Utilizing InceptionV3 for Categorizing Cervical Spine Fractures and Assessing Accuracy Against a Convolutional Neural Network

Kaviya V. H and P. V. Parimala
Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, 602105, India

Keywords: Cervical Spine Fracture, Convolutional Neural Network, Deep Learning, Health, Novel InceptionV3, CT.

Abstract: The intent of this study is to compare the accuracy of Novel InceptionV3 and Convolutional Neural Networks in detecting cervical spine fractures from CT images. The two groups of learning models proposed in this study are the Novel InceptionV3 deep learning model and the Convolutional Neural Network (CNN). Cervical fracture is the dataset taken for the analysis which is obtained from the open source Kaggle repository with a sample size of 4200 CT images. In which 3800 images were given to train the model and 400 images to evaluate the model. With the value of G power = 0.8 with 95% confidence interval the experiment is iterated tenfold. The classification accuracy yielded by the proposed algorithm Novel InceptionV3 is 94.56% while CNN obtained an accuracy of 77.32%. The T-test ($p<0.001$, two tailed) shows that Novel InceptionV3 appears to have more significance than CNN. Conclusion: The study investigated the performance of two deep learning models in predicting cervical spine fractures with higher accuracy. The outcome indicates that Novel InceptionV3 is more effective in comparison with convolutional neural networks.

1 INTRODUCTION

Cervical fractures if untreated can result in lifelong paralysis, which is highly dangerous and potentially fatal (Campagnolo et al. 2011). In the US, there are an estimated 12,000 new cases of cervical spine injuries each year, 42% of which are caused by automobile accidents. Of these cervical spine injuries, sports account for roughly 8% (Belval 2015). In North America, injuries to the cervical spine result in more than a million visits to emergency departments annually (Milby et al. 2008). The initial step in treating them is finding the fracture (Dambhare and Kumar 2022). Therefore, an automated cervical spine fracture detection system is highly important for early diagnosis in today’s world. The applications of deep learning have made incredible progress in the early diagnosis of Cervical fractures and are revolutionizing the healthcare industry and enabling clinicians to treat patients well using clinical data (Davenport and Kalakota 2019).

Around 1100 papers in Research gate and 600 Science direct articles in over the preceding five years have been published that are pertinent to the detection of cervical spine fractures. Various techniques were used to improve the model's performance. One in which Hojjat used an approach by equipping deep sequential learning techniques for identifying fractures on the cervical part of the spinal column on CT scans, with a 70.92% accuracy rate (Salehinejad et al. 2021a). Guillermo et al. proposed two deep learning models VGG16 and ResNet18 to precisely predict fractures on sagittal radiographic images. The accuracies obtained were 88%(ResNet18) and 84%(VGG16) (Rosenberg et al. 2022). A study to assess the accurate estimation and error rate analysis of a Deep neural network to discover the presence of fractures on the cervical spine was performed and attained an accuracy of 54.9% (Hodler, Kubik-Huch, and von Schulthess 2020; Voter et al. 2021).The ResNet 50 and Bidirectional Long Short-Term Memory (BLSTM) models were combined using an ensemble methodology by Hojjat and others, demonstrating the effectiveness of deep neural networking models in tackling this issue (Salehinejad et al. 2021b). Earlier algorithms were unable to identify severe fracture locations due to the dataset's imbalance. The most significant limitations of earlier studies that reduced the generalizability of the results were research design and selection bias, hence had an impact on classification performance. The intent of this research is to contrast the effectiveness and functionality of a CNN and a deep learning model called Novel InceptionV3 using CT to ascertain which model is preferable at classifying cervical spine fractures.

