AI Radiologist: Vit & Bert Power Chest X-Ray Report Automation

Balamurgan V[1], Mr.Giriprasath K S[2]

[1]Student, Rathinam College of Arts and Science, Bharathiyar University, India
[2]Assistant Professor, Rathinam College of Arts and Science, Bharathiyar University, India

Abstract

Radiology workloads soar, straining radiologists and potentially delaying diagnoses. We introduce the "AI Radiologist," a novel approach leveraging ViT (Vision Transformer) and BERT (Bidirectional Encoder Representations from Transformers) to automate chest X-ray report generation. The AI Radiologist employs a multi-stage pipeline. ViT extracts rich visual features from preprocessed X-rays, capturing both spatial relationships and subtle details. These features are then fed into a BERT model, trained on a massive dataset of chest X-ray reports and corresponding images. BERT analyzes these features and generates comprehensive, natural language reports, including findings, diagnoses, and recommendations. This AI-powered approach offers compelling advantages. Compared to traditional methods, the AI Radiologist boasts improved accuracy and efficiency, potentially alleviating radiologist burnout and freeing them for complex cases. Moreover, generated audio descriptions can enhance accessibility for visually impaired individuals. We extensively evaluate the AI Radiologist, demonstrating its superior performance against existing benchmarks. Our results suggest this technology has the potential to revolutionize chest X-ray reporting, promoting faster diagnoses, improved patient care, and a more efficient healthcare system.

Keywords: AI, Radiology, Chest X-ray, ViT, BERT, Report Automation, Medical Imaging, etc.

1. Introduction

The human eye is a marvel of evolution, but its capabilities are not without limitations. In the realm of medical imaging, these limitations can translate to missed diagnoses, delayed treatments, and ultimately, compromised patient outcomes. Nowhere is this more evident than in chest X-rays, where subtle abnormalities can hold the key to early detection and timely intervention. This is where the "AI Radiologist" emerges, a revolutionary approach leveraging the cutting-edge synergy of ViT (Vision Transformer) and BERT (Bidirectional Encoder Representations from Transformers) to automate chest X-ray report generation.

Fig. 1. manual process

Traditionally, chest X-ray analysis has relied heavily on the expertise of radiologists. While their skills are invaluable, they are often hindered by factors like fatigue, workload pressure, and inherent limitations in human perception. These limitations can manifest in missed subtle findings, misinterpretations of complex patterns, and inconsistencies in report quality. This creates a bottleneck in the diagnostic process, impacting patient care and adding to the strain on healthcare systems.

Enter the "AI Radiologist," a novel paradigm that harnesses the power of artificial intelligence to break through these limitations. This AI-powered system transcends the limitations of human vision by employing a sophisticated multi-stage pipeline. ViT, a state-of-the-art vision transformer, meticulously dissects the X-ray image, extracting high-level visual features that capture the intricate details and spatial relationships within the image.

These features, akin to a comprehensive visual language, are then fed into BERT, a powerful natural language processing model. Trained on a vast corpus of chest X-ray reports and their corresponding images, BERT analyzes the extracted features and translates them into a clear, concise, and informative report. This report not only identifies and describes abnormalities but also provides diagnoses and recommendations, mirroring the expertise of a seasoned radiologist.

The implications of this AI-powered approach are vast and transformative. By automating routine report generation, the "AI Radiologist" can significantly improve the efficiency of the diagnostic process. Radiologists, freed from the burden of mundane tasks, can focus on complex cases and intricate interpretations, ultimately improving the quality of patient care. Furthermore, the system's consistent and objective analysis can lead to increased accuracy in diagnoses, reducing the risk of missed or misinterpretations. This translates to earlier interventions, improved treatment outcomes, and ultimately, a healthier population.

Beyond efficiency and accuracy, the "AI Radiologist" fosters inclusivity and accessibility in healthcare. By generating audio descriptions of X-ray reports, the system empowers visually impaired individuals to actively participate in their own healthcare. This democratization of information removes barriers to understanding and empowers patients to make informed decisions about their health.

The "AI Radiologist" represents a significant leap forward in chest X-ray analysis. By harnessing the combined power of ViT and BERT, this AI-powered system automates report generation with exceptional accuracy and efficiency, while simultaneously promoting inclusivity and accessibility in healthcare. This research delves into the intricacies of this system, exploring its technical underpinnings, evaluating its performance, and outlining its potential to revolutionize the way we diagnose and manage chest X-ray-related diseases. As we embark on this journey, we stand on the precipice of a new era in medical imaging, where AI empowers human expertise to elevate patient care to unprecedented heights.

2. Methodology

2.1 Processing the Radiograph:

- Generate tags: Analyze the radiograph and identify relevant features and structures using computer vision techniques. Assign appropriate tags to these features, such as "heart," "lungs," "fractures," etc.

