Cervical Spine Fracture Localization using Semi-Supervised Learning

Birzhan Abdikhan, BEng in Biological Sciences

Student ID: 201549559

Submitted in fulfilment of the requirements for the degree of

Masters of Science in Biomedical Engineering

School of Engineering and Digital Sciences

Department of Chemical and Materials Engineering

Nazarbayev University

53 Kabanbay Batyr Avenue, Nur-Sultan, Kazakhstan, 010000

Supervisor: Dana Akilbekova

Co-supervisor: Zhanibek Kozhirbayev

April 2023

DECLARATION

I hereby declare that this manuscript, entitled "*Cervical Spine Fracture Localization using Semi-supervised Learning",* is the result of my own work except for quotations and citations which have been duly acknowledged.

I also declare that, to the best of my knowledge and belief, it has not been previously or concurrently submitted, in whole or in part, for any other degree or diploma at Nazarbayev University or any other national or international institution.

Name:

Date:

Abstract

Cervical spine fracture localization in medical images is a challenging task that requires a large amount of labeled data for accurate diagnosis. However, obtaining labeled data is time-consuming and difficult, which limits the application of supervised learning methods. In this thesis, we propose a semi-supervised learning approach to improve the accuracy of cervical spine fracture localization by combining a small amount of labeled data with a larger amount of unlabeled data.

Our approach leverages semi-supervised learning techniques to learn patterns and features in the larger set of unlabeled CT scans, which improves the model's ability to generalize to new and unseen cases. Additionally, our approach is more robust to noisy or inaccurate labeled data, as the model can learn to ignore or weight the labeled data based on its confidence in the label.

To increase the amount of labeled data available for training, we also explore data augmentation techniques, such as rotation, flipping, cropping. We demonstrate the effectiveness of our approach through experiments on a dataset of CT scans for cervical spine fracture localization.

Our results show that our semi-supervised learning approach improves the accuracy of cervical spine fracture localization compared to traditional supervised learning methods, even when trained on a limited amount of labeled data. Overall, our approach has the potential to improve the diagnosis of CSFs in medical images, which can ultimately lead to better patient outcomes.

Acknowledgement

Firstly, I would like to thank my supervisor Dr. Dana Akilbekova for her constant support and guidance throughout this project. I am proud of being under her supervision during this research project and for being part of Tissue Engineering laboratory. I also want to express my gratitude to my thesis co-supervisor Zhanibek Kozhirbayev for his help.

This work was carried out with support from Grant Award Number under the “Cervical Spine Fracture Localization and Vertebrae segmentation using Semi-Supervised Learning” project from Nazarbayev University.

We acknowledge the online competition RSNA 2022 for providing the dataset and metadata.

Table of Contents

List of Abbreviations & Symbols

CNN	Convolutional Neural Network
CSF	Cervical Spine Fracture
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
FOV	Field of view

List of Figures

Chapter 1 - Introduction / Literature Review

After experiencing blunt trauma or vehicular collisions most patients are stable and around 1% were reported to have cervical spine fracture [1]. On Emergency department, patients with potential CSF are immobilized and wait for radiologist examination in resuscitation rooms. Afterwards, these patients are kept till diagnostic imaging results are finished which burdens crowded ED space [1]. The standard radiographic examination for the Cervical Spine Fractures (CSFs) includes 3 projections: anteroposterior, cross-table lateral and open mouth odontoid views [2]. 50% of fractures, are overlooked under radiographic identification. In Emergency departments, failure to find fractures on plain radiography is the widespread diagnostic error [3].

CSFs occur for 2-5% patients with spinal fractures [4]. They generally occur due to motor vehicle accidents, followed by falls and sports injuries. Accurate and timely diagnosis of CSF's is critical for determining appropriate treatment options and preventing further damage to the spinal cord. However, current methods for cervical spine fracture localization, such as radiographic imaging techniques, can be time-consuming, expensive, and may not always be readily available in emergency situations. Therefore, there is a need for an efficient and accurate method for cervical spine fracture localization that can improve the speed and accuracy of diagnosis.

This research works on developing EfficientNET algorithm to detect vertebral positions and localize fractures by providing better input to CNN. Therefore, following Research Questions were posed:

1. How can different semi-supervised learning models be used to improve the accuracy of and cervical spine segmentation and fracture localization in medical images?
2. What are the limitations of using supervised learning methods for cervical spine segmentation and fracture localization, and how can these be overcome using semi-supervised learning?
3. How can image pre-processing and data augmentation techniques be used to increase the amount of labeled data available for training a semi-supervised learning model for detection of vertebral position and cervical spine fracture localization?

The objective of this thesis is to develop a cervical spine fracture localization system using semi-supervised learning. In this chapter, we will review the relevant literature on the general problem of cervical spine fracture detection and localization, as well as the specific topic of semi-supervised learning.

The literature review will be structured around the following three areas related to the problem:

1. Vertebrae segmentation: This section will review the literature on vertebrae segmentation, and deep learning architectures that have shown promising results on various computer vision tasks. We will focus on studies that have used machine learning for medical image analysis and discuss its potential for cervical spine segmentation.
2. Fracture detection: This section will review the existing research on cervical spine fracture detection and localization using machine learning techniques. It will include a discussion of different approaches that have been used to address this problem, including supervised, semi-supervised and unsupervised learning.

This literature review will provide a comprehensive overview of the state-of-the-art in cervical spine fracture detection and localization using machine learning techniques, as well as the potential of semi-supervised learning in this area.

1.1 Traditional Vertebrae fracture screening

The cervical spine is located in the upper part of the spine and connects head to the shoulders. Cervical Spine Fractures have initial symptoms of "neck pain, torticollis, altered mental status, sensory loss, motor loss and respiratory arrest" [5].

Multiple image modalities are being in use by the physicians for diagnosis of CSFs: Plain chest radiography, Computed Tomography (CT) and MRI:

- Plain radiography is initially performed for its readily availability and low radiation dosage. It has detection capability for bony CSFs of around 90% and is used for patients who are have no neurological abnormalities [5].

- In order to rule out neck fractures, a radiograph or, in exceptional circumstances when the radiograph does not show sufficient results, a CT-scan of the cervical spine is typically performed. CT has 100% detection ability for bony CSF [2]. The use of CT imaging might decrease the number of diagnostic errors due to the high sensitivity and accuracy. However, it introduces from 90 to 200-times increase in radiation to the patient. In cases of children, it increases risks of thyroid cancer.
- MRI has not been in wide use due to its time and financial costs [5].

Fractures occurring in complex anatomical regions are known to be challenging to identify using plain radiographs, which are the primary imaging method used in emergency departments [3]. Misinterpretation of fractures can lead to delay in appropriate treatment and unfavorable future performance of patients.

1.2 Machine learning

Machine learning refers to a collection of techniques that can identify patterns in data automatically and subsequently use these patterns for various purposes, such as predicting future data or making decisions under uncertain conditions. This could also include devising strategies for gathering more data. Machine learning can be classified into two main categories [6]. The first is predictive or supervised learning, where the objective is to learn a extracting features from inputs to outputs, given a labeled set of inputs and outputs that is called the training set. The training set contains some examples. The input vector can be a simple any dimensional vector, representing attributes such as height and weight, or a complex structured object such as a medical image. The output variable can be a categorical variable, such as presence or absence of a fracture. The second type of machine learning is descriptive or unsupervised learning, where only the inputs are given, and the goal is to discover interesting patterns in the data. This task is more loosely defined than supervised learning, as there is no clear objective function or error metric to optimize.

1.2.1 Deep learning

Machine learning models usually have a simplistic two-layer architecture with input-output relation. Scientists have aspired for the multiple connectivity of brain neuronal function for the processing of decision-making tasks where each neuron (in machine learning called perceptron) learns features and myelination of dendritic connections define the importance of learned information. This has aspired scientists to develop deep learning that has multiple input-output pairs for each layer and weight vectors that would define their importance [6]. The layers are Fully Connected which means that output from previous layers is received by every layer is input and weight of their connection is calculated. For that reason, this fundamental architecture was called Neural Network.

Convolutional Neural Network (CNN) introduce mathematical operation of convolution for matrices where one matrix, in our case multiple pixel matrices from 1 image, is treated by another called filter, where each element of the image piece is added to each element in the kernel.

For deep learning, A 2D CNN is used for processing 2D images. It applies convolutional operations to the image pixels, typically in 2 dimensions (height and width), to extract features and learn representations that can be used to classify or detect objects in the image. A 3D CNN is used for processing 3D volumes of data, such as medical images or video frames. It applies convolutional operations in 3 dimensions (height, width, and depth), allowing it to capture spatial and temporal features in the data. A 2.5D CNN is a type of CNN that applies 2D CNN operations to a stack of 2D slices from a 3D volume. This approach can be used to process 3D volumes with less computational cost than a full 3D CNN, while still capturing some of the 3D spatial features.

Pre-processing and image augmentation are important steps in preparing data for use in deep learning models [7]. Pre-processing refers to the manipulation of input data to transform it into a more suitable format for use in the model. This can involve tasks such as normalization, resizing, and cropping. Normalization involves scaling the pixel values of an image to a specific range to improve the efficiency of the training process. Resizing involves changing the dimensions of an image to fit the input size of the model. Cropping involves selecting a portion of the image that contains the most relevant information.

Image augmentation refers to the process of artificially increasing the size of the training dataset by applying transformations to the input images. This can involve techniques such as rotation, translation, flipping, and zooming. These transformations help to introduce variability into the training data and improve the robustness of the model to variations in input images. For example, by rotating an image by a small angle, the model can learn to recognize the same object from different orientations.

In addition to improving the quality of the input data, pre-processing and image augmentation can also help to prevent overfitting. Overfitting occurs when a model learns to recognize specific examples from the training data too well, but does not generalize well to new examples. By introducing variability into the training data, pre-processing and image augmentation can help to reduce the risk of overfitting and improve the overall performance of the model.

Most of the times clinical image datasets have problem of uneven distribution of classes of images. It makes the process of outlier, rare cases, detection difficult. Class imbalance is a situation in machine learning where one or more classes in the training data are significantly underrepresented compared to the others [8]. This can lead to biased models that perform poorly on the underrepresented classes. For example, if a binary classification problem has 90% positive and 10% negative examples, a model that always predicts positive will achieve 90% accuracy, even though it is not useful for the negative class. There are several methods to handle class imbalance, including [9]:

1. Oversampling: This involves randomly duplicating examples from the minority class to balance the distribution of the classes.
2. Undersampling: This involves randomly removing examples from the majority class to balance the distribution of the classes.
3. Synthetic sampling: This involves generating new examples from the minority class using techniques such as SMOTE (Synthetic Minority Over-sampling Technique).

4. Cost-sensitive learning: This involves assigning different misclassification costs to the different classes, such that the cost of misclassifying the minority class is higher than that of the majority class.
5. Ensemble learning: This involves combining multiple models, each trained on a different subset of the data, to improve performance on the underrepresented class.

1.2.2 Model architecture

EfficientNet is a family of neural network architectures that are designed to achieve state-of-the-art performance with high efficiency, meaning they can achieve high accuracy while using fewer parameters and less computational resources compared to other models. These models have been shown to be effective in various computer vision tasks, including medical image analysis. In the context of cervical spine fracture localization, EfficientNet can be used as a backbone network for a semi-supervised learning approach. This involves training a neural network on a small labeled dataset of cervical spine fracture images and a large unlabeled dataset of similar images. In the pioneering research for EfficientNET, it has been observed empirically that scaling dimensions are not independent. Specifically, it has been noted that for higher resolution images, network depth should be increased to allow for larger receptive fields, which can capture similar features that include more pixels in larger images [8]. Additionally, it has been suggested that network width should also be increased for higher resolution images in order to capture more fine-grained patterns with more pixels. These observations imply that a coordination and balance of different scaling dimensions is necessary, rather than the conventional approach of single-dimension scaling.

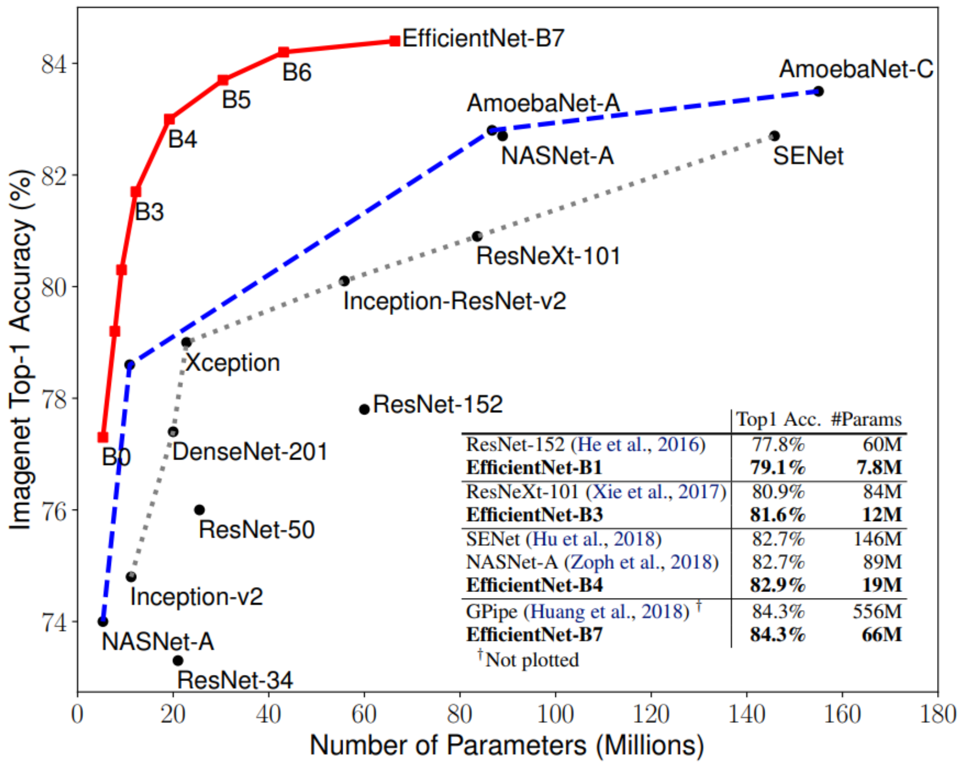


Figure 1-1. Comparison of model architectures [10]

Deep CNNs have been used for many years and are known for having multiple convolutional layers that extract hierarchical features from image data. They often have many parameters, which can make them challenging and time-consuming to train. On the other hand, EfficientNet is a newer architecture that prioritizes computational efficiency and high accuracy. It uses a compound scaling method that ensures a balance between the model's depth, width, and resolution at each layer.

Here's an example of a CNN model architecture that could be used for the task of cervical fractures detection using convolutional neural networks:

- Input Layer: The input layer takes as DICOM images from the cervical spine CT images.
- Convolutional Layers: The first layer of the CNN applies filters to the input image to extract features such as edges, corners, and other low-level features.
- Pooling Layers: After each convolutional layer, a pooling layer is used to reduce the spatial dimensions of the output feature maps.
- Fully Connected Layers: After several convolutional and pooling layers, the output is flattened and passed through one or more fully connected layers. These layers perform a classification task, using the extracted features to classify the input image as either having a fracture or not.

- Output Layer: The output layer produces a single output, which represents the predicted class label for the input image.
- Activation Functions: Nonlinear activation functions such as ReLU, sigmoid, or tanh can be used between the layers to introduce nonlinearity into the model.

The detection of fractures on CT involves several stages, each with its own specific purpose and role in the overall process. These stages include segmentation, feature extraction, aggregation, and classification. Segmentation is the process of identifying and separating the relevant structures in the CT images, such as the vertebrae and any potential fractures.

After training all the predictors, the algorithm can make predictions for new instances by combining previous predictions of all predictors. The typical aggregation function used is the statistical mode for classification or the average for regression [4]. Although each individual predictor has a higher bias than if it were trained on the original training set, aggregating reduces both bias and variance.

1.3 Segmentation

In the context of CSF localization, the segmentation task plays an important role of identifying and isolating the regions of interest (ROI) in the medical images [11]. It involves dividing an image into multiple regions, where each region represents a particular object or structure within the image. In this case, it helps in isolating and enumerating cervical vertebrae and separate them from the surrounding tissues that may be the source of noise or irrelevant information. After segmentation task is complete, CNN can focus on assigned ROIs and perform fracture detection tasks.

A CT image's dimensions are determined by its length and width, which can be influenced by factors such as the scanner type, scanning protocol, and reconstruction settings [8]. During a CT scan, the scanner captures X-ray images of the body or organ being examined, which are typically square or rectangular in shape. The image size is determined by the number of pixels or voxels along each dimension. The most common CT image size is 512 x 512 pixels, but other sizes may be used depending on the scanner and the preferences of the radiologist or technologist. Slightly larger image sizes may be used when the scanner acquires images with a larger field of view or when the reconstruction settings are adjusted. However, these minor variations do not impact the diagnostic quality of the images.

Some authors skipped segmentation of the spine before extraction and provides a CNN with mapping slices and give feedback for fracture types [12]. Authors used a 2D CNN for analyzing of sagittal batches that were combined prior along the spine with a Recurrent Neural Network to receive results of multiple batches from the same patient. Some authors used a same approach but performed Long Short-Term Memory (LSTM) units [13].

1.4 Fracture Detection

After segmentation task which outputs the ROI, the fracture detection performs feature extraction, aggregation, and classification tasks.

Fracture localization requires software to be able to discern contrast between vertebrae and fracture. The pre-processing steps would require such techniques as contrast enhancement, histogram equalization, and filtering that can highlight features of fractures. Image augmentation techniques can include random shifts, zooms, rotations, as well as flipping and mirroring.

On the following Table 1-1, the model architectures were summarized for segmentation, extraction, aggregation, and classifier tasks.

Table 1-1. Summary of model architectures used for segmentation, extraction, aggregation and classifier tasks

Author	Dataset	Segmentation	Extraction	Aggregation	Classifier
Bar et al. 2017	Compression fractures	-	2D CNN for sagittal view	RNN	-
Inoue et al. 2022	Chest, abdomen, and pelvis fractures	Inception-V2	Faster R-CNN		
Tomita et al. 2018	Osteoporotic vertebral fractures	-	ResNet34	LSTM	ResNet34
Thian et al. 2019	Wrist fracture radiographs	Faster R-CNN			
Xue et al.	Hand fractures	Region Proposal Network combined with ROI Pooling	Guided anchoring with Faster R-CNN	GA module	Fast R-CNN
Sato et al. 2021	Hip fracture radiographs	EfficientNet-B4			
Ukai et al. 2021	Pelvic fractures	YOLOv3			
Zhou et al. 2020	Rib fractures	YOLOv3	Faster R-CNN		

Most of the publications were working with developing CNN in 2D space. Faster R-CNN was used by most authors who reported its exceptional performance under different sizes of input images [14]. Bar et al. developed a method of detection for vertebral compression fractures based on Long Short-term Memory (LSTM). In this method, CNN calculates a probability vector with Deep CNN and classifies them with LSTM which was reportedly achieving accuracy, sensitivity and specificity of 89.1%, 83.9% and 93.8%, respectively [12]. Faster R-CNN was used by some authors who reported its exceptional performance under different sizes of input images [14].

Ukai et al. has developed a 2.5-dimensional YOLOv3 detection of pelvic fractures with the accuracy of 96.4% validated on 93 fracture subjects and 121 no-fracture subjects [15]. Authors have used multiple 2D-CNNs in which each image is evaluated in different orientation.

In the commercial sector, Aidoc, a software company that provides AI-powered solutions, developed BriefCase, an FDA-approved platform of CNN powered software for radiologists to analyze CT images for various medical conditions, including vertebral fractures. The C-spine software was developed to detect CSF's which performs 2 tasks: regional proposal and false-positive reduction based on Residual Network architecture [16]. In the official patent of the software, authors have indicated sensitivity of 91.7% (05% CI: 82.7%, 96.9%) and specificity of 88.6 (95% CI: 81.2%, 93.8%) [17]. Aidoc has developed several software solutions for analyzing CT images, including its cervical spine fracture detection solution.

Previously, ResNet architectures with only 1 FC layer has shown efficient feature extraction [18]. The number of FC layer can be increased for the improved accuracy of the model. Additional layers can potentially, learn complex patterns and relationships [19]. However, there are risks of overfitting model from training data and increasing computational cost and training time. Tomita et al has introduced 2 FC layers for the purpose of reduction of dimensionality extracted features [13]. Each of FC have batch normalization layer and ReLU function followed by them. Adams et al, has compared DCNN architecture with GoogleNet and AlexNet which resulted in giving preference for AlexNet [4].

After successful fracture detection, the representation of the results is important. Sato et al. have used gradient-weighted class activation mapping (Grad-CAM) to create a heatmap with the probabilities that would aid diagnosticians with evaluating the performance of the model in clinical setting [20].

Chapter 2 - Methods

This chapter describes the methods used in this thesis to develop a cervical spine fracture localization system using semi-supervised learning. We will first provide an overview of the dataset used in this study, followed by a description of the actual procedures, measurement instruments, and data analysis used in this study.

2.1 Materials

2.1.1 Dataset

The dataset is provided by Radiological Society of North America (RSNA) with the American Society of Neuroradiology (ASNR) and the American Society of Spine Radiology (ASSR) [21]. The given data consists of four distinct CSV files and three folders containing scan slices (1318 images) and segmentations (87 files).

Table 2-1. Training data

- StudyInstanceUID: Study ID which is unique for each patient scan
- Patient_overall: Patient overview, meaning whether any vertebrae has a fracture
- C1-C7: Indicating whether a vertebrae is fractured or not.

Figure 2-1. Example of Image slices

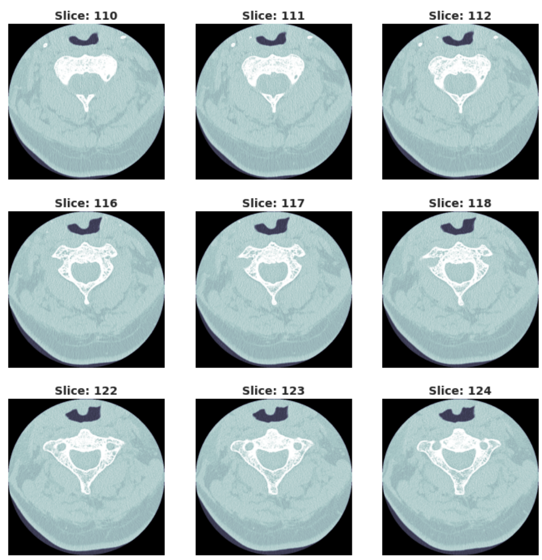
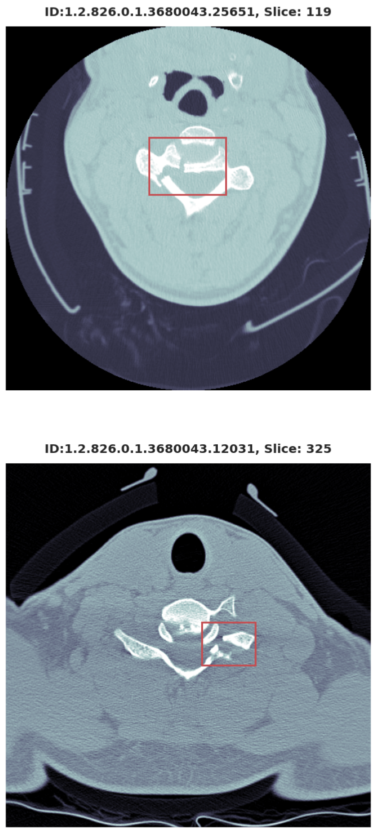


Table 2-2. Bounding Box Data	Figure 2-1. Example of bounding boxes
• StudyInstanceUID: Study ID which is unique for each patient scan • X: x coordinate of the bottom left corner of the bounding box • Y: y coordinate of the bottom left corner of the bounding box • Width • Height • Slice_number: slice number of the scan	

From Figure 2-3, it is observable that overall target sites are roughly balanced with a 52/48 split, which means that the distribution of the target variable, which could be a binary outcome or a categorical variable, is relatively evenly split between the two categories. This suggests that there is no significant bias or imbalance in the distribution of the target variable. C7 has the highest proportion of fractures (19%), while C3 has the lowest (4%). This suggests that there may be differences in the susceptibility of different cervical vertebrae to fractures, and that some vertebrae may be more prone to fractures than others. Several patients have more than one fracture, which suggests that some patients may be at a higher risk of fractures than others, and that the occurrence of one fracture may increase the likelihood of additional fractures. if multiple fractures occur on a single patient, they tend to occur in vertebrae that are close together. For example, C4 and C5 are more likely to be affected than C1 and C7. This information is useful for understanding the patterns of fracture occurrence and may be relevant for developing prevention or treatment strategies.

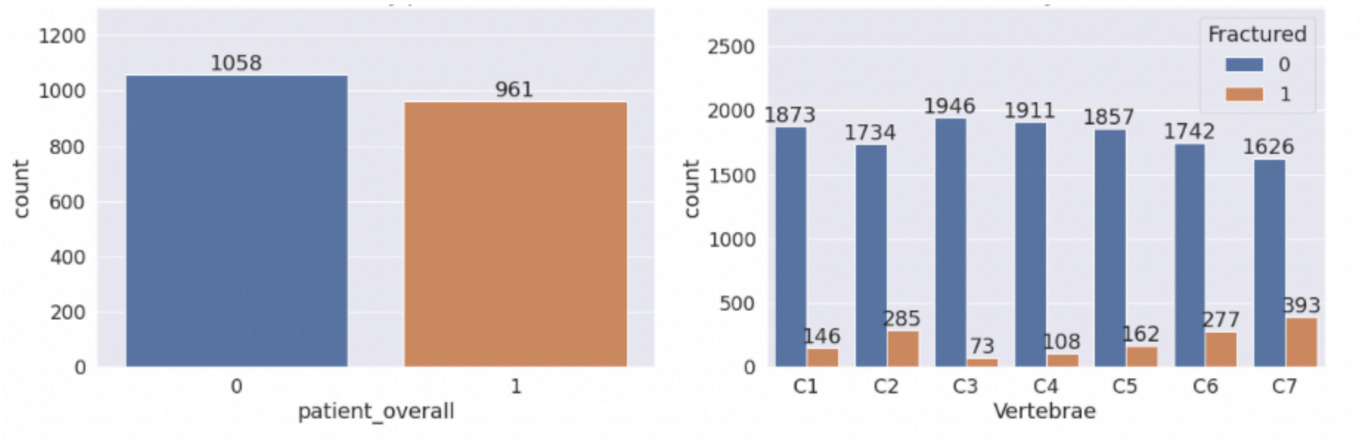


Figure 2-3. Number of fractures by patients and cervical vertebrae

From Figure 2-4, it can be observed that almost all images have size *512 x 512*, while 245 images have size of *512 x 519* and 782 have size of *768 x 768*. It is important to resize the rest of the images to *512 x 512* to standardize the batch of images for CNN as input data.

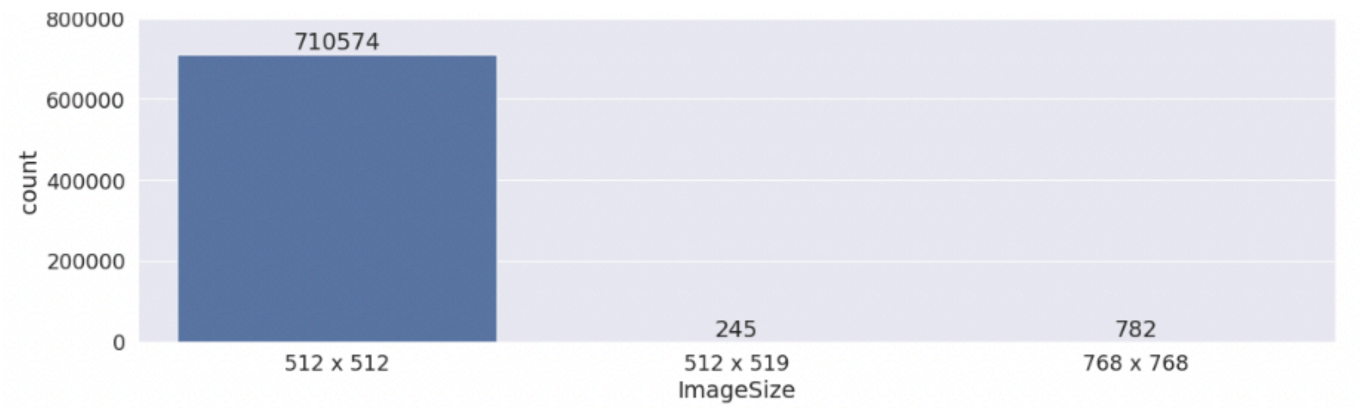


Figure 2-2. Image sizes in train images

The dataset contains metadata with segmentation files for semi-supervised learning for each patient that has provided refined metadata by removing image slices that have low contrast [22].

3.1.2 Hardware

Official Google Colab online software was used. GPU oriented runtime has following specifications: "Intel Xeon CPU @2.20 GHz, 13 GB RAM, Tesla K80 accelerator, and 12GB GDDR5 VRAM".

2.2 Segmentation

2.2.1 Baseline model

The first code was adjusted from existing baseline model from Slaykovskiy that uses EfficientNet-V2 model to predict presence of vertebrae on the image slice. It allows for faster segmentation compared to pixel-by-pixel segmentation [23].

2.2.2 Data analysis

To analyze the performance of the cervical spine fracture localization system, we used a confusion matrix to visualize the model's predictions and calculate the evaluation metrics. We also conducted statistical analyses to compare the performance of the semi-supervised learning approach with that of supervised learning and unsupervised learning.

Grouped k-folds cross validation was used to test the results of the segmentation. The output files were grouped by their respective StudyInstanceUID.

2.3 Fracture detection

We used a semi-supervised learning approach to develop a cervical spine fracture localization system. Specifically, we employed a pre-trained EfficientNet model to identify the presence of a vertebra and classify whether there is a fracture. We fine-tuned the EfficientNet model using the labeled data in the training set, and used the semi-supervised learning approach to improve the model's performance by leveraging the unlabeled data in the training set. We also used data augmentation techniques to increase the size of the training set and improve the model's robustness.

2.3.1. Baseline model

The code has been adjusted from the Slaykovskiy baseline notebook that uses EfficientNet-v2 backbone for fracture detection [23].

The EfficientNet backbone is used to extract features from the images, and then a semi-supervised learning algorithm is used to learn a mapping between the extracted features and the location of the fracture. By using EfficientNet as the backbone network, the semi-supervised learning approach can achieve high accuracy while using fewer parameters and less computational resources compared to other models. This can significantly reduce the time and resources required for training and inference, making it a more efficient solution for cervical spine fracture localization.

The images are first loaded from the train folder and then transformed into *3 x 384 x 384* tensors using the transformations that were utilized for pretraining EfficientNet_V2_S on ImageNet 1000. Afterward, the pre-trained encoder of EfficientNetV2 is used to process the images, and the final classification layer is disregarded because it is not relevant to the current task. The flattened layer with a shape is used as the base for the final classification layer, which is transformed into 2 parallel linear layers followed by a sigmoid. The loss function utilizes logits to enhance numerical stability. Only visible fractures are predicted on the current slice by masking the fracture targets with visible vertebrae targets, and the visible vertebrae targets are used in the loss function without any modifications. The predictions and targets are then passed to the BCELoss function, which optimizes seven independent binary classification targets for C1-C7. A weighted loss is also applied for the fracture targets.

The end result is a model that detects fractures and visible C1-C7 vertebrae in a single image. To estimate the final result with one record per patient instead of one record per scan, a non-parametric model is utilized to combine the predictions of the base models. The predictions for each of C1-C7 vertebrae are first aggregated for each patient by weighted averaging the fracture predictions, with the probabilities of vertebrae used as weights. For instance, if there is uncertainty that C3 is in the slice but there is high probability of C3 being fractured it will be added to the final aggregate with low weight.

2.3.2 Solver parameters

OneCycleLR is a type of learning rate scheduler used in neural network training. It is designed to increase the learning rate for the first few epochs and then gradually decrease it, following a "one cycle" pattern [24]. The learning rate schedule is typically modeled after a triangular wave, where the learning rate starts low, gradually increases to a maximum value, and then decreases back to the starting value. The OneCycleLR scheduler has several advantages over other learning rate schedulers, including faster convergence, improved generalization, and reduced sensitivity to the initial learning rate. By gradually increasing and decreasing the learning rate, the OneCycleLR scheduler enables the model to quickly converge to a good solution and then fine-tune the parameters to achieve better performance.

To use the OneCycleLR scheduler, the user specifies the initial learning rate, the maximum learning rate, and the total number of epochs. During training, the learning rate is adjusted according to the triangular wave pattern. The exact shape of the triangular wave can be modified using additional hyperparameters such as the percentage of the cycle spent increasing the learning rate and the percentage spent decreasing it.

2.3.3 Data Analysis

To measure the performance of the cervical spine fracture localization system, we used several evaluation metrics, including accuracy, precision, recall, and F1 score. These metrics were calculated on the validation and test sets to assess the performance of the model on unseen data.

The work of the algorithm will be evaluated using weighted multi-label logarithmic loss. Therefore, there should be a prediction of a probability that a fracture is present for each vertebrae.

$$L_{ij} = -w_j(y_{ij}\log(p_{ij}) + (1 - y_{ij})\log(1 - p_{ij}))$$

Chapter 3 - Results

3.1 Segmentation

3.1.1 EfficientNet vertebrae detection

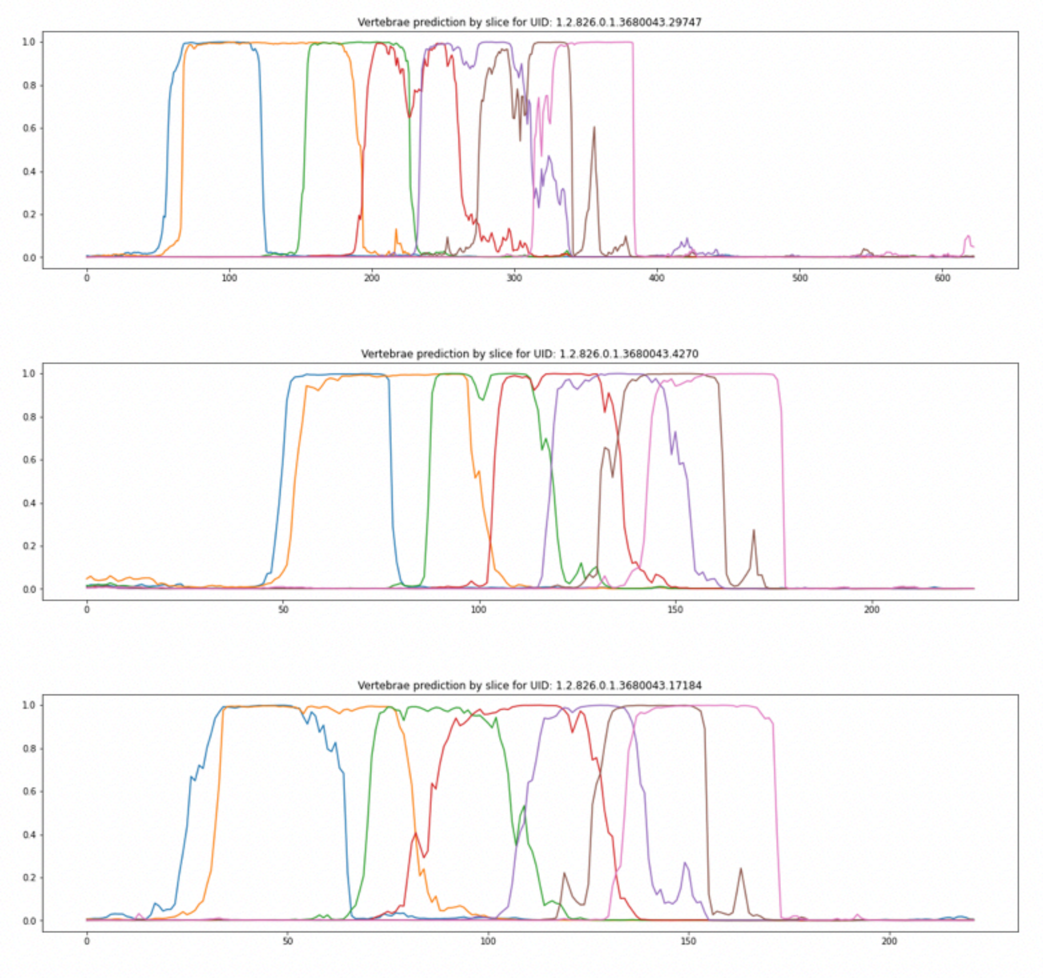


Figure 3-1. Vertebrae prediction by slices for 3 patients

On Fig 3-1, it is possible to observe the performance of the segmentation task of the algorithm where it provides the vertebrae prediction. Each color represents the vertebrae and in most cases it is able to discern vertebrae from one another.

The confusion matrices provide a detailed view of the model's performance for each vertebra. Based on the true positive and false positive values, we can calculate the precision of the model, which is the ratio of true positives to the total number of positive predictions made by the model. On the other hand, based on the true positive and false negative values, we can calculate the recall of the model, which is the ratio of true positives to the total number of actual positive cases.

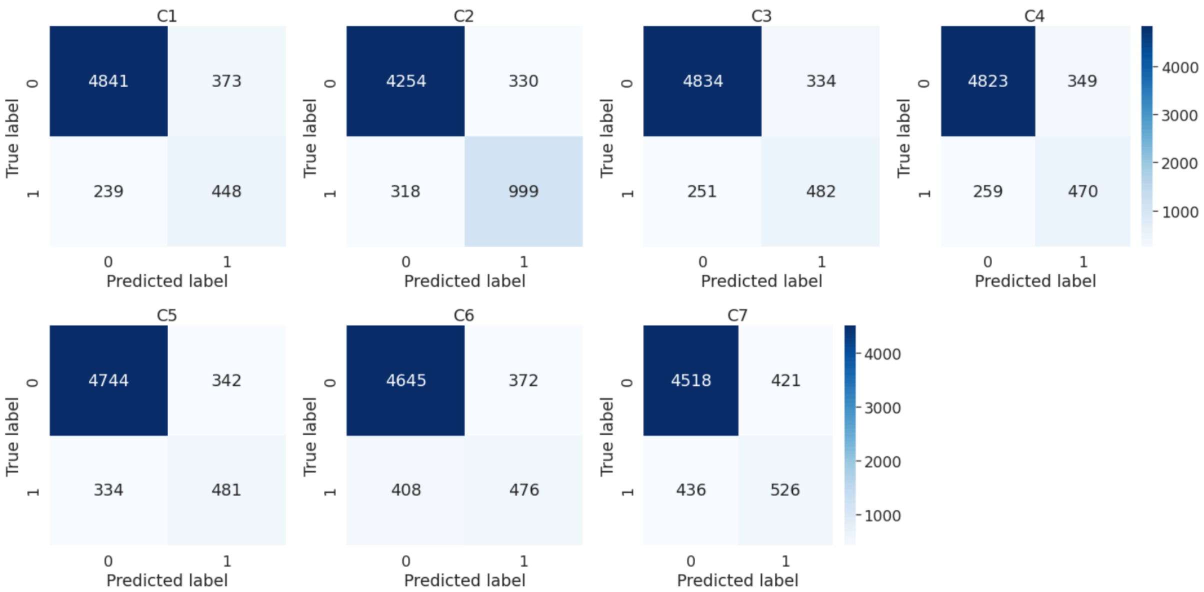


Figure 3-2. Confusion matrix for predicted labels

3.2 Fracture detection

The following figures showcase the results of the performance of the model. The loss score on confusion matrix reveal that vertebrae segmentation provides good differentiation of the observed vertebrae. The higher score (deep blue) demonstrates significant difference between vertebrae found by algorithm. In fig 4, fracture prediction and vertebrae prediction example for 1 patient is provided. It can be observed that fracture was predicted to be found on C7. The algorithm differentiates well between vertebrae as it provides different scores for the slices.

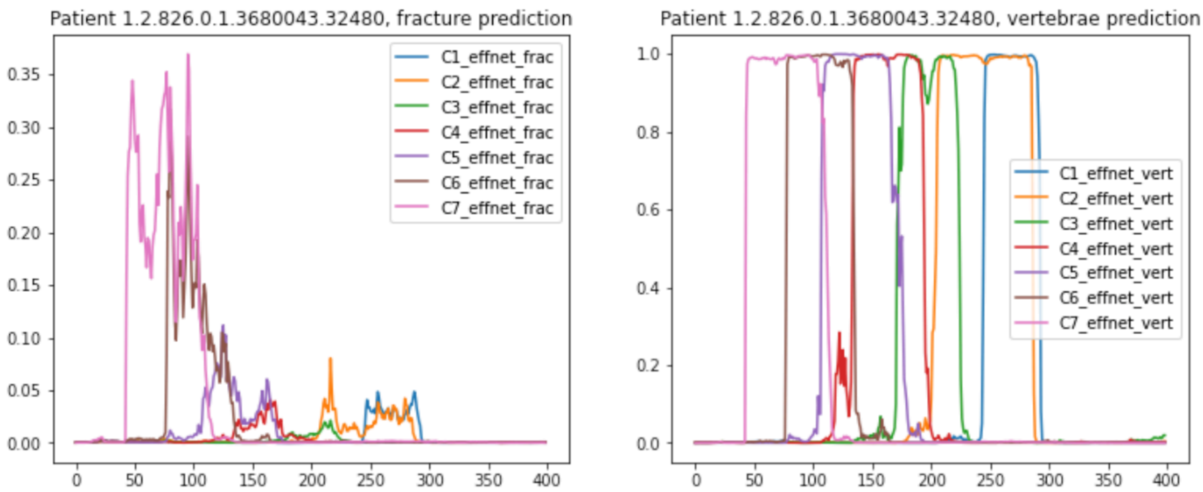


Figure 3-5. Weights for fracture prediction and vertebrae prediction

Chapter 6 Conclusion

Semi-supervised learning can be used to improve the accuracy of cervical spine fracture localization in medical images by combining a small amount of labeled data with a larger amount of unlabeled data. This approach allows the model to learn from both the labeled and unlabeled data, improving its ability to generalize to new and unseen cases. For example, a semi-supervised model could be trained on a small set of labeled CT scans of CSF’s and then use unsupervised learning techniques to learn patterns and features in the larger set of unlabeled CT scans, ultimately improving the model's accuracy in identifying CSF’s in new cases.The main barrier of using supervised learning methods for cervical spine fracture localization is the need for a large amount of labeled data, which can be difficult and time-consuming to obtain. Semi-supervised learning can overcome this limitation by leveraging a smaller amount of labeled data in combination with a larger amount of unlabeled data, ultimately improving the accuracy of the model. In addition, semi-supervised learning can be more robust to noisy or inaccurate labeled data, as the model can learn to ignore or weight the labeled data based on its confidence in the label.

Data augmentation techniques can be used to increase the amount of labeled data available for training a semi-supervised learning model for cervical spine fracture localization. For example, CT scans can be rotated, flipped, or cropped to create additional labeled examples for the model to learn from. In addition, synthetic images can be generated using generative adversarial networks (GANs) or other methods, which can provide additional variation and diversity to the training dataset.

The main limitation of this study is that it was examining only axial slices without sagittal or coronal slices. In future works, it is important to perform analysis for all 3 planes. The second limitation is that this study was not validated in clinical conditions or with physicians. There has been only validation under research purposes and therefore, there is yet no ability to scale this study to clinic. The third limitation is that this study does not consider relevant clinical history of the patients and only performs analysis on supplemented images.

Bibliography

[1] I. G. Stiell *et al.*, “A Multicenter Program to Implement the Canadian C-Spine Rule by Emergency Department Triage Nurses,” *Ann Emerg Med*, vol. 72, no. 4, pp. 333–341, Oct. 2018, doi: 10.1016/j.annemergmed.2018.03.033.

[2] M. Scaglione, F. Iaselli, G. Sica, B. Feragalli, and R. Nicola, “Errors in imaging of traumatic injuries,” *Abdominal Imaging*, vol. 40, no. 7. Springer New York LLC, pp. 2091–2098, Oct. 29, 2015. doi: 10.1007/s00261-015-0494-9.

[3] A. Pinto *et al.*, “Errors in imaging patients in the emergency setting,” *British Journal of Radiology*, vol. 89, no. 1061. British Institute of Radiology, 2016. doi: 10.1259/bjr.20150914.

[4] M. Adams, W. Chen, D. Holcdorf, M. W. McCusker, P. D. L. Howe, and F. Gaillard, “Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures,” *J Med Imaging Radiat Oncol*, vol. 63, no. 1, pp. 27–32, Feb. 2019, doi: 10.1111/1754-9485.12828.

[5] A. Slaar *et al.*, “Triage tools for detecting cervical spine injury in pediatric trauma patients,” *Cochrane Database of Systematic Reviews*, vol. 2017, no. 12, Dec. 2017, doi: 10.1002/14651858.CD011686.pub2.

[6] A. Geron, *Hands-on Machine Learning with SciKit-Learn, Keras & Tensorflow*, 2nd ed. France: O’Reilly, 2018.

[7] “Medical Image Preprocessing,” *MathWorks*. https://www.mathworks.com/help/medical-imaging/ug/overview-medical-image-preprocessing.html (accessed Apr. 07, 2023).

[8] L. Gao, L. Zhang, C. Liu, and S. Wu, “Handling imbalanced medical image data: A deep-learning-based one-class classification approach,” *Artif Intell Med*, vol. 108, Aug. 2020, doi: 10.1016/j.artmed.2020.101935.

[9] C. P. Langlotz *et al.*, “A roadmap for foundational research on artificial intelligence in medical imaging: From the 2018 NIH/RSNA/ACR/The Academy workshop,” *Radiology*, vol. 291, no. 3, pp. 781–791, 2019, doi: 10.1148/radiol.2019190613.

[10] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[11] B. Sahiner *et al.*, “Deep learning in medical imaging and radiation therapy,” *Medical Physics*, vol. 46, no. 1. John Wiley and Sons Ltd, pp. e1–e36, Jan. 01, 2019. doi: 10.1002/mp.13264.

[12] A. Bar, L. Wolf, O. B. Amitai, E. Toledano, and E. Elnekave, “Compression Fractures Detection on CT,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.01671

[13] N. Tomita, Y. Y. Cheung, and S. Hassanpour, “Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans,” *Comput Biol Med*, vol. 98, pp. 8–15, Jul. 2018, doi: 10.1016/j.compbiomed.2018.05.011.

[14] Q. Q. Zhou *et al.*, “Automatic detection and classification of rib fractures on thoracic ct using convolutional neural network: Accuracy and feasibility,” *Korean J Radiol*, vol. 21, no. 7, pp. 869–879, Jul. 2020, doi: 10.3348/kjr.2019.0651.

[15] K. Ukai *et al.*, “Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-91144-z.

[16] J. E. Small, P. Osler, A. B. Paul, and M. Kunst, “Ct cervical spine fracture detection using a convolutional neural network,” *American Journal of Neuroradiology*, vol. 42, no. 7, pp. 1341–1347, Jul. 2021, doi: 10.3174/ajnr.A7094.

[17] J. J. Smith, “K190896 Radiological computer aided triage and notification software,” 2019 [Online]. Available: https://www.fda.gov/CombinationProducts/GuidanceRegulatoryInformation/ucm597488.htm

[18] J. E. Burns, J. Yao, and R. M. Summers, “Vertebral body compression fractures and bone density: Automated detection and classification on CT Images,” *Radiology*, vol. 284, no. 3, pp. 788–797, Sep. 2017, doi: 10.1148/radiol.2017162100.

[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[20] Y. Sato *et al.*, "Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study," *BMC Musculoskelet Disord*, vol. 22, no. 1, Dec. 2021, doi: 10.1186/s12891-021-04260-2.

[21] "RSNA 2022," *Radiological Society of North America*, Oct. 2022. https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection (accessed Apr. 07, 2023).

[22] S. Cortinhas, "Extracting Vertebrae C1, ..., C7," *Kaggle*, Nov. 2022.

[23] C. Slaykovskiy, "PyTorch-EffNetV2 vertebrae detection (acc: 0.95)," *Kaggle*, Oct. 2022.

[24] "ONECYCLELR," *Pytorch*. https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.OneCycleLR.html (accessed Apr. 07, 2023).