

# **CSE4099 Capstone Project**

## **Project Proposal**

**Reg.No:** 17BCE0136

**Name:** R.S. Rahul Sai

**Title:** Medical Report Generation on Chest X-Ray Images

### **Abstract:**

Computer Aided Diagnosis (CAD) and medical imaging systems have evolved in the past decade to a point where they can partially mimic radiologists and doctors. These systems can learn and differentiate the features and abnormalities in medical images, and provide objective evidence with a higher diagnostic confidence and faster inference. In this project, we focus on generating medical reports on chest X-ray images, which can be adapted later to work with other diagnostic tools such as ultrasounds and mammograms. The Indiana University dataset provides us with CXR images corresponding to various lung and heart ailments, along with well defined reports and findings. The generation of medical reports mainly consists of two broad tasks. The first task is to treat the problem as a multi-label classification task to obtain the accurate tags for a particular image from the visual features. The second task is to generate the reports using these aforementioned tags, which requires the use of recurrent neural networks such as hierarchical LSTMs.

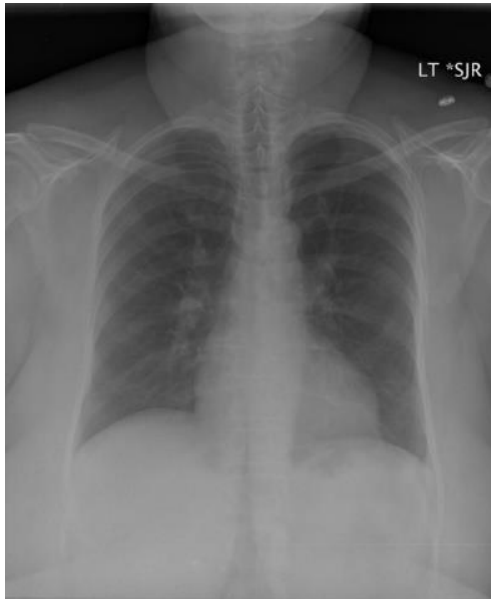
### **Dataset Description:**

The dataset to be used in this project is the Indiana University Chest X-Ray (CXR) Image dataset [1]. It is a high resolution CXR dataset with multiple views i.e. frontal, side and posterior views. There are 7,470 images accompanying 3,955 well written reports encoded in XML. These XML reports have references to the CXR images, the findings, impressions and the indication from the CXR images. These CXR images are obtained from patients diagnosed with tuberculosis, pneumonia and various heart ailments

### **Workflow:**

1. Preparing an NLP Pipeline for the original findings, impressions and indication
  - a. Removing Contractions, Punctuations and Numbers
  - b. Tokenization
  - c. Representing the reports in word embeddings
2. Obtain the image features using a Convolutional Neural Network (acts as the encoder here). [2]
3. Generate the text using the labels and tags obtained from the encoder using a Sequence-to-Sequence Model (Hierarchical LSTMs) (acts as the decoder here).
4. Substituting Attention mechanisms [3] to improve the encoder-decoder approach and compare them.
5. Utilizing newer NLP techniques such as BERT [4] and Transformers [5] in order to generate better reports

## Example from the Dataset



**Findings:** Heart size and mediastinal contour are within normal limits. There is no focal airspace consolidation or suspicious pulmonary opacity. No pneumothorax or large pleural effusion. Mild degenerative change of the thoracic spine

**Indication:** Evaluate for infection

**Impression:** No acute cardiopulmonary findings

## References

- [1] D. Demner-Fushman *et al.*, "Preparing a collection of radiology examinations for distribution and retrieval," *J. Am. Med. Informatics Assoc.*, vol. 23, no. 2, pp. 304–310, 2016.
- [2] B. Jing, P. Xie, and E. Xing, "On the Automatic Generation of Medical Imaging Reports," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2577–2586.
- [3] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1*, 2015, pp. 577–585.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "{BERT:} Pre-training of Deep Bidirectional Transformers for Language Understanding," *CoRR*, vol. abs/1810.0, 2018.
- [5] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998–6008, 2017.