Cervical Spine Fracture Detection through Two-stage Approach of Mask Segmentation and Windowing based on Convolutional Neural Network

Doyeon Kim†, Xujia Ning†, Kaicheng Liang†, Yi Ni†, Duan Wang†, Mingyuan Li†, Yichuan Wang†,
Erick Purwanto*, Ka Lok Man
Xi'an Jiaotong-Liverpool University
Suzhou, China
e-mail: {D.Kim19, Xujia.Ning21, Kaicheng.Liang21, Yi.Ni21, Duan.Wang21, Mingyuan.Li21,
Yichuan.Wang21}@student.xjtlu.edu.cn,
{Erick.Purwanto, Ka.Man}@xjtlu.edu.cn

***Abstract*—Neck pain may be caused by cervical bone fracture, which must be promptly detected and treated, as severe cases can lead to paralysis or even death. The diagnostic precision of radiologists in the identification of cervical spine fractures depends on the clinical manifestation of the patient. Current fracture detection accuracy among radiologists stands at only 73.98% for alert blunt traumatic patients. To address this concern, this paper presents an approach based on deep learning models that can quickly analyze CT scans and diagnose cervical spine fracture. The approach includes two stages: Stage 1 utilizes UNet-EfficientNet for CT image segmentation, while Stage 2 incorporates CrackNet-LSTM to achieve spinal injury detection. Notably, the models excel in accurately identifying fractures. Implementing these strategies with the aforementioned models yields impressive results: 99.91% accuracy for Stage 1, 94.9% accuracy for Stage 2, and a combined accuracy of 94.9% for the overall examination process. This approach significantly improves the accuracy and the efficiency, thus proving to be highly qualified in assisting radiologists and alleviating their workload in detecting cervical spine fractures.**

Keywords*—*cervical spine; fracture detection; computer vision; semantic segmentation

I. Introduction

An injury such as break and dislocation in the neck bone is known as cervical spine fracture, which occurs with bimodal age distribution: ages between 15 and 24 years, and patients over 55 years [1]. The cervical spine fracture is problematic on several occasions, given the potential for severe consequences such as paralysis or fatality, and thus treatments such as removal surgery are urgently required to address the nerve damage caused by bone fragments. However, certain fragments of cervical spine fractures are unnoticeable, thus resulting in additional complications such as malunion and delayed union.

Computerized Tomography (CT) is a general modality employed during clinical trials for the purpose of medical image examination. This imaging technique is utilized to acquire highly detailed cross-sectional images, which enables the visualization of the internal anatomical structures. Cervical spine fractures are diagnosed by analyzing the CT scans. The diagnostic precision of radiologists in the identification of cervical spine fractures depends on the clinical manifestation of the patient. When a patient presents with obvious symptoms or the fractures are of an acute nature, the diagnostic accuracy of radiologists reaches 95% [2]. However, the accuracy decreases to only 80% when assessing blunt trauma patients [3]. Due to the inherent limitation of human capability, enhancing the diagnostic accuracy of radiologists is a formidable challenge. Consequently, an efficacious cervical spine fracture detection technique is needed to rapidly identify the fractures.

Deep learning has been employed in numerous tasks within the medical and healthcare field, including fall detection, physiological signal-based analysis, medical image analysis [4][5][6]. Deep neural networks such as Convolutional Neural Networks and Transformers have achieved breakthroughs in computer vision owing to their ability of extracting features from the training data in semantic segmentation and image classification.

This paper aims to achieve vertebrae images analysis for cervical spine fracture detection that is laborious by radiologists, and thus help for expeditious treatment and reliable results. To achieve this aim, we built a two-stage deep learning model to detect cervical spine fracture from CT scans. This work focuses exploring effective models for developing new procedures in medical image detection or enhancing existing ones.

II. Related works

A. Semantic Segmentation

Krizhevsky et al. [7] designed AlexNet, an image classification model that experienced rapid development. Semantic segmentation is perceived as an advanced task of image classification since it annotates an image pixel-to-pixel. Using Fully Convolutional Network (FCN), Long et al. [8] achieved a breakthrough and become predominant in semantic segmentation tasks. Thereafter, numerous works were built

† Indicates equal contribution.

** Represents the corresponding author.*

around FCN and had major development. U-Net, proposed by Ronneberger et al. [9] is extended from fully convolutional network that has the ability to learn locational information, and it is able to excel in various medical image segmentation tasks even based on very few training images. However, there is a trade-off between localization accuracy and the use of context. In addition, it is computationally demanding. DeepLab, proposed by Chen et al. [10], applies AtrousSpatial Pyramid Pooling (ASPP) to capture contextual information at different scales. EfficientNet proposed by Tan and Le [11] is a carefully-balanced Convolutional Neural Network, which achieves both higher accuracy and better efficiency than other ConvNets. In particular, EfficientNet is a scaled-up neural network architecture, where the models scale all dimensions with a compound coefficient, which is a newly proposed method known as compound scaling.

B. Cervical Spine Fracture Detection

Though deep learning has been used in various medical image analysis tasks, only a limited number of techniques have been developed for diagnosing fractures in cervical spine CT scans. Dunsker et al. [12] first used 3D ResNet-101 DCNN (Deep Convolutional Neural Network) to obtain an accuracy of 0.85 and 0.82, respectively, on the Area Under the Receiver Operating Characteristics and Area Under Precision-Recall Curve metrics at the case level. However, the scarcity of publicly available annotated cervical spine data, totaling only 1212 cases, poses a significant obstacle to the advancement of automated cervical spine fracture identification. To address this, Salehinejad et al. [13] utilized a combination of 3D ResNet-50 DCNN and BiLSTM on a dataset consisting of 2,937 normal and 729 fracture cases, totaling 3,666 cases in total. The proposed framework achieved a performance of 79.18% after training on the dataset. Although it resolved the issue of the limited dataset, the three approaches have not yet addressed the interpretation of selection bias. This motivates leveraging a Convolutional Neural Network to manage this bias by extrapolating the dataset. Despite the structure's performance reaching up to 92%, it still misses severe fracture-dislocations.

An alternative approach is to adopt transformers in the classification procedure. ViT (Vision Transformer) is a state-of-the-art deep learning architecture that applies the Transformer model to image classification tasks. Based on which Chłąd and Ogiela [14] built a model that detects spinal bone fracture, and achieved an accuracy of 98%.

III. METHODOLOGY

A. Dataset

The experiment used data obtained from the RSNA Competition held in 2022, specifically the Cervical Spine Fracture Detection dataset on Kaggle [15]. The dataset contains 2019 CT scans from patients in 9 different nations. Patient information is protected by encrypting the Hospital Number, which is known only by the principal investigator. The paper obtained permission from the Kaggle Competition rules as an individual.

The data is incorporated in three parts. The first part consists of backbone CT images in three different planes: axial, sagittal, and coronal. These images are in DICOM (Digital Imaging and Communications in Medicine) format and contain detailed patient information, such as patient ID and slice thickness. While the focus of the dataset is on the cervical spine (C1 to C7), it also includes some thoracic labels (T1 to T12). The metadata for the train and test sets are stored in a .csv file and include the Study Instance UID, study ID, patient level outcome, and single vertebrae fracture target information. The second part of the data retrieval involves the segmentation ground truth, which includes the identification of fractures and numbering of neck bones. This segmentation is performed using a 3D UNet model and validated by radiologists. Pixel level annotations are provided for a subset of the training set. It is important to note that these segmentation files are in the sagittal plane, and simply retrieving the images would not be sufficient for training. The last part of the retrieved data includes the locations of the bounding boxes for the fractures. The dataset is divided into an 80:20 ratio, with 80% used for training and 20% for testing. The separation occurs randomly during the segmentation process, while the classification process is executed sequentially.

B. CT Image Preprocessing

The proposed optimized model combines a UNet + EfficientNet for preprocessing with a CNN + LSTM model for classifying CT vertebrae images. Prior to proceeding into Stage 1, a windowing preprocessing technique is applied, as depicted in Fig. 1. Windowing involves rescaling CT image slices into Hounsfield Units, which enables the measurement of radiodensity [16]. According to Parin and Thitirat [16], HU scale is accomplished with the following equation:

$$HU = (rescale\_slope * CT\_number) + rescale\_intercept \quad (1)$$

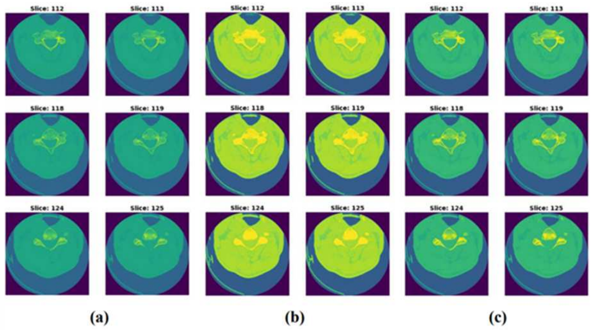


Fig. 1. Windowing preprocessed images with three Hounsfield regions: (a) head and neck temporal bones 1, (b) head and neck temporal bones 2, and (c) spine bone.

Based on the HU scale, the selected regions are: (a) head and neck temporal bones 1, (b) head and neck temporal bones 2, and (c) spine bone as described by Baba [17]. Each image slice is windowed with the corresponding Hounsfield Units: (a) (w = 2800, l = 600), (b) (w = 4000, l = 700), and (c) (w = 1800, l = 400). The combination of these three units is applied to each RGB channel, resulting in a 3D input image that is prepared for training in Stage 1.

The experiment also involves image cropping. Specifically, Yolov5 model is used to extract the mask region. The model is trained to predict the segmentation section and draw a bounding box around it. By observing the maximum and minimum values for the segmentation image bounding boxes and cropping around the obtained location, the dataset with the vertebrae centered is created, as shown in Fig. 2.

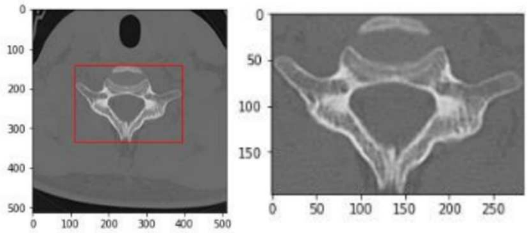


Fig. 2. Windowing Pre-cropping Images with the Bounding Boxes Generated around Segmentation Images.

The first step is to convert the windowed image from the .npz file to a .jpg file. Then, using the provided masked dataset, the bounding box position around the segmentation mask is extracted and the position information is stored. Finally, the pictures are predicted along with their respective locations. In addition, voxel cropping is also utilized. A voxel represents a volume component in the patient. Since the fracture information is unknown for each image but can be identified through the vertebrae, images of individual cervical bones are compiled to form each spine. However, the number of collections may not be consistent. Thus, an average of 30 slices, which represents the average slice for each vertebra, is selected and combined.

C. 3D U-Net + EfficientNet-B5 for Cervical Spine Segmentation

When the cervical spine needs to be segmented continuously from a series of CT image slices, 3D image segmentation algorithms outperform the 2D models because of smooth identification across the sequence. This implies that 3D image segmentation is considered the most effective technique. However, building a very deep neural network from scratch is challenging due to issues such as disappearing gradients, high computational costs, and large GPU memory usage. To overcome these challenges, this research article proposes using windowed-3-channeled 3D images, which incorporate partial 3D information, serving as a compromise between accuracy and computing resources.

The Stage 1 prototype consists of the 3D UNet and EfficientNet models, which are specifically used to detect vertebrae. The architecture of this model can be seen in Fig. 3. It is trained on a dataset that has been preprocessed and randomly transformed using the 3D windowing technique. The input data is 3-channels, and the axial ground truth image is used for training. However, the axial dataset does not include the cervical label, so the spine bones cannot be categorized into each vertebra. To overcome this problem, a multilabel segmentation approach is introduced. A total of 9 channels of ground truth are determined based on the types of vertebrae. The 0th label represents the background of the mask image, labels 1 to 7 indicate the cervical spine from C1 to C7, and label 8 represents the thoracic bones from T1 to T12. The combined 9-channels ground truth, along with the windowed 3D input, is trained using the proposed U-Net-EfficientNet model.

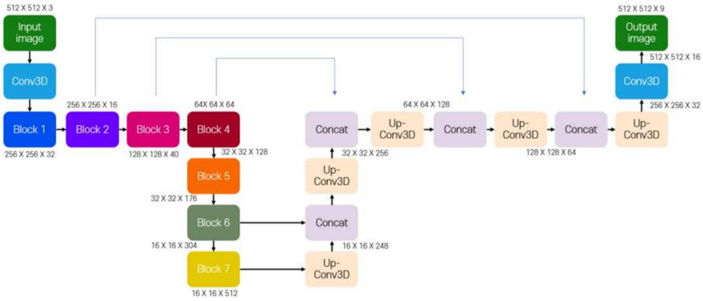


Fig. 3. 3D UNet-EfficientNet B5 Model Architecture for Semantic Segmentation.

The Stage 1 segmentation model utilizes the EffUNet architecture, with the EfficientNet-B5 serving as the encoder backbone. In particular, EfficientNet is a scaled-up neural network architecture, where the models scale all dimensions with a compound coefficient. The model makes use of pretrained ImageNet weights. The architecture follows the design of a typical U-Net, consisting of a contracting path and an expansive path [18]. Notably, the pooling operators in the contracting path are replaced with up sampling operators. The encoder backbone, EfficientNet-B5, is represented by blocks 1 to 7 in the architecture ae seen in Fig. 4.

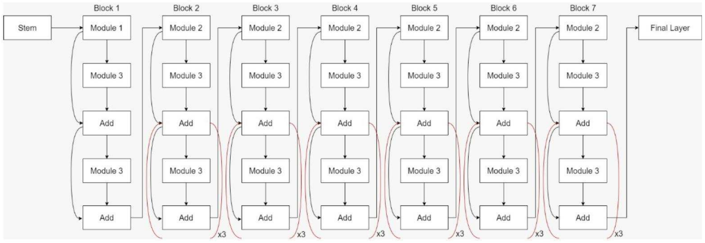


Fig. 4. The Total Architecture of EfficientNet-B5 [19].

D. CrackNet + BiLSTM for Cervical Spine Fracture Classification

Since there are strong correlation between adjacent bone injuries, it is necessary to sequentially detect cervical spine fractures for classification purposes [20]. Therefore, it is recommended to combine the cropped voxel images of each vertebra to create a comprehensive representation of the entire cervical bones. However, this approach is challenging due to limitations in GPU memory. To address this issue, Yolov5 is used to crop each windowed image, focusing on the vertebrae of interest. However, the dataset also contains other features such as dislocations. To detect these characteristics, it is crucial to avoid cropping each individual image. Instead, bounding boxes are recorded for each neck bone and clipped based on the maximum distances found. This method also helps avoid image stretching and enables the machine to learn about the corresponding cervical spine accurately.

CrackNet is a paradigm for models that can be tailored to specific datasets and application scenarios [21]. It offers advantages such as a lightweight model with the ability to adapt to different application circumstances and a small number of parameters. The network architecture is illustrated in Fig. 5. CrackNet functions as an image feature extractor, converting raw data into numerical features. The extracted image information is then fed into the LSTM model, enabling the machine to learn about the entire vertebrae.

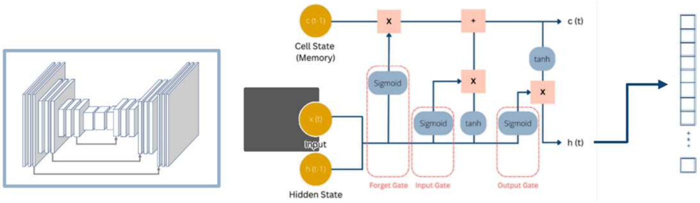


Fig. 5. CrackNet-LSTM Model Architecture.

The Stage 2 classification model is a fusion of two models: CrackNet and LSTM. CrackNet is a hierarchical feature aggregation model that incorporates ResNet and UNet. It helps in obtaining the final semantic segmentation outcomes. UNet is used for feature extraction in this model. The encoder backbone of ResNet-101 is initialized with CrackNet. Similar to the segmentation model, the pretrained encoder weights from ImageNet are included. The input channel is set to 30 due to cropping voxel. After segmentation, the class number is 7, representing vertebras. The two-layered LSTM model with bidirectional enabled in the architecture uses information sequences from ResUNet, and helps in resolving the vanishing gradient problem and learns temporal long-term dependencies. Finally, the output is linearized to a single dimension, and a linearization process is applied for single fracture prediction.

IV. Experiment

A. *Parameter Selection and Training Computational Time*

Although most of the images in the dataset have a size of 512 × 512 pixels, there are some datasets with different sizes. In order to validate the segmentation result, the image is rescaled to be 512 × 512 for the entire dataset. In both stages, AdamW optimizer and cosine annealing learning rate scheduler are adopted. The input channel of Stage 1 UNet-EfficientNet is 3, and the output channel is 9. Augmentations such as random horizontal flip, vertical flip, and shift scale rotation are also performed. Additionally, one of the following augmentations is selected: grid distortion, optical distortion, or elastic transform. Lastly, a random coarse dropout is applied. Using the aforementioned setup, the proposed model was trained using randomly initialized weights and tested using a 5-fold-cross-validation scheme.

In Stage 2, the training image is augmented to be 512 × 512. The Yolov5 clipped dataset significantly reduces the image size by half or more. The input channel of Stage 2 CrackNet-LSTM is 30, and the output channel is 1. Augmentations such as random horizontal flip, vertical flip, transpose, random brightness, and shift scale rotation are also performed. Additionally, one of the following augmentations is selected: motion blur, median blur, Gaussian blur, or Gaussian noise. A choice is made between grid distortion and optical distortion, and the image is sharpened for improved feature extraction.

B. *Classification Metrics*

To measure the loss performance value, the project utilizes both BCE logits and IoU coefficient. This is done because of the ground truth segmentation and multilabel conversion in conjunction with the binary imaging approach. Due to the presence of a large false positive rate from a limited number of labeled images, the IoU (Jaccard index) is used to balance them.

Their formulas are shown below:

1) *BCE logits coefficient*

The loss function in this project combines a Sigmoid layer and BCELoss into a single class. This approach is more stable in numerical computations compared to using a simple Sigmoid layer followed by BCELoss, and log-sum-exp method is used during training to improve numerical stability [22].

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (2)$$

2) *The Jaccard index*

The Jaccard index or Jaccard coefficient is used to measure the size of the union of two label sets divided by the spatial overlap of the intersection [23], as shown in (3).

$$J = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

3) *Dice coefficient*

The dice coefficient measures the overlap between two binary images. As stated by Kelly et al. [24], the coefficient values range from 0 indicating no overlap to 1 representing complete agreement. It is also used to detect false positives when comparing predicted and actual values, and the equation can be derived using (4).

$$DICE = \frac{2|A \cap B|}{2|A| + |B|} \quad (4)$$

C. *Experimental Results*

2.5D preprocessing involves combining n neighboring slices to create consecutive two-dimensional slices. In this study, 3 sections of images were selected, with the RGB channels being filled with the previous slice (n-1), the current slice (n), and the next slice (n+1). The results can be seen in Fig. 6.

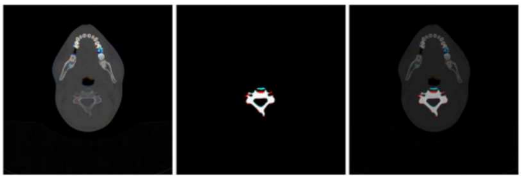


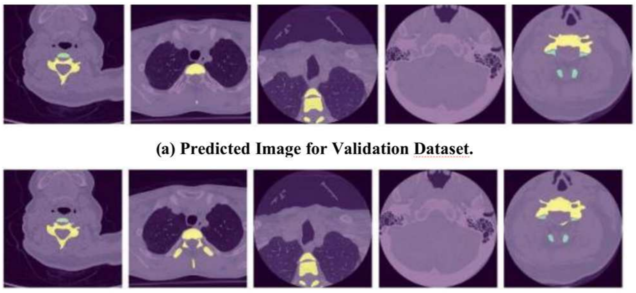
Fig. 6. 2.5D Preprocessed Images.

The EfficientNet-UNet model was evaluated on a CT image dataset and underwent numerous tests. The proposed model achieved a DICE score of 100%, which is a significant improvement compared to other frameworks. Additionally, the validation accuracy for Stage 1 was 99.91%, as shown in Table I. The model performs similarly to the equivalent model with 2.5D preprocessing, with only a 1.9% difference. These results demonstrate the effectiveness of the suggested framework, especially when compared to the 3D ResNet-101 prototype, which has a noticeable difference of 17.91%.

TABLE I. STAGE 1 MODEL PERFORMANCE COMPARISON

Stage 1 Model	Accuracy	Dice precision
Unet-EfficientNet with Windowing	0.999	1.000
Unet-EfficientNet with 2.5D	0.980	1.000
3D ResNet-101 [25]	0.82	0.52
EfficientNet-V2	0.953	-

Using the EfficientNet encoder backbone and UNet decoder model, Fig. 7 showcases five samples of axial cervical spine images along with their corresponding ground-truth masks and predicted segmentation. The prediction demonstrates that UNet-EfficientNet successfully captures the majority of attributes present in diverse cervical spines without producing any false positives. In this experiment, the likelihood of a cervical spine fracture is solely based on the spatial features contained in each axial image, without considering temporal information. This results in a segmentation accuracy of 99.94% for the training phase and 99.91% for the validation phase.


(a) Predicted Image for Validation Dataset.

(b) Ground Truth of Validation Dataset.

Fig. 7. Five Samples of Axial Cervical Spine Images with Corresponding Ground Truth Masks and Predicted Segmentation Image.

Table II presents the performance results of stage 2 models using different preprocessing techniques. The dataset with and without Yolov5 shows almost no difference, with a mean similarity of 94.9% and 94.3% with the ground truth, and a loss of 0.21 and 0.20, respectively. However, using ResNet-50 and the no cropping voxel method resulted in a significant reduction in accuracy, with only 40%. Additionally, the loss did not converge during the image studies.

TABLE II. STAGE 2 MODEL PERFORMANCE COMPARISON

Stage 2 Model	Loss	Accuracy
CrackNet-LSTM without YOLOv5	0.21±0.01	0.943
CrackNet-LSTM with YOLOv5	0.20±0.01	0.949
YOLOv5	0.64	0.94
ResNet-50 without Cropping Voxel	1.6	0.4
Simple CNN-LSTM	0.32	0.87

Finally, the combination of the two stages, UNet-EfficientNet and CrackNet-LSTM, along with windowing, Yolov5, and Cropping Voxel preprocessing methods, resulted in an accuracy of 94.9%. Additionally, the results from Table III indicate that the model's performance is significantly higher compared to other prototypes.

TABLE III. STAGE 1 AND 2 COMBINATIONAL MODEL PERFORMANCE COMPARISON

Stage 1+2 Model	Accuracy
Unet-EfficientNet + CrackNet-LSTM	0.949
Unet-EfficientNet + YOLOv5	0.94
Unet-EfficientNet + ResNet-50 no cropping voxel	0.4
Unet-EfficientNet + Simple CNN-LSTM	0.87
AlexNet + LSTM [25]	0.71
ResNet + LSTM [25]	0.71
VGGNet + LSTM [25]	0.84
CNN + SVM [26]	0.70
ResNet-50 + BLSTM-256 [27]	0.792

V. CONCLUSION

In conclusion, detecting fractures in cervical spine CT scans accurately is a difficult task to accomplish automatically. Nevertheless, the results of the experiments, both in terms of quantity and quality, clearly demonstrate that the newly proposed method surpasses other traditional methods in terms of accuracy, correct error segmentation and classification, as well as detection performance. Throughout the process, the deep neural networks used for image segmentation and classification have achieved results that are on par with multiple techniques, showcasing their capability to tackle the challenges of the project. Real medical data, with file extensions including DICOM and NIFTI, was employed to test the suggested algorithm. In the preparation phase, the data underwent preliminary processing steps such as windowing, cropping the image with Yolov5, and voxel clipping. Evaluation was done using similarity metrics like BCE logits coefficient, Jaccard index, Dice coefficient, and accuracy, which yielded results of 0.20 ± 0.01 for the two loss coefficients, 100% for dice coefficient, and 94.9% for accuracy.

ACKNOWLEDGMENT

This work is partially supported by the XJTLU AI University Research Centre and Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation at XJTLU. Also, it is partially funded by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (SZS2022004) as well as funding: SURF-2023-0123, XJTLU-RDF-22-01-062, XJTLU-REF-21-01-002 and XJTLU Key Program Special Fund (KSF-A-17).

References

[1] R. M. Marcon, A. F. Cristante, W. J. Teixeira, D. K. Narasaki, R. P. Oliveira, and T. E. P. d. Barros, “Fractures of the cervical spine,” Clinics, vol. 68, pp. 1455–1461, 2013.

[2] J.E. Small, P. Osler, A.B. Paul, and M. Kunst, “CT Cervical Spine Fracture Detection Using a Convolutional Neural Network,” in Am J Neuroradiol, vol. 42, pp. 1341-1347.

[3] M. Hussain and G. Javed, “Diagnostic Accuracy of Clinical Examination in Cervical Spine Injuries in Awake and Alert Blunt Trauma Patients,” in Asian Spine J., vol. 5, no. 1, pp. 10-14, Accessed Apr. 30, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047893/

[4] S. Yu, Y. Chai, H. Chen, R. A. Brown, S. J. Sherman, and J. F. Nunamaker Jr, “Fall detection with wearable sensors: A hierarchical attention-based convolutional neural network approach,” Journal of Management Information Systems, vol. 38, no. 4, pp. 1095–1121, 2021.