H., K. and Parimala, P.
Utilizing InceptionV3 for Categorizing Cervical Spine Fractures and Assessing Accuracy Against a Convolutional Neural Network.
DOI: 10.5220/0012543000003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 583-588
ISBN: 978-989-758-661-3

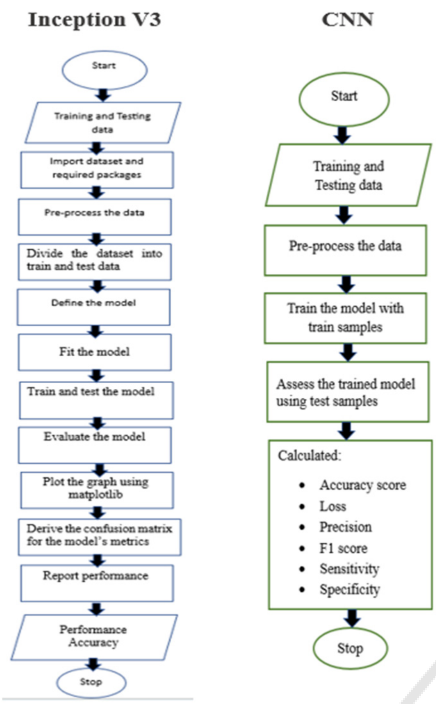

Figure 1: Displays the model's flowchart performed in Novel InceptionV3 and Convolutional Network.

2 METHODS & MATERIALS

The comparative analysis was performed in the Programming Laboratory of Saveetha School of Engineering, Saveetha Institute of medical and technical Sciences (SIMATS). Novel InceptionV3 and Convolutional Neural Network are the two neural networks taken for the analysis with size of (N=10) samples. Using the aforementioned techniques, the experiment was iterated over ten times. The G-power was established to be 0.8 with 95% confidence interval for the supplied data samples.

The HP Pavilion Laptop 14ec with Windows 11 Home had an AMD Ryzen 5 series processor, 8GB DDR4-3200 MHz RAM, a 64-bit OS, x64 based processor and AMD Radeon Graphics served as the procedure's infrastructure. Google Colaboratory served as the platform to train the model. Adam is the optimizer which was used to compile the model.

Table 1: Accuracy values for the groups: Inception V3 and CNN model.

S.NO	ACCURACY	
	Inception V3	CNN
1	74.95	72.08
2	92.80	73.3
3	94.38	77.50
4	96.09	74.17
5	96.17	76.25
6	96.81	79.58
7	97.59	77.08
8	97.59	82.92
9	97.24	79.58
10	97.97	80.83
Average	94.56	77.32

Table 2: Values obtained from the performance metrics on evaluating the models.

S.NO	METRICS	InceptionV3	CNN
1	Accuracy	94.5%	83.2%
2	Precision	100%	83%
4	F1-score	94.7%	83.2%
5	Sensitivity	90%	83.4%
6	Specificity	100%	83.08%

The Cervical fracture dataset used for the analysis was obtained from kaggle suggested by V3 and CNN model. (Sairam 2022) and consists of 4,200 cervical spine CT images sized 224x224. They were subdivided into fracture and normal samples. The model was trained on 80% of the image samples and evaluated on 20% of the image samples. The CT images of the train and test were then categorized into normal and fracture samples.

2.1 Inception

InceptionV3 is a transfer learning model for image analysis. This network is an improved version of the InceptionV1 model. There are 48 layers in total as shown in Fig. 1. It is more efficient, has deeper networks than the Inception V1 and V2 models, but its speed is not compromised. It is less expensive in terms of computation. It consists of Convolutional layers which are factored, smaller, and asymmetric to lower the computational efficiency. During training, an auxiliary classifier serves as a regularizer between layers, and the loss it incurs is added to the primary network loss. Feature maps are concatenated in parallel with a stride two convolution layer and a max-pooling layer.

