Radiology Report Generation of Spine MRI images Using Machine Learning Techniques

Srinivas P, Sivasai M, Anmol B, Harinath C, Venkata Sesha Reddy

Abstract: The automatic generation of radiology reports given medical radiographs has significant potential to operationally and improve clinical patient care. A number of prior works have focused on this problem, employing advanced methods from computer vision and natural language generation to produce readable reports but these are not verbose & were limited to only X-Rays. However, these works often fail to account for the nuances of the radiology domain, and the critical importance of clinical accuracy in the resulting generated reports. In some cases, for less-experienced radiologists and pathologists, especially those working in the rural areas where the quality of healthcare is relatively low, writing medical-imaging reports is demanding or on the other hand for experienced radiologists and pathologists, writing imaging reports can be tedious and time consuming. Plan is to programmatically extract the labels & the images (MRI/CT) & associated reports from the sourced web data link then apply Computer Vision models to understand the deeper objects/insights in the images & detailed report is generated using NLP.

Keywords: Image Processing, CV, NLP, CNN, Transfer Learning

1. Introduction

A critical task in radiology practice is the generation of a free-text description, or report, based on a clinical radiograph (e.g., a SPINE MRI/CT). Providing automated support for this task has the potential to ease clinical workflows and improve both the quality and standardization of care. However, this process poses significant technical challenges. Many traditional image captioning approaches are designed to produce far shorter and less complex pieces of text than radiology reports. Further, these approaches do not capitalize on the highly templated nature of radiology reports. Additionally, generic natural language generation (NLG) methods prioritize descriptive accuracy only as a byproduct of readability, whereas providing an accurate clinical description of the radiograph is the first priority of the report. Prior works in this domain have partially addressed these issues, but significant gaps remain towards producing high-quality reports with maximal clinical efficacy.

2. Problem Statement/Clinical Relevance

This work focuses on generating a clinically useful radiology report from a Spine X-ray/MRI image. This task has been explored multiple times, but directly trans-planting natural language generation techniques onto this task only guarantees the reports to look real rather than to predict right. A more immediate focus for the report generation task is thus to produce accurate disease profiles to power downstream tasks such as diagnosis and care providing. Our goal is then minding the language fluency while also increasing the clinical efficacy of the generated reports.

3. Literature Survey

Generating automated radiology reports from spine MRI images of different views is still a challenging task. With an advancement in machine learning algorithms in image processing and natural language processing have given a hope that accurate automated radiology reports will be possibility. Earlier work on radiology report generation is very limited and that too curtailed to specific issue or part of spine MRI images.

In [1], authors are automatically generated unified reports of lumbar spinal MRIs in the field of radiology, i.e., given an MRI of a lumbar spine. This work is focused on lumbar spines images only. They have achieved via a weakly supervised framework that combines deep learning and symbolic program synthesis theory to overcome four inevitable tasks: semantic segmentation, radiological classification, positional labeling, and structural captioning. This work is limited to lumbar spines only but other type of MRI images are not considered to train the model. In [3], authors have used NLP to extract the information from unstructured free text breast MRI images and tried to classify final assessment categories. In [5], authors have doe the work related to automatic image captioning with detailed captioning. To enable better and more detailed caption generations, they proposed a dense captioning architecture which first extracts and describes the objects of the image which in turn is helpful in generating dense and detailed image captions.

In [7], authors have focused to generate automated radiology reports for chest X-rays with below three step model. (1) build a multi-task learning framework which jointly performs the prediction of tags and the generation of paragraphs, (2) propose a co-attention mechanism to localize regions containing abnormalities and generate narrations for them, (3) develop a hierarchical LSTM model to generate long para- graphs.

Even other references also have focused specific disease predictions of spine, brain, breast MRI images than generating complete radiology report by feeding all images generated from spine MRI.

We have tried here to feed all possible views of spine MRI images like lumbar, cervical views into the model to get radiology report impression of the patient.

4. Data Collection

In this paper we will working on the real datasets getting generated from one of the private hospitals.

4.1 Sample Data

Fig:1: Sample Spinal images

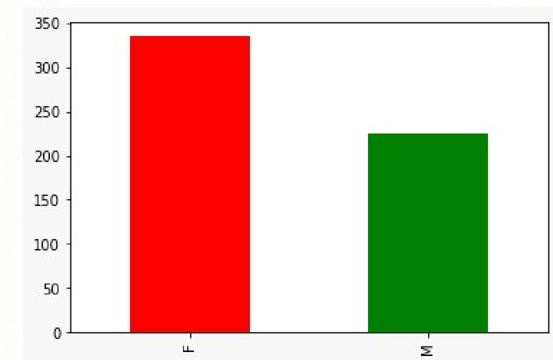
4.2 Data Processing

- All MRI images are grey scale images and analysis is done in those grey scale images only.
- Data cleansing has done to remove the special characters i.e.!()[]{};:'"\,<>/?@#$%^&*~
- Also added <Start> & <End> tags.
- Data has been Masked to protect the PII.
- Data augmentation is used to balance the dataset.

Data Split. Data has been split in **75:25** ratio i.e Overall 560 Images from the dataset out of which 420 has been used for training & 140 for the testing.

*Exploratory Data Analysis (EDA)***.** This is Medical Imaging dataset covering Spine related MRIs from one of the private hospitals which is covering across genders cervical & lumbar spines images. This data contains both Problematic & Non-Problematic patients. But data for Non-Problematic was limited so need to do the data augmentation to balance the dataset.

