Final Paper.pdf

By f f

MEDICAL REPORT GENERATION USING CHEST X-RAY

Dr.M.Ashok Kumar,
Panitini Monica, Vemulapalli Saiesh, Motamarri Jaya Naga Venkata Sai, Department of Information Technology,
Velagapudi Ramakrishna Siddhartha Engineering College,
KanuruVijayawada, Andhra Pradesh, India, 520007
ashokkumar.munnangi@gmail.com
monicapanitini2002@gmail.com, saieshvemulapalli07@gmail.com,
sai.doc.45@gmail.com

Abstract*— The diagnostic x-ray examination is carried out using the chest x-ray. It is the responsibility of the radiologist to analyze the x-rays and draw conclusions from them in order to prescribe the proper care. To obtain comprehensive medical report from these x-rays, it is frequently time-consuming. Images of the heart, lungs, airways, spine, and chest bones can be seen in a chest x-ray. A radiologist may see thousands of x-ray images in populous nations. The goal of this project is to present a collection of the best deep learning techniques for producing medical reports from X-ray pictures automatically. Deep learning algorithms have been used with models to handle this difficult task and produce correct results. Therefore, a lot of work and time can be saved if a properly trained deep learning model can generate these medical reports automatically. In this research, the text report is produced using an encoder and decoder with attention model, while the image features are obtained using a pretrained CheXnet model. The BLEU score is used to evaluate the resulting text report.

***Keywords:* Chest x-rays, Radiologist, Encoder, Decoder, Attention, BLEU, Chexnet**

I INTRODUCTION

1.1 Origin of the problem

One of the most significant and difficult tasks in deep learning is captioning images. The creation of a written description for an image is what it entails. We can create a caption or description based on the photographs. This is the project's main motivation. And numerous deep learning models have already been developed to complete this task. The model creates the matching textual description after comprehending the contents of the image. When radiologists need to describe medical X-ray images, it is highly beneficial in the medical profession. It is a difficult task to summarise an X-ray in a radiology report, so extra attention should be given while creating reports. The radiology report is an exhaustive examination of X-ray images describing both normal and abnormal circumstances, which helps the patient choose the appropriate course of treatment. The Radiologist is therefore expected to carefully describe a report. Writing medical reports typically takes 5 to 10 minutes per report, but the issue here is that in a day, doctors must write hundreds of reports, which can take a lot of their time. Writing medical imaging reports is demanding for less experienced radiologists and pathologists, especially those working in rural areas where the quality of healthcare is relatively low, or on the other hand, writing imaging reports can be tedious and time consuming for experienced radiologists and pathologists. Therefore, the goal of this project is to develop a deep learning model that automatically generates the text description of chest X-rays in order to save doctors' time and lessen some of their workload. To do this, we are using a publicly available dataset from Indiana University that consists of chest X-ray images and reports (in XML format), which contain information about the findings and impression of the X-ray. Predicting how the medical report related to the photographs will be received is the objective.

1.2 Basic definitions and background

1.2.1 Parsing XML file

XML contains a wealth of data, including the xray picture id, indication, findings, impression, etc. Since they are more helpful for the medical report, we will take the findings and impressions from these files and consider them reports. In order to obtain the x-rays associated with each report, we must additionally extract the picture id from these files.

1.2.2 Stuctured Data

Although there are only two image types—Front and Lateral—each patient is associated with a number of x-rays. A report may have up to five photos attached to it, with a minimum of one. Two photographs are the most frequently connected to a report. Each data point is accompanied by more than two photos, and occasionally fewer than two photographs as well. We must think of a way to format the data points so that only two photographs can be included in each data point.

1.2.3 Evaluation Metric(BLEU Score)

Bilingual Evaluation Understudy is the abbreviation for this score. In this case, the BLEU score will be the metric. The BLEU score determines how many words were predicted that were in the original sentence by comparing each word in the predicted sentence to the reference sentence (also done in n-grams). It gives back a number between 0 and 1. The two are quite comparable if the metric is close to 1.

$$\text{BLEU}= \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Formula 1.1:BLEU Score

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

Formula: 1.2 BP Score

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}.$$

Formula: 1.3 Modified Precision

1.2.4 Tokenization

When dealing with text, the first thing we need to do is come up with a method to turn strings into numbers (or "vectorize" the text) before feeding it to the model. We'll use Tokenizer to turn text input into numerical data. Tools for carrying out this technique are provided by the Tensorflow deep learning library.

