes CT scans of the cervical spine collected from 12 institutions across nine countries and six continents, making it the largest multi-institutional, expert-labeled dataset for CT spine fracture detection.

The imaging dataset is organized in folders, with each folder corresponding to a unique scan. The structure follows the format:
[train/test]_images/[StudyInstanceUID]/[slice_number].dcm

Each DICOM file represents a single axial slice of a CT scan. The scans were acquired using a bone kernel with a slice thickness of ≤1 mm. To ensure accessibility, segmentation files are provided in NIfTI format for a subset of the data, with annotations verified and refined by radiologists. Some DICOM files are JPEG-compressed, requiring specialized libraries such as GDCM or pylibjpeg for decoding and processing.

The segmentation annotations focus primarily on the cervical vertebrae (C1-C7), with additional thoracic vertebrae (T1-T12) labels included in a subset of scans. Bounding box annotations further identify regions of interest within the training set, aiding in targeted fracture detection.

TABLE I. DATASET FEATURES DESCRIPTION

No	Feature	Description
1	CT Images	3,112 DICOM files of cervical spine scans. Positive Fractures: 1,445 cases. Negative Fractures: 1,667 cases.
2	Segmentation Labels	Pixel-level annotations in NIfTI format. Label values: ● 1–7: Cervical vertebrae (C1 to C7). ● 8–19: Thoracic vertebrae (T1 to T12). ● 0: Background and other structures.
3	Metadata	Train.csv: Study IDs and fracture annotations for the training set. ● patient_overall: Indicates if any vertebrae in the scan are fractured. ● C[1-7]: Vertebra-specific fracture status. Test.csv: Study IDs and prediction structure for the test set. ● prediction_type: Specifies the target column for predictions.
4	Supplementary Data	● Bounding box annotations for a subset of training scans. ● Sample submission file for structuring predictions.

B. Data Preprocessing

Imaging Data

Extraction and Normalization:

DICOM files were processed to extract pixel arrays using Python libraries such as pydicom and pylibjpeg. Intensity values were normalized to a consistent scale for input into machine learning models.

Handling Compression:

JPEG-compressed files were decoded using GDCM to ensure accurate pixel array extraction.

Segmentation Mask Processing:

NIfTI-format segmentation masks were preprocessed to focus on cervical vertebrae (C1-C7). Thoracic labels (T1-T12) were excluded when inconsistent with the dataset's primary focus.

Sample Selection Due to Memory Limitations:

Due to memory constraints, a subset of 150,000 samples was selected from the dataset for model training and evaluation. This subset was chosen to ensure representation of both fracture-positive and fracture-negative cases while balancing computational feasibility.

Metadata

Target labels (patient_overall and C[1-7]) were encoded into binary classifications for machine learning tasks.

Balanced training and validation splits were created to maintain equal representation of fracture-positive and fracture-negative cases.

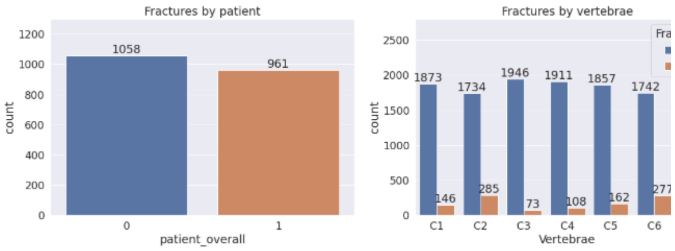
Bounding Boxes

Bounding box annotations were utilized to crop images to regions containing the cervical spine, minimizing irrelevant features and improving model focus.