Please find the detailed EDA of the dataset below



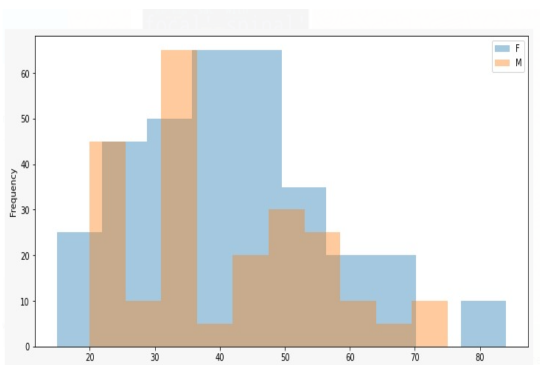
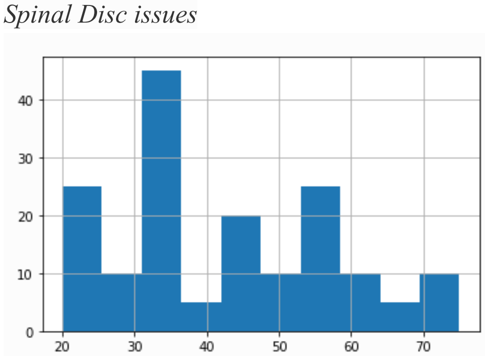