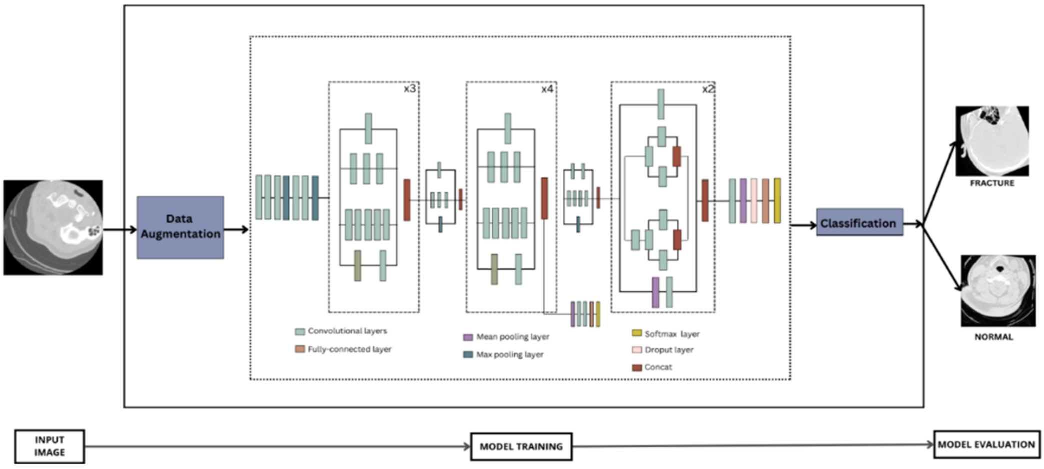

Figure 2: Architectural representation of the model InceptionV3.

2.2 Convolutional Neural Network

Convolutional Neural Network, or CNN has three layers: convolutional, pooling, and fully connected (FC).

Convolutional is the first layer to extract features and FC is the second which ensures every input of the input vector influences every output of the output vector. From the CL to the FC layer, CNN increases with complexity. With increasing complexity, the CNN can capture more intricate, larger portions of an image till discovering the complete object.

2.3 Statistical Analysis

The comparison was conducted using Statistical Package for the Social Sciences software 26. This tool was designed to undertake statistical analysis for the given data. In its stage of development, it offers a vast library of AI statistics, open-source scalability, and the ability to evaluate the mean accuracy of different algorithms (SPSS Software, n.d.). Adequacy is the dependent variable, and the accuracy of Novel InceptionV3 is the independent variable. The comparison of the two independent groups InceptionV3 and Convolutional Neural Network was done using the Independent Samples T-test to see if there is a proof that the means of the corresponding populations are significantly dissimilar.

Figure 3a demonstrates the accuracy of the model trained and validated of both the models using a line graph proving the accuracy of InceptionV3 (94.56%) is higher compared to CNN (77.32%). The training and validation losses of InceptionV3 and CNN are compared in Fig. 3b. From the confusion matrix of InceptionV3 in Fig. 4a &b and CNN, it can be seen that the values of True.

Figure 1: Flowchart Inception V3 and CNN. Positive and True Negatives are greater in InceptionV3 in contrast to CNN inferring that InceptionV3 has detected the cases more accurately than CNN. Fig. 4c demonstrates the mean accuracy and loss of InceptionV3 vs CNN with the group plotted on the X axis, and Y axis showing the mean accuracy and loss. Novel InceptionV3 seemed to have a higher accuracy in detecting cervical fractures using CT images compared to CNN.

3 DISCUSSION

Novel InceptionV3 obtained an accuracy of 97% while CNN attained an accuracy of 80%, proving that Novel InceptionV3 is much more accurate. With p=0.001, it is noted that Novel InceptionV3 functioned better than expected and was more effective in detecting cervical spine fractures.

According to research, the prevalence of undetected fractures in the spine lies from 19.5% - 45%. Pranata and others created two CNN-based models for calcaneal fracture classification using CT radiographic images. The included model is a potential tool for future usage in automated diagnosis with accuracy of 79%, and 72.9% of specificity (Pranata et al. 2019). In one study, a computer-aided technique was suggested for identifying fractures in calcaneus on Computed Tomography scan images. They opted for the Sanders fracture classification system, which makes use of color segmentation to identify and classify calcaneus fragments. The model has an accuracy of 86% (Zhang et al. 2018). Some preliminary studies have demonstrated CNNs are a suitable tool for fracture prediction on radiographic images. (Olczak et al. 2017) used a network trained on a variety of hand, wrist, and ankle radiographs to achieve an accuracy of 83% in fracture detection. (Kim and MacKinnon 2018) achieved an area under the curve (AUC) of 0.954 with a model trained on 1389 lateral wrist radiographs. According to studies, the prevalence of undiagnosed spine fractures ranges from 19.5% to 45%. (Muehlematter et al. 2019) was using lumbar and thoracic CT images to classify, detect, and locate vertebral spine fractures, as well as assess lumbar vertebral bone density. The accuracy of healthy/unhealthy vertebrae was poor, with an AUC of 0.5. The sensitivity for compression fracture identification and localization was 0.957, with a falsified rate of 0.29 per patient.