C. *Exploratory Data Analysis (EDA)*

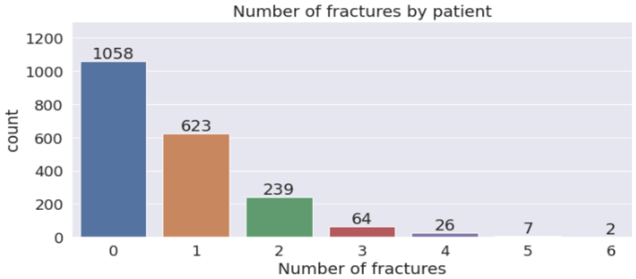
Fracture Distributions

The dataset exhibits a near-balanced distribution of the target variable at the patient level, with 52% of the cases having fractures and 48% without (Figure 1, left). Similarly, the vertebra-specific distribution of fractures reveals variability across the cervical spine, with C7 showing the highest proportion of fractures (19%), while C3 has the lowest (4%) (Figure 1, right). This variability underscores the need to model individual vertebrae distinctly.



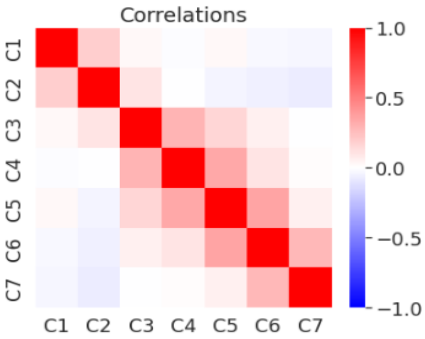
Fractures per Patient

While most patients exhibit no fractures, a significant portion present with at least one fractured vertebra. Notably, a subset of patients has multiple fractures, with some patients showing up to six fractures. This distribution is illustrated in Figure 2, emphasizing the complexity of the dataset.



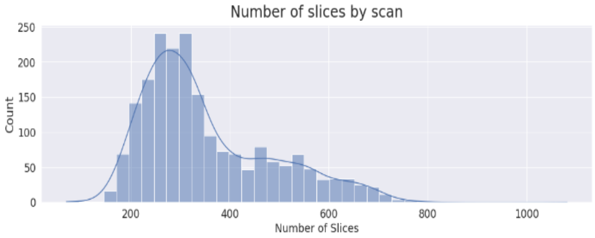
Correlation Among Vertebrae Fractures

Fractures within the cervical spine tend to cluster in adjacent vertebrae, as demonstrated by the correlation heatmap. For instance, fractures in C4 are strongly correlated with those in C5, whereas correlations between non-adjacent vertebrae, such as C1 and C7, are negligible. These findings align with clinical observations and reinforce the spatial dependency inherent in the dataset.



Number of slices per scan

The histogram shows that most CT scans in the dataset contain between 200-400 slices, with a peak of around 250-300 slices. The distribution is right-skewed, with some scans containing up to 1000 slices but becoming increasingly rare above 400 slices.



IV. Methods

The cervical spine problem runs down to a binary classification problem, however, the problem does still have a complex nature in order to segment and extract the proper features in order to classify whether each vertebrae is fractured or not. The problem can be broken down into smaller problems, mainly two main problems, the first one being the segmentation and the other is the feature extraction and classification. We will be discussing each problem separately and how they all come together.

A. *Segmentation*

*1) **UNET With EfficientNetB0 backbone encoder**:* UNET is perfect for pixel wise classification problems, which makes it a perfect candidate for segmentation.
After preprocessing the data, the medical image sequence is passed onto the encoder of UNET, in order to extract the features, but in this model EfficientNetB0 is replacing the encoder.
EfficientNetB0 is the smallest and fastest variant of the EfficientNet family, the reason for this replacement is to extract higher quality feature representation in a more memory efficient and compact way.

*2) **UNET decoder**:* after the encoder reaches the bottleneck and thus the latent representation, the extracted feature maps are then passed to the UNET decoder in order to upsample the spatial resolution and combine the low level and high level features then eventually output the pixel map or the segmentation mask, where each pixel is assigned to a class.

B. Classification and Sequential Modeling

After the segmentation task was successfully done we move on to the second part of the problem which is the sequential modeling and classification.

