e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

MEDICAL IMAGE CAPTIONING USING DEEP LEARNING

Tripti Gupta*1, Ankita Kumari*2, Ankita Agarwal*3

*1,2,3Department of Computer Science, Shri Ramswaroop Memorial College of Engineering & Management, Lucknow, Uttar Pradesh, India.

DOI : https://www.doi.org/10.56726/IRJMETS40909

ABSTRACT

Medical images are widely used in the medical domain for the diagnosis and treatment of diseases. Reading a medical image and summarizing its insights isa routine, yet Medical image captioning is an emerging field that involves generating textual descriptions for medical images using deep learning techniques. This review paper provides a comprehensive analysis of the state-of-the-art techniques in medical image captioning using deep learning. The paper discusses different deep learning models, evaluation metrics, challenges, and future directions in this field. Nonetheless time-consuming task, which often represents a bottleneck in the clinical diagnosis process.

Automatic report generation can relieve the issues. However, generating medical reports presents two major challenges: (i) it is hard to accurately detect all the abnormalities simultaneously, especially the rare diseases;

(ii) a medical image report consists of many paragraphs and sentences, which are longer than natural image captions. We present a new framework to accurately detect the abnormalities and automatically generate medical reports.

I. INTRODUCTION

Medical image captioning is a rapidly growing field in the area of computer vision and natural language processing. It involves the generation of textual descriptions for medical images, which can aid healthcare professionals in diagnosis and treatment planning. In recent years, deep learning techniques have shown great promise in this area, particularly with the use of Encoder-Decoder models, Attention models and Context Aggregation Modules. This review paper presents a comprehensive analysis of the various deep learning models that have been proposed for medical image captioning, with afocus on the three models implemented in this project the Encoder-Decoder model, Attention model and Context Aggregation module. These models are evaluated based on their performance in generating accurate and meaningful captions for a range of medical images.

The paper begins with an overview of medical image captioning, including itsimportance in the field of healthcare and the challenges associated with it. This is followed by a detailed description of the three deep learning models employed in the project, highlighting their strengths and limitations.

Next, the paper presents a review of theexisting literature on medical imagecaptioning, focusing on the use of deep learning techniques. The review covers arange of studies that have implemented different deep learning models andprovides insights into the state-of-the-art techniques used in this field.Finally, the paper concludes with a discussion of the results obtained from the implementation of the three deep learning models in this project. The results areanalyzed in terms of their accuracy, efficiency, andoverall performance, andpotential future work in this area is identified.

Overall, this review paper provides a valuable contribution to the field of medicalimage captioning, highlighting the potential of deep learning models andproviding insights into their effectiveness for generating accurate andmeaningful descriptions for a range of medical images.

II. LITERATURE SURVEY

The paper [1]proposesa multimodal recurrent neural network(m-RNN) for image captioning. The m-RNN model is trained on a large-scaledataset of image-caption pairs and is capable of generating captions formedical images.

The paper [2] proposes a novel encoder-decoder model for radiology reportgeneration. Also applied visual attentions in a late fusion fashion, and enrichesthe semantics involved in the hierarchical LSTM decoder with medical concepts.

The paper [3] proposes a new framework to learn to detect disease, and generate medical reports from the initial images. This paper conduct medical annotation generation experiments on IU X-Ray dataset. Author introduce GLP mechanism, matching mechanism, context and semantic attention to hierarchical RNN.

The paper [5] presents an overview of applied approaches and their performance,as well as the task description, participation and distributed data set for the ImageCLEF 2020 concept detection task.

In this paper [8] author introduce visual attention mechanism based on the encoder-decoder structure, a novel show, attend, and tell model has been designed and implemented. In this paper, SPEA-II has been used to tune the initial attributes of an SPEA-II-based.

The paper [11] proposes a novel Knowledge-driven Encode, Retrieve, Paraphrase (KERP) approach and retrieval-based methods. Basically KERP is used to decomposes medical report generation into explicit medical abnormality graph learning.