Table 3: Independent Sample T-Test is applied for the data set fixing confidence interval as 95% and Significance as p<0.001 (p<0.05) (2-tailed).

		Leven's Test of Equality of Variances		T-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std Error difference	Lower	Upper
Accuracy	Equal Variance assumed	0.06	0.03	7.17	18	<.001	15.901	2.215	11.24	20.58
	Equal Variance not assumed			7.17	16.34	<.001	15.901	2.215	11.21	20.58
Loss	Equal Variance assumed	0.302	0.589	-6.55	18	<.001	-3.153	0.4814	-4.164	-2.141
	Equal Variance not assumed			-6.55	15.94	<.001	-3.153	0.4814	-4.173	-2.132

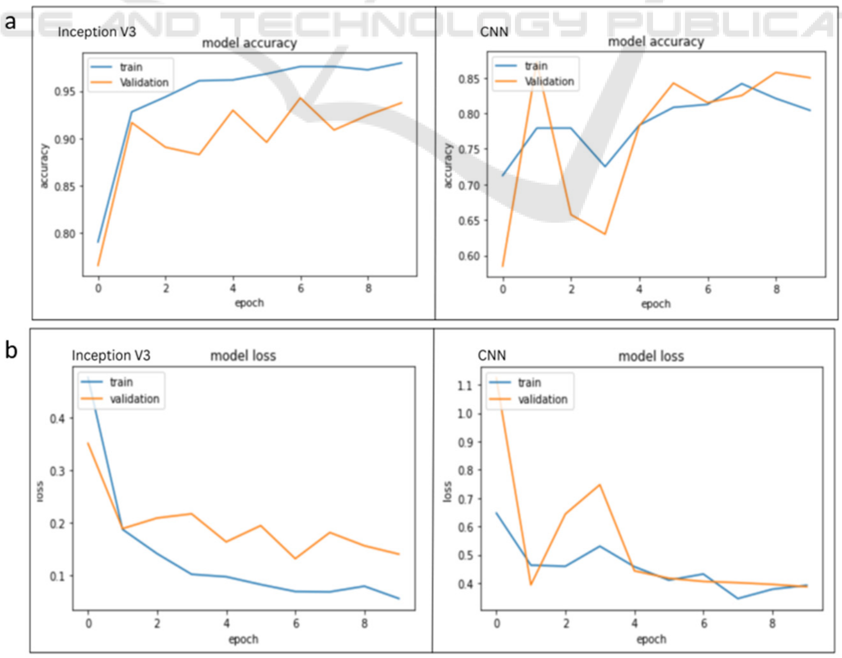

Figure 3: (a) Comparison of Training and Validation Accuracies of InceptionV3 and CNN (b) Comparison of Training and Validation Losses of InceptioNV3 and CNN.