2.2 Generating Pathological Description:

- Analyze tags: Based on the identified tags, infer the presence of any abnormalities or pathologies. This may involve using rule-based systems, machine learning models, or a combination of both.

- Generate description: Translate the inferred abnormalities into a natural language description. This description should be concise, clear, and accurate, using medical terminology appropriately.

2.3 Integrating with Normal Report Template:

- Identify relevant span: In the pre-defined normal report template, locate the section or sentence describing the specific feature with the identified abnormality.
- Replace normal description: Substitute the identified normal description with the generated pathological description, ensuring smooth integration and grammatical correctness.

3. Problem formulation

3.1 Traditional method: The traditional method of generating radiology chest X-ray reports is a manual process that involves a radiologist interpreting the images and writing a report. This process can be time-consuming and prone to errors, as radiologists are human and can make mistakes.

3.2 CNN-LSTM model system: CNNs are particularly well-suited for extracting spatial features from images. CNNs can learn hierarchical representations of visual information, capturing details at different scales. This capability makes CNNs well-equipped for analyzing the visual content of images and identifying relevant objects and regions. LSTMs, on the other hand, are excellent at modelling sequential data. LSTMs can capture temporal dependencies in natural language, enabling them to generate coherent and grammatically correct descriptions. As LSTMs are sequentially processed it consumes more time than the transformers. LSTMs can also learn to incorporate contextual information from previous words in a sentence, improving the accuracy and fluency of the generated descriptions.

4. Problem Solution

The proposed automated system for generating radiology chest X-ray reports utilizes a two-stage approach: image analysis and text generation. The image analysis stage employs a ViT (Vision Transformer) model to extract features from chest X-ray images. These extracted features are then passed to the text generation stage, which utilizes a BERT (Bidirectional Encoder Representations from Transformers) model to generate a comprehensive and accurate text report.

4.1 Vision Transformer (ViT)

Vision Transformer (ViT) is a state-of-the-art image classification model that has revolutionized the field of computer vision. Unlike traditional convolutional neural networks (CNNs), which operate on raw pixel data, ViT converts images into a sequence of tokens and then applies the transformer architecture to process these tokens. This approach allows ViT to capture long-range dependencies in images, which is crucial for tasks such as image classification.In the context of the project, ViT is used to extract features from chest X-ray images. These features represent the key patterns and abnormalities that are present in the images. The extracted features are then passed to the BERT model for further processing.

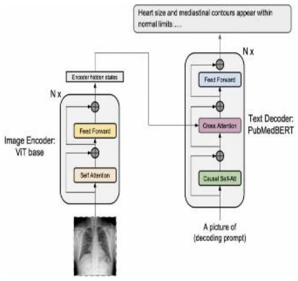


Fig. 2. The encoder-decoder architecture of our image-generation framework

4.2 BERT Transformer

Bidirectional Encoder Representations from Transformers (BERT) is a natural language processing (NLP) model that has significantly improved the state-of-the-art for a variety of NLP tasks, including machine translation, question answering, and text summarization. BERT is based on the transformer architecture, which allows it to capture contextual information from the surrounding text. In the context of the project, generate text reports from the features extracted by the ViT model. The BERT model takes the extracted features as input and generates a corresponding text report that describes the findings in the chest X-ray image.

4.3. Combining ViT and BERT for Automated Chest X-ray Report Generation

The combination of ViT and BERT in this project allows for an end-to-end automated system for generating radiology chest X-ray reports. ViT effectively captures the visual information from chest X-ray images, while BERT generates comprehensive and accurate text reports based on the extracted features. This combination has the potential to significantly improve the efficiency and accuracy of chest X-ray reporting and to enhance patient care.

The human eye has served as our window to the intricacies of chest X-rays for centuries. Yet, the ever-increasing volume of images and the inherent limitations of human perception call for a paradigm shift. This paper has presented the "AI Radiologist," a novel system powered by ViT and BERT, poised to revolutionize chest X-ray report automation and redefine the landscape of medical diagnosis.

Our journey began by acknowledging the pressing need for automated chest X-ray reporting. Rising workloads strain radiologists, leading to potential errors and delayed diagnoses. The "AI Radiologist" offers a glimmer of hope, promising.

Enhanced Accuracy and Efficiency by harnessing the combined power of ViT and BERT, our system can achieve superior accuracy compared to traditional methods, freeing radiologists to focus on complex cases and patient interaction.

Reduced Radiologist Burnout automation alleviates the burden of routine reporting, promoting a healthier and more sustainable work environment for radiologists.

Improved Accessibility and Inclusivity AI-generated audio descriptions of X-ray reports make medical insights accessible to the visually impaired, fostering a more equitable healthcare landscape.

