

Automatic radiology report generator using transformer with contrast-based image enhancement

R.Lenitha ¹, L. Kebila Anns Subi ², D.Deiva Thaya Sweetlin ³

1PG Student, M.E Communication and networking, Ponjesly college of engineering, Kanyakumari, Tamil Nadu, India.

2Professor M.E Communication and networking, Ponjesly college of engineering, Kanyakumari, Tamil Nadu, India.

3 Assistant Professor M.E Communication and networking, Ponjesly college of engineering, Kanyakumari, Tamil Nadu, India.

Abstract — It takes a lot of time and requires the knowledge of experienced radiologists to write radiology reports based on radiographic pictures. Therefore, it would be beneficial to incorporate technology that can generate reports automatically. The primary difficulty with automatic report generating is creating a logical predictive text. Techniques that can improve the relevance of features in generating predictive text must be developed. This study used the transformer approach and image enhancement technology to build a model for generating medical reports. An methodology to improve the medical image's noise-proneness is investigated in this study, along with the transformers way to produce a radiologist report based on chest X-ray images, in order to take advantage of the visual and semantic elements. The impact of image improvement techniques on the radiology report generator was examined using four contrast-based image enhancement approaches. The encoder-decoder model is employed with a pre-trained model ChexNet and Multi-Head Attention (MHA) mechanism for visual feature extraction and Bidirectional Encoder Representation from Transformer (BERT) for text feature embedding. With a 0.412 value, MHA outperforms the baseline model by 15% as well. This approach can perform better than the baseline model and other earlier studies. It can be said that BERT and the transformer MHA encoder layer work well for utilizing textual and visual information. Furthermore, it has been discovered that incorporating an image enhancement technique improves the model's performance.

Key Words: Bidirectional Encoder Representation from Transformer (BERT). Multi-Head Attention (MHA)

1. INTRODUCTION

As medical imaging technology advances, medical image diagnostics become increasingly complex, necessitating

the necessity for medical professionals. According to data published in the Medical Journal of Radiology in 2015, radiologists' workload increased by 26% over the preceding ten years. [1] Due to medical advancements, radiologists now have to compare a lot more information, including many factors, in order to make comprehensive and reliable diagnosis based on medical images. The goal of this project is to create a system that employs cutting-edge natural language processing (NLP) models, specifically transformers, in conjunction with image enhancement techniques to automatically generate radiology reports from medical pictures, such as X-rays, scans, or MRIs. [2]. Its goal is to make the process of creating radiology reports from medical images as efficient as possible. The method improves the visibility of important anatomical features and abnormalities in medical imaging including X-rays, MRIs, and CT scans by employing sophisticated contrast enhancement techniques. Following image processing, a transformer-based natural language model is used to evaluate the improved images' attributes and produce precise, thorough radiologist reports. [3] Sections like as patient information, an evaluation of the image quality, comprehensive findings, and a conclusion with suggestions are usually included in these reports. Structured, context-aware medical reports in natural language can be produced by integrating a transformer model, such GPT or MedBERT. By automating regular report generation, this method not only saves radiologists time but also guarantees full reports, lowering the possibility of oversight. Additionally, in both developed and underprivileged locations, the system could improve accessibility to timely medical interpretations, help healthcare providers, and increase the efficiency of diagnostics. [4] Enhance medical image quality to make pertinent information (such tumors, fractures, or other anomalies) easier to see and identify. To improve image features and highlight important