1.2.5 Tranfer Learning

Images and excerpts from reports serve as the model's inputs. Every image must be transformed into a fixed-size vector before being used as input by the model. For this, transfer learning will be used.
The largest publicly accessible chest X-ray dataset at the moment, ChestX-ray14 has over 100,000 frontal-view X-ray pictures with 14 diseases. CheXNet is a 121-layer convolutional neural network trained on this dataset. Our goal is to simply obtain the bottleneck features for each image and not to categorise the images. As a result, this network's final classification layer is not required.

1.2.6 Encoder Decoder Architecture

A deep learning model called a "sequence-to-sequence" takes a sequence of items (in this case, the features of an image) and produces another sequence of objects (reports). Each item in the input sequence is processed by the encoder, which then gathers the data it has collected into a vector known as the context. The encoder transmits the context to the decoder after processing the full input sequence, For each of the decoder time steps, the decoder calls the one-step attention layer and computes the scores and attention-weights. The "all-outputs" variable contains the results of each time step. The subsequent word in the sequence is what each decoder step outputs. Our final output will be 'all-outputs'.

1.3 Problem Statements with Objectives and Outcomes

The most typical diagnostic x-ray procedure is a chest x-ray. Images of the heart, lungs, airways, blood vessels, and spine and chest bones can be seen on a chest x-ray. It is frequently the responsibility of a radiologist to interpret these x-rays so that patients can receive the proper care. Obtaining comprehensive medical reports from these x-rays is frequently time-consuming and laborious. Therefore, a lot of work and time can be saved if a properly trained machine learning model can generate these medical reports automatically.
In order to automatically generate medical reports from X-Ray images, this project seeks to present a compendium of the best deep learning techniques. Deep learning algorithms have been used with models to address this difficult task, with encouraging results. the final medical report that was produced using the model.

II LITERATURE REVIEW

[1] This research proposes a method to detect diseases using continuous integration of convolutional neural networks and long short-term memory, followed by an attention mechanism to generate sequences based on these ailments.

[2] They described a method based on a CNN-RNN architecture with an attention mechanism in their trials. predicted the outcomes of their model after being trained on the data.

[3] This study focuses primarily on the models using a single LSTM decoder, which perform much worse than those with a hierarchical LSTM decoder.

[4] They developed a novel deep learning

approach that conditions a pre-trained transformer using visual and semantic data and then adds semantic similarity metrics in addition to word-overlap metrics for the quantitative analysis.

[5] This research proposes a unique paradigm for the creation of medical reports that takes into account both diagnostic and language fluency.

[6] They presented a contractive attention model to accurately capture and describe aberrant regions.

A method for an application to generate automated textual reports for CXR was proposed in one of the aforementioned articles with the intention of helping medical professionals produce reports more quickly and effectively. It is built on an RNN decoder that produces corresponding sentences based on the learnt image characteristics, which is followed by a CNN feature extraction model that serves as an encoder that turns an image into a fixed-size vector representation.

III METHODOLOGY

3.1 Design Methodology

Chest X-ray images and XML reports are the initial data set that this project gathers from Indiana University. There are about 3955 XML reports with sample data visualisations and 7471 X-ray images. Following the data extraction from the XML report using the Python library, preprocessing is performed on the data. The next step is to create a new data set, a csv file, and separate it into train, validation, and test data. Then, in order to import two photos per patient and tokenize train data, we require structured data. Here, we load the Chexnet model, a 92.03% accurate pretrained model with over 14000 photos. The encoder/decoder model is then employed. The encoder uses the LSTM method, and the decoder uses the GRU algorithm. A user-defined loss function is created, the model is trained on the validation dataset, and then predictions are made on the test dataset.

3.2 Algorithms

3.2.1 LSTM Algorithm

In this algorithm the input is in an ordered sequence and the previous information is also important for predicting.