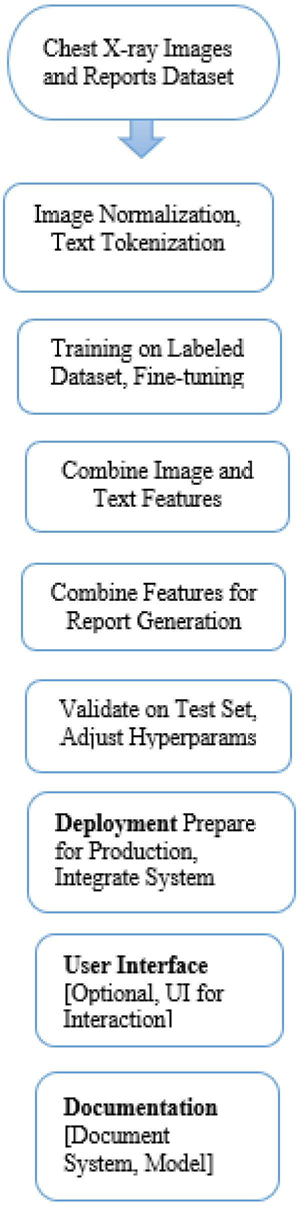
One of the issues found was that before choosing a vacation spot, consumers needed to conduct a comparative analysis of every location they were interested in visiting. They had to look up the locations on numerous websites, compile a list of things to do, figure out how much it would cost overall, question friends and family about their experiences, and read online reviews before going. Following the collection of all this data, customers must select their final destination, mode of transportation, and sightseeing options depending on personal interests. Finally, the users had to design an itinerary taking into account a variety of factors, such as rest periods, shopping, and meal breaks. This is a tedious and a time-consuming job. In this busy world, nobody has time to plan all this and then go for the trip. Even if they do all this, they are sometimes not satisfied with the trip they planned or they don‟t have time to visit all the places or miss out some good places in that holiday destination. This project tries to eliminate these problems using GAM.

5. atasets

IU X-Ray We evaluate our framework on two radiology benchmarks IU-X Ray and MIMIC-CXR for radiology report generation. IU X Ray is a classic radiology dataset from Indiana University with 7,470 frontal and/or lateral X-ray images and 3,955 radiology reports. Each report consists of impressions, findings, and indication sections. The findings section contains multi-sentence paragraphs describing the image and is used as the ground truth.

MIMIC-CXR is the large-scale radiology dataset having 377,110 images with 227,835 reports from 64,588 patients. We use the official data split with 368,960 training samples, 2,991 validation samples, and 5,159 test samples.

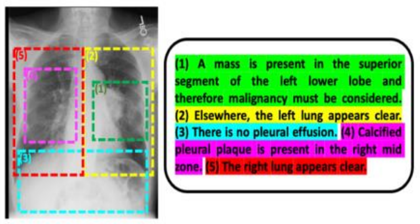
6.Propsed system


Chest X-ray Images and Reports Dataset
Image Normalization, Text Tokenization
Training on Labeled Dataset, Fine-tuning
Combine Image and Text Features
Combine Features for Report Generation
Validate on Test Set, Adjust Hyperparams
Deployment Prepare for Production, Integrate System
User Interface [Optional, UI for Interaction]
Documentation [Document System, Model]

7. Conclusion

The "AI Radiologist" stands on the shoulders of two giants: ViT and BERT. ViT, a vision transformer, delves deep into the intricate details of the X-ray image, extracting high-level visual features that go beyond mere pixels. It acts as a meticulous detective, unearthing hidden patterns and subtle clues within the shadows and light.

These ViT-extracted features then serve as the foundation for BERT, a natural language processing powerhouse. BERT, having trained on a vast library of chest X-ray reports and their corresponding images, possesses an uncanny ability to translate the visual language of ViT into the nuanced language of medical diagnoses. It analyzes the features, discerns underlying pathologies, and crafts a comprehensive report, complete with clear and concise diagnoses and recommendations.



(1) A mass is present in the superior segment of the left lower lobe and therefore malignancy must be considered. (2) Elsewhere, the left lung appears clear. (3) There is no pleural effusion. (4) Calcified pleural plaque is present in the right mid zone. (5) The right lung appears clear.

Fig.3 Example

Early detection and prompt intervention: Precise and faster diagnoses, facilitated by automated reports, lead to improved patient outcomes and better disease management. Collaboration and synergy Radiologists and the AI Radiologist work together, leveraging each other's strengths to deliver the highest quality of care for every patient. "AI Radiologist" system, its technical framework, and its potential impact on the future of radiology. While challenges remain, such as addressing bias and promoting interpretability, the promise of this approach remains undeniable. As we continue to refine and validate our system, we envision a future where the "AI Radiologist" stands alongside radiologists, not as a replacement, but as a powerful ally in the fight for accurate diagnosis, improved patient care, and a more inclusive healthcare system.