structures, use contrast enhancement techniques like adaptive contrast enhancement and histogram equalization. This is particularly helpful when photos are taken in less-than-ideal conditions or when certain details are difficult to see. [5] To provide more individualized and precise results, make sure the generated reports are context-aware, which means they consider not just the results but also the clinical context supplied in the image (such as the location of the abnormality, the patient's age, and gender) [6]. Using medical picture data, automatically produce thorough and contextually accurate radiology reports by utilizing transformer-based natural language processing (NLP) models. Include recommendations for additional imaging, testing, or follow-up appointments in the prepared report, along with actionable insights and next steps for radiologists [7]. Make the system flexible enough to accommodate various radiology imaging modalities (e.g., brain MRIs, abdomen CT scans, or chest X-rays) as well as the various report formats and language used in various medical specialties. Increase the availability of high-quality medical reports and interpretations in underserved or isolated areas where radiologists may be hard to find. Improve the interpretability of reports produced by AI and make sure that medical experts may examine and validate them before using them to make therapeutic decisions [8]. In order to make an appropriate diagnosis, medical image quality is essential. Important details may be obscured by poor image quality, which could cause misunderstandings and postpone treatment. This project makes sure that even difficult images (such as low-contrast scans or scans with artifacts) are enhanced for greater visibility of anomalies including cancers, fractures, and lesions by implementing contrast-based image enhancement algorithms. More accurate diagnostic results are a direct result of this improvement. Despite their experience, radiology reports are typically written by human specialists, human error and variability in interpretation can occur [9].

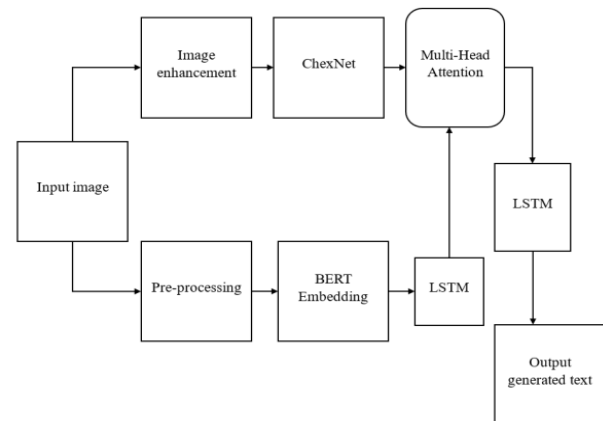


Fig 1: Block diagram

2.METHODOLOGIES

Radiology reports are produced from medical pictures utilizing a structured procedure by the Automatic Radiology Report Generator employing Transformer with Contrast-Based Image Enhancement technology. Using transformer-based natural language processing, the system creates written reports based on the improved photos after integrating image enhancement algorithms to increase image quality. A medical image is first fed into the system. This image may be in one of the regularly used radiology diagnostic formats, such as X-ray, CT scan, or MRI. Preprocessing is the process of applying contrast-based image enhancement algorithms to the input image. This can use techniques such as Adaptive Histogram Equalization (AHE) or Histogram Equalization. Enhancing the visibility of critical features in the image—like lesions, tumors, or abnormalities—that may be challenging to spot in low-contrast pictures is the aim. The model finds it easier to extract pertinent elements from enhanced photos. Important visual aspects that will be used to explain the findings in the radiology report are identified by the CNN. These characteristics are essential for making the diagnosis. After that, the image's features are mapped to the appropriate textual descriptions. A transformer model that has been trained is used in this step. The retrieved features and their mapped descriptions are used by the transformer model to create a textual radiological report.

2.1 Dataset Collection

The suggested approach makes use of the ChestX-ray dataset. When creating autonomous radiology report generating systems, particularly those that use sophisticated models like transformers with contrast-based image augmentation, the ChestX-ray dataset is

essential. Models are trained using ChestX-ray datasets, which provide image pairings together with the radiologist reports that go with them. These datasets aid in the development of models that are able to link X-ray visual patterns to diagnostic and medical terminology. Duplication data for comparable diagnoses is minimized in normal diagnoses, and minor diagnoses are collected in normal and abnormal to lessen imbalances. Prior to being implanted, the diagnoses, which are medical reports, undergo pre-processing. Word deconstruction, character and number removal, and letter conversion to lowercase are all part of the medical report's pre-processing. After that, it is trained using BERT to extract features from text data and obtain text embedding. Additionally, the medical report is filtered according to the data's occurrence. The data was divided into training, validation, and test data according to the percentage of occurrences. The highest frequency group is undersampled, and the lowest frequency group is oversampled, according to the number of recorded occurrences used to segment the data. The percentage of each minor and major data is then used to split the data into train, validation, and test data.

