Recent Advances in Artificial Intelligence for Radiology Report Generation: A Brief Review

Caitlyn Tran, BS,[1] Basudha Pal[2], Thomas Stirrat, BS,[1] Victoria Shi, BS,[3] Muhammad Umair, MD[4]
[1]Georgetown University School of Medicine, Washington, DC 20007, USA
[2]Johns Hopkins University, PhD Program, Department of Electrical and Computer Engineering, Baltimore, MD 21218, USA
[3]Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
[4]Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA

Introduction

Rising imaging volumes combined with radiologist workforce shortages have created significant workflow bottlenecks → potential solution: artificial intelligence (AI)

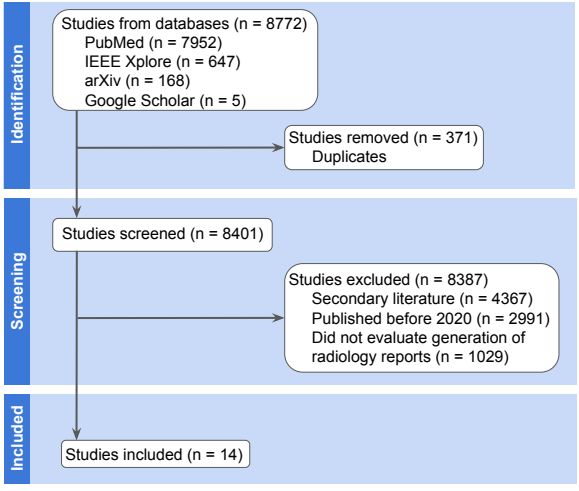
+85% imaging utilization increase among Medicare enrollees (2003-2016)

+66% imaging utilization increase among commercially insured adults (2003-2016)

+13% radiologist workforce growth over the same period

This review evaluates recent advancements in AI vision-language models (VLMs) for radiology report generation

Methods



Results

- 11 of 14 VLMs trained primarily on chest radiographs (MIMIC-CXR) with limited representation of CT, MRI, and ultrasound
- No single optimization strategy dominates → each has distinct trade-offs between between linguistic fidelity and clinical accuracy

Common Optimization Strategies:

Strategy	Strengths	Weaknesses	Example VLMs
Masked Modeling	↑ Natural language generation performance (METEOR, CