UNIVERSITY PARTNER

6CS014: Complex Systems

Literature Review on Medical Report Generation

<Individual Task>

Student Id: 220495
Student Name: Aashutosh Dahal
Group: <L4CG10>
Submitted to: Mr. Shiv Yadav
Word Count : 1527
Submitted on: <19-12-2022>

1. Introduction

Medical Report Generation is a process of automated generation of medical report using medical images, patients' data or using electronic medical reports using an artificial intelligence (AI) and machine learning algorithms which can properly analyze and interpret medical data and produce actual report which can be summary of the findings, diagnoses and treatment recommendations.

In the technological era many hospitals and clinics have been digitalized and the records all are being stored digitally into massive servers or databases. In some medical institutions medical treatments are being done using precise technologies. Using the patient's data, we can generate an automated medical report which can help reduce human errors and inefficiency in manually generated report by human doctors or nurses who might not be as experiences.

2. Aims

- To utilize patients' data including medical images, scans and different vitals to generate automated medical report

3. Objectives

- Using AI, ML and exploring the advancements in automate medical report generation
- Examining techniques for analyzation of medical data, including images, scans and EHRs.
- Evaluating benefits such as enhanced accuracy, reduced human error and efficiency
- To address Challenges like data privacy, algorithmic biases and technical limitations
- Comparing existing systems in terms of performance, scalability and adaptability
- To identify gaps, including patient centered designs and real time data involvement
- To provide recommendations for improving automated medical reporting systems and their adoption in healthcare.

4. Literature Review

4.1. Auxiliary Signal-Guided Knowledge Encoder-Decoder for Medical Report Generation

This paper by Mingjie Li, Fuyu Wang, Xiaojun Chang and Xiaodan Liang conducts rigorous experiments using different datasets to validate the robustness and generalization of the Auxiliary Signal-Guided Knowledge Encoder-Decoder (ASGK) framework. This experiment uses the CX-CHR and COVID-19 CT Report datasets. CX-CHR includes 45,598 images from 35,609 patients, marked with 173 medical tasks comprising 155 abnormal and 28 normal terminologies. In the case of the COVID-19 datasets it contains 728 images which are split into COVID-positive and COVID-negative cases, which includes 68 medical tags. The reports were tokenized and preprocessed to build a vocabulary covering over 98.7% of the corpus. Three key metrics were used for the evaluation which are the ROUGE-L, CIDER-D and the BLEU alongside human assessment and the dataset were split into training validation and tests set with a 7:1:2 split. The ASGK model achieved a CIDER-D score of 324.5 on CX-CHR and in the COVID-19 CT reports with a score of 68.4 which was an improvement of 35 points over the HRGR-Agent baseline and significantly outperforming state-of-the-art models. Ablation studies revealed that internal auxiliary signals boosted classification accuracy by 4.5%, while external signals increased report generation metrics, notably improving CIDER-D by 27.5 points on CX-CHR. The approach was further validated by the radiologist preferring ASGK's reports in 20-27% cases for their alignment with clinical observations. The model was consisted with the radiologists' workflows as it focused on diagnostically relevant regions. Hence, these findings highlight ASGK's potential to generate accurate, coherent medical reports, making it a potentially valuable tools for clinical applications. (Mingjie Li, 2020)

4.2. Cross-Modal Causal Intervention for Medical Report Generation

This paper by Weixing Chen, Yang Liu, Ce Wang, Jiarui Zhu, Shen Zhao, Guanbin Li, Cheng-Lin Liu, Liang Lin all fellow IEEE (Institute of Electrical and Electronics Engineers) evaluates its Visual-Linguistic Casual Intervention (VCLI) framework through experiments on the IU-Xray and MIMIC-CXR datasets, which focuses on the precision and reliability of medical report generation. This integrates a Visual Deconfounding Module (VDM) and Linguistic Deconfounding Module (LDM) which enhances the alignment of visual and linguistic features through casual front-door intervention to mitigate cross-modal biases. To improve multi-modal learning the pre-training employs Prefix Language Modeling (PLM) and Masked Image Modeling (MIM). The experiments results show that the VLCI surpasses the state-of-the-art methods. VLCI achieves significant improvements in BLEU-4 (0.190 vs 0.178) and CIDEr (0.592 vs 0.382) scores on the IU-Xray datasets comparing to the leading methods. Similarly, VLCI demonstrates superior clinical efficacy which achieves significant boost in F1 score (40.1% vs 37.3% by previous models) which shows a marked improvement in precision for abnormality detection for the larger and more comple (Mingjie Li, 2020)x MIMIC-CXR datasets. The evaluation particularly for complex abnormalities like edema and cardiomegaly reveals that VLCI excels at generating precise, unbiased and detailed reports. Even under complex casual conditions, outperforming non-casual models that struggle with spurious correlations qualitative analyses shows VLCI’s ability to accurately identify and to describe abnormalities. The frameworks lightweight architecture and reliance on casual inference enable efficient computation without external knowledge bases, proving its potential for scalable and reliable medical applications. (Weixing Chen, 2023)

4.3. Medical Report Generation for Retinal Images via Deep Models and Visual Explanation

This paper evaluates a DeepOpht framework through comprehensive experiments on a large-scale retinal disease dataset, DeepEyeNet (DEN) and smaller datasets manually annotated by ophthalmologists. The DEN datasets which comprised of 15,706 images and diverse labels, was used to train and validate the Retinal Disease Identifier (RDI) and Clinical Description Generator (CDG) modules. The RDI, leveraging ImageNet pre-trained deep models, outperformed baselines, achieving a Prec@5 accuracy of 80.75%, effectively narrowing down retinal disease candidates in real-world scenarios. Similarly, using a keyword-driven approach with improvements across BLEU-4, CIDEr and ROUGE metrics compared to non-keyword driven models the CDG demonstrated superior performance. For instance, reflecting improved text generation quality the BLEU-model increased from 0.012 to 0.032 with the keyword reinforcement. Qualitative evaluations confirmed that the activation maps aligned well with clinically relevant image features which were supported by visual explanation modules like CAM. Highlighting the framework’s ability to capture domain specific insights the alignment was validated by comparing model-generated maps to ground truth annotations. Additionally, to prove the framework’s utility in aiding ophthalmological diagnosis the manually labeled dataset es employed to showcase the clinical relevance of generated caption and explanations. The results emphasize on how DeepOpth’s potential to automate and enhance the traditional retinal disease treatment process through precise, interpretable and efficient report generation. (Jia-Hong Huang, 2020)

4.4. Automated Radiology Report Generation by Learning with Increasing Hard Negatives

This study by Bhanu Prakash Voutharoja, Lei Wang and Luping Zhou evaluates its framework for radiology report generation on the IU-Xray and MIMIC-CXR datasets using state-of-the-art backbones like R2GenCMN and XproNet. A new and innovative approach was introduced by this framework as it trained the model by synthesizing increasing harder negative samples, which challenged the model to identify fine-grained differences between image-report pairs, significantly improving the diversity and accuracy of the reports that were generated. The method achieved substantial improvements and showcased enhanced capability in capturing clinical nuances in key metrics, including boost in BLEU-4 from 0.165 to 0.183 and CIDEr from 0.426 to 0.459. Similarly, the framework demonstrated significant improvements in clinical efficacy with the F1 score rising from 0.255 to 0.137 which reflected higher [precision and recall in identifying thoracic abnormalities on the larger and more complex MIMIC-CXR datasets. The framework generated missing clinical observations such as ‘hilar contours are stable’ and ‘lungs are hyperinflated’ which were overlooked by baseline models was highlighted by qualitative analysis. Ablation studies confirmed the crucial role of the synthesized harder negatives, revealing improvements in CIDEr by over 24%, emphasizing the importance of aligning features from both visual and textual modalities. The frameworks robustness and generalizability are validated by the consistent gains across the datasets and metrics. The model advances automated medical report generation producing clinically accurate, diverse and interpretable outputs that hold significant potential for medical applications by addressing subtle differences and integrating multi-modal features effectively. (Bhanu Prakash Voutharoja, 2023)

4.5. Automated radiology report generation using conditioned transformers

The paper investigates automating radiology generation by using a conditioned transformer model (CDGPT2). The methodology in composed of three stages which are extracting visual features from chest X-rays using fine-tuned ChexNet model, computing weighted semantic features through pre-trained word embeddings and generating reports by conditioning a pre-training word embeddings and generating reports by conditioning a pre-trained DistilGPT2 transformer on these features. The architecture to combine visual and semantic cues to enhancer accuracy and the model is trained on the IU-Xray dataset. In experiments to better evaluate the reports quality CDGPT2 achieved notable improvements in word-overlap metrics like BLEU and introduced novel semantic similarity metrics. The CDGPT2 demonstrated faster training and superior semantic alignment compared to traditional hierarchical recurrent models and prior transformer-based approaches, although it lagged behind some models in specific such as CIDEr. It was revealed by radiologists that qualitative evaluation that 61.7% of the reports were highly precise with minimal false reports (10.2%). The study also highlighted limitations including a bias towards normal cases and a tendency to miss details in abnormal findings, likely stemming from dataset constraints. It was emphasized that the need for larger and more diverse datasets to refine the model’s performance. Overall, this work underscores the potential of pre-trained transformers in medical report generation. (Omar Alfarghaly, 2021)

4.6. Interactive and Explainable Region-guided Radiology Report Generation

The study by Tim Tanida, Philib Muller, Georgios Kaissis, Daniel Rueckert introduces a Region-Guided Radiology Report generation (RGRG) method that enhances transparency and interactivity in radiology report automation. RGRG employs object detection to focus on 29 distinct anatomical regions of chest X-rays unlike the traditional models that generate reports using image-level features. The pipeline consists of four key modules such as an object detector, a region selection classifier, an abnormality classifier, and a transformer-based language model fine-tuned on medical abstracts. The methodology enabled radiologists to manually or automatically select anatomical regions for targeted descriptions by allowing region-specific sentence generation. The abnormality classifier embeds strong abnormal cues, which aids in precise report generation. MIMIC-CXR dataset was used to evaluate the system and compared against state-of-the-art models across natural language generation (NLG) and clinical efficacy (CE) metrics. Results tend to show that RGRG achieves competitive or superior performance on NLG metrics like BLEU and METEOR with significant improvements in CE metrics, particularly in recall. The interactive tools, like sentence generation based on manually drawn bounding boxes, were designed to handle small changes in aspect ratio and size, making them reliable to use in a clinic setting. However, dataset dependency and the exclusion of non-anatomical objects from generation process are the challenges of the study. Overall, the work automates the radiology reporting by integrating explainability and adaptability making it a promising tool for clinics and hospitals while acknowledging further research on the topic. (Tim Tanida, 2023)

4.7. Factual Serialization Enhancement: A Key Innovation for Chest X-ray Report Generation

This study focuses on the factual serialization enhancement (FSE) method addressed key challenges in automated chest X-ray report generation by focusing on aligning medical images with factual descriptions and integrating relevant historical cases. FSE is operated in two stages, in the first stage meaningful and textual features while excluding presentation-style vocabulary from reports are extracted by factuality-guided contrastive learning. Using the Structural Entities (SE) approach, factual content is isolated, ensuring that only medically relevant details guide the alignment. The textual features are extracted using SciBERT and the visual features are obtained through a RestNet101 model. Both at instance and token levels, a contrastive learning framework aligns these features maximizing semantic correspondence between, X-rays and their factual descriptions. In the second stage, bypassing the reliance on disease labels the model retrieves similar historical cases from training set using factual serialization by the evidence-driven report generation. A cross-modal fusion network integrates the input X-ray features with the retrieved evidence, combining the patient’s radiological data with insights from similar cases. The transformer-based text decoder processes the fused features to generate clinically accurate and contextually rich reports. This methodology aligns the radiographs with factual content and utilizing historical data for context which enhances diagnostic precision and report completeness. Experiments on the MIMIC-CXR and the IU X-ray datasets demonstrates the FSE’s effectiveness, which achieves improvements in both natural language generation metrics and clinical impact. Hence, the automated radiology reports are generated by prioritizing factual serialization, FSE which ensures accurate, concise and clinically useful reports. (Kang Liu, 2024)

5. Analysis of Findings

The selected studies demonstrate significant advancements in automatic report generation and leveraging different methodologies to enhance accuracy, interpretability and clinical relevance. In the papers, innovative frameworks like ASGK, VCLI, DeepOht and FSE shows different approaches tailored to specific medical datasets and clinical requirements, showcasing outcomes in various evaluation metrics. The ASGK framework exemplifies the benefits of leveraging auxiliary signals, achieving notable improvements in report generation. Similarly, the VCLI frameworks adopts casual intervention techniques to mitigate cross-modal biases, demonstrating superiority I precision and excelling at identifying complex abnormalities like cardiomegaly and edema. These findings underscore importance of integrating multi-modal learning and casual methods to improve clinical relevance. DeepOpht with its focus on retinal image analysis which highlights the utility of keyword-driven models in achieving superior natural language generation (NLG). The alignment with domain-specific features underscores the significance of leveraging pre-trained models and activation maps for capturing clinically relevant data. On the other hand, the FSE method introduces factual serialization demonstrating that combining radiological data with insights from historical cases enhances diagnostic precision and report finishing.

Across all studies, the common trend is the emphasis on integrating domain-specific insights, leveraging advanced deep learning techniques, and addressing clinical nuances through ablation studies. While significant progress has been made, challenges such as computational efficiency, dataset biasness, and scalability persists signaling the need for further refinement. Regardless these innovations collectively can mark a huge innovative step towards a reliable and clinically valuable automated medical report generation.

6. Conclusion

Hence, we can conclude that using different machine learning (ML) and AI models which takes different parameters according to the different parameters, which perform better than the existing module for different purposes like retinal report generation, X-ray report generation and many more. This can be a revolutionary technology as it can reduce human errors and save time generating medical report.

7. References

Bhanu Prakash Voutharoja, L. W. a. L. Z., 2023. Automatic Radiology Report Generation by Learning. 7 August.

Jia-Hong Huang, C.-H. H. Y. F. L. M. T. Y.-C. L. T.-W. W., 2020. DeepOpht: Medical Report Generation for Retinal Images via Deep Models and Visual Explanation. 1 November.

Kang Liu, Z. M. M. L. Z. J. X. K. Q. M. K. X., 2024. Factual Serialization Enhancement: A Key Innovation for Chest X-ray Generation. 12 September.

Mingjie Li, F. W. X. C. X. L., 2020. Auxiliary Signal-Guided Knowledge Encoder-Decoder for Medical Report Generation. 6 June.

Omar Alfarghaly, R. K. A. E. M. H. A. F., 2021. Automated radiology report generation using conditioned transformers. 26 March.

Tim Tanida, P. M. G. K. D. R., 2023. Interactive and Explainable Region-guided Radiology Report Generation. 17 April.

Weixing Chen, Y. L. C. W. J. Z. S. Z. G. L. C.-L. L. L. L., 2023. Cross-Modal Causal Intervention for Medical Report Generation. 28 February.