[5] R. Raman, J. Bhalani, B. Sumathy, S. Bothe, A. Gehlot, and M. K. Chakravarthi, “Review of deep learning for physiological signal-based healthcare applications,” in 2022 5th International Conference on Contemporary Computing and Informatics (IC3I). IEEE, 2022, pp. 1688–1693.

[6] B. Pandey, D. K. Pandey, B. P. Mishra, and W. Rhmann, “A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions,” Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 8, pp. 5083–5099, 2022.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Advances in neural information pro-cessing systems, vol. 25, 2012.

[8] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[9] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234–241.

[10] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[11] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in International conference on machine learning. PMLR, 2019, pp. 6105–6114.

[12] Stewart B Dunsker, Michael Zhang, Lily Kim, Robin Cheong, Ben Cohen-Wang, Katie Shpanskaya, et al., “Deep-learning artificial intelligence model for automated detection of cervical spine fracture on computed tomography (CT) imaging”, in Journal of Neurosurgery, vol. 131, 2019.

[13] H. Salehinejad, E. Ho, H. M. Lin, P. Crivellaro, et al., (13-16 Apr. 2021). Deep Sequential Learning For Cervical Spine Fracture Detection On Computed Tomography Imaging. Presented at 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France. [Online]. doi: 10.1109/ISBI48211.2021.9434126.

[14] P. Ch lad and M. R. Ogiela, “Deep learning and cloud-based computation for cervical spine fracture detection system,” Electronics, vol. 12, no. 9, p. 2056, 2023.

[15] Adam Flanders, 2022, “RSNA 2022 Cervical Spine Fracture Detection,” Radiological Society of North America (RSNA), Kaggle. [Online]. Available: https://kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection.

[16] P. Kittipongdaja and T. Siriborvornratanakul, “Automatic kidney segmentation using 2.5D ResUNet and 2.5D DenseUNet for malignant potential analysis in complex renal cyst based on CT images,” Journal on Image and Video Processing, vol. 2022, no. 5, pp. 1-15.

[17] Y. Baba, “Windowing (CT),” radiopaedia.org. https://radiopaedia.org/articles/windowing-ct (accessed Nov. 14).

[18] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv:1505.04597v1 [cs.CV], May 2015.

[19] T. Hussain et al., “CNN Based Detection of the Severity of Diabetic Retinopathy from the Fundus Photography using EfficientNet-B5,” in 2020 11th IEEE Annual information Technology, Electronics and Mobile Communication Conference (IEMCON). 2020, pp. 147-150.

[20] Y. Ma, Y. Luo, “Bone fracture detection through the two-stage system of Crack-Sensitive Convolutional Neural Network,” Informatics in Medicine Unlocked 22, vol. 22, no. 100452, Oct. 2020.

[21] K. H. Zou et al., “Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index1: scientific reports,” National Institutes of Health, vol. 11, no. 2, pp. 178-189, Mar. 2006.

[22] Pytorch, “BCEWithLogitsLoss,” pytorch.org. https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html (accessed May. 6, 2023)

[23] E. Trucco et al., “Validation,” Computational retail Image Analysis, E. Trucco, T. MacGillivray and Y. Xu, UK: American Press, 2019, ch. 9, pp. 157-170.

[24] K. H. Zou et al., “Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index1: scientific reports,” National Institutes of Health, vol. 11, no. 2, pp. 178-189, Mar. 2006.

[25] I. Shahzadi, Y. B. Tang, F. Meriadeau and A. Quyyum, “CNN-LSTM: Cascaded Framework For Brain Tumour Classification,” in 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018, pp 633-637.

[26] V. Kalbhor, S. Mishra, Y. Walke and Prof. M. Pawar, “Bone Fracture Detection using CNN and SVM,” International Journal of Innovative Research in Technology, vol. 8, no. 3, pp. 152183-492–152183-496, Aug. 2021.

[27] H. Salehinejad et al., “DEEP SEQUENTIAL LEARNING FOR CERVICAL SPINE FRACTURE DETECTION ON COMPUTED TOMOGRAPHY IMAGING,” arXiv:2010.13336v4 [eess.IV], Feb. 2021.