Proceedings of the 7th International Conference on Intelligent Computing and Control Systems (ICICCS-2023)
IEEE Xplore Part Number: CFP23K74-ART; ISBN: 979-8-3503-9725-3

Report Generation on Chest X-rays using Deep Learning

Dr.M.Ashok Kumar
Information Technology
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
ashokkumar.munnangi@gmail.com

Suresh Ganta
Information Technology
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
sureshganta5555@gmail.com

Gopiswara Rao Chinni
Information Technology
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
chinnigopiswara123@gmail.com

2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) | 979-8-3503-9725-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICCS56967.2023.10142637

Abstract—**Nowadays the chest X-ray is the most frequently used diagnostic procedure. A radiology specialist can understand the detailed information contained in an X-ray of a chest regarding the human body's heart and lungs. Specialists in radiology are typically tasked with reviewing chest X-rays so that patients can get the right treatment. Because a doctor may see more than 100 X-rays each day in larger cities and thorough inspection necessitates skilled doctors, obtaining a thorough medical diagnosis from such X-rays is frequently laborious and time-consuming. This project's goal is to demonstrate deep learning techniques for autonomously extracting clinical data from X-ray images. Deep learning approaches have been combined with algorithms to tackle this difficult challenge, with promising performance. If such reports can be generated automatically by a trained model, a lot of time and effort can be saved. In this project, some deep learning techniques such as an encoder and decoder and a pre-trained chexnet model have been used. The task of obtaining visual features from an x-ray is performed by the chexnet model which is passed to an encoder which then sends its result to a decoder these types of techniques are mainly used in image captioning which aims to produce text from an image Here, LSTM is employed as the encoder while GRU and Bi GRU is used as the decoder. Both of these are recurrent neural networks. The generated text report is evaluated by the BLEU score.**

Keywords—Radiology specialist, chexnet model, encoder, decoder, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), BiLingual Evaluation Understudy (BLEU)

I. INTRODUCTION

Captioning photos is one of the most important and challenging tasks in deep learning. This is the process of translating an image into textual form. A text or summary can be created based on the images. This is the main motivation for the proposed research work. And a number of deep-learning models were previously created to carry out this duty. After understanding an image's content, the model generates descriptions that are compatible; in this case, the images are just chest X-rays. Further, this study develops a model to automatically generate the text report for a given X-ray. In order to help medical practitioners, analyze the images, a technology called automatic description creation for medical images creates natural language explanations of the images. By using deep learning models, the technique can recognize important features and patterns in medical images and generate accurate captions. One of the most often employed diagnostic methods in medicine is chest X-rays. They are used to detect various lung diseases such as pneumonia, tuberculosis, and lung cancer. However, interpreting chest X-rays can be a challenging task that requires trained medical professionals to identify and diagnose abnormalities accurately. Using deep learning techniques has shown great potential in assisting medical professionals in the interpretation of chest X-rays. Deep learning algorithms can analyze large amounts of data, learn from patterns, and accurately classify images, making them promising tools for improving the accuracy and speed of chest X-ray analysis.

In this study, we'll talk about how chest X-ray reports can be produced using the latest deep-learning techniques. This study also investigates the performance of the algorithms, the algorithm architecture, and the data utilized to train the algorithms.

Importance of this project

Interpretation of chest X-rays can be challenging, requiring trained medical professionals to accurately identify and diagnose abnormalities. Using deep learning algorithms can help to improve the accuracy and speed of chest X-ray analysis, potentially improving patient outcomes by providing proper medication plans. The use of deep learning techniques for report generation can help to automate the analysis of X-rays, reducing the workload of professionals and enabling more efficient patient care.

In conclusion, the importance of producing observation notes on chest X-rays with the help of deep learning lies in its potential to increase the precision and quickness of chest X-ray analysis, automate the analysis of medical data, and reduce the risk of medical errors, ultimately improving patient outcomes.