- A forget gate removes information that is no longer available in the cell state.
- Adding useful information about the state of the cell is done by the input gate.
- The task of extracting useful information from the current cell state and presenting it as an output is done by output gates.

3.2.2 GRU Algorithm

- Update Gate determines how much past knowledge is transferred into the future. This is similar to the output gate in LSTM.
- Reset Gate(r) determines how much past knowledge is forgotten. This is analogous to combining input and forget gates in an LSTM repeat unit.
- The current Memory Gate is integrated into the Reset Gate, just as the Input Modulation Gate is a sub-part of the Input Gate, and is used to introduce a non-linearity into the input and release the input.

3.3 System Architecture Diagram

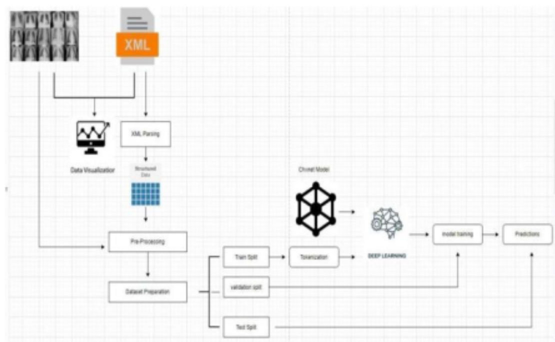

Fig: 3.1

3.4 Data Set

The Indiana University hospital network is the source of the information for this issue. There are two sections to the data. Chest x-ray images can be found in the X-rays. The reports include the x-medical ray's report. There are 7471 total X-ray images and 3955 total XML reports.

3.4.1 User Interface

The Google Colab, a user-friendly Python Graphical User Interface, serves as this system's user interface.

3.4.2 Hardware Interface

To enable user interaction with the console, Python features are employed.

3.4.3 Software Interface

The Python environment has been imported with the necessary modules (ChexNet model).

3.4.4 Hardware Requirements

1. Pentium-IV processor
2. RAM - 4GB (Minimum)
3. 256GB HDD/SSD (Minimum)

IV RESULT ANALYSIS

4.1 Stepwise Description of Results

A Loading the pre trained cheXnet model

The largest publicly accessible chest X-ray dataset at the moment, ChestX-ray14 has over 100,000 frontal-view X-ray pictures with 14 diseases. CheXNet is a 121-layer convolutional neural network trained on this dataset.

B Training the model

With the input dataset, the encoder decoder model with attention is successfully trained, and each epoch takes up to 600 ms. We obtained an average loss rate of 0.0010 after training the model.

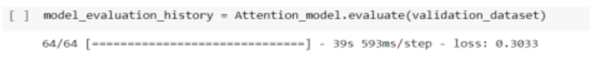

Fig: 4.1

The Encoder Decoder Model with Attention was successfully tested using the validation dataset, as shown in Figure 4.3, and it received a value loss of 0.3033.

4.2 Test Case Results

Fig: 4.2

The size of the heart and the pulmonary vascularity, as shown in figure 4.4, are both within normal ranges. There is no localised airspace disease in the lungs. With a BLEU score of 0.79, no pneumothorax or pleural effusion is seen.

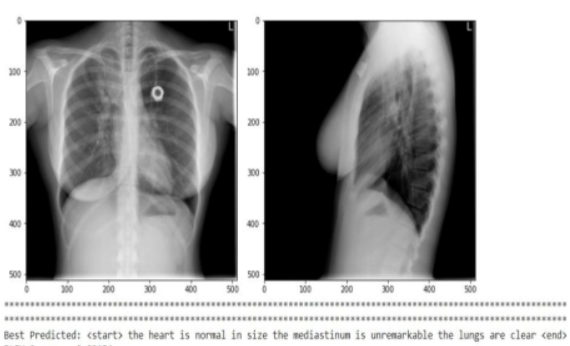

Fig: 4.3

The report that the lungs are clear, the mediastinum is unremarkable, and the heart is of typical size is shown in Fig. 4.5. Here, the BLEU score is 0.82154.

