International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Automated Medical Report Generation Using Deep Learning

Anupama Phakatkar[1], Advait Kulkarni[2], Mayuresh Khankale[3], Aniket Kolte[4], Samiran Kulkarni[5]

[1,2,3,4,5]Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India

Abstract: Recent advancements in deep learning have dramat- ically transformed how we interpret medical data. In particular, the automated generation of detailed medical reports from imaging and textual information has emerged as a promising tool to support clinical decision-making. Drawing inspiration from image captioning techniques, this paper presents an extensive survey of various methodologies—including hierarchical RNN architectures, attention mechanisms, and reinforcement learning strategies—that have been developed for medical report gen- eration. We discuss the datasets, methods, real-world applica- tions, and evaluation metrics used in this field, address current challenges like data imbalance and limited interpretability, and explore exciting future research directions that could ultimately enhance patient care.

Index Terms: Medical Imaging, Automatic Report Generation, Image Captioning, Deep Learning, CNN, RNN, AI in Healthcare.

1. INTRODUCTION

Medical imaging modalities such as X-rays, CT scans, and MRIs are indispensable in modern diagnostics and treatment planning. Traditionally, expert radiologists and pathologists invest considerable time in preparing detailed reports that describe their observations and recommend further actions. However, the rapidly growing volume of imaging data and the need for prompt, accurate diagnoses have prompted the exploration of automated reporting systems.

The convergence of artificial intelligence (AI) and deep learning has opened up new avenues for developing systems that can generate comprehensive medical reports. By adapting techniques from image captioning, where Convolutional Neu- ral Networks (CNNs) extract image features and Recurrent Neural Networks (RNNs) generate textual descriptions, these systems promise to streamline report creation, reduce manual workload, and enhance diagnostic accuracy. In this paper, we survey the state-of-the-art methods in automated medical report generation, discussing their benefits, challenges, and future potential.

2. Motivation and Background

The motivation behind automating medical report gener- ation is both practical and transformative. In high-pressure clinical environments, reducing the time spent on routine documentation allows healthcare professionals to devote more attention to complex cases. By automating report generation, clinicians can benefit from consistent, high- quality preliminary drafts that serve as a basis for final reports, reducing the risk of oversight and human error.

At the heart of these systems lies the image captioning paradigm. Originally developed for general-purpose image description tasks, image captioning employs CNNs to derive meaningful features from images and RNNs (or LSTMs) to articulate these features in natural language. Researchers have extended these models to handle the nuances of medical im- ages by incorporating advanced techniques such as hierarchical RNNs and attention mechanisms, which better capture the subtle details crucial for accurate diagnosis.

3. Related Work

A wealth of research in recent years has explored the automatic generation of medical reports. One of the pioneering works in this field was presented by Shin et al. [3], who proposed a CNN-RNN framework for generating descriptive reports from chest X-rays. Their approach laid the groundwork for subsequent innovations, including the integration of atten- tion mechanisms that enable models to focus on diagnostically significant regions within an image.

Other studies have combined reinforcement learning with traditional deep learning architectures to iteratively refine report quality based on feedback [4]. Furthermore, several researchers have embraced multi-modal approaches, merging imaging data with additional clinical information to produce richer and more personalized reports. These efforts underscore the promise of automated systems in standardizing diagnostic reporting and enhancing clinical workflows.

4. Deep Learning Techniques in Medical Report Generation

This section provides an in-depth look at the key deep learning techniques employed in automated medical report generation.

A. Convolutional Neural Networks (CNNs)

CNNs serve as the backbone for feature extraction in medical imaging. They are particularly adept at identifying and learning hierarchical patterns—such as edges, textures, and shapes—from complex images. In the context of med- ical report generation, CNNs convert raw pixel data into structured feature maps that encapsulate crucial diagnostic information [1], [2].

B. Recurrent Neural Networks (RNNs)

After extracting image features, RNNs, especially LSTM networks, generate descriptive text that aligns with the visual information. These networks excel in processing sequences, ensuring that the narrative produced is coherent and contextu- ally relevant. By integrating visual cues sequentially, RNNs help create reports that mirror the logical flow of clinical observations [1], [2].

C. Attention Mechanisms

Attention mechanisms have become an integral part of modern deep learning models. They allow the network to con- centrate on the most significant parts of an image during the report generation process. By dynamically weighting different image regions, attention mechanisms ensure that even subtle abnormalities are not overlooked, thereby enhancing both the accuracy and interpretability of the generated reports.

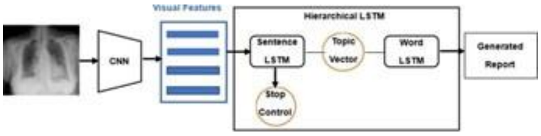

Fig. 1. An attention-based framework for report generation, illustrating how the model selectively focuses on key regions within a medical image [2].

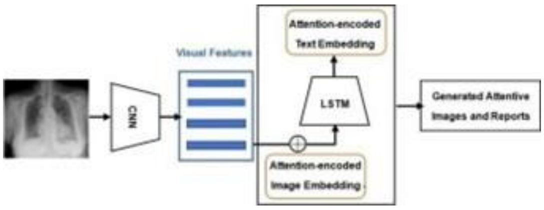

Fig. 2. A hierarchical RNN-based framework that processes image features in stages to produce coherent diagnostic narratives [3].

5. Practical Applications and Real-World Impact

Automated medical report generation is not merely an academic pursuit—it has profound practical implications. In clinical settings, these systems can significantly reduce the time required to generate diagnostic reports, allowing physi- cians to focus on critical decision-making. For instance, by automating the initial drafting of reports, radiologists can verify and refine the output, ensuring that final reports are both precise and comprehensive.

Moreover, automated reporting can help standardize doc- umentation across institutions, making it easier to compare and interpret diagnostic results. In large hospital networks or multi-center studies, such consistency is invaluable. The integration of these systems into daily clinical workflows promises to enhance efficiency, minimize human error, and ultimately improve patient outcomes.

6. Challenges and Limitations

Despite their immense potential, automated medical report generation systems face several challenges. A major hurdle is the scarcity of large-scale, high-quality annotated datasets. Due to the sensitive nature of medical data and strict privacy regulations, compiling diverse and representative datasets re- mains a significant challenge.

Another critical issue is the interpretability of deep learning models. Although these systems can generate highly accurate reports, the underlying decision-making processes are often opaque, which can impede clinical trust and acceptance. Ad- ditionally, variability in imaging quality and patient-specific factors demands that these models be robust and adaptable. Addressing these challenges calls for further research into transparent model design, improved data augmentation tech- niques, and effective transfer learning strategies.

7. Survey Table

The following table summarizes various approaches in auto- mated medical report generation. The content below preserves the original technical details while rephrasing the language to be more accessible.

8. Performance Evaluation Metrics

The quality of automatically generated medical reports is commonly assessed using several well-established metrics:

A. BLEU

BLEU (Bilingual Evaluation Understudy Score) measures the overlap of n-grams between the generated and reference reports. A higher BLEU score signifies closer alignment with human-authored text, making it essential for evaluating translation and text generation tasks.

B. ROUGE

ROUGE (Recall-Oriented Understudy for Gisting Evalua- tion) focuses on the recall of word sequences, such as bigrams and longer phrases, between generated and reference texts. It is particularly useful for ensuring that critical details are retained in the final report.

C. METEOR

METEOR (Metric for Evaluation of Translation with Ex- plicit Ordering) extends beyond exact word matches by con- sidering synonyms, stemming, and paraphrasing, while also accounting for word order. This metric offers a more flexible assessment of semantic content, which is important in the context of medical report generation.

D. CIDEr

CIDEr (Consensus-based Image Description Evaluation) uses a weighted n-gram similarity measure based on TF-IDF, emphasizing rare yet significant terms. In specialized fields like medical reporting, CIDEr is valuable for ensuring that critical technical terminology is accurately represented.

TABLE I Survey of Approaches in Automated Medical Report Generation

Author Name	Algorithm Used	Research Summary	Scalability/Dataset	Features
Guangyi Liu, Yinghong Liao, Fuyu Wang, Bin Zhang, Lu Zhang, Xiaodan Liang, Xi- ang Wan, Shaolin Li, Zhen Li, Shuixing Zhang, Shuguang Cui	Medical-VLBERT (BERT variant with transfer learning and DenseNet-121)	Utilizes a transfer learning strategy with fine-tuning on the CX-CHR dataset.	Scalable through the use of transfer learning and external datasets.	Automatic report generation, abnormality classification, ter- minology prediction, alternate learning strategy.
Maram Mahmoud A. Monshi, Josiah Poon, Vera Chung	Deep Learning (CNN-RNN)	Focuses on capturing patient- specific conditions and imag- ing details.	Employs large public datasets such as IU X-ray and ChestX- ray14.	Radiology image analysis, dis- ease description, report gener- ation, text-to- image alignment.
Leslie Pack Kaelbling, Michael L.	Reinforcement Learning (Q- learning, TD	Uses trial-and- error methods to adjust agent	Capable of scaling to larger state spaces with model-	Balances exploration vs. ex-

Littman, Andrew W. Moore	learning, policy iteration)	behavior.	based methods.	ploitation; integrates model-free and model-based meth- ods; utilizes Q-learning and TD learning.
X. Chen, Y. Zhang, Q. Ai, H. Xu, J. Yan, and Z. Qin [7], 2017	Collaborative Filtering, LSTM (Long Short-Term Memory) Networks	Leverages time-synchronized user input to tailor results based on individual engage- ment.	A unified framework that han- dles multi-modal data (visual and textual) using deep learn- ing techniques like LSTMs.	Combines visual and textual data; incorporates time-synchronized comments.
Alistair E. W. Johnson, Tom J. Pollard, Seth J. Berkowitz	Computer and Deep Vision Learning Models	Developed to enhance radiol- ogy workflow.	Utilizes 377,110 images from 227,835 studies.	Provides comprehensive chest radiographs; supports semi-structured reporting and de-identification protocols.
Philipp Harzig, Yan-Ying Chen, Francine Chen, Rainer Lienhart	Dual Word LSTM with Hier- archical LSTM Model	Tailors report generation for both normal and abnormal findings.	Operates on a dataset compris- ing 7,470 images and reports.	Facilitates hierarchical sen- tence generation, abnormal- ity prediction, and multi-task learning.
Jianbo Yuan, Haofu Liao, Rui Luo, Jiebo Luo	Generative Encoder-Decoder with Multi-view CNN and Hierarchical LSTM	Fine-tunable for domain-specific imaging requirements.	Processes 224,316 multi-view chest X-ray images.	Enables multi-view fusion, medical concept enrichment, cross-view consistency, and hierarchical generation.
Hyebin Lee, Seong Tae Kim, Yong Man Ro	Deep Learning Network with VGG16 and Justification Gen- erator	Customizable for different medical imaging contexts.	Effective with limited data through augmentation tech- niques.	Supports multimodal output, visual word constraints, LSTM-based text

				generation, and channel-wise attention.
Xin Rui Li, Dongxiao Cao, Zhu	DenseNet + LSTM with Attention Mechanism	Implements a standardized approach across diverse cases.	Utilizes transfer learning from larger datasets.	Facilitates disease classification, localization, visual support, and natural language reporting.
William Gale, Luke Oakden-Rayner, Gustavo Carneiro, Lyle J. Palmer, Andrew P. Bradley	RNN with Two LSTM Layers and Visual Attention	Produces human-style explanations for clinicians.	Requires minimal additional labeling.	Offers model-agnostic interpretability, natural language descriptions, and simplified grammar.
Pablo Pino, Denis Parra, Pablo Messina, Cecilia Besa, Sergio Uribe	CNN-LSTM Variants with Different Architectures	Trained using the IU X-ray dataset.	Evaluated on standard medical imaging datasets.	Uses multiple evaluation metrics; integrates disease classification and attention mechanisms.
Wenting Xu, Chang Qi, Zhenghua Xu, Thomas Lukasiewicz	ResNet-101 + LSTM with X-linear Attention	Adopts a general training approach without explicit personalization.	Handles variable-length reports up to 184 tokens.	Incorporates X-linear attention, reinforcement learning, repetition penalty, and improved coherency.

