a

By a a

WORD COUNT	3214	TIME SUBMITTED	29-DEC-2022 12:42PM
		PAPER ID	94998496

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. Radiologists and pathologists from the rural and small villages are very low and less experienced when they are related to the health care, or on the other hand, It will take more huge time for image writing or analyzing it for experienced radiologists and pathologists. Therefore, the goal of this project is to develop a classified efficient deep learning model that automatically generates the text report description of chest X-rays in order to save the doctor's time and lesson some of their workload. To do this, we have taken a dataset which is available in the public domain from Indiana University which consists of a chest X-ray images and XML reports, which contains the information about the findings and impressions attributes of an 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 the how many words are predicted from the original sentence by comparing each word in the model predicted sentence to the reference sentence in dataset (also done in ngrams format). 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 that case, the features of an image) and then it produces another sequence of objects (i.e.; 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 layer will transmits the context to the decoder layer after processing the full input sequence, For each of the decoder time steps, Then the decoder layer calls the one-step attention layer and then computes the scores and attention - weights. The "alloutputs" 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 'alloutputs'.

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] After training their model on the data, they predicted the results of their trials using a strategy based on a CNN-RNN architecture with an attention mechanism. They have created two systems for producing reports. The Encoder-Decoder model is one. RNN served as the decoder in this model while CNN served as the encoder. Encoder-decoder model with attention is the second method. The dataset utilized for the two mechanisms is the OpenI dataset. The encoder-decoder model with attention, which is the second technique with the best results of the two models, has an accuracy of 94%.

[2] This paper proposes a unique paradigm for the development of medical reports that

takes into account both language proficiency and diagnostic precision. From chest radiograph pictures, the encoder extracts visual cues and a variety of medical concepts. The hierarchical decoder then adds the ideas at the sentence and word levels to provide reports. However, the introduction of adversarial reinforcement learning (ARL) into the training process of medical report production is more significant. Both the encoder-decoder and the reward modules are seen as generators and discriminators, respectively. While the generator is taught using reinforcement learning, discriminators are tuned during training iterations using maximum-likelihood estimation. Finally, the reward modules provide incredibly accurate awards, and the generator produces superior reports. In the tests, the excellent performance of the proposed entire model is first demonstrated by performance comparison with a number of traditional or newly presented models from various aspects on two significant chest X-ray datasets.

[3] In order to provide full-text reports for chest X-ray pictures, this article developed a new conditioning approach that trains a DistilGPT2 model on visual and semantic data. In comparison to most earlier non-hierarchical recurrent models and transformer-based methods, conditioning a pre-trained transformer model demonstrated three advantages: (1) faster training; (2) the elimination of the need to specify a vocabulary for the model; and (3) better word-overlap based quantitative performance measures. As this thinks that conventional word-overlap based approaches are insufficient for sensitive systems, the paper also offer the first study to incorporate semantic similarity quantitative measures that act on the embedding level to assess medical reports. Additionally, this paper incorporate a qualitative analysis for our study from a radiologist from Egypt's national institute of cancer, which demonstrated encouraging outcomes.

[4] It established the Contrastive Attention model for creating chest X-ray reports, which compares the input images to typical images to spot trouble areas. The experiments on two open datasets demonstrate the effectiveness of the technique, which can be easily included into existing models to enhance their performance on crucial metrics. The assertions are supported by clinical efficacy ratings, and human evaluations show how effectively their approach helps existing models accurately gather and portray abnormalities. They specifically found the most cutting-edge results on the two datasets with the highest human preference, which should make radiologists' clinical judgements easier and minimize their workload.

This work employs an encoder-decoder model with attention based on the findings from the papers mentioned above. Results from this have outperformed those from earlier articles that were published. In this model, GRU served as the decoder and LSTM the encoder. The LSTM model converts the characteristics that are extracted from the pictures into vectors. The model now produces the best outcomes from training what it was given based on these features.

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 visualizations 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.

RNNs are constrained in that they can't retain inputs from earlier levels for very long. This can be overcome with LSTM, an RNN enhancement. It is mostly employed to address the issue of long-term dependency. A cell state, which is employed for information flow, is a component of LSTM. A LSTM model goes through these three steps:

1. The LSTM should decide in the first step what data it needs to keep from the previous inputs and what data it should discard. The Tanh

function, a sigmoid function, will be employed in this. For each piece of information in the cell state, this results in 1s and 0s. In other words, if the model multiplies the information by 0, it means that it is no longer necessary, and if it multiplies the information by 1, it means that it should be maintained.

2. This phase must determine what data from the new state ought to be saved in the cell state. In this stage, there are two layers: the sigmoid layer and the tanh layer. The candidate values will be created by the tanh layer, and the sigmoid layer will update the values that need to be changed. Vector.
3. Selecting the output is the final step. This has two layers. The cell state is first tested against the sigmoid layer. This will determine what output should be produced. After that, it will go via a tanh function and multiply with a sigmoid gate to only provide the output that the sigmoid layer determines.

3.2.2 GRU Algorithm

Another addition to RNN is GRU. It is used to address issues where RNN falls short. Either LSTM or GRU are employed, depending on the size of the application. Like the LSTM, GRU also has the ability to retain information for a long time. It is an inherent component of GRU. Two gates are present in this: an update gate and a reset gate.

How much historical data should be used to make predictions about the future is decided by the update gate.