4.3 Observation from work

The number of photographs that can be included in a report is shown below. The total number of images per count is 2-3208,1-446,3-181,0-104,4-15, and 5-1, respectively.

Fig:

The validation loss, a statistic used to assess how well the deep learning model performed on the validation set, is shown in the loss graph in Figure 4.8.

V CONCLUSION AND FUTURE WORK

5.1 Conclusion

In order to help medical practitioners prepare reports more quickly and effectively, the proposed model is an application to generate automated text reports for CXR. Its foundation is an LSTM feature extraction model that serves as an encoder, turning an image into a fixed-size vector representation. An RNN decoder then uses the learned image characteristics to produce related words. The CXR dataset was used to conduct quantitative and qualitative analyses of the model's performance.

5.2 Future Work

The performance of this project is predicted to improve in the future by training on more images and using a larger dataset. For any of the models, no significant hyperparameter adjustment was performed. As a result, improved hyperparameter tweaking might result in better outcomes. Utilizing slightly more sophisticated methods, such as BERT or transformers, may produce superior outcomes.

References

[1] Sirshar M, Paracha MFK, Akram MU, Alghamdi NS, Zaidi SZY, Fatima T (2022) Attention based automated radiology report generation using CNN and LSTM.

[2] Bustos A., Pertusa A., Salinas J. M., & de la Iglesia-Vayá M. (2020). Padchest: A large chest

x-ray image dataset with multi-label annotated reports. Medical image analysis.

[3] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., et al. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison.

[4] Wang, X., Peng, Y., Lu, L., Lu, Z., & Summers,

[5] R. M. (2018). Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays.

[6] Li, Y., Liang, X., Hu, Z., & Xing, E. P. (2018). Hybrid retrieval-generation reinforced agentfor medical image report generation. In Advances in neural information processing systems.

[7] Zhang, Y. Xie, Z. Liao, G. Pang, J. Verjans,(2020) W. Li, Z. Sun, J. He, Y. Li, C. Shen, et al., Viral pneumonia screening on chest x-ray images using confidence-aware anomaly detection.

[8] Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U. Rajendra Acharya,(2019) Automated detection of covid-19 cases using deep neural networks with x-ray images.

[9] .E.-D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images, arXiv preprint arXiv

[10] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.(2018): Bottom-up and top-down attention for image captioning and visual question answering.

[11] Jing, B., Xie, P., Xing, E(2018): On the automatic generation of medical imaging reports.arXiv preprint arXiv

[12] Li, Y., Liang, X., Hu, Z., Xing, E.P (2018): Hybrid retrieval-generation reinforced agent for medical image report generation.

[13] Vinyals, O., Toshev, A., Bengio, S., Erhan, D.: Show and tell: A neural image caption generator. In: Proceedings of the IEEE conference on computer vision and pattern recognition.

Final Paper.pdf

ORIGINALITY REPORT

18%
SIMILARITY INDEX

PRIMARY SOURCES

1	towardsdatascience.com Internet	137 words — 6%
2	www.ncbi.nlm.nih.gov Internet	80 words — 3%
3	origin.geeksforgeeks.org Internet	62 words — 3%
4	"Advances in Computer Science for Engineering and Education", Springer Science and Business Media LLC, 2019 Crossref	43 words — 2%
5	Fazli Subhan, Xiaoxue Wu, Lili Bo, Xiaobing Sun, Muhammad Rahman. "A deep learning - based approach for software vulnerability detection using code metrics", IET Software, 2022 Crossref	25 words — 1%
6	prezi.com Internet	17 words — 1%
7	assets.researchsquare.com Internet	15 words — 1%
8	deepai.org Internet	14 words — 1%

9	"Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2020 Crossref	10 words — < 1%
10	Sathyabama Kaliyapillai, Saruladha Krishnamurthy. "Differential Evolution based Hyperparameters Tuned Deep Learning Models for Disease Diagnosis and Classification", Advances in Science, Technology and Engineering Systems Journal, 2020 Crossref	9 words — < 1%
11	link.springer.com Internet	9 words — < 1%
12	www.ijraset.com Internet	9 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 9 WORDS