The Preliminary work

The basic concepts and vocabulary utilized in this work are defined in this section.

Image Captioning:

The basic concept of image captioning [9] is to prepare a deep-learning model to associate images with their corresponding captions or descriptions. When given an image as input, the model creates a natural language sentence that explains the image's content. Image Encoder, Language Model, Attention Mechanism, Loss Function, and Evaluation Metric are the typical components of an image captioning system.

There are some image captioning models like the Encoder-Decoder model, Encoder-Decoder with Attention, Template-based models, and Reinforcement learning-based models.

Attention: The attention approach enables the model to concentrate on a single, distinct, and interesting area of a picture while coming up with a new phrase to finish a description. Thus, it merely indicates that a decoder will

produce some descriptions using a limited set of detail. Consequently, using this new focus on architecture, we can anticipate the following sentence of the medical note.

Encoder: the image is encoded into fixed-sized vectors using the encoder model for retrieving the visual features in this analysis, we took into account the chexnet architecture we concatenate each picture characteristic because there are two images for each patient to obtain a vector with a smaller dimension we can now pass this vector across dense layers the encoders ultimate output will be this vector.

Decoder: The output of the encoder must now be converted into text. We do this by using LSTM networks, which are excellent for handling text data. As a sequence-to-sequence model, in this case, we employ LSTM. One word is produced at a time from the LSTM networks' inputs, which are provided in time increments.

BLEU score: Bilingual Evaluation Understudy Score, or BLEU score, is a frequently used statistic for assessing the quality of the computer-generated text, especially in machine translation tasks in deep learning. It measures the degree of overlap between the machine-generated text and a collection of reference works, assigns a grade between 0 and 1, and, with higher scores indicating better performance.

LSTM: [13] LONG SHORT-TERM MEMORY Recurrent neural networks (RNNs) of the Long Short-Term Memory (LSTM) type are frequently employed in applications involving sequence modeling and processing of natural languages (NLP). Standard RNNs have a vanishing gradient problem that LSTMs are made to solve, which enables them to collect dependencies that persist in input patterns**.**

GRU: [13] GATED RECURRENT UNIT A sort of recurrent neural network (RNN) architecture known as a gated recurrent unit (GRU) employs gating methods to regulate the input flow. It has shown success in various applications, including Picture description, recognition of voices, and processing of natural languages due to its ability to handle sequential data efficiently

II. PREVIOUS WORK

This part of the paper concentrates primarily on earlier research that used a variety of machine learning as well as deep learning techniques to produce reports on chest x-rays.

In 2018, Allaouzi et al. [1] probably the first in-depth analysis of computerized picture captioning in the medical field. They covered the majority of the currently utilized techniques, benchmark datasets for medical picture captions, and assessment criteria applied in the reviewed literature. However, the survey was too brief and did not accurately examine cutting-edge techniques.

Rahul Sai [2] In this study, the authors develop a multi-task learned model that links paragraph production and tag prediction, they suggest a co-attention method to locate regions with anomalies and provide narrations for them, and they construct a hierarchy LSTM network to produce long texts.

Daibing Hou [3] in this study, For chest x-ray pictures, a framework for adversarial strengthened report production is suggested. It comprises linguistic proficiency and diagnostic accuracy. Additionally, it has accuracy and fluency discriminators, who act as evaluators. It is suggested to use a fresh framework for creating medical reports that take linguistic and diagnostic fluency into account.

Fenglin Liu [4] In this study, they suggested a commutative attention model to accurately capture and explain aberrant regions. For the purpose of creating a chest X-ray report, a contrastive attention model is used to acquire the image as input and regular images.

Mehreen Sirshar [5] This research offers a solution to this problem based on the constant deployment of deep networks of neurons for disease detection, followed by a method of attention for series analysis based on such ailments. A number of images were examined using three techniques: The VGG, LSTM, and GRU. The outcomes of the learning procedure for the aforementioned models demonstrate that the VGG + LSTM model is clearly more accurate than the other approaches.