This workproposes a knowledge distillation-based approach to improve the efficiency ofneural network models for medical image captioning. The model is trained on adataset of medical images and their corresponding captions and is capable ofgenerating accurate and descriptive captions for medical images. In this paper [12] author proposed an efficient system to learn, detect disease, and explain their settings. This framework uses Medical Subject Headings (MeSH) annotation.By summarising the CNN/RNN outputs and their states for each of the image/text instances, author proposed a strategy to mining joint contexts from a collection of images and their associated text.

The paper [14] author demonstrate an effective approach for categorising and retrieving images from medical image databases, particularly huge radiograph archives. Also demonstrated the optimum performance using dense sampling of straightforward features with spatial information and an SVM classifier with a nonlinear kernel.

The paper [16] proposes that key findings in chest radiology reports can be captured using a narrowly controlled vocabulary and post-coordination of terms for later indexing and retrieval in a search engine. The terms that is used in this paper are MeSH and Radlex terms to built a vocabulary.

PROBLEM STATEMENT

The problem we are going to solve in this case study is Medical image captioning or report generation.

We have to extract features (bottleneck) from images. Then use these extracted features to predict the captions.

Our input will be a medical image and the output will be sequence of character.

III. PROPOSED METHODOLOGY

The proposed methodology for medical image captioning using deep learning typically involves the following steps:

Data preprocessing: The first step is to preprocess the medical image dataset to extract relevant features, such as image intensity, texture, and shape, and to prepare the dataset for training the deep learning models. This step may involve data augmentation techniques, such as rotation, scaling, and cropping, to increase the diversity of the dataset.

Model architecture design: The next step is to design a deep learning architecture that can effectively extract features from medical images and generate clinically relevant captions. Commonly used architectures include convolutional neural networks (CNNs) for image feature extraction and recurrent neural networks (RNNs) for sequence generation. Attention mechanisms can also be incorporated to focus on important regions of the image.

Training and evaluation: The model is then trained on the preprocessed dataset using appropriate optimization techniques, such as backpropagation and gradient descent. The model is evaluated using metrics such as BLEU, ROUGE, and CIDE are to assess the quality and relevance of the generated captions.

Fine-tuning: The model can be fine-tuned on new datasets or with additional training samples to improve its accuracy and generalization ability.

Interpretation: The final step is to interpret the generated captions and assess their clinical relevance and accuracy. This involves comparing the generated captions to ground truth annotations and evaluating their usefulness for clinical applications.

Deployment: Once the model has been trained and optimized, it can be deployed in clinical settings to assist medical professionals in analyzing medical images and making accurate diagnoses.

Overall, the proposed methodology of medical image captioning using deep learning involves a combination of computer vision and natural language processing techniques to generate accurate and informative captions for medical images.

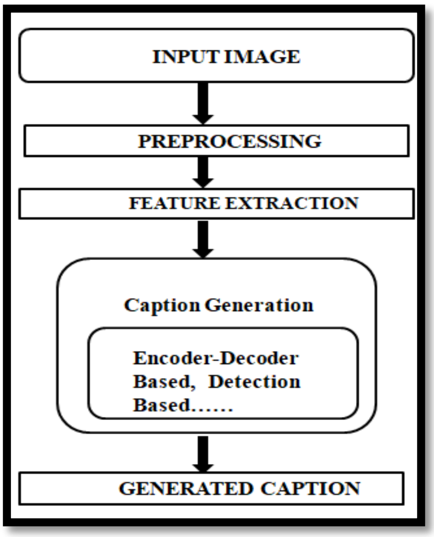
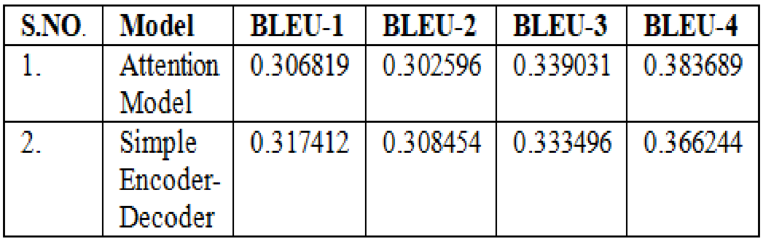


Fig 1: Flowchart of our proposed methodology

In this, Encoder-decoder model (CNN+LSTM) as well as RNN for generating a caption related to chest X-ray and trained on a Indiana University datasets.

