DEEP SEQUENTIAL LEARNING FOR CERVICAL SPINE FRACTURE DETECTION ON COMPUTED TOMOGRAPHY IMAGING

Hojjat Salehinejad[1,2,3,*], *Edward Ho*[4], *Hui-Ming Lin*[4], *Priscila Crivellaro*[4], *Oleksandra Samorodova*[4],
Monica Tafur Arciniegas[4], *Zamir Merali*[5], *Suradech Suthiphosuwan*[4], *Aditya Bharatha*[4],
Kristen Yeom[6], *Muhammad Mamdani*[1,2,7], *Jefferson Wilson*[5], *and Errol Colak*[1,4]

[1]LKS-CHART, St. Michael's Hospital, Toronto, Canada
[2]St. Michael's Hospital, Unity Health Toronto, Toronto, Canada
[3]Department of Electrical & Computer Engineering, University of Toronto, Toronto, Canada
[4]Department of Medical Imaging, University of Toronto, Toronto, Canada
[5]Division of Neurosurgery, Department of Surgery, University of Toronto, Toronto, Canada
[6]Department of Radiology, Stanford University, Stanford, California, USA
[7]Temerty Faculty of Medicine, University of Toronto, Toronto, Canada
Email: hojjat.salehinejad@unityhealth.to

ABSTRACT

Fractures of the cervical spine are a medical emergency and may lead to permanent paralysis and even death. Accurate diagnosis in patients with suspected fractures by computed tomography (CT) is critical to patient management. In this paper, we propose a deep convolutional neural network (DCNN) with a bidirectional long-short term memory (BLSTM) layer for the automated detection of cervical spine fractures in CT axial images. We used an annotated dataset of 3,666 CT scans (729 positive and 2,937 negative cases) to train and validate the model. The validation results show a classification accuracy of 70.92% and 79.18% on the balanced (104 positive and 104 negative cases) and imbalanced (104 positive and 419 negative cases) test datasets, respectively.

Index Terms— Cervical spine, deep learning, fracture detection.

1. INTRODUCTION

The cervical spine consists of seven stacked bones called vertebrae, labeled C1 through C7, intervertebral discs, and ligaments. The top of the cervical spine articulates with the skull, and the bottom connects to the thoracic spine. Trauma to the cervical spine results in over 1 million emergency department visits per year in North America [1]. Severe trauma from falls, motor vehicle accidents, and sports injuries can result in fractures and/or dislocations of the cervical spine [2]. Cervical spine fractures may lead to permanent paralysis from spinal cord injury or even death [2]. A computed tomography (CT) scan of the cervical spine is the most common and preferred imaging modality used to detect fractures.

Despite advances in machine learning, very limited methods have been proposed for fracture detection in cervical spine CT images. A 3D ResNet-101 [3] deep convolutional neural network (DCNN) was utilized in [1], which was trained on 990 normal and 222 fracture cases. Performance of this method with respect to Area Under the Receiver Operating Characteristic (AUROC) and Area Under Precision-Recall Curve (AUPRC) metrics on the validation dataset (98 normal and 37 with fracture cases) at image level was AUROC=0.87 and AUPRC=0.52 and at case level was AUROC=0.85 and AUPRC=0.82 [1].

Automated fracture detection in cervical spine images is a very challenging problem. Cervical spine CT scans with a fracture are highly imbalanced in terms of normal images and those with a fracture. Different methods such as spatial transformations [4], spatial-temporal transformations [5], and generative adversarial models [6] can be used to reduce this imbalance. In addition, publicly available annotated cervical spine data is very limited, making advances to automated cervical spine fracture detection even more difficult. To address this, our team has annotated a dataset of cervical spine CT images (see Subsection 3.1) and we are in process of releasing this dataset to the research community[1]. In this paper, we model fracture detection as a classification problem and propose a DCNN with bidirectional long short-term memory (BLSTM) layer as a baseline model to address this problem on axial cervical spine CT images.

2. FRACTURE DETECTION IN CERVICAL SPINE

A cervical spine CT has N number of axial image slices along the cranio-caudal axis where we represent each image with a vector $\mathbf{x}_n$. Therefore, we can model the cervical spine as the set of input images $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ with the corresponding image level labels $\mathbf{y} = (y_1, y_2, ..., y_N)$, where

---

[1]For updates visit dila.ai.
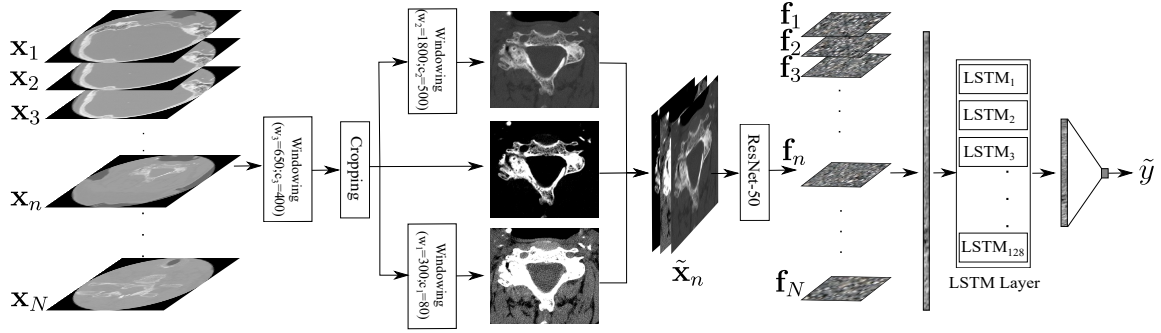
Fig. 1: Proposed model for fracture detection on cervical spine scans. The input scan has N axial images and the BLSTM layer has 128 LSTM units.
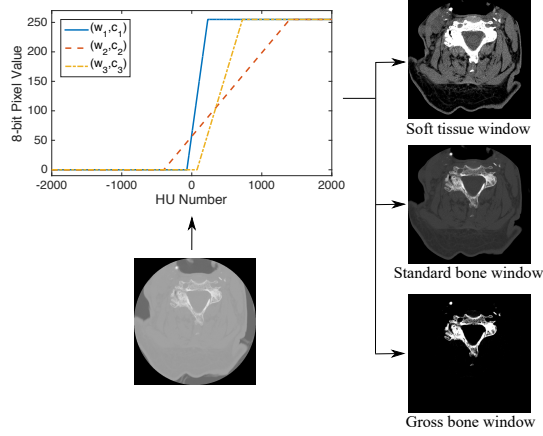
Fig. 2: Windowing of a given cervical spine CT image (in Hounsfield units (HU)) according to the following rules: 1) Soft tissue window ($w_1 = 300; c_1 = 80$); 2) Standard bone window ($w_2 = 1,800; c_2 = 500$); 3) Gross bone window ($w_3 = 650; c_3 = 400$).

$y_n = 1$ means the image contains at least one fracture and $y_n = 0$ is the opposite. We can also define a case level label $y \in \{0, 1\}$, where if $y = 1$ at least one image contains a fracture and if $y = 0$ none of the images contain a fracture. Figure 1 shows different steps of the proposed model. It has two major steps, which are prepossessing of the input images $\mathbf{X}$ and learning a mapping function from the preprocessed images to the target y. Different steps are discussed as follows.

2.1. Preprocessing

Each pixel $x_i \in \mathbf{x}_n$ is a quantitative representation of the radiodensity of different substances in the scanning area where $-1,000 \lesssim x_i \lesssim 3,000$, [7]. Fractures of the cervical spine are heterogeneous in their location, type, and composition. They are defined as a break in at least one bone in the cervical spine. The cervical spine can be malaligned as a result of a fracture and/or a ligamentous injury. Both fractures and ligaments injuries can be associated with localized bleeding and soft tissue swelling. In consultation with a panel of radiologists, three window width and center schemes are chosen to enhance cervical spine bone and surrounding tissues, which are soft tissue window ($w_1 = 300; c_1 = 80$), standard bone

Table 1: Distribution of negative and positive cervical spine fracture cases based on the gender and age.

	Positive	Negative
Female	6.74%	28.20%
Male	13.15%	51.91%
Age	59.42 ± 22.91	50.40 ± 21.7

window ($w_2 = 1,800; c_2 = 500$), and gross bone window ($w_3 = 650; c_3 = 400$). Figure 2 shows the corresponding functions to the proposed windowing schemes to map x_i to three different values $\tilde{x}_i^{(1)}$, $\tilde{x}_i^{(2)}$, $\tilde{x}_i^{(3)}$. The represented image with the gross bone window $\tilde{\mathbf{x}}_n^{(3)}$ is then used to crop the images with 5% margin from each side of detected cervical spine using Otsu's method [8]. The cropped images are normalized and resized (padded with zero if not square) to 384×384.
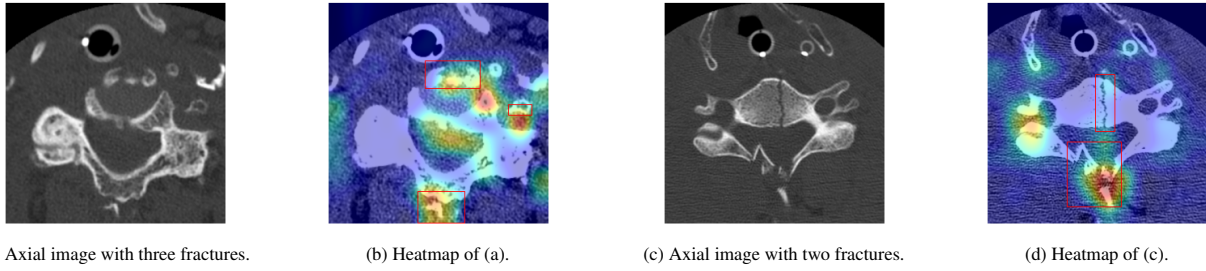
2.2. Learning and Inference

For the image $\mathbf{x}_n \in \mathbf{X}$, the preprocessed images are concatenated as $\tilde{\mathbf{x}}_n = (\tilde{\mathbf{x}}_n^{(1)} \oplus \tilde{\mathbf{x}}_n^{(2)} \oplus \tilde{\mathbf{x}}_n^{(3)})$ where $\oplus$ is the concatenation operator. The cervical spine images have spatiotemporal dependency. First, ResNet-50 [9] is trained using randomly selected batches of images from all training cases. The objective is to train ResNet-50 as the function $\mathbf{F} = \phi(\tilde{\mathbf{X}})$ where $\mathbf{f}_n \in \mathbf{F}$ is the extracted feature map from $\tilde{\mathbf{X}}_n \in \mathbf{X}$. As Figure 1 shows, the feature maps are vectorized and concatenated as $\tilde{\mathbf{f}} = (\mathbf{f}_1 \oplus \mathbf{f}_2 \oplus, ..., \mathbf{f}_N)$. The second phase is learning the temporal dependency among axial images using a bidirectional network of LSTM units to map $\tilde{\mathbf{f}}$ to the case label y. The loss is calculated using binary cross-entropy function with respect to the target label y. In the inference mode, the feature extractor $\phi(\cdot)$ and BLSTM layer work together for a given case $\mathbf{X}$ to generate the inference $\tilde{y}$.

3. RESULTS

3.1. Data

The dataset contains cervical spine CT scans of 3,666 cases (729 positive and 2,937 negative cases) which corresponds to 1,174,335 images (20,392 fracture positive and 1,153,943 fracture negative). This dataset is pixel-level annotated by

(a) Axial image with three fractures.   (b) Heatmap of (a).   (c) Axial image with two fractures.   (d) Heatmap of (c).

Fig. 3: (a),(c): Samples of cervical spine axial images with fracture; (b),(d): Grad-CAM [10] heatmap from the last layer of ResNet-50 corresponding to (a) and (c), respectively. Ground-truth is represented with red boxes.

Table 2: Image level performance results on imbalanced (Imblcd.) (2,909 positive cases; 164,349 negative images; based on the distribution of training dataset) and balanced (Blcd.) (2,909 positive images; 2,909 negative images) test datasets with 7-fold cross-validation using ResNet-50. TPR: Sensitivity; TNR: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; F1: F1 score; Acc: Accuracy; MCC: Matthews correlation coefficient; AUC: Area under the curve. All the values are in percent.

Data	TPR	TNR	PPV	NPV	F1	Acc	MCC	AUC
Imblcd.	77.21 ± 4.1	80.06 ± 3.0	06.47 ± 0.6	99.50 ± 0.1	11.92 ± 1.0	80.01 ± 2.9	18.47 ± 1.0	78.63 ± 1.2
Blcd.	77.21 ± 4.1	77.62 ± 2.7	13.78 ± 1.0	98.67 ± 0.2	23.35 ± 1.3	77.61 ± 2.5	26.11 ± 1.0	77.42 ± 1.1

Table 3: Case level performance results on imbalanced (Imblcd.) (104 positive cases; 419 negative cases; based on the distribution of training dataset) and balanced (Blcd) (104 positive cases; 104 negative cases) test datasets with 7-fold cross-validation. TPR: Sensitivity; TNR: Specificity; PPV: Positive predictive value; NPV: Negative predictive value; F1: F1 score; Acc: Accuracy; MCC: Matthews correlation coefficient; AUC: Area under the curve. All the values are in percent.

Model	Data	TPR	TNR	PPV	NPV	F1	Acc	MCC	AUC
ResNet-50 + BLSTM-96	Imblcd.	64.19 ± 5.7	78.67 ± 6.6	43.62 ± 6.3	89.83 ± 1.5	51.66 ± 5.5	75.79 ± 5.2	37.84 ± 7.6	71.43 ± 3.9
ResNet-50 + BLSTM-128		62.28 ± 6.0	80.84 ± 2.9	44.83 ± 4.8	89.62 ± 1.6	52.06 ± 4.9	77.15 ± 2.9	38.54 ± 6.7	71.56 ± 3.7
ResNet-50 + BLSTM-256		59.01 ± 5.7	84.12 ± 4.9	48.54 ± 6.7	89.34 ± 1.5	52.92 ± 4.6	$\mathbf{79.18 \pm 3.8}$	40.36 ± 6.5	71.57 ± 3.1
ResNet-50 + BLSTM-96	Blcd.	64.19 ± 5.7	77.11 ± 7.3	74.14 ± 5.4	68.36 ± 3.1	68.58 ± 3.8	70.65 ± 3.5	41.90 ± 7.1	70.65 ± 3.5
ResNet-50 + BLSTM-128		62.28 ± 6.0	79.84 ± 3.1	75.55 ± 3.0	68.06 ± 3.5	68.17 ± 4.2	$\mathbf{71.06 \pm 3.1}$	42.86 ± 5.9	71.06 ± 3.1
ResNet-50 + BLSTM-256		57.75 ± 4.9	84.09 ± 5.3	78.87 ± 4.9	66.63 ± 1.8	66.44 ± 2.8	70.92 ± 1.9	43.62 ± 4.3	70.92 ± 1.9

three expert radiologists and the intersection of annotated regions was reviewed by a neuroradiologist. Table 1 shows the distribution of gender and age of cases over the fracture positive and negative classes. This tables shows that from a gender perspective, men are more prone to cervical spine fractures. From an age perspective, the average of positive cases is 9 years greater than negative cases which suggests this injury is more common among the elderly [11].

3.2. Training Setup

Training setup of the BLSTM [12] model was as follows: Adam optimizer [13] with a learning rate of 10^{-6}; 100 training epochs; Batch size of 4; Input size of 2,048. Setup of the ResNet-50 [3] model was as follows: Adam optimizer [13] with a learning rate of 10^{-3} and step decay of 5 with gamma 0.2; Number of input channels was 3; Random horizontal flip and rotation (10 degrees) augmentation method; Batch size of 16 and 50 training epochs. Early stopping and dropout regularization methods were applied. Given above setup, the proposed model was trained from scratch with randomly initialized weights and tested using a 7-fold cross-validation scheme.