Iqra Naz [6] in this study, they used the cnn-rnn model with and without attention and proposes a novel approach for generating reports of lung diseases from chest X-ray images using natural language processing (NLP). It was found the encoder-decoder approach was less accurate than the attention approach.

Omar alfarghaly [7] suggested a brand-new deep learning technique that conditions a transformer with prior training using graphical and semantics data before adding similarity measure metrics in addition to measures for text overlap in the qualitative evaluation.

Guanxiong Liu [8] In this paper, they offer an autonomous system for developing an x-ray report that is domain aware and first predicts the report's subjects, then conditionally generates phrases that relate to those topics. The ultimate method is tuned via reinforcement training while taking into account fluency and diagnostic correctness as measured by the suggested Clinically Coherent Reward.

Yuan [11] The paper proposes a novel approach that combines retrieval and generation methods, enhanced by a reinforced learning component, to generate medical image reports. This approach shows promising results, outperforming existing methods while also providing interpretable results. Additionally, it has potential applications in improving medical diagnosis and treatment, making it a valuable contribution to the field.

Jing et al [10] The paper reviews the current techniques for auto creation of reports from diagnostic imaging. It emphasizes the benefits of report generating that is automated, including reducing workload and making medical evaluation and therapy more effective. The paper outlines various approaches for report generation, highlighting their strengths and limitations. The paper also identifies several challenges and future research directions in the field, such as developing more robust models and addressing data privacy and security concerns.

III. SYSTEM ARCHITECTURE

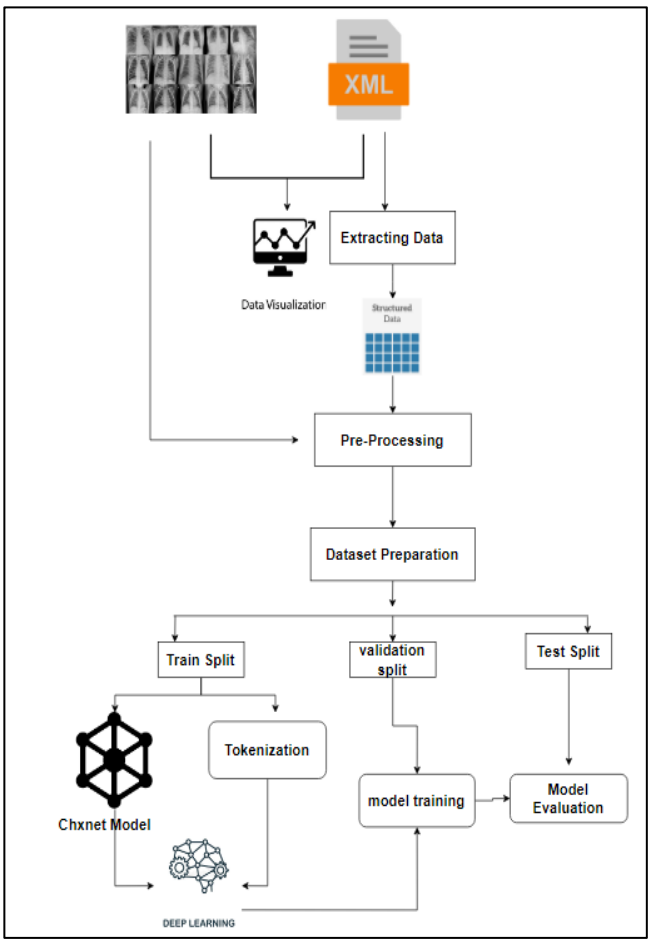

Fig. 1. System architecture of the proposed model

IV. DESIGN METHODOLOGY

An encoder-decoder model is a pre-owned model for generating reports on x-rays using deep learning, to retrieve significant data from the x-ray and give a matching response. This model mainly consists of two parts "an encoder" and "a decoder". A feature vector created by the encoder is sent into the decoder, another deep networks of neurons, which then produces the matching report and it is nothing more than a deep network of neurons that retrieves the most important features from the chest x-ray image as input. In conclusion, the encoder-decoder model is a well-liked method for creating clinical information from chest images. The decoder creates a corresponding report using the key features that the encoder has extracted from the image. The model can learn to produce precise and informative medical reports by being trained on a series of reports including chest scan images that go along with them.