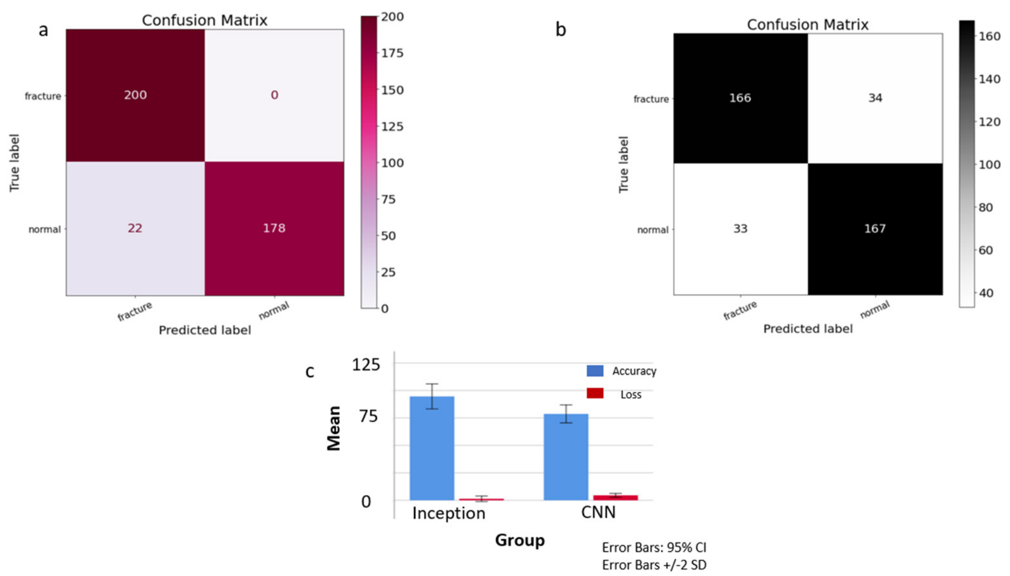

Figure 4: Visualization of the true and predicted cases using confusion matrix of (a) InceptionV3 (b)CNN (c) Bar chart representing the comparison of Mean accuracy of InceptionV3 & CNN for CT scans.

The amount of data consumed is indeed low when adopting models from different multi-layer neural networks, and the test dataset's modest size had an impact on classification accuracy when evaluating the model. Future research could be primarily focused on developing deep learning models that can fit in and get trained more quickly while utilizing smaller datasets. **Fig. 4c.** Bar chart representing the comparison of Mean accuracy of InceptionV3 and Convolutional neural network model in cervical spine fracture detection using CT scans. InceptionV3 appears to produce better results with standard deviation. X Axis: InceptionV3 vs Convolutional neural network (CNN) and Y axis: Mean Accuracy of detection SD = ±2 and confidence interval of 95%.

Table 4: Statistical computation of independent samples tested among InceptionV3 and CNN deep learning models. The mean accuracy of InceptionV3 is 94.5 and CNN is 77.329 Standard Deviation of InceptionV3 is 5.68 and CNN is 4.09. The T-test for comparison for InceptionV3standard error mean is 1.79 and CNN is 1.29.

	Group	N	Mean	Std. Deviation	Std. Mean Error
Accuracy	InceptionV3	10	94.56	5.687	1.7986
	CNN	10	77.329	4.091	1.293
Loss	InceptionV3	10	1.361	1.254	0.3967
	CNN	10	4.515	0.862	0.2725

4 CONCLUSIONS

The experimental finding demonstrates that the Novel InceptionV3 performs better than CNN, with an accuracy of 94.56% as opposed to the value of 77.32%. In terms of workflow effectiveness, the proposed approach has a wide range of potential applications in medical imaging. The analysis exhibits the applicability of deep learning models that can handle the issue of large datasets with improvised accuracy.

REFERENCES

Belval, Luke. (2015). “Cervical Spine Injury.” Korey Stringer Institute. March 4, 2015. https://ksi.uconn.edu/emergency-conditions/cervical-spine-injury/.

Campagnolo, Denise I., Steven Kirshblum, Mark S. Nash, Robert F. Heary, and Peter H. Gorman. (2011). Spinal Cord Medicine. Lippincott Williams & Wilkins.

Dambhare, Shruti, and Sanjay Kumar. (2022). “Machine Learning in Healthcare.” Machine Learning and Deep Learning in Efficacy Improvement of Healthcare Systems. https://doi.org/10.1201/9781003189053-1.

Davenport, Thomas, and Ravi Kalakota. (2019). “The Potential for Artificial Intelligence in Healthcare.” Future Healthcare Journal 6 (2): 94–98.