Fig. 2. Total Males and Females

Fig. 3. Gender wise Age distribution

Spinal Disc issues



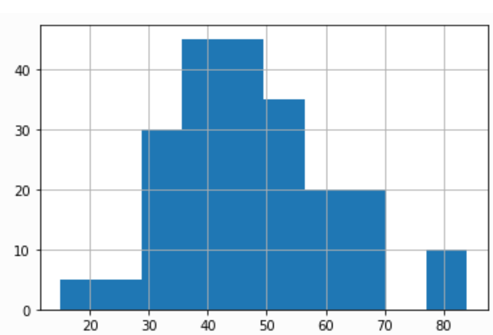


Fig. 4. Age Distribution of Males

Fig. 5. Age Distribution of Females

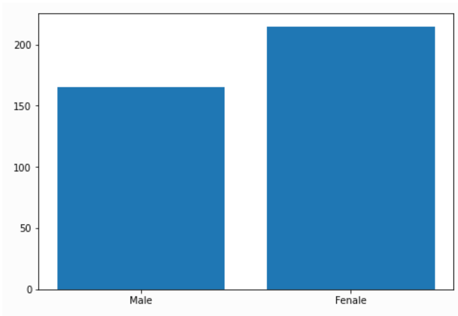


Fig. 6. Females are facing more spinal disc issues in our case

EDA on the Images

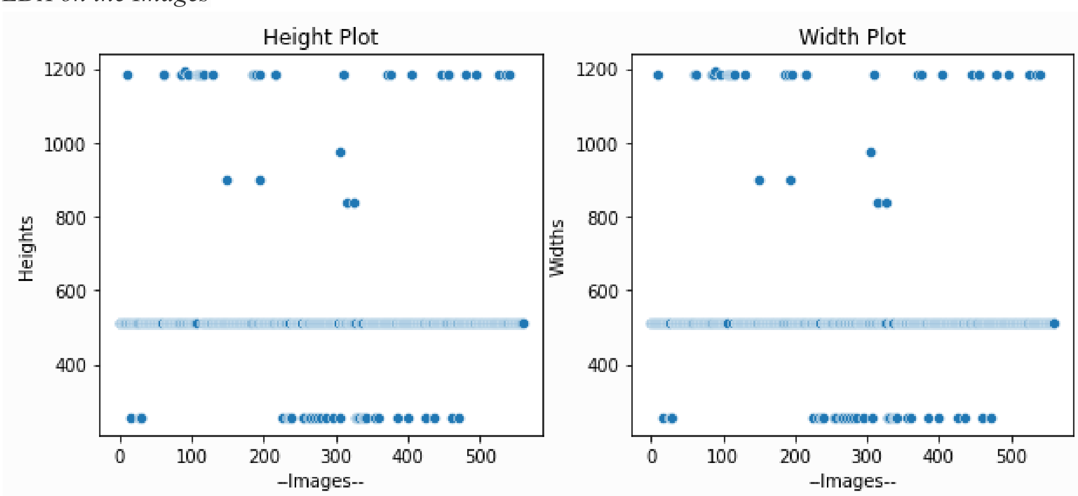


Fig. 7. Height and width of MRI images

EDA on the Diagnosis Reports/Impressions

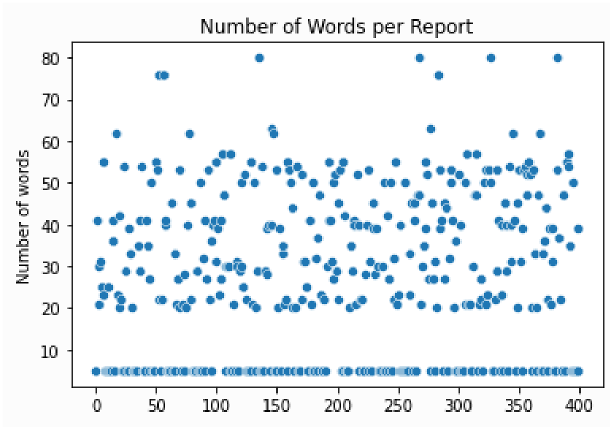


Fig. 8. Number of words per report

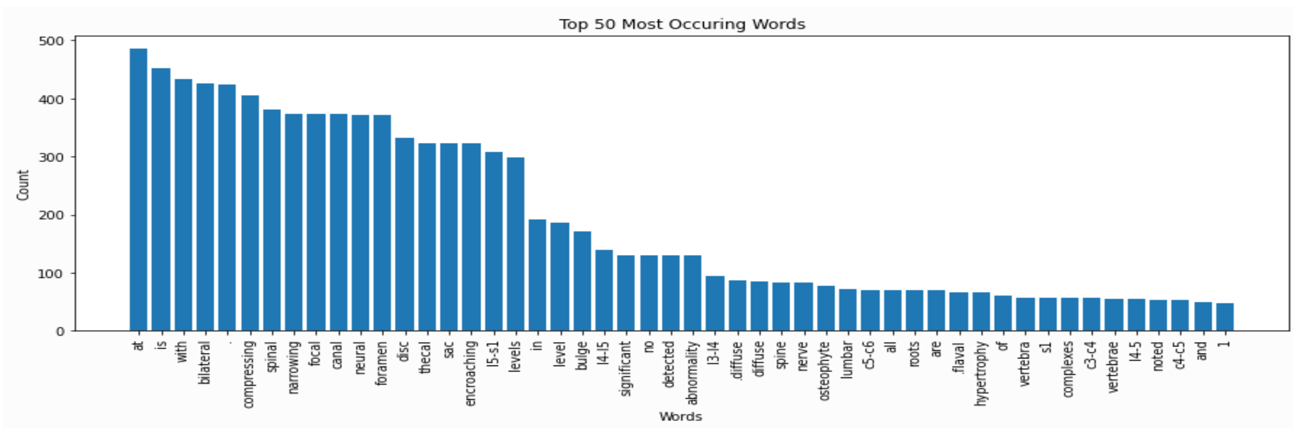


Fig. 9. Top 50 Most Occuring Words

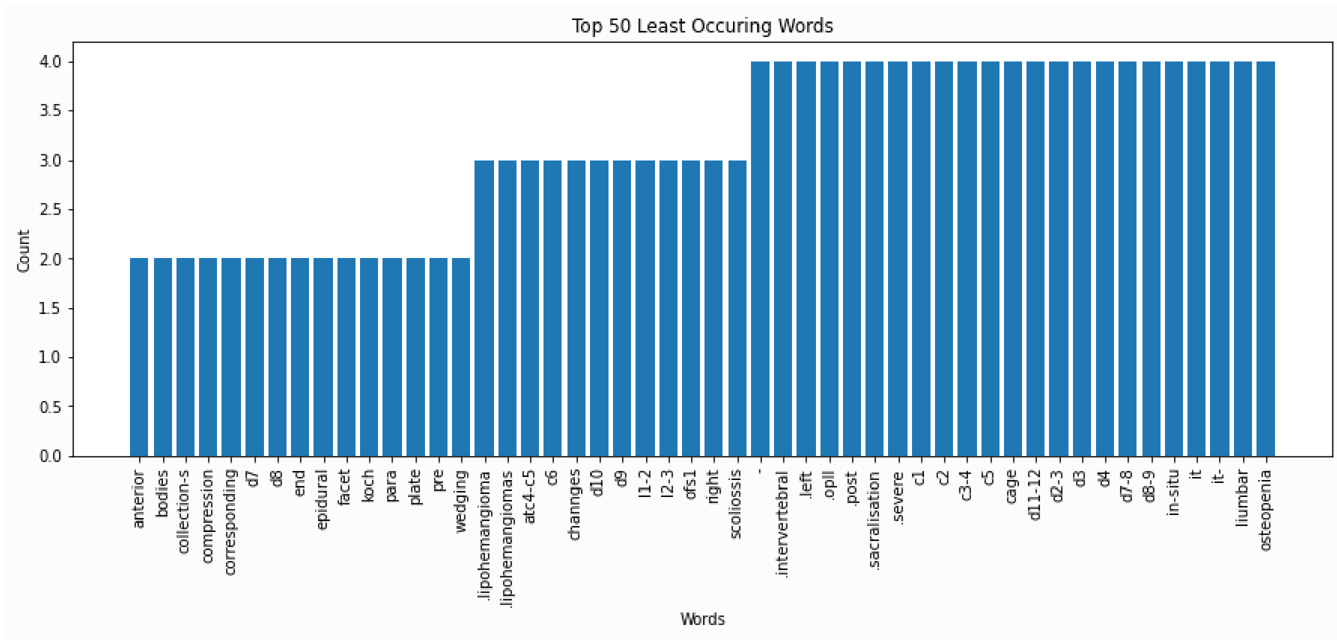


Fig. 10. Top 50 Least Occurring Words

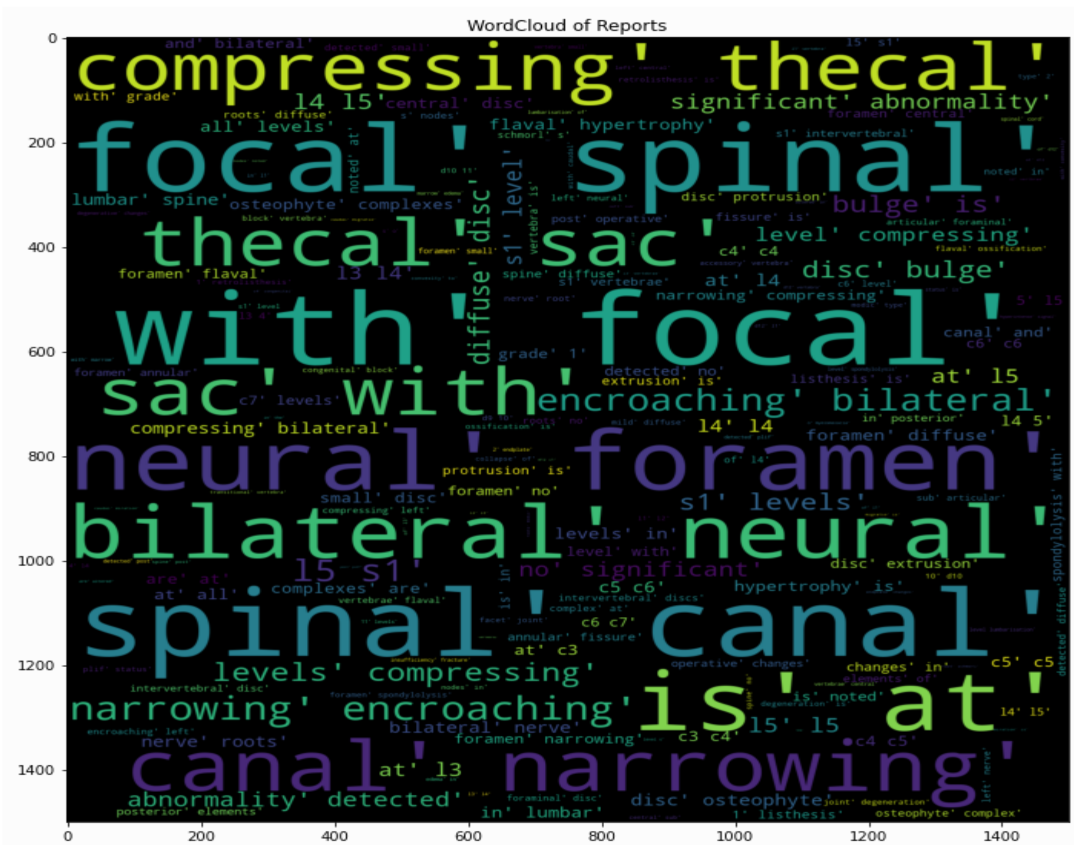


Fig. 11. Word Cloud Reports

5. Methodology

In this work we opt to focus on Deep learning using Computer vision for feature extractions by using some of the existing transfer learning models like CheXNet, Resnet50 & Encoder-Decoder using LSTM for the report generation.

5.1 Base model & Architecture

Diagram showing the model architecture from input to output

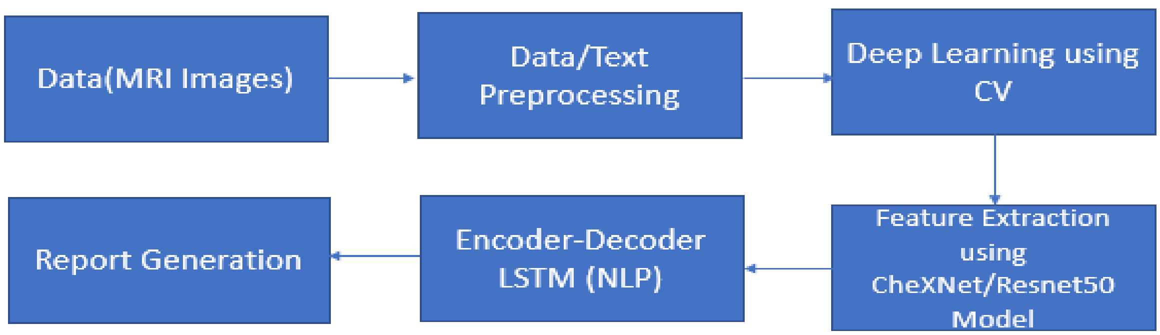


Fig. 12. Model Architecture

5.2 Feature Extraction

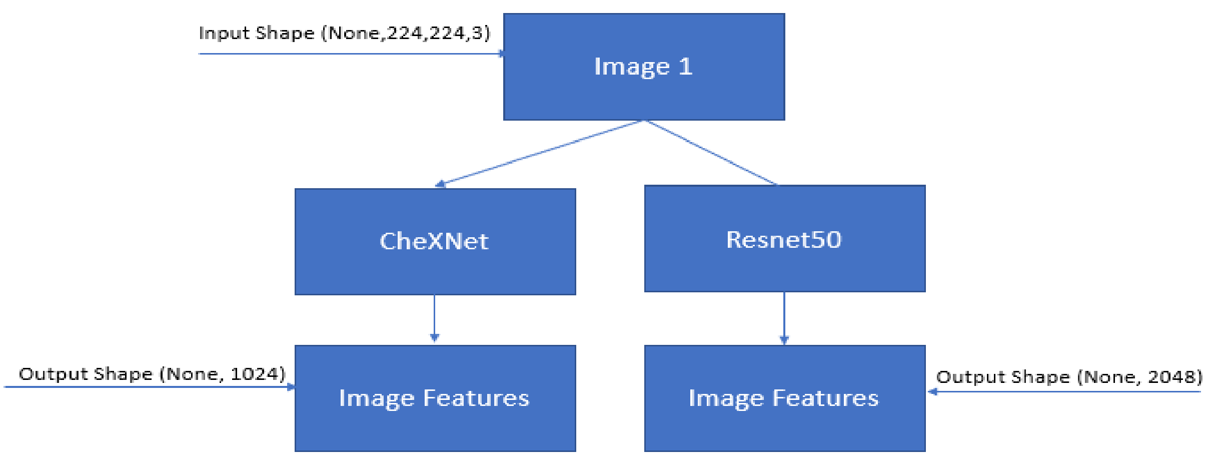


Fig. 13. Feature extraction process

5.3 Report Generation Method 1: Uni-Directional LSTM

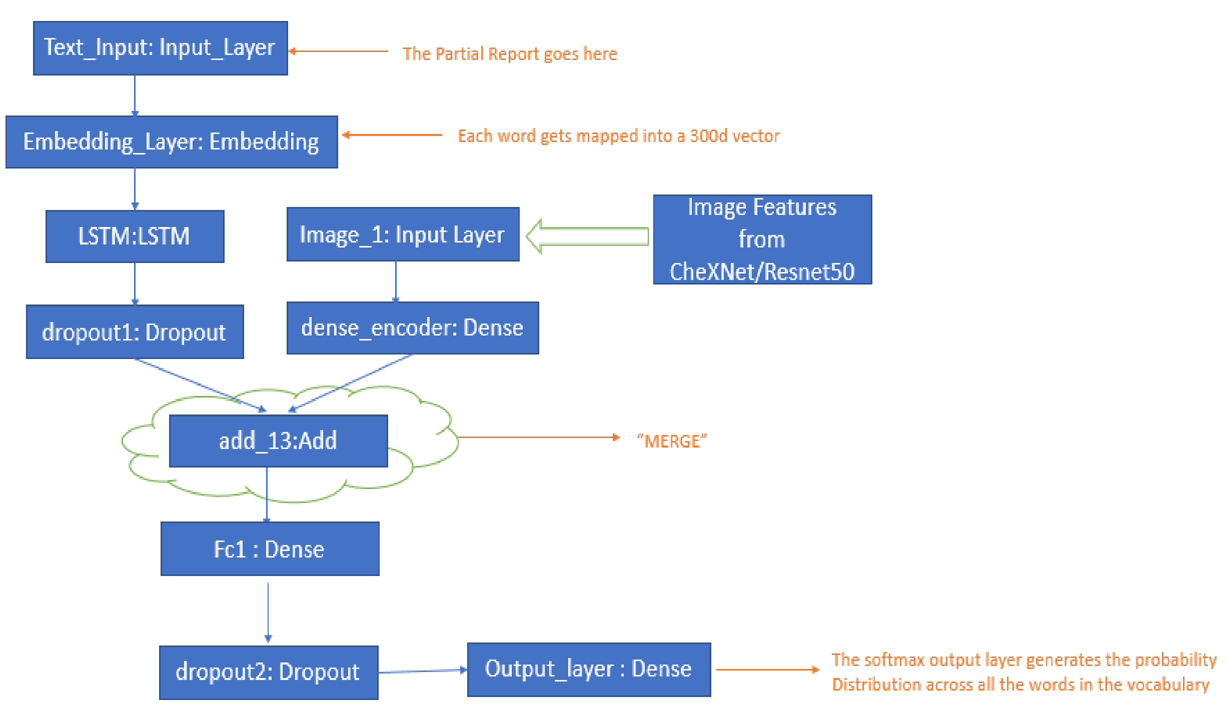


Fig. 14. Uni-Directional LSTM

Report Generation Method 2: Bi-Directional LSTM

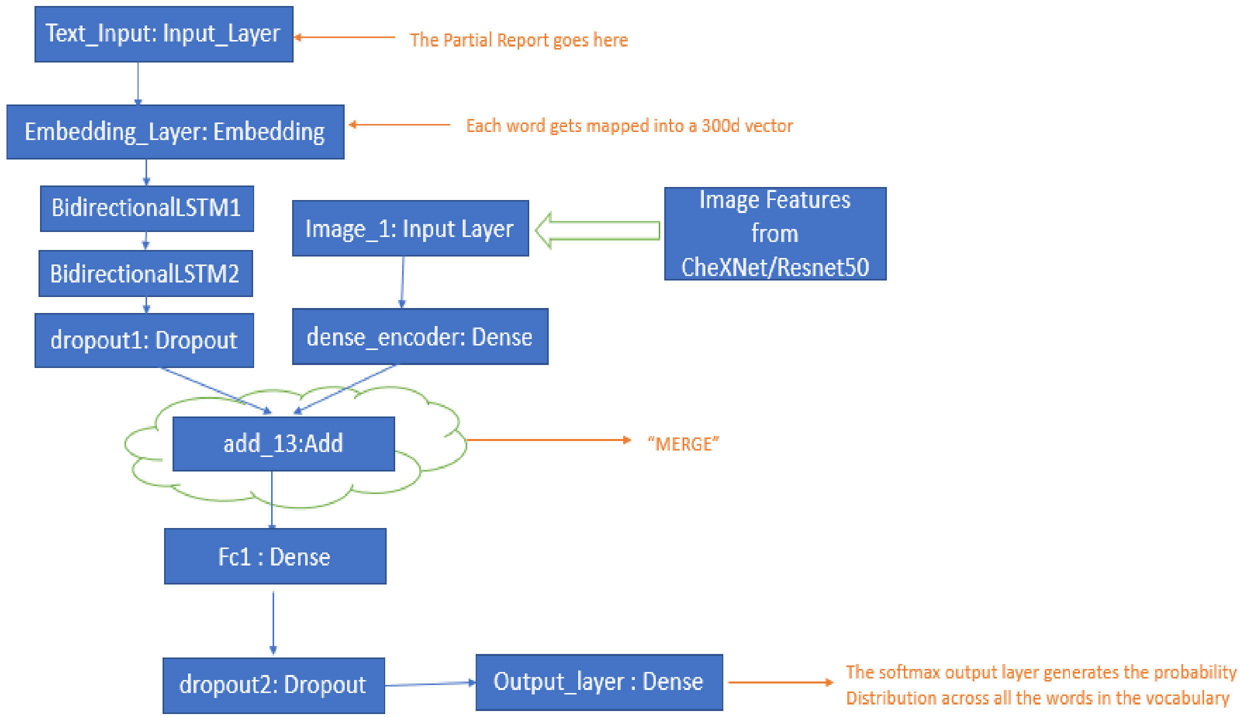


Fig. 15. Bi-Directional LSTM

6. Algorithms

- Pretrained Transfer Learning Model like CheXNet, Resnet50
- Encoder-Decoder Using LSTM
- Uni-Directional & Bi-Directional LSTMs.
- Dense Layers
- NLTK Library.
- Greedy/BEAM Search for BLEU scores.

6.1 Training Process

Resnet50 has already been trained on the weights from ImageNet & Similarly CheXNet has been trained on the weights i.e brucechou1983_CheXNet_Keras_0.3.0_weights.h5. We are using these predefined models with our input dataset & fine tuning it for this business case. Last 2 layers from these models has been used as output layers. Activation is Sigmoid

We used the high-level programming language Keras for creating the model. Tensorflow acted as the backend for Keras. BLEU is a precision focused metric that calculates n-gram overlap of the reference and generated texts. This n-gram overlap means the evaluation scheme is word-position independent apart from n-grams' term associations. One thing to note in BLEU — there is a brevity penalty i.e. a penalty applied when the generated text is too small compared to the target text. We have been training the model using Google Colab.

6.2 Encoder-Decoder Summary

Model Summary: CheXNet Bi-Directional

```
Model: "model"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
Text_Input (InputLayer)         [(None, 82)]         0
__________________________________________________________________________________________________
Embedding_layer (Embedding)     (None, 82, 300)      54900       Text_Input[0][0]
__________________________________________________________________________________________________
bidirectional (Bidirectional)   (None, 82, 512)      1140736     Embedding_layer[0][0]
__________________________________________________________________________________________________
Image_1 (InputLayer)            [(None, 1024)]       0
__________________________________________________________________________________________________
bidirectional_1 (Bidirectional) (None, 256)          656384      bidirectional[0][0]
__________________________________________________________________________________________________
dense_encoder (Dense)           (None, 256)          262400      Image_1[0][0]
__________________________________________________________________________________________________
dropout1 (Dropout)              (None, 256)          0           bidirectional_1[0][0]
__________________________________________________________________________________________________
add (Add)                       (None, 256)          0           dense_encoder[0][0]
                                                                 dropout1[0][0]
__________________________________________________________________________________________________
fc1 (Dense)                     (None, 256)          65792       add[0][0]
__________________________________________________________________________________________________
dropout2 (Dropout)              (None, 256)          0           fc1[0][0]
__________________________________________________________________________________________________
Output_layer (Dense)            (None, 183)          47031       dropout2[0][0]
==================================================================================================
Total params: 2,227,243
Trainable params: 2,172,343
Non-trainable params: 54,900
```

Model Summary: Resnet Unidirectional LSTM

```
Model: "model_2"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
Text_Input (InputLayer)         [(None, 82)]         0
__________________________________________________________________________________________________
Embedding_layer (Embedding)     (None, 82, 300)      54900       Text_Input[0][0]
__________________________________________________________________________________________________
LSTM1 (LSTM)                    (None, 82, 256)      570368      Embedding_layer[0][0]
__________________________________________________________________________________________________
Image_1 (InputLayer)            [(None, 2048)]       0
__________________________________________________________________________________________________
LSTM2 (LSTM)                    (None, 256)          525312      LSTM1[0][0]
__________________________________________________________________________________________________
dense_encoder (Dense)           (None, 256)          524544      Image_1[0][0]
__________________________________________________________________________________________________
dropout1 (Dropout)              (None, 256)          0           LSTM2[0][0]
__________________________________________________________________________________________________
add (Add)                       (None, 256)          0           dense_encoder[0][0]
                                                                 dropout1[0][0]
__________________________________________________________________________________________________
fc1 (Dense)                     (None, 256)          65792       add[0][0]
__________________________________________________________________________________________________
dropout2 (Dropout)              (None, 256)          0           fc1[0][0]
__________________________________________________________________________________________________
Output_layer (Dense)            (None, 183)          47031       dropout2[0][0]
==================================================================================================
Total params: 1,787,947
Trainable params: 1,733,047
Non-trainable params: 54,900
```

7. Results

During the process we have learnt how to get the data from different sources using Python Automation APIs like Selenium & also converted the medical imaging from DCM to JPG format. Learnt how to process the images & how image size would influence the overall performance of the model. Understood working of encoder decoder models using LSTM.

Resnet50 is doing far better than CheXNet on the Spine MRIs which is clear from the Bleu scores. ChexNet Bleu scores are low as those were trained on Chest X-Rays.