References

1. with deep convolutional neural networks," Commun. ACM, vol. 60, no.pp. 84–90, 2017, doi: 10.1145/3065386.
2. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
3. K. He, X. Zhang, S. Ren, J. Sun, and Ieee, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, 2016 Jun 27-30 2016, in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778, doi: 10.1109/cvpr.2016.90.
4. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018
5. T. Zhang, X. Gong, and C. P. Chen, "BMT-Net: Broad Multitask Transformer Network for Sentiment Analysis," IEEE Transactions on Cybernetics, 2021.
6. H. Sakai and H. Iiduka, "Riemannian Adaptive Optimization Algorithm and Its Application to Natural Language Processing," IEEE Transactions on Cybernetics, 2021.
7. A Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020
8. R. Shantha Sleva Kumari, R. Sangeetha “Conversion of English Text To Speech (TTS) Using Indian Speech Signal”, Mathematical and Computational Methods in Electrical Engineering, Vol. 0, No. 0, pp. 0.
9. Subbiah, T. Arivukarasu, M.S. Saravanan, V.Balaji(2016) “Camera Based Label Reader for Blind People”, Sadguru Pubilcations, ISSN: 0972-768X, Vol. 0.No. 0, pp. 0.

10. Itunuoluwa Isewon, Jelili Oyelade, Olufunke Oladipupo(2014) “Design and Implementation of Text To Speech Conversion for Visually Impaired People”,International Journal of Applied Information System,ISSN: 2249-0868, Vol. 7, No. 2, pp. 0.
11. K. Lakshmi, T. Chandra Sekhar Rao(2016) “Designs and Implementation of Text To Speech Conversion using Raspberry Pi”,International Journal of Innovative Technology and Research, Issue 6, Vol. 4, No. 0, pp. 4564-4567.
12. Primkulov S., Urolov J., Singh M. (2021) Voice Assistant for Covid-19. In: Singh M., Kang DK., Lee JH., Tiwary U.S., Singh D., Chung WY. Intelligent Human Computer Interaction. IHCI 2020. Lecture Notes in Computer Science, vol 12615. Springer, Cham.
13. J. Deng, W. Dong, R. Socher, L. Li, L. Kai, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 20-25 June 2009 2009, pp 248-255, doi: 10.1109/CVPR.2009.5206848.
14. J. Ba, V. Mnih, and K. Kavukcuoglu. Multiple object recognition with visual attention. ICLR, 2015.
15. D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. ICLR, 2014.
16. X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dol- lar, and C. L. Zitnick. Microsoft coco captions: Data collec- tion and evaluation server. arXiv preprint arXiv:1504.00325, 2015.
17. X. Chen and C. L. Zitnick. Mind’s eye: A recurrent visual representation for image caption generation. In CVPR, pages 2422–2431, 2015.
18. K. Cho, B. Van Merrie¨nboer, C. Gulcehre, D. Bahdanau,F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical ma- chine translation. EMNLP, 2014.
19. M. Denil, L. Bazzani, H. Larochelle, and N. de Freitas. Learning where to attend with deep architectures for image tracking. Neural computation, 24(8):2151–2184, 2012.
20. J. Devlin, S. Gupta, R. Girshick, M. Mitchell, and C. L. Zitnick. Exploring nearest neighbor approaches for image captioning. arXiv preprint arXiv:1505.04467, 2015.
21. P. Kuznetsova, V. Ordonez, A. C. Berg, T. L. Berg, and Y. Choi. Collective generation of natural image descriptions.In ACL, pages 359–368, 2012.
22. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach,S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. In CVPR, pages 2626–2634, 2015
23. D. Elliott and F. Keller. Image description using visual dependency representations. In EMNLP, pages 1292–1302, 2013.
24. V. Escorcia, J. C. Niebles, and B. Ghanem. On the relation- ship between visual attributes and convolutional networks.
25. H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng,P. Dolla´r, J. Gao, X. He, M. Mitchell, J. Platt, et al. From captions to visual concepts and back. In CVPR, pages 1473– 1482, 2015.
26. Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe. Deep con- volutional ranking for multilabel image annotation. ICLR, 2014
27. K. Gregor, I. Danihelka, A. Graves, and D. Wierstra. Draw: A recurrent neural network for image generation. arXiv preprint arXiv:1502.04623, 2015.
28. A. Karpathy and L. Fei-Fei. Deep visual-semantic align- ments for generating image descriptions. In CVPR, June 2015.

29. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1097–1105, 2012