Classification of Cervical Spine Fracture using Deep Learning

Arunesh Tiwari[1], Swapnil Singh[2] , Adarsh Pandey[3], Brijendra Pratap Singh[4], Dinesh Kumar[5] and Dharmendra Kumar[6]

[1] KIET Group of Institutions, Ghaziabad, India
[2] KIET Group of Institutions, Ghaziabad, India
[3] KIET Group of Institutions, Ghaziabad, India
[4] KIET Group of Institutions, Ghaziabad, India
[5] KIET Group of Institutions, Ghaziabad, India
[6] KIET Group of Institutions, Ghaziabad, India

aruneshtiwari186@gmail.com
swapnilsingh.cms@gmail.com
adarshpandeyy00@gmail.com
brijendra.2024cse1066@kiet.edu
dinesh.kumar@kiet.edu
dharmendra.kumar@kiet.edu

Abstract. For many years cervical spine fractures and dislocations are major cause of paralysis and death is some cases. This makes it important for the proper diagnosis and treatment of these injuries to reduce fatal injuries. There are many methods to do this, in this paper we present a computer-aided cervical spine injury diagnosis model that uses deep learning approaches like AlexNet and GoogleNet. The proposed model can be used by doctors for faster identification of cervical injuries. To train our model we have used 772 CS fractures and 707 normal images. The model came up with the accuracy of 99.67 percent, which is higher than accuracy of radiologists; we have also used saliency maps to check degree of instance for a given class. This paper has both clinical and research-based applications.

Keywords: Medical image analysis, Computer-aided diagnosis, Artificial intelligence in healthcare, Deep Learning.

1 Introduction

The first seven vertebrae situated below the skull and above the thoracic spine make up the cervical spine (CS). Figure 1 shows that the CS has two groups: the axis and atlas (C1 and C2) and the sub axial cervical vertebrae (C3- C7). High CS injury rates harm health and longevity [4,5,6].

According to Indian data, a preformed proforma was used to analyze demographics, epidemiological data, and neurological status of consecutive SCI cases admitted between January 2000 and December 2008. Out of 2716 cases, 1,400 were cervical and 1,316 were thoracolumbar. The male-to-female ratio was 4.2:1, and 71% were between 20 and 49. Vehicle collisions caused 28% of injuries, while falls from height caused 53%.[7]

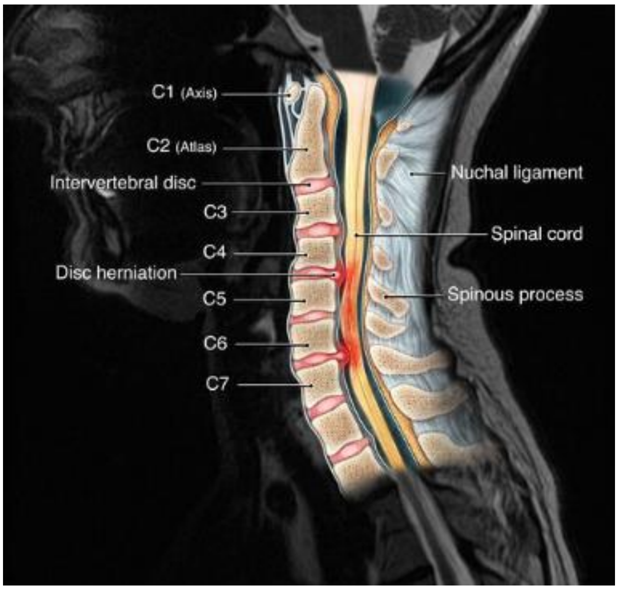

Fig. 1. Lateral View of Cervical Spine

We noticed that visual identification of CS is laborious and risky task. Eighty percent of emergency department diagnostic errors are caused by radiologists who do not possess expert knowledge or are overworked [11]. Understanding these circumstances, we are proposing a deep learning model to help doctors to easily interpret CS X-rays in order to minimize human error.

In modern times, artificial intelligence is widely being used in medicine for treatment of various diseases like cancer, injured organs etc. but very few methods exist for classification of cervical spine fractures and dislocations using deep learning algorithms. In this paper we try to solve this problem.

Convolutional neural networks detected CS fractures with 92% accuracy (95% CI, 90-94%), 76% sensitivity, and 97% specificity. We studied various sources and came to know that average radiologist accuracy lies between 90% to 95%. Radiologists and convolutional neural networks missed similar fractures. The anterior osteophytes, transverse processes and spinous processes were fractured, as was the lower part of cervical spine, which CT beam attenuation obscures.

In this paper, authors proposed a CNN based model to detect cervical spine fractures and dislocations. However, CS dislocation may be fatal [6]. Unfortunately, the studies' datasets are unavailable.

With an accuracy of 99.6% our model efficiently classifies CS X-ray images into normal, fracture, and dislocation.

The proposed model works, requires little setup, and can run on PCs or cheap embedded systems.

2 PROCEDURE

2.1 Trends in Recent Years

In recent years deep learning algorithms are useful for analysis of image, spectral and textual data efficiently. Artificial Intelligence makes it easier to diagnose medical images with the help of various deep learning algorithms. These algorithms can efficiently analyze data and recognize unique features. To recognize digital images each digital image is a 2D array of values each value represents greyscale code 0 to 255. These pixel values are passed through convolutional layer and pooling layers then fed to dense layers, during this process weights are updated based on degree of dissimilarity between output and true label. In the following sections we will discuss methods used in this study.

2.2 Deep Neural Networks

CNN or ConvNet is popular deep learning architecture, it is capable of learning from input data without need of human intervention for feature extraction. Figure shows example of basic CNN model including convolution, pooling, and fully connected layers. Unlike traditional neural networks, CNN can deal with overfitting problem by using dropout. CNNs are used to deal with 2D shapes and are useful in medical image analysis, image segmentation and many more. This makes CNN more powerful than traditional neural networks. There are various CNN based models such as Visual Geometry Group (VGG), AlexNET, Xception, Inception, ResNet; these models are useful in various applications.

There are various methods to apply deep learning models, transfer learning is one of them. In transfer learning we use a already learned model to a new problem set, this is done when there is lack of training data. We can train a model with previous saved weights using transfer learning approach on large datasets. In this way a pre-trained model can be improved using transfer learning

In our study we used AlexNet and GoogleNet weights in our model to run dataset we were able to correctly distinguish between three states (dislocation, fracture, and normal). We used transfer learning method to train our model and then made predictions out of it. AlexNet model is trained on ImageNet dataset. Three fully connected and five convolutional layers make up the total of eight layers in the AlexNet model. This model connects the final three convolutional layers to the fully connected layers, and the initial pair of convolutional layers to the overlapping max-pooling layers. Rectified Linear Units (ReLu) are utilized at each output of a convolutional layer and fully connected layer for nonlinear functions. The final output layer is linked to a SoftMax activation layer, which produces a category from a set of 1,000 class labels. AlexNet model is trained over 15 million images. The AlexNet model's complete architecture is seen in Figure 2.

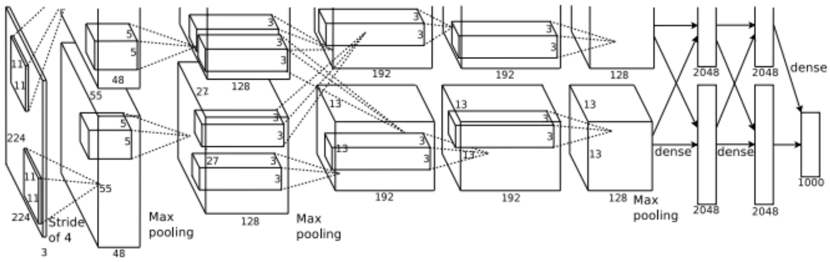

Fig. 2. AlexNet Architecture [14]

With a minimal link between activations, the GoogleNet architecture was built. Put otherwise, pruning algorithms will not allow all 512 output channels to be linked to all 512 input channels. As a result, GoogleNet is an initial module that roughly approximates a conventionally dense, sparse CNN. Convolutions of different sizes (5 × 5, 3 × 3, 1 × 1) are used by GoogleNet to gather data at different scales (Figure 3).

type	**patch size/ stride**	**output size**	**depth**	#1×1	**#3×3 reduce**	#3×3	**#5×5 reduce**	#5×5	**pool proj**	**params**	**ops**
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig. 3. GoogleNet incarnation of the Inception architecture [13]

2.3 Proposed Algorithm

In this section we will understand the characterization process for cervical spine (dislocation, fracture and normal) images. This part has some limitations due to resemblance of different condition (dislocation, fracture and normal) images in dataset. Due to this a precise computer aided model is very important for the successful completion of this task.

We conducted image preprocessing on the dataset. Images in dataset are in grayscale color code and have various dimensions (width and heights). AlexNet and GoogleNet use three input channels corresponding to red, green and blue color codes, input dimensions for GoogleNet is (224 X 224) and AlexNet is (227 X 227).

We performed image processing in two steps. First all images are resized to conform to the input layer dimensions of AlexNet and GoogleNet. Second, Original image is duplicated three times for input channels (Red, Green and blue).

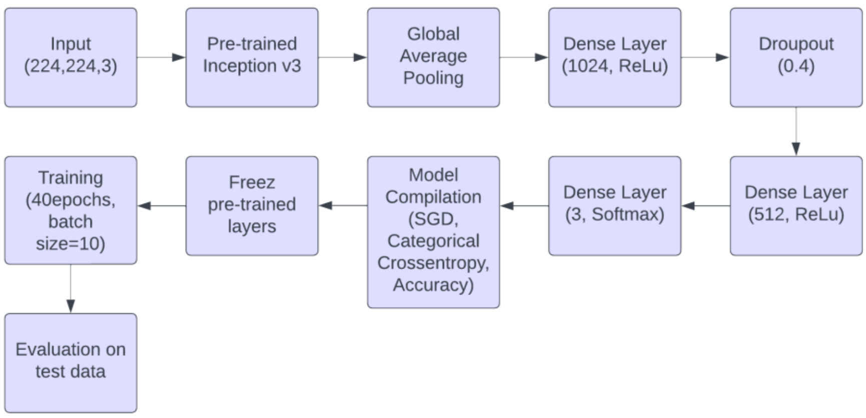

Fig. 4. Diagram of proposed method

We have used pre-trained transfer learning model to extract important information from the dataset images by identifying key parts. The construction of our model involved the consecutive stacking of numerous convolutional neural networks (CNNs). We have created a model by utilizing two pre-configured model AlexNet and GoogleNet, this is a multidimensional version of logistic function. After dropout with regularization, we have replaced bottom layer of the model by two fully connected layers this help us to combine the information derived from the previous layers to generate output. We have used a softmax layer to convert a vector of K-real values transformed into a probability distribution with number of possible outcomes. At the end a softmax layer is used to normalize the output.

2.4 Dataset

Accessible at Kaggle, this dataset offers a valuable collection of medical imaging data dedicated to cervical spine fractures; The dataset is divided into three groups these are 530 CS dislocation images (Row 1), 772 CS fracture images (Row 2). and 707 normal images (Row 3) as shown in Figure 5.

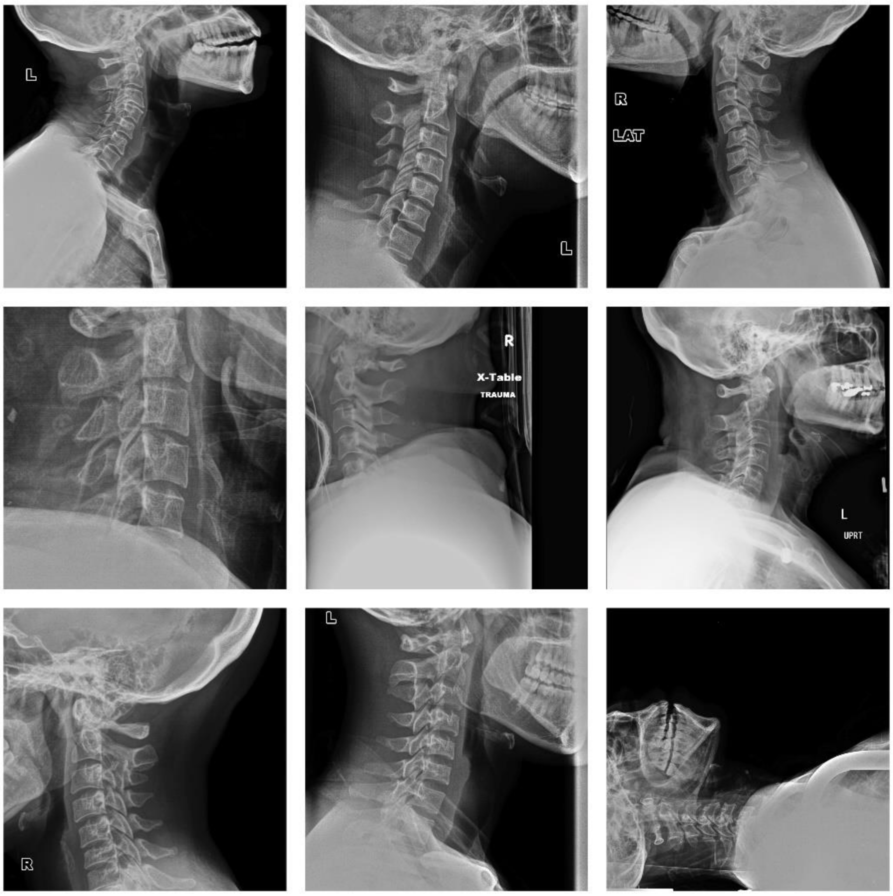

Fig. 5. Sample images of different groups from dataset.

This dataset plays a pivotal role in our study, enabling us to examine and understand the complexities associated with cervical spine fractures, their diagnosis, treatment, and outcomes. By utilizing this dedicated dataset, we aim to contribute significantly to the body of knowledge surrounding cervical spine fractures, ultimately advancing medical research and improving patient care in this critical area of healthcare.

2.5 Performance Analysis

Our proposed model was trained on 70%,15% and 15% of dataset for training, validation and testing respectively. We have trained our model as per the above data distribution and we have assessed model's performance based on accuracy, sensitivity, precision,F1-score and specificity, and Receiver Operating Characteristic curve, as per below given equations:

Table 1. List of abbreviations

Abbreviation	Definition
fp	False Positive
fn	False Negative
tp	True Positive
tn	True Negative

$$Accuracy = (tp + tn) / (tp + fp + fn + tn)$$

$$Sensitivity = tp / (tp + fn)$$

$$Specificity = tn / (fp + tn)$$

$$Precision = tp / (tp + fp)$$

$$F1 - score = (2 * tp) / (2 * tp + fp + fn)$$

3 Results

We have trained our proposed model for 30 iterations with each batch containing 10 samples. To update model's parameters and minimize training error we have used Stochastic Gradient Descent (SGD) with learning rate of 0.001, we have used Cross Entropy Loss Function (CCELF).

To enhance the convergence of the optimizer and minimize errors, various techniques can be employed to update deep learning parameters beyond the use of SGD alone. One widely adopted strategy involves gradually decreasing the learning rate. In this a higher learning rate is utilized for the initial iterations, in following iterations lower rates are used. An annealing method, integrated with SGD, expedites optimizer convergence, facilitating the attainment of minimal errors. We have used scheduled annealing to mitigate the risk of getting trapped in the vicinity of local minima and saddle points, supporting the achievement of the global optimum solution through convergence. The scheduled annealing directly modulates noise. We have implemented a reduction learning rate i.e. the size of the steps taken during the optimization process every four epochs, this balances computational efficiency with starting learning rate.

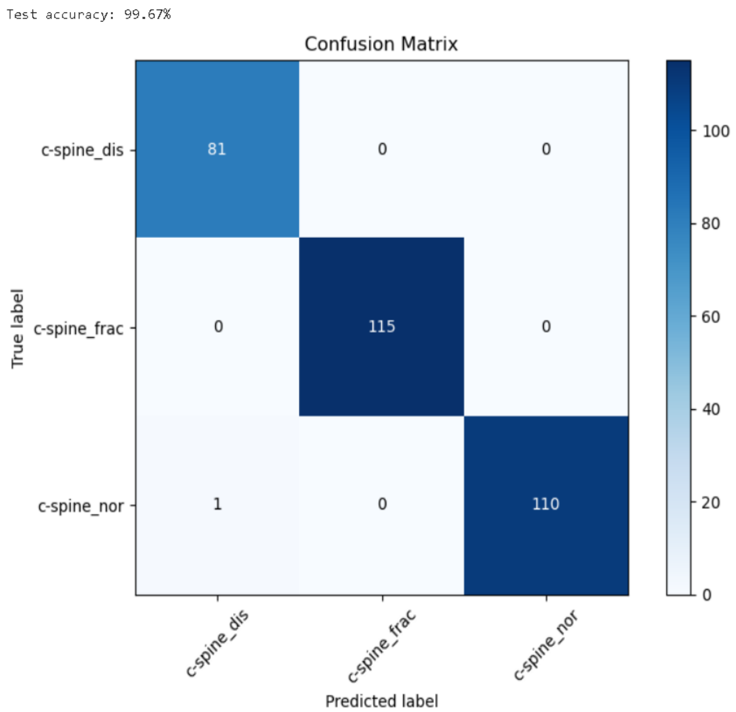

Fig. 6. Confusion Matrix

In our study, we implemented a LR reduction every four epochs to balance computational efficiency with a higher initial LR. This strategy establishes a scheduled annealing framework within the Stochastic Gradient Descent optimization process.

We performed comparison between our proposed model and radiologists. We came to know that radiologists has accuracy of 92%. Figure 6 shows the receiver operating characteristic (ROC) for various methods.

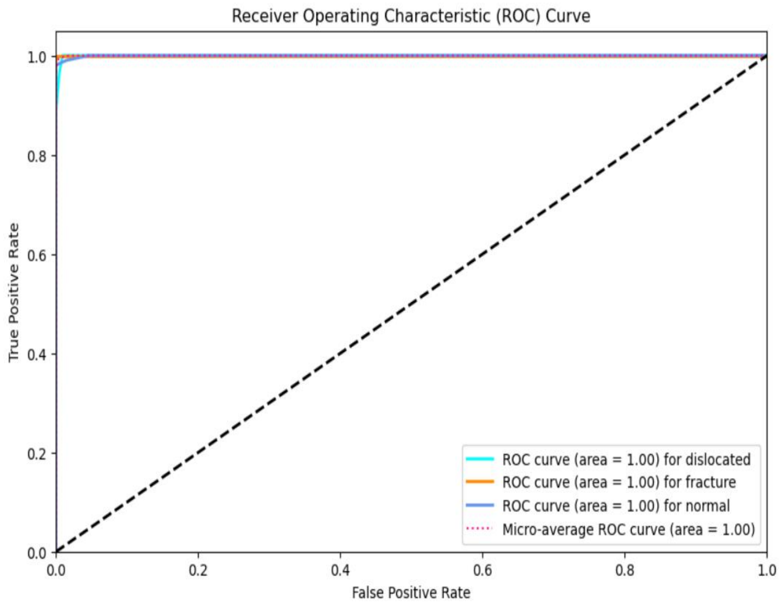

Fig. 7. ROC Curve

By the Figure 5 we can clearly see a comprehensive comparison between the proposed method and the radiologist. We have used an improved GoogleNet for evaluating CS fractures. We have studied various materials and research papers on this and came up with refined GoogleNet model with an accuracy of 99.34%. We have exceeded radiologists' accuracy rate of 92%.

Analyzing Table 1 further reveals that the least favorable measurements were associated with the proposed method utilizing AlexNet. AlexNet's limitations, including a restricted number of layers and filters per convolutional layer, as well as direct serial connections between layers, make it susceptible to overfitting. Consequently, AlexNet struggles with classifying images within categories containing fewer instances. On the contrary, the proposed method based on GoogleNet outperformed both AlexNet and the radiologist. We analyzed GoogleNet's architecture which features diverse convolutional layers with varying filter sizes and intricate interconnections, making it to accurately extract features while working with a relatively limited dataset.

To assess the effectiveness of our proposed model in cervical spine injuries (normal, fracture and dislocation), we acquired 80 unlabeled cervical spine X-ray images from the radiology center. We asked radiologist to perform classification of these unlabeled images. Radiologist identified 65 as normal, 9 as fractured and 6 dislocated. Fractured images were identified through the presence of fracture patterns, disruption of bone integrity, vertebral collapse, and other fracture-related features. Dislocated images were characterized by the disturbance in spinal alignment and facet joint alignment.

These images serve as representative examples, illustrating the diversity within each category and showcasing the challenges and variations present in clinical X-ray data. The orthopedic surgeons' expertise ensured a robust and accurate classification of these images, forming the basis for the evaluation of the proposed method's capability in distinguishing between normal, dislocated, and fractured CS X-ray images.

We have used our proposed model for classification of X-rays and the results are shown in the below table.

Table 2. Comparison of proposed model and radiologists

Method	No. of correctly classified	No. of wrongly classified
Proposed Method	78	2
Radiologists	71	9

As depicted in Table 2, the radiologist achieved a correct identification rate of 71 out of 80 images. The misclassifications involved two standard X-ray scans, erroneously categorized as fractures and dislocations. The lower diagnostic accuracy can be attributed to the inherent variability in medical images, variations in imaging parameters between the training dataset and real clinical cases, and the limitations imposed by small datasets. Enhancing diagnostic precision could be achieved by incorporating more extensive databases and employing multiparametric CT.

However, our study acknowledges certain limitations. The documentation of the deep convolutional neural networks (DCNNs) was limited to images and related diagnostic outcomes without detailed feature descriptions, constituting a significant drawback. The "black box" mechanism inherent in DCNNs processing medical images presents a contradiction, as the logic employed by these networks differs from both human and conventional computer logic processes. While our investigation yielded promising results, the specific features utilized remain unknown.

In addressing these limitations, a saliency map was employed in our study to identify key areas within the images, focusing the analysis on those regions. By rotating the images and applying the proposed method, we successfully highlighted the region of interest (ROI) corresponding to the spine, as illustrated in Figure 15. This approach enhances interpretability by revealing the specific areas contributing to the network's decision-making process, providing valuable insights into the model's functioning, and aiding in addressing the "black box" nature of DCNNs.

4 Conclusion

In our experiment accuracy of our model reached 99.67%. We came to the conclusion that our proposed method is capable of more quickly predicting cervical spine fractures than radiologists with more accuracy. This shows high potential of deep learning in medical industry. By using this method orthopedic surgeons and radiologists will be able to diagnose CS spine injuries much faster. This method has potential for future studies in medical applications of deep learning. This will help patients by reducing diagnose time.

In our study we proposed a deep learning method for diagnosing cervical spine fracture injuries. We worked on a dataset having three class labels i.e. normal, dislocation and fracture. Deep Learning Algorithms have high potential in rapid and accurately diagnosing medical injuries. We proposed a solution for cervical spine fracture injuries, this will reduce diagnose time and improve overall quality of medical services. The result of our proposed method achieved:

Precision: 0.9878, Recall: 1.0000, F1 Score: 0.9939 for Dislocation
Precision: 1.0000, Recall: 1.0000, F1 Score: 1.0000 for Fracture and
Precision: 1.0000, Recall: 0.9910, F1 Score: 0.9955 for Normal
with an accuracy of 0.9967. From medical perspective our proposed method has future development in medical field to classify cervical spine injuries more specifically.

The proposed model works well with real time data in detection of fractures of dislocated X-ray images of cervical spine. The method has high accuracy, but it is to be noted that any type of clinical decision should be based on clinical examination under doctor supervision. Our proposed model has some limitations due to potential selection bias and limited generalizability of our findings. Dataset is taken from a single source. In future more broader data source could be used for model training for more practical applicability. Dataset has high number of fracture images this may have increased sensitivity of the model.

References

1. Joshi, D., & Singh, T. P. (2020, January 1). A survey of fracture detection techniques in bone X-ray images. Artificial Intelligence Review, 53(6), 4475–4517.
2. Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., & Potter, H. (2018, October 22). Deep neural network improves fracture detection by clinicians. Proceedings of the National Academy of Sciences, 115(45), 11591–11596.
3. Kalmet, P. H. S., Sanduleanu, S., Primakov, S., Wu, G., Jochems, A., Refaee, T., Ibrahim, A., Hulst, L. V., Lambin, P., & Poeze, M. (2020, February 5). Deep learning in fracture detection: a narrative review. Acta Orthopaedica, 91(3), 362–362.

4. Guan, B., Zhang, G., Yao, J., Wang, X., & Wang, M. (2020, January). Arm fracture detection in X-rays based on improved deep convolutional neural network. Computers & Electrical Engineering, 81, 106530.
5. Voter, A., Larson, M., Garrett, J., & Yu, J. P. (2021, June 11). Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures. American Journal of Neuroradiology, 42(8), 1550–1556.
6. Wang, J., Eltorai, A. E., DePasse, J. M., Durand, W., Reid, D., & Daniels, A. H. (2018, June). Variability in Treatment for Patients with Cervical Spine Fracture and Dislocation: An Analysis of 107,152 Patients. World Neurosurgery, 114, e151–e157.
7. Nishida, N., Tripathi, S., Mumtaz, M., Kelkar, A., Kumaran, Y., Sakai, T., & Goel, V. K. (2022, August). Soft Tissue Injury in Cervical Spine Is a Risk Factor for Intersegmental Instability: A Finite Element Analysis. World Neurosurgery, 164, e358–e366.
8. Roth, H. R., Wang, Y., Yao, J., Lu, L., Burns, J. E., & Summers, R. M. (2016, March). Deep convolutional networks for automated detection of posterior-element fractures on spine CT. In Medical imaging 2016: computer-aided diagnosis (Vol. 9785, pp. 165-171). SPIE.
9. Murata, K., Endo, K., Aihara, T., Suzuki, H., Sawaji, Y., Matsuoka, Y., Nishimura, H., Takamatsu, T., Konishi, T., Maekawa, A., Yamauchi, H., Kanazawa, K., Endo, H., Tsuji, H., Inoue, S., Fukushima, N., Kikuchi, H., Sato, H., & Yamamoto, K. (2020, November 18). Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. Scientific Reports, 10(1).
10. Suzani, A., Seitel, A., Liu, Y., Fels, S., Rohling, R. N., & Abolmaesumi, P. (2015). Fast automatic vertebrae detection and localization in pathological ct scans-a deep learning approach. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 678-686). Springer International Publishing.
11. AlGhaithi, A., & Al Maskari, S. (2021). Artificial intelligence application in bone fracture detection. Journal of Musculoskeletal Surgery and Research, 5(1), 4.
12. Salehinejad, Ho, Lin, Crivellaro, & Samorodova. (2021, February 5). DEEP SEQUENTIAL LEARNING FOR CERVICAL SPINE FRACTURE DETECTION ON COMPUTED TOMOGRAPHY IMAGING. IEEE International Symposium on Biomedical Imaging (ISBI) 2021.
13. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
14. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May 24). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90.