2.2 Image Enhancement Method

To increase the communication range and dependability, put in place a relay-based network. Relay Node Placement and Selection: Based on signal strength, distance, or energy, this algorithm establishes the best relay node placement and selection criteria. Relay Handover Mechanism: Maintains the strongest communication link by controlling relay switching. An essential part of improving underwater wireless optical communication (UWOC) systems' communication range and dependability is the relay network module. This module ensures strong and long-lasting underwater communication lines by reducing signal attenuation and scattering effects by adding relay nodes to the communication network. use algorithms to find the best location for relay nodes, guaranteeing maximum coverage and little signal loss. To guarantee effective network topology, variables including water turbidity, node distance, and environmental factors are taken into account. uses variables such as distance measurements, energy efficiency, and signal intensity to dynamically choose the best relay node. Link quality monitoring in real time aids in network adaptation to shifting circumstances.

optimizes the number and placement of relay nodes to reduce energy consumption. This is particularly important for underwater gadgets that run on batteries. Continuous data transmission is ensured by seamless switching, which controls relay node changeover without interfering with

the communication channel. Handover to the path with the best link quality is carried out by utilizing a channel-aware algorithm that assesses the quality of alternative relay paths. Load balancing: To avoid overloading and guarantee steady network performance, data traffic is divided across several relay nodes. By linking several relay nodes, multi-hop communication allows for multi-hop data transmission and greatly increases the communication range. The network is adaptable for bigger underwater deployments thanks to its scalable design, which can accommodate the addition of more relay nodes. Cluster-Based Relay Networks: Enhance network efficiency and structure by implementing clustering techniques in which nodes in a particular region connect with one another via a specified cluster head relay. Latency Reduction: Minimizes the number of hops needed for data transmission and optimizes relay placement to reduce communication delay.

2.3 Image Encoder

Data from 14 chest X-ray datasets were used to train ChexNet, a 121-layer dense convolutional network (DenseNet). DenseNet can be used to optimize existing networks by enhancing gradients and information across the network. More X-ray pictures were used to train the DenseNet model on top of the pre-trained ChexNet. ChexNet is used as an encoder in this study to extract picture attributes as convolutional features from the data. Through the model's last unfrozen layer, the pretrained ChexNet will send the weight to the image. We eliminate the top layer on ChexNet in order to modify the output into 1024 as a dimensional picture beginning features value. A batch size of 100, a dropout rate of 0.2, a learning rate of 10⁻², and sigmoid activation were the parameters used in this investigation. After that, the output vector size of 512 and ReLU activation will be used to forward the pooling of the initial weight initiation. In order to obtain context information and text features in the multi-head attention, this dimension will be modified via BERT embedding.

2.4 Multi-Head Attention

In order to align the data with the semantic features as a query in a parallel process, the encoder's initial weight will be sent as an input to multi-head attention as a key and value. Technically speaking, attention is the process of mapping a query to the keys from data that has a value as an output. It is made up of three parts: value, key, and query. Multiple single attentions that function frequently and concurrently are referred to as multi-head attention. The multi-head attention function can be represented as $\text{multihead}(k,q,v)=\text{concat}(\text{head}1..n) W$

In this case, k , q , and v stand for the report embedding and image features. Mapping a query using key-value pairs to obtain a new value is known as attention. In which the values, keys, and query are all vectors. The compatibility function between the query and keys determines the weights given to each value. The sum of the weighted values is used to calculate the final output. This method is particularly effective at correlating textual and picture information to provide context. The context vector, which is then supplied as input to the decoder, is the attention weight that is produced.

2.5 Bert Embedding

The pre-trained BERT embedding is used to extract the semantic characteristics. Google released a training method called Bidirectional Encoder Representations from Transformers (BERT), which is based on neural networks. A whole set of words from a sentence or inquiry can be used by BERT to train a language model in both directions. BERT enables language models to examine a word's context by considering the words that surround it, rather than only the ones that come before or after it. BERT examines the contextual relationships between words in the medical report by using an attention mechanism during the data encoding phase. To understand the word's context, the encoder will read the entire word.