3.3. Performance Analysis

Figure 3 shows two samples of axial cervical spine images with corresponding ground-truth masks and heatmaps generated using Grad-CAM [10] from the last layers of ResNet-50. The heatmaps show ResNet-50 can capture most fracture areas with a relatively high false positive rate. Classification performance of the ResNet-50 at image level fracture detection in cervical spine axial images is presented in Table 2. In this experiment, cervical spine fracture is predicted just based on the spatial features in each axial image without counting temporal information. The results show 80.01% and 77.61% classification accuracy for the imbalanced and balanced datasets, respectively. However, the main drawback of this approach is the high level of false positives which leads to an inaccurate prediction for the entire case. Such that even if a single image in a case is false positive, the entire case will be false positive and vice-versa. These results show the importance of incorporating temporal features in training and inference. Table 3 shows performance results of the combined BLSTM and ResNet-50 model at the case level for different number of LSTM units. The results regarding the imbalanced dataset show as the number of units increases from 96 to

256, the classification accuracy also increases from 75.79% to 79.18%. However, the accuracy is approximately 71% for the balanced dataset and it is less dependent on the number of LSTM units.

Cervical spine imaging cases from a fracture detection perspective are highly biased from two perspectives. First is the larger number of negative cases when compared to positive cases. The second is generally a greater number of negative images than positive images within a positive case. These results clearly show the importance of considering these biases. For the imbalanced dataset, which reflects the natural distribution of negative and positive cases, as more LSTM units are utilized the imbalanced accuracy also increases. However, we observed no significant change in balanced accuracy. The higher performance on the imbalanced dataset is mainly due to the bias of the dataset toward negative cases and images and bias of the BLSTM layer in capturing the dependency between negative cases.

4. CONCLUSIONS

Automated fracture detection in cervical spine CT images is a very challenging task. In this paper, we proposed a machine learning model based on ResNet-50+BLSTM layer, which demonstrates capability of deep neural networks to address this challenge. We encourage the research community to work actively on this problem and we are in the process of releasing our large labeled dataset for research purposes.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was approved by our Institutional Review Board with a waiver for informed consent.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Stewart B Dunsker, Michael Zhang, Lily Kim, Robin Cheong, Ben Cohen-Wang, Katie Shpanskaya, Jessica Wetstone, Nidhi Manoj, Pranav Rajpurkar, and Kristen Yeom, "Deep-learning artificial intelligence model for automated detection of cervical spine fracture on computed tomography (ct) imaging," in *Journal of Neurosurgery*, 2019, vol. 131.

[2] John H Bland and Dallas R Boushey, "Anatomy and physiology of the cervical spine," in *Seminars in arthritis and rheumatism*. Elsevier, 1990, vol. 20, pp. 1–20.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] Hojjat Salehinejad, Shahrokh Valaee, Tim Dowdell, and Joseph Barfett, "Image augmentation using radial transform for training deep neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3016–3020.

[5] Hojjat Salehinejad, Sumeya Naqvi, Errol Colak, Joseph Barfett, and Shahrokh Valaee, "Cylindrical transform: 3d semantic segmentation of kidneys with limited annotated images," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 539–543.

[6] Hojjat Salehinejad, Errol Colak, Tim Dowdell, Joseph Barfett, and Shahrokh Valaee, "Synthesizing chest x-ray pathology for training deep convolutional neural networks," *IEEE transactions on medical imaging*, vol. 38, no. 5, pp. 1197–1206, 2018.

[7] Anil Kalra, "Developing fe human models from medical images," in *Basic Finite Element Method as Applied to Injury Biomechanics*, pp. 389–415. Elsevier, 2018.

[8] Sunil L Bangare, Amruta Dubal, Pallavi S Bangare, and ST Patil, "Reviewing otsu's method for image thresholding," *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 21777–21783, 2015.

[9] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016.

[10] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[11] Anna-Lena Robinson, Anders Möller, Yohan Robinson, and Claes Olerud, "C2 fracture subtypes, incidence, and treatment allocation change with age: a retrospective cohort study of 233 consecutive cases," *BioMed research international*, vol. 2017, 2017.

[12] Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee, "Recent advances in recurrent neural networks," *arXiv preprint arXiv:1801.01078*, 2017.

[13] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.