Reset Gate determines how much historical data must be discarded rather than carried forward.

Additionally, in GRU, tanh and sigmoid functions are utilized in the gates to determine which information should be carried and which should be discarded.

3.3 System Architecture Diagram

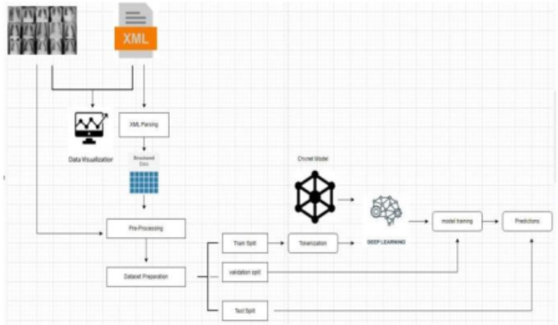


Fig: 3.1

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)

3.4.5 Software Requirements

1. Python
2. Keras
3. Tensorflow
4. Python libraries

IV RESULT ANALYSIS

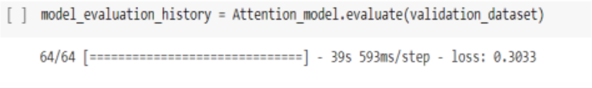
4.1 Stepwise Description of Results

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. An average loss rate of 0.0010 is obtained after training the model.



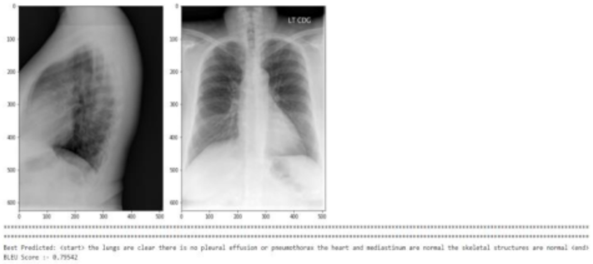
```
[ ]  model_evaluation_history = Attention_model.evaluate(validation_dataset)

     64/64 [==============================] - 39s 593ms/step - loss: 0.3033
```

Fig: 4.1

The Encoder Decoder Model with Attention was successfully tested using the validation dataset, as shown in Figure 4.1, and it received a value loss of 0.3033.

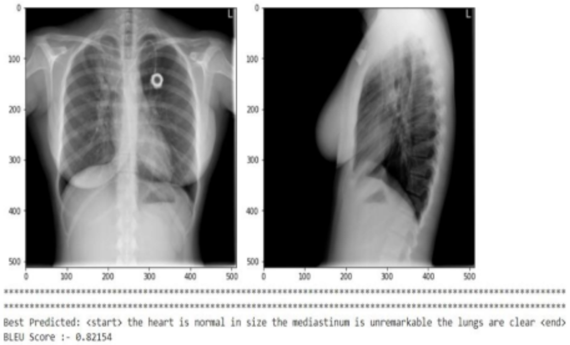
4.2 Test Case Results



```
************************************************************************************************************
************************************************************************************************************
Best Predicted: <start> the lungs are clear there is no pleural effusion or pneumothorax the heart and mediastinum are normal the skeletal structures are normal <end>
BLEU Score :- 0.79542
```

Fig: 4.2

The size of the heart and the pulmonary vascularity, as shown in figure 4.2, 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.



```
**********************************************************************************************
**********************************************************************************************
Best Predicted: <start> the heart is normal in size the mediastinum is unremarkable the lungs are clear <end>
BLEU Score :- 0.82154
```

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. 4.3. 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.

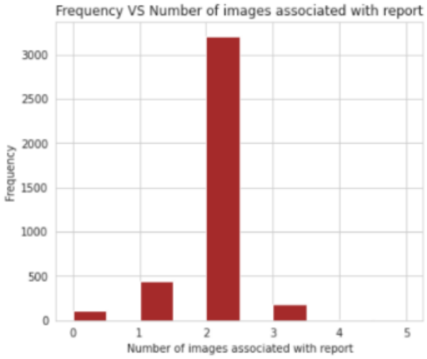


Fig: 4.4

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 4.4.

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 works 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 hyper parameter adjustment was performed. As a result, improved hyper parameter 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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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

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

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

[11] 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.

a

ORIGINALITY REPORT

7%

SIMILARITY INDEX

PRIMARY SOURCES

1	Omar Alfarghaly, Rana Khaled, Abeer Elkorany, Maha Helal, Aly Fahmy. "Automated Radiology Report Generation using Conditioned Transformers", Informatics in Medicine Unlocked, 2021 Crossref	62 words —	2%
2	"Advances in Computer Science for Engineering and Education", Springer Science and Business Media LLC, 2019 Crossref	43 words —	1%
3	journals.plos.org Internet	25 words —	1%
4	prezi.com Internet	17 words —	1%
5	assets.researchsquare.com Internet	15 words —	< 1%
6	web.archive.org Internet	10 words —	< 1%
7	"Advances in Information and Communication", Springer Science and Business Media LLC, 2020 Crossref	9 words —	< 1%
8	"Neural Information Processing", Springer Science and Business Media LLC, 2017	9 words —	< 1%

	Crossref	
9	"Trends and Innovations in Information Systems and Technologies", Springer Science and Business Media LLC, 2020 Crossref	9 words — < 1%
10	link.springer.com Internet	9 words — < 1%

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