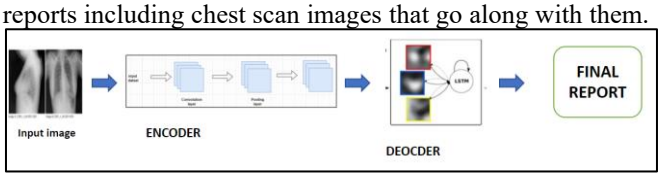

Fig. 2. Flow Diagram of the Model

The encoder in a sequence-to-sequence Transformer model makes a vector of fixed length depicting an input sequence by taking the sequence as input. The decoder takes this context vector and produces the output sequence one token at a time, conditioned on the previous tokens it has generated. Together, the encoder and decoder form a model that can handle variable-length input and output sequences and capture the dependencies between them. A step-by-step process for developing an encoder-decoder model for sequence-to-sequence prediction in Keras is Prepare the Data, Define the Encoder Model, Define the Decoder Mode, Define the Training Model, Train the Model, Evaluate the Model, and Make Predictions. All these things are done by using the Keras functional API. In an encoder-decoder model, attention mechanisms [5] aid the decoder in producing each output token by directing its attention to particular segments of the input sequence. This is important because the input sequence can be long and the context vector generated by the encoder may not contain all the necessary information needed to generate the output sequence.

Overview of Dataset

To understand the data we'll be working with, we'll first get and view it. The chest scan dataset created by Indiana University will be used by us. This dataset includes developed by Indiana University. This dataset consists of

- The count of X-ray images is 7471
- The count of XML reports corresponding to the above x-ray images is 3955.

Fig.3 XML report

Many patient-related data are included in this XML document that includes image id and text captions with terms like "comparison," "indication," "findings," "impression," and more. As they are more relevant for a medical report, we will retrieve these features from all of these documents and treat them as reports. We also need to get the picture id from these files in order to get the X-rays related to every XML report.

V. ALGORITHM DESCRIPTION

1. LSTM algorithm

This method[13] requires an orderly sequence of input, and earlier data is also essential for the prediction..

RNNs are limited to this so they are unable to store inputs from earlier levels for an extended period of time. Using LSTM, an RNN improvement can be avoided. It is mostly used to deal with the problem of long-term reliance. A part of an LSTM is a cell state, which is used for flow of data. In an LSTM model, the following three steps are taken:

1. In the first step, the LSTM should choose which data from the earlier data to preserve and which

information to reject. Here, a sigmoid function called the Tanh function will be used. For every bit of data stored in the cell state, this produces 1s and 0s.

2. During this stage, Which data from the latest state should be stored in the state of the cell has to be chosen. The potential values will be generated by the tanh layer. and the values entered will be modified by the sigmoid layer which requires updating.
3. The final step is choosing the output. Two layers make up this. The sigmoid layer is used to test the cell state initially. What should be produced as an output will be decided by this. Following that, it will only produce the output that is determined by the layer of sigmoid after being processed by a tanh and multiplied by a sigmoid gate.

The general steps for using LSTM as an encoder are as follows:
Establish your LSTM model:
Get your input data ready.
Encode the input you provide such that it signifies Your input sequence is sent on to the LSTM model.
Process the encoder output for further processing

2. GRU algorithm

GRU [13] is another addition to RNN. It is employed to address problems when RNN fails. Based on the application this GRU is used. The GRU has a similar long-term memory as the LSTM. It is a fundamental part of GRU. Two gates are present in this.
One gate tells the amount of past data used for producing predictions. The amount of previous data that must be destroyed rather than carried forward is determined by Reset Gate.
Additionally, in GRU, the gates use the tanh and sigmoid functions to decide which pieces of input should be transmitted and which should be ignored.