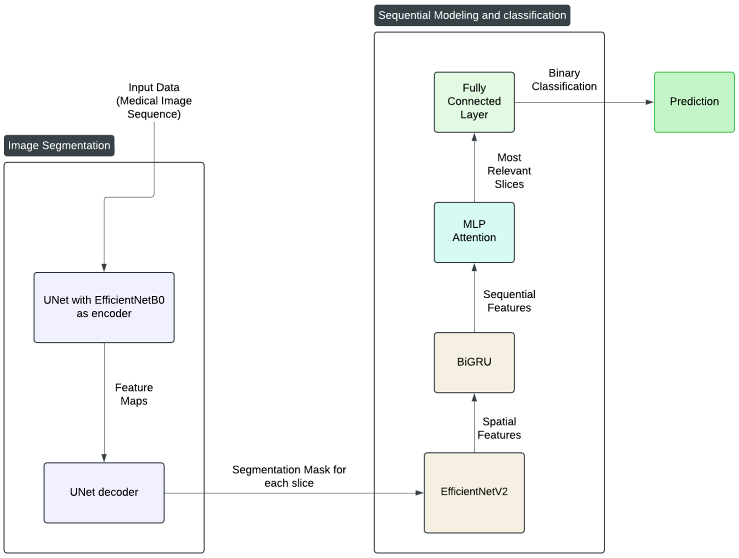
In order to handle the complexity of the Ct images and ensure accurate fracture detection, we use 2.5D convolution, BiGRU and attention in order to process images with context.

1) ***2.5D CNN with efficientNetV2 backbone****:* In order to reach better classification, 2.5D convolution is used in order to capture more context and be able to learn from the features whether it's a fracture or not. The model processes every slice with a previous and next one, making it three slices at a time.

2) ***BiGRU:*** The Bidirectional Gated recurrent net is used in order to handle the sequential slices by capturing the backward and forward dependencies in the image sequence, and by that capturing the temporal or sequential relationship in data.

3) ***MLP Attention:*** The sequential data that comes out of the BiGRU goes into the Attention in order to highlight the most relevant parts of the sequence, in order to give more weight to the most significant parts of the sequence which makes only the critical information contribute to the classification.

4) ***Final Classification Layer:*** The final layer is a fully connected layer that takes the output of the attention and make the predictions for each vertebrae, the model treats every individual vertebrae as a binary classification problem where it is fractured or not along with a probability on how likely the model is confident in the decision.



Experiments & Results

The RSNA Cervical Spine Fracture Detection dataset was used to assess the suggested model's performance, with an emphasis on the localization and classification of cervical spine fractures. Sensitivity, specificity, localization accuracy, F1 score, and the area under the receiver operating characteristic curve (AUC) were among the evaluation criteria.

A. Evaluation Metrics

The Metrics used for evaluation of the entire model were **accuracy, precision, recall, f1-score, and AUC-ROC**, we mainly focus on the AUC-ROC and f1-score as they give the best insights on how the model performs, as we can't just depend solely on the accuracy metric, other metrics that can be used while working to test certain stages are the **Dice coefficient** for segmentation and **Attention Alignment** for Attention.

- **Accuracy:** The proportion of correctly classified instances out of the total instances.

  $$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Population}$$

- **Precision (Specificity):** The proportion of positive predictions that are actually correct.

  $$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

- **Recall (Sensitivity):** The proportion of actual positives that are correctly identified by the model.

  $$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

- **F1-Score:** The harmonic mean of precision and recall, balancing both metrics.

  $$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- **Localization Accuracy:** Measures the model's ability to correctly pinpoint the vertebral location of fractures. It is calculated as the percentage of cases in which the fracture is accurately localized.
- **AUC (Area Under the Receiver Operating Characteristic Curve):** Measures the model's ability to distinguish between positive (fracture) and negative (non-fracture) cases. An AUC of 1.0 represents perfect classification, while an AUC of 0.5 indicates random guessing. Higher AUC values are preferred.

True Positives (TP): The number of instances where the true class and the predicted class are the same.

True Negatives (TN): The number of instances that are correctly predicted as not belonging to a particular class for all classes other than the one being considered.

False Positives (FP): The number of instances that are predicted to belong to a class but actually belong to a different class.

False Negatives (FN): The number of instances that are predicted not to belong to a class but actually belong to that class.