2.6 Evaluation

The model's ability to describe the image will be evaluated using the Bilingual Evaluation Understudy (BLEU). The BLEU technique methodology counts word occurrences in the model-generated text using real data as a guide. The ngram calculation is the process of figuring out how well the words generated by the model match the real data. The value of the sentence's n-gram words will be computed in order to compare the expected text quality value with the reference data.

$$BLEU = BP \cdot \exp \sum_{n=1}^N w_n \log p_n$$

p_n : Precision of predicted text for nn -grams.

w_n : Weight assigned to the nn -gram word.

N : Gram value used (typically 4 for candidate sentences).

BP : Brevity penalty (BP penalizes errors in predicted text length).

In contrast, BP is a shortness penalty for predicted text faults and is a critical component that is carefully crafted to assess and penalize anticipated text length errors.

Together, the precision value p of the anticipated text, the weight given to words w , and the selected gram value N help to calculate blood pressure in a subtle manner. When comparing candidate sentences to reference sentences, it is common practice to assign candidate sentences a gram value of 4. This indicates that the four words in the forecast are analyzed to determine how closely they resemble the reference sentence. The calculation of the shortness penalty is made more complex by the gram value, represented by the letter N , which is crucial in establishing the scope of the comparison. Fundamentally, the n-gram value turns into a crucial factor in this analysis, directing the matching procedure and influencing the accuracy evaluation of the predicted text in comparison to the reference text. This thorough comprehension of the relationship between precision, brevity penalty, and gram values provides a thorough awareness of the complexities involved in evaluating the accuracy and concision of predicted text in natural language processing tasks.