G. Ramkumar, R. Thandaiah Prabu, Ngangbam Phalguni Singh, U. Maheswaran, (2021), Experimental analysis of brain tumor detection system using Machine learning approach, Materials Today: Proceedings,ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2021.01.246.

Hodler, Juerg, Rahel A. Kubik-Huch, and Gustav K. von Schulthess. (2020). Diseases of the Brain, Head and Neck, Spine 2020–2023: Diagnostic Imaging. Springer Nature.

Kim, D. H., and T. MacKinnon. (2018). “Artificial Intelligence in Fracture Detection: Transfer Learning from Deep Convolutional Neural Networks.” Clinical Radiology 73 (5): 439–45.

Kumar M, M., Sivakumar, V. L., Devi V, S., Nagabhooshanam, N., &Thanappan, S. (2022). Investigation on Durability Behavior of Fiber Reinforced Concrete with Steel Slag/Bacteria beneath Diverse Exposure Conditions. Advances in Materials Science and Engineering, 2022.

Milby, Andrew H., Casey H. Halpern, Wensheng Guo, and Sherman C. Stein. (2008). “Prevalence of Cervical Spinal Injury in Trauma.” Neurosurgical Focus 25 (5): E10.

Muehlematter, Urs J., Manoj Mannil, Anton S. Becker, Kerstin N. Vokinger, Tim Finkenstaedt, Georg Osterhoff, Michael A. Fischer, and Roman Guggenberger. (2019). “Vertebral Body Insufficiency Fractures: Detection of Vertebrae at Risk on Standard CT Images Using Texture Analysis and Machine Learning.” European Radiology. https://doi.org/10.1007/s00330-018-5846-8.

Olczak, Jakub, Niklas Fahlberg, Atsuto Maki, Ali Sharif Razavian, Anthony Jilert, André Stark, Olof Sköldenberg, and Max Gordon. (2017). “Artificial Intelligence for Analyzing Orthopedic Trauma Radiographs.” Acta Orthopaedica 88 (6): 581–86.

Rosenberg, Guillermo Sánchez, Andrea Cina, Giuseppe Rosario Schiró, Pietro Domenico Giorgi, Boyko Gueorguiev, Mauro Alini, Peter Varga, Fabio Galbusera, and Enrico Gallazzi. (2022). “Artificial Intelligence Accurately Detects Traumatic Thoracolumbar Fractures on Sagittal Radiographs.” Medicina 58 (8). https://doi.org/10.3390/medicina58080998.

S. G and R. G, "Automated Breast Cancer Classification based on Modified Deep learning Convolutional Neural Network following Dual Segmentation," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1562-1569, doi: 10.1109/ICESC54411.2022.9885299.

Sairam, Vuppala Adithya. 2022. “Spine Fracture Prediction from C.T.” https://www.kaggle.com/vuppalaadithyasairam/spine-fracture-prediction-from-xrays.

Salehinejad, Hojjat, Edward Ho, Hui-Ming Lin, Priscila Crivellaro, Oleksandra Samorodova, Monica Tafur Arciniegas, Zamir Merali, et al. (2021)a. “Deep Sequential Learning For Cervical Spine Fracture Detection On Computed Tomography Imaging.” 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). https://doi.org/10.1109/isbi48211.2021.9434126.

Voter, A. F., M. E. Larson, J. W. Garrett, and J-P J. Yu. (2021). “Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures.” AJNR. American Journal of Neuroradiology, June. https://doi.org/10.3174/ajnr.A7179.

Yadollahi, Mahnaz, Shahram Paydar, Haleh Ghaem, Mohammad Ghorbani, Seyed Mohsen Mousavi, Ali Taheri Akerdi, Eimen Jalili, et al. (2016). “Epidemiology of Cervical Spine Fractures.” Trauma Monthly 21 (3): e33608.

Zhang, Zhirui, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen. (2018). “Bidirectional Generative Adversarial Networks for Neural Machine Translation.” Proceedings of the 22nd Conference on Computational Natural Language Learning. https://doi.org/10.18653/v1/k18-1019.