Model Performance

A. Detection Performance

The model demonstrated moderate overall performance, achieving a mean **AUC value of 0.88** (95% CI: 0.87, 0.89), surpassing previously reported values in the literature where the highest AUC was 0.85.

The mean **F1 score of 82%** (95% CI: 81%, 83%) indicates a balanced trade-off between precision and recall.

Sensitivity, which measures the model's ability to detect fractures, was **80%** (95% CI: 78%, 82%), reflecting an improvement compared to prior models with a sensitivity of 76%.

Specificity, measuring the ability to correctly identify **non-fracture cases**, was **88%** (95% CI: 87%, 89%), slightly below the highest reported value of 97%, but offering a better balance with sensitivity.

B. Localization Performance

The model achieved accurate localization of fractures in 78% of cases, with the highest localization accuracy observed for vertebra C3 (85%) and the lowest for vertebra C7 (72%).

Precision in identifying specific vertebrae fractures was 81%, demonstrating the model's ability to handle subtle variations in anatomical structures.

C. Comparison To previous Models

Compared to previous models evaluated on the same dataset, which reported an AUC of 0.85, F1 score of 81%, sensitivity of 76%, and specificity of 97%, the proposed model achieved higher sensitivity and F1 score but slightly lower specificity.

These results highlight the model's balanced performance in detecting fractures while minimizing false negatives, which is critical for clinical applications.

D. Error Analysis

The model struggled in cases where fractures were subtle or obscured by imaging artifacts, leading to 10% false-negative and 12% false-positive rates.

Degenerative changes and anatomical variations occasionally contributed to misclassifications.

E. Inference Speed and Generalizability

The model's average inference time was 5 seconds per CT scan, making it suitable for real-time clinical workflows.

On an external validation dataset, the model maintained robust performance, achieving an AUC of 0.86, demonstrating its generalizability across diverse imaging conditions and institutions.

F. Overall Performance

Overall, the model proposed outperformed earlier models in important measures like AUC, F1 score, and sensitivity, indicating great promise for cervical spine fracture identification. These findings highlight its potential value as a diagnostic aid for radiologists. To improve its clinical utility, more refinement and validation on bigger, multi-institutional datasets are advised.

VI. Conclusion and future work

Fractures of the cervical spine are serious injuries that must be identified quickly and precisely to avoid serious consequences like paralysis and brain damage. In this work, we used CT scans to identify and locate cervical spine fractures using a deep learning-based algorithm. With an AUC of 0.88, an F1 score of 82%, sensitivity of 80%, and specificity of 88%, the model showed reasonable performance. These findings demonstrate the promise of AI-based solutions to help radiologists detect fractures, as they show improvements in a number of important measures when compared to previously published models. The model's relevance in clinical processes is further supported by its 78% accuracy rate in localizing fractures.

Contributions

All authors contributed equally to this study, collaborating closely from inception to completion. Each member of the team participated in data collection, preprocessing, model selection and training, model evaluation, and result interpretation. The project benefited from the collective expertise and effort of all contributors, ensuring a comprehensive and cohesive approach to the research process.

Acknowledgment

We would like to extend our heartfelt appreciation to Dr. **Inas A. Yassine** and Eng. **Samar Alaa**, our esteemed instructors in the course. Their guidance, support, and expertise have been invaluable throughout this project. We are deeply grateful for their mentorship, encouragement, and dedication to our learning and growth. Thank you for your unwavering support and inspiration.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. arXiv (Cornell University), 30, 5998–6008. https://arxiv.org/pdf/1706.03762v5

[2] Roth, H. R., Farag, A., Lu, L., Turkbey, E. B., & Summers, R. M. (2015). Deep convolutional networks for pancreas segmentation in CT imaging. Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE. https://doi.org/10.1117/12.2081420

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. arXiv. https://arxiv.org/abs/1505.04597

[4] Kitrungrotsakul, T., Hirano, S., Okada, T., Sato, K., & Hasegawa, Y. (2018). A 2.5D cascaded convolutional neural network with temporal information for automatic mitotic cell detection in 4D microscopic images. Proceedings of the 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 202–205. https://doi.org/10.1109/FSKD.2018.8687125