RESULT & DISCUSSION

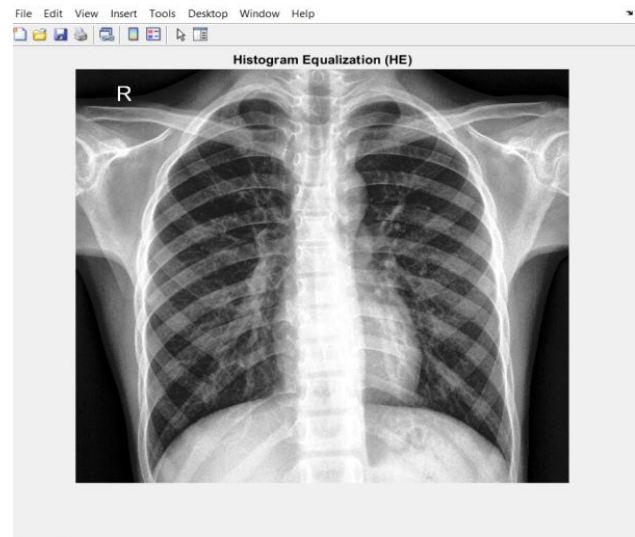


Fig.2. Histogram Equalization

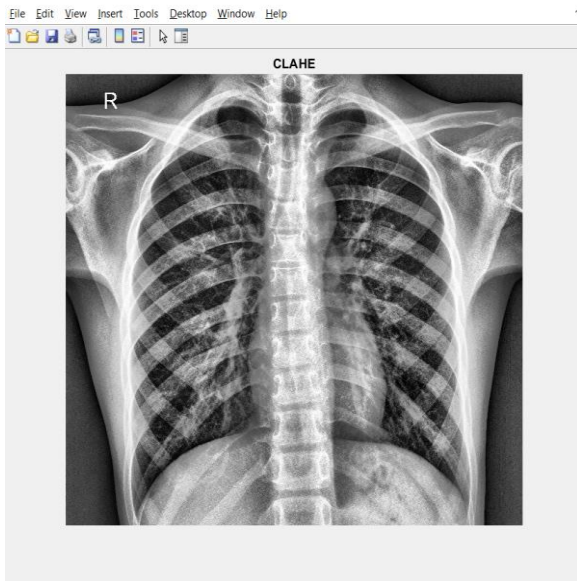


Fig 3: CLAHE

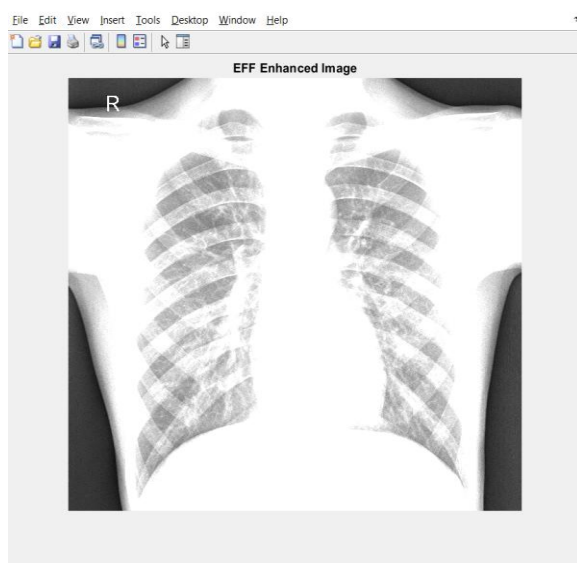


Fig 4: EFF Enhanced Image

CONCLUSION

An automated radiological diagnosis generator that makes use of a transformer architecture's transforming powers. Our main goal is to clarify the complex relationships among model design, evaluation measures, and image improvement procedures in the context of medical picture captioning. The suggested approach uses BERT embedding to extract complex text features and integrates a Multi-Head Attention (MHA) mechanism. The investigation of

different enhancement procedures and their effects on radiological pictures was then carried out in order to improve the model's overall effectiveness. The Gamma Correction technique proved to be the most successful of the four enhancement techniques used, all of which shown performance gains over the original photos. This sophisticated comprehension of picture improvement methods and how they affect enhancing model performance. Contrast noise is a problem that frequently affects radiography images and has a big impact on the evaluation results of generated captions. Our results demonstrate how important it is to improve visual contrast in order to allow the model to extract richer and more contextually relevant data. This subtle realization is in line with current debates about how crucial preprocessing procedures are to maximizing model performance for medical image analysis.

REFERENCES

- [1] L. I. T. Lee, S. Kanthasamy, R. S. Ayyalaraju and R. Ganatra, "The Current State of Artificial Intelligence in Medical Imaging and Nuclear Medicine," *British Institute of Radiology Journal*, vol. 1, no. 1, 2019.
- [2] J. Ker, L. Wang, J. Rao and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Special Section On Soft Computing Techniques For Image Analysis In The Medical Industry Current Trends, Challenges And Solution*, pp. 9375-9389, 2018.
- [3] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. A. Maadeed, S. M. Zughair, M. S. Khan and M. E. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19," *Computers in Biology and Medicine*, vol. 132, 2021.
- [4] H. Tsaniya, C. Fatichah and N. Suciati, "Exposure Fusion Framework in Deep Learning-Based Radiology Report Generator," *IPTEK Journal of Science and Technology*, vol. 33, 2022.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and L. Polosukhin, "Attention Is All You Need," *Neural Information Processing System*, 2017.
- [6] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren and A. Y. Ng, "Chexnet: radiologistlevel pneumonia detection on chest x-rays with deep learning," *Stanford ML Group Project: CheXNet*, 2017.

[7] X. Wang, Y. Peng, L. Lu and Z. Lu, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.

[8] M. Ivašić-Kos and I. Hrga, "Deep Image Captioning: An Overview," in International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2019.

[9] Adriyendi, "A Rapid Review of Image Captioning," Journal of Information Technology and Computer Science, vol. 6, pp. 158-169, 2021.

[10] Z. Khong, Y. Cui, Z. Xia and H. Lv, "Convolution and Long Short-Term Memory Hybrid Deep Neural Networks for Remaining Useful Life Prognostics," Applied Science, vol. 9, no. Machine Fault Diagnostics and Prognostics, 2019.