3. BI-GRU Algorithm

1. For every time step, Bi-GRU [13] contains two hidden states: one for the forward pass and one for the backward pass. These concealed states, which are present in both the past and future contexts of the input sequence, respectively, capture the pertinent data.
2. The forward pass hidden state is updated every time by accounting for the current data and the earlier state, whereas the backward pass hidden state is updated by accounting for the present input and the subsequent hidden state.
3. Concatenating the forward and backward pass hidden states. This combined state can be utilized for making predictions or categorizing data by combining information from the input sequence's past and future contexts.
4. Bi-GRU has output states in addition to concealed states. These output states are formed by applying a linear transformation to the final hidden state, followed by an activation function such as a SoftMax or a sigmoid.
5. The output states can be used to classify a sequence of images or anticipate the next word in a sentence using the output states.
6. A method known as t-SNE allows for the visualization of the hidden states and output states (t-Distributed Stochastic Neighbor Embedding). By converting high-dimensional data into a lower-dimensional space, this method enables us to see the connections between the hidden or output states at various time points.

VI. IMPLEMENTATION AND RESULTS

In this project, we will be working with X-rays and XML reports for building a model using deep learning that is able to generate reports. We start by importing the necessary libraries and datasets and visualizing the sample data. After that, we preprocess the data by decontracting the text, extracting the data from XML files, removing null values, and preprocessing text features. We then display sample images with text features and prepare the dataset by splitting it and tokenizing the text.
We will load the CHXNet model and to produce the reports, we will construct a model of an encoder-decoder with attention. We will also create a custom loss function to prepare the model and the next step is to train this model on the prepared dataset and determines how well it does on test data. Finally, we will generate predictions on the sample test data to see how well our model performs. By the end of this project, We'll get a model using deep learning capable of automatically producing observations from X-rays, potentially saving medical professionals time and resources.

.

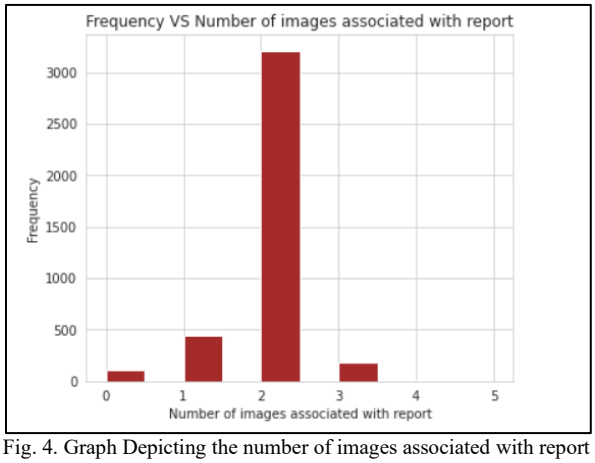

Fig. 4. Graph Depicting the number of images associated with report

To normalize this graph we preprocess our data so that a constant number of X-ray images will be associated with XML reports. This Graph shows the number of x-rays associated with the XML reports. It shows that the maximum number of images associated with a report is and the minimum number of images associated with a report is 0

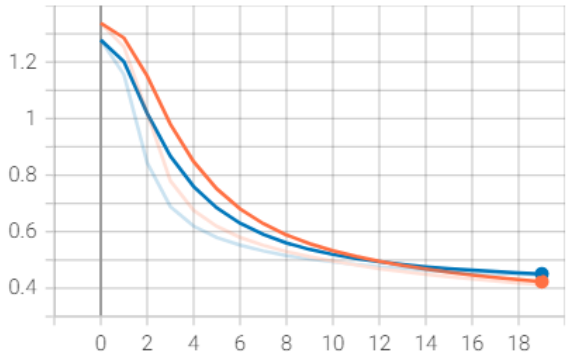

Fig.5 Train Vs Validation Loss

The above Graph shows the train loss and validation loss rate.

Phase 1: LSTM-GRU WITH ATTENTION
In this phase, We employed GRU as the decoder and LSTM as the encoder. It gives a report for a set of chest X-rays.

Phase-2: LSTM-BI GRU with Attention
In this phase, We employed BI GRU as the decoder and LSTM as the encoder. It gives a report for a set of chest X-rays.

TABLE I. LOSS RATE COMPARISON OF MODELS

Model	Loss rate
LSTM-GRU	0.3036
LSTM-BI GRU	0.2416

Result 1:

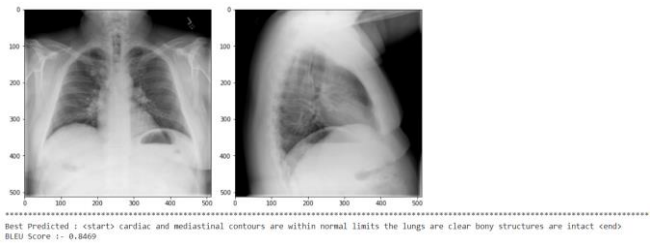

Fig 6: Result 1

Predicted output:
We can see that an automatic text report was generated when we gave a set of chest x-rays as an input and we can clearly see that we have a BLEU Score of 0.84 and generated report is "cardiac and mediastinal contours are within normal limits the lungs are clear bony structures are intact."

Result 2:

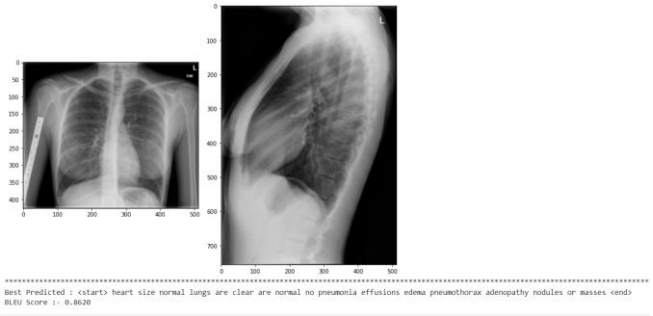

Fig 7 Result 2

We can see that an automatic text report was generated when we gave a set of chest x-rays as an input and we can clearly see that we have a BLEU Score of 0.86 and generated report is "heart size is normal lungs are clear are normal no pneumonia effusions edema pneumothorax adenopathy nodules or masses"

VII. CONCLUSION AND FUTURE WORK

Conclusion:
In conclusion, the presented model is a promising solution for automated textual reporting in CXR. By leveraging the power of LSTM feature extraction and RNN decoding, the model can effectively convert image data into descriptive sentences that medical professionals can use for diagnosis and treatment planning. The model's effectiveness has been demonstrated through both quantitative and qualitative analysis on a CXR dataset. However, there is still room for improvement, particularly in terms of model performance and hyperparameter tuning. Overall, this model shows a crucial improvement in the development of automated reporting tools for medical imaging, and it has the potential to revolutionize the way medical professionals work.

Future Work:
In the future, there are several areas of improvement that can be explored for this model. Firstly, the model's efficiency may be measured by improving training it on a larger dataset of CXR images. This enables the model to acquire new knowledge on diverse image properties and enhances its ability to generate accurate and informative reports. Secondly, better hyperparameter tuning can be performed to fine-tune the model's architecture and optimize its performance. Advanced techniques like transformers or BELU can be explored to improve the model's accuracy and efficiency. Additionally, integrating the model with other medical imaging tools and technologies can help streamline the diagnostic process further. Overall, there is a lot of scope for further research and development in this area, and this model provides an excellent foundation for future work. Achieving high accuracy and improving the model's performance in obtaining medical reports can be achieved through a combination of high-quality annotated data, preprocessing and augmentation techniques, appropriate model architecture, transfer learning, effective training and optimization, and appropriate evaluation metrics

VIII. REFERENCES

[1] Allaouzi, Imane, and Mohamed H. Ahmed. "A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases." IEEE Access 7 (May 22, 2019): 64279–88. https://doi.org/10.1109/access.2019.2916849.

[2] Sai, Rahul, Kather, Sharmila, Tripathy, B.K. "Automated Medical Report Generation on Chest X-Ray Images using Co-Attention Mechanism", 2021/12/11. https://doi.org/10.13140/RG.2.2.26061.15843

[3] Hou, Daibing, Zijian Zhao, Yu-Ying Liu, Faliang Chang, and Sanyuan Hu. "Automatic Report Generation for Chest X-Ray Images via Adversarial Reinforcement Learning." IEEE Access 9 (February 1, 2021): 21236–50. https://doi.org/10.1109/access.2021.3056175

[4] Liu, Fenglin, Changchang Yin, Xian Wu, Shen Ge, Ping Zhang, and Xu Sun. "Contrastive Attention for Automatic Chest X-Ray Report Generation." Meeting of the Association for Computational Linguistics, August 1, 2021. https://doi.org/10.18653/v1/2021.findings-acl.23

[5] Sirshar, Mehreen, Muhammad Majid Paracha, Muhammad Akram, Norah Saleh Alghamdi, Syeda Ramsha Zaidi, and Tatheer Fatima. "Attention Based Automated Radiology Report Generation Using CNN and LSTM." PLOS ONE 17, no. 1 (January 6, 2022): e0262209. https://doi.org/10.1371/journal.pone.0262209

[6] Iqra Naz1, Shagufta Iftikhar1, Anmol Zahra1, Syed Zainab "Report Generation of Lungs Diseases From Chest X-Ray Using NLP." International Journal of Innovations in Science and Technology 3, no. 5 (February 26, 2022): 223–33. https://doi.org/10.33411/ijist/2021030518

[7] Alfarghaly, Omar, Rana Khaled, Ahmed S. Elkorany, Maha Helal, and Aly A. Fahmy. "Automated Radiology Report Generation Using Conditioned Transformers." Informatics in Medicine Unlocked 24 (January 1, 2021): 100557. https://doi.org/10.1016/j.imu.2021.100557.

[8] Liu, Guanxiong, Tzu-Ming Harry Hsu, Matthew B. A. McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, and Marzyeh Ghassemi. "Clinically Accurate Chest X-Ray Report Generation." Machine Learning for Healthcare Conference, October 28, 2019, 249–69. http://proceedings.mlr.press/v106/liu19a/liu19a.pdf.

[9] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, "Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering." Computer Vision and Pattern Recognition, June 18, 2018. https://doi.org/10.1109/cvpr.2018.00636.

[10] Jing, Baoyu, Pengtao Xie, and Eric P. Xing. "On the Automatic Generation of Medical Imaging Reports." ArXiv (Cornell University), July 1, 2018. https://doi.org/10.18653/v1/p18-1240.

[11] Li, Yuan, Liang Lin, Zhiting Hu, and Eric P. Xing. "Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation." Neural Information Processing Systems 31 (May 1, 2018): 1530–40. https://arxiv.org/pdf/1805.08298.pdf.

[12] Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and Tell: A Neural Image Caption Generator." Computer Vision and Pattern Recognition, June 7, 2015. https://doi.org/10.1109/cvpr.2015.7298935.

[13] Bensghaier, Mabrouka, Wided Bakari, and Mahmoud Neji. "Investigating the Use of Different Recurrent Neural Networks for Natural Language Inference in Arabic." Research Square (Research Square), March 20, 2023. https://doi.org/10.21203/rs.3.rs-2693608/v1.