The Accuracy of Attention model and Simple Encoder-decoder model are in the below table.



S.NO.	Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
1.	Attention Model	0.306819	0.302596	0.339031	0.383689
2.	Simple Encoder-Decoder	0.317412	0.308454	0.333496	0.366244

Fig 2: Accuracy of the model

IV. CONCLUSION

In conclusion, medical image captioning using deep learning is an emerging area of research with significant potential for improving healthcare outcomes. Encoder-decoder models, attention models, and context aggregation module shave been proposed as effective techniques for generating accurate and informative captions for medical images. However, the development of better evaluation metrics and the integration of clinical data and patient history remains a challenge for this field.

V. REFERENCES

[1] Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang, Alan Yuille: ” Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN)”, in conference on ICLR 2015.

[2] Jianbo Yuan1, Haofu Liao1, Rui Luo2, and JieboLuo:, “Automatic Radiology Report Generation basedon Multi-view Image Fusion andMedical Concept Enrichment.”, 2019.

[3] Changchang Yin, Buyue Qian, Jishang Wei, Xiaoyou Li, Xianli Zhang, Yang Li, Qinghua Zheng: "Automatic Generation Of Medical Imaging Diagnostic Report with Hierarchical Reccurenrt Neural Network", in IEEE Conference on Data Mining, 2019.

[4] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics, 2002

[5] Obioma Pelka, Christoph M. Friedrich, Alba G. Seco de Herrera, Henning Mulle: "Overview of the ImageCLEFmed 2020 ConceptPrediction Task: Medical Image Understanding", 2020.

[6] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. Journal of Machine Learning Research, 2008.

[7] S. Bird, E. Klein, and E. Loper. Natural language processing with Python. " O'Reilly Media, Inc.", 2009.

[8] Arjun Singh, Jaya Krishna Raguru, Gaurav Prasad, Surbhi Chauhan, Pradeep Kumar Tiwari, Atef Zaguia, Mohammad Aman Ullah: "Medical Image Captioning Using Optimized Deep Learning Model", 2022.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2011.

[10] C. Y. Li, X. Liang, Z. Hu, and E. P. Xing, " Knowledge-driven encode, retrieve, paraphrase for medical image report generation," arXiv preprint arXiv: 1903.10122,2019.

[11] G. Kulkarni, V. Premraj, V. Ordonez, S. Dhar, S. Li, Y. Choi, A. C. Berg, and T. Berg. Babytalk: Understanding and generating simple image descriptions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(12):2891– 2903, 2013

[12] Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, Ronald M Summers: "Learning to read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation", in IEEE Conference on Computer Vision and Pattern Recognition, 2016

[13] Y. Shi, H.-I. Suk, Y. Gao, and D. Shen. Joint coupled-feature representation and coupled boosting for ad diagnosis. In CVPR, 2014.

[14] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. Medical Imaging, IEEE Transactions on, 2011.

[15] T. A. Ngo and G. Carneiro. Fully automated non-rigid segmentation with distance regularized level set evolution initialized and constrained by deep-structured inference. In CVPR, 2014.

[16] D.Demner-Fushman, S. E. Shooshan, L. Rodriguez, S. Antani, and G. R. Thoma. Annotation of chest radiology reports for indexing and retrieval. Multimodal Retrieval in the Medical Domain (MRMD) 2015.

[17] H.-C. Shin, L. Lu, L. Kim, A. Seff, J. Yao, and R. M. Summers. Interleaved text/image deep mining on a very largescale radiology database. In CVPR, 2015

[18] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko. Sequence to sequence– video to text. ICCV, 2015

[19] L. Van der Maaten and G. Hinton. Visualizing data using t-sne. Journal of Machine Learning Research, 2008.

[20] S. Liang, X. Li, Y. Zhu et al., "ISIA at the image clef 2017 image caption task," in Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017., 2017.

[21] C. Eickhoff, I. Schwall, A. G. S. de Herrera, and H. Muller, ¨ "Overview of imageclefcaption 2017 - image caption prediction and concept detection for biomedical images," in Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, Dublin, Ireland, September 11-14, 2017., 2017.