9. Discussion and Future Directions

The ongoing progress in deep learning is poised to further revolutionize medical report generation. Future research will likely focus on several key areas:

- **Multi-modal Integration:** Combining imaging data with additional clinical information (such as electronic health records) to create more personalized and context-rich reports.
- **Improved Interpretability:** Developing transparent models that not only generate accurate reports but also explain the reasoning behind their outputs, thereby fos- tering clinical trust.
- **Data Augmentation and Transfer Learning:** Ad- dressing the challenge of limited high-quality annotated datasets by leveraging transfer learning and sophisticated data augmentation techniques.
- **Ethical and Regulatory Considerations:** Ensuring that AI systems are developed within robust

ethical frame- works, with a focus on patient privacy, data security, and minimizing algorithmic bias.
- As these advancements unfold, the integration of automated report generation into clinical workflows is expected to stream-line diagnostic processes, reduce the burden on healthcare professionals, and ultimately improve patient outcomes.

10. Ethical Considerations and Implications

With the increasing adoption of AI in healthcare, ethical concerns have become paramount. Ensuring patient privacy, securing sensitive data, and mitigating biases are critical factors that must be addressed during the development and deployment of these systems. Transparent models and clear regulatory guidelines will be essential in building trust and ensuring that AI-driven solutions are used responsibly in clinical practice.

Tian J, Li C, Shi Z, Xu F. A diagnostic report generator from CT volumes on liver tumor with semi-supervised attention mechanism. In: International Conference on Medical Image Computing and Computer- Assisted Intervention; 2018:702–710.

[12] Zhang Z, Chen P, Sapkota M, Yang L. TandemNet: Distilling knowledge from medical images using diagnostic reports as optional semantic references. Cham: Springer; 2017.

[13] Pang T, Li P, Zhao L. A survey on automatic generation of medical imaging reports based on deep learning. 2023.

11. Conclusion

Automated medical report generation using deep learning represents a transformative step forward in clinical diagnos- tics and patient care. By harnessing sophisticated algorithms and leveraging extensive datasets, these systems can rapidly produce detailed, human-like diagnostic reports. In this paper, we have surveyed the state-of-the-art techniques, discussed practical applications, outlined current challenges, and ex- plored future directions. As research in this field progresses, it is expected that such systems will become integral to clinical workflows—enhancing efficiency, reducing errors, and ultimately leading to better patient outcomes.

Acknowledgment

The authors gratefully acknowledge the support provided by the Department of Computer Engineering at Pune Institute of Computer Technology.

REFERENCES

1. Liu G, Liao Y, Wang F, et al. Medical-VLBERT: Medical visual language BERT for COVID-19 CT report generation with alternate learning. IEEE Trans Neural Netw Learn Syst. 2021;32(9):3786–3797.
2. Monshi MMA, Poon J, Chung V. Deep learning in generating radiology reports: a survey. Artif Intell Med. 2020;106:101878.
3. Shin HC, Roberts K, Lu L, Demner-Fushman D, Yao J, Summers RM. Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:2497–2506.
4. Kaelbling LP, Littman ML, Moore AW. Reinforcement Learning: A Survey. J. Artif. Intell. Res. 1996;4:237–285.
5. Jing B, Xie P, Xing E. On the automatic generation of medical imaging reports. arXiv preprint

arXiv:1711.08195; 2017.
6. Harzig P, Chen YY, Chen F, Lienhart R. Addressing data bias problems for chest X-ray image report generation. arXiv preprint arXiv:1908.02123; 2019.
7. Yuan J, Liao H, Luo R, Luo J. Automatic radiology report generation based on multi-view image fusion and medical concept enrichment. arXiv preprint arXiv:1907.09085; 2019.
8. Huang FR, Zhang XM, Zhao ZH, Li ZJ. Bi-directional spatial-semantic attention networks for image-text matching. IEEE Trans Image Process. 2019;28(4):2008–2020.
9. Chang YC, Hsing YC, Chiu YW, et al. Deep multi-objective learning from low-dose CT for automatic lung-RADS report generation. JPM. 2022;12(3):417.
10. Wu F, Yang H, Peng L, et al. AGNet: Automatic generation network for skin imaging reports. Comput Biol Med. 2022;141:105037.
11. Han Z, Wei B, Xi X, et al. Unifying neural learning and symbolic reasoning for spinal medical report generation. Med Image Anal. 2021;67:101872.