Attempt	Model	Epochs	LSTM Direction	Bleu1 Score	Bleu2 Score	Bleu3 Score	Bleu4 Score
1	CheXNet	20	Uni-Direction	0.10112044	0.07870352	0.08140587	0.10171706
2	CheXNet	20	Bi-Directional	0.09046569	0.06346033	0.06494746	0.06440563
3	Resnet50	20	Uni-Direction	0.26162158	0.23196109	0.21278055	0.20326725
4	Resnet50	20	Bi-Directional	0.23780004	0.20692362	0.18668749	0.17907181

Table. 1. Comparative Summary

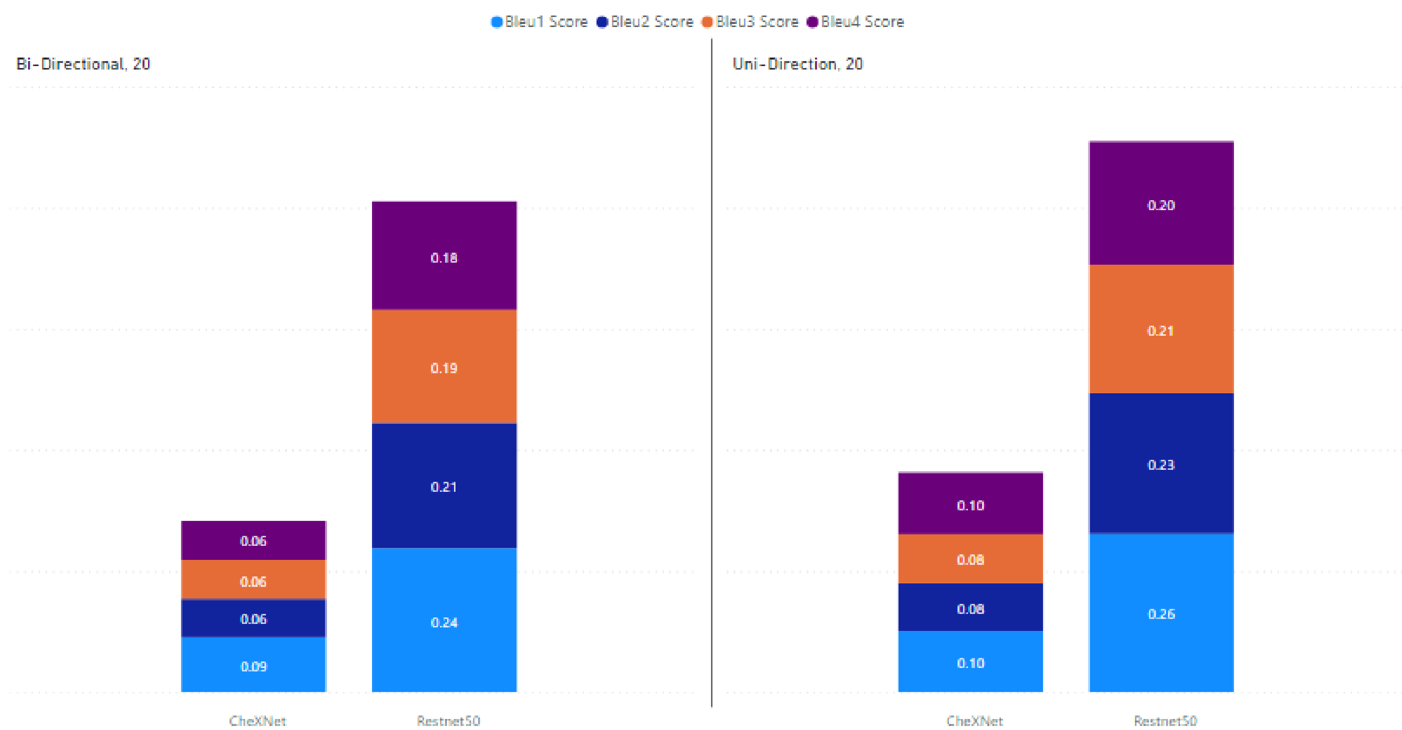


Fig. 16. Bi-Directional & Uni Directional Bleu Scores of CheXNet and Resnet50 models

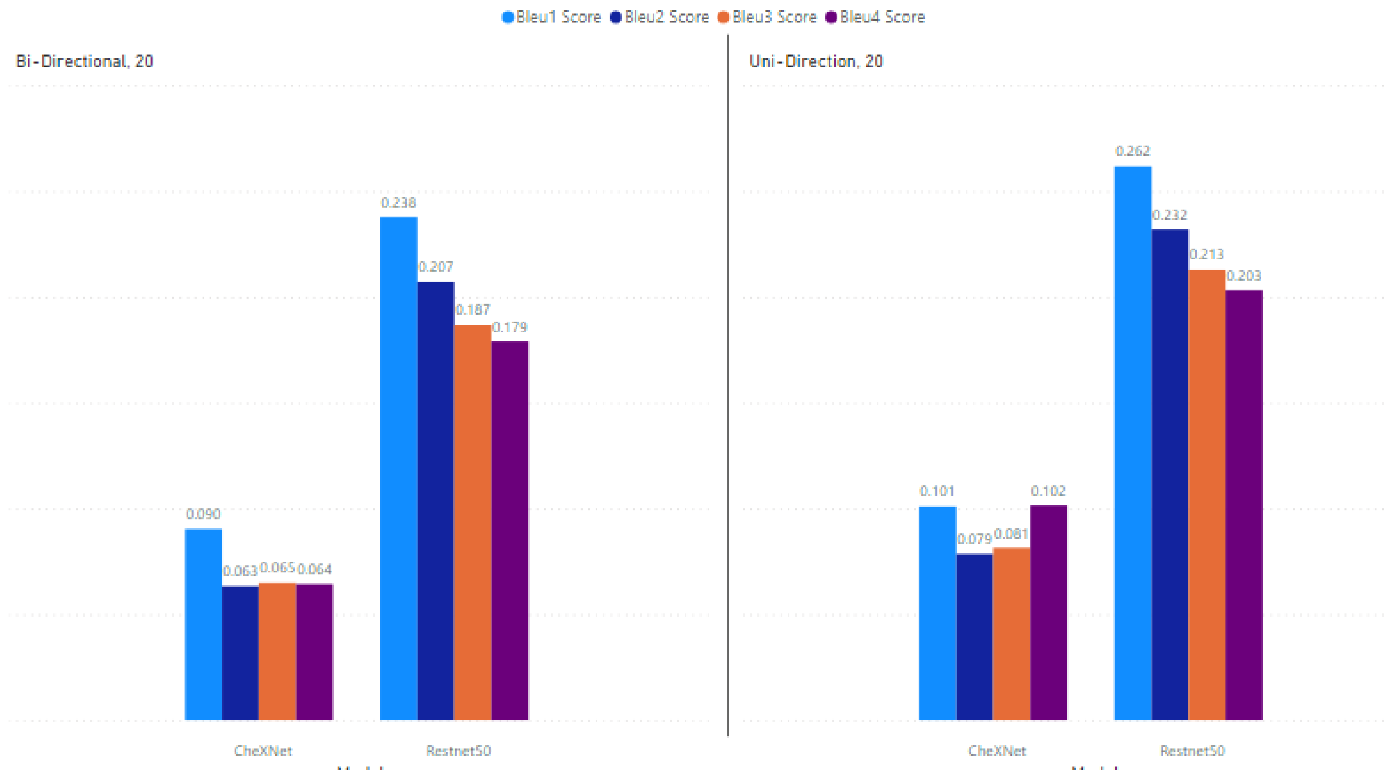


Fig. 17. Bi-Directional & Uni Directional Bleu Scores of CheXNet and Resnet50 models

8. Next Steps

- Multiple MRI images are combined per patient and providing input to the model.
- Using Other transfer learning models like AlexNet, VGG etc
- Due to time & resource constraints we had to limit to 20 epochs, but we can train on more numbers of epochs to get more higher performance of the model.
- Exploring the Attention Mechanism & Transformer Model for the report generation.

9. References

1. Towards Automatic Report Generation in Spine Radiology Using Weakly Supervised Framework. Zhongyi Han, Benzheng Wei, Stephanie Leung,Jonathan Chung, Shuo Li
2. Caton, M.T., Wiggins, W.F., Pomerantz, S.R. *et al.* The Composite Severity Score for Lumbar Spine MRI: a Metric of Cumulative Degenerative Disease Predicts Time Spent on Interpretation and Reporting. *J Digit Imaging* (2021). https://doi.org/10.1007/s10278-021-00462-1
3. Automatic extraction of imaging observation and assessment categories from breast magnetic resonance imaging reports with natural language processing. Yi Liu, Li-Na Zhu, Qing Liu, Chao Han, Xiao-Dong Zhang, Xiao-Ying Wan
4. Krishna, R., et al.: Visual genome: connecting language and vision using crowdsourced dense image annotations. Int. J. Comput. Vis. **123**(1), 32–73 (2017)
5. Detailed Sentence Generation Architecture For Image Semantics Description. Advances in Visual Computing, 2018, Volume 11241. ISBN : 978-3-030-03800-7. Imran Khurram, Muhammad Moazam Fraz, Muhammad Shahzad
6. Hassanpour S, Langlotz CP. Unsupervised Topic Modeling in a Large Free Text Radiology Report Repository. J Digit Imaging. 29(1):59–62, 2016.

7. Pavel Kisilev, Eugene Walach, Ella Barkan, Boaz Ophir, Sharon Alpert, and Sharbell Y Hashoul. 2015. From medical image to automatic medical report generation. IBM Journal of Research and Development, 59(2/3):2–1.
8. Wang, X., Peng, Y., Lu, L., Lu, Z., Summers, R.M.: TieNet: text-image embedding network for common thorax disease classification and reporting in chest X-rays. arXiv preprint arXiv:1801.04334 (2018)
9. Automated Diagnosis of Neural Foraminal Stenosis Using Synchronized Superpixels Representation. Yilong Yin, Manas Sharma, Gary Brahm, Ashley Mercado, Shuo Li
10. Natural Language Processing for Automated Quantification of Brain Metastases Reported in Free-Text Radiology Reports. Joeky T. Senders1; Aditya V. Karhade1; David J. Cote1; Alireza Mehrtash, MSc1; Nayan Lamba1; Aislyn DiRisio1;
11. Clinically Accurate Chest X-Ray Report Generation
12. Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, Marzyeh Ghassemi
13. Jonathan Krause, Justin Johnson, Ranjay Krishna, and Li Fei-Fei. 2017. A hierarchical approach for generating descriptive image paragraphs. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
14. Pengfei Liu, Xipeng Qiu, Xinchi Chen, Shiyu Wu, and Xuanjing Huang. 2015. Multi-timescale long short-term memory neural network for modelling sentences and documents. In Proceedings of the 2015 conference on empirical methods in natural language processing, pages 2326–2335.
15. Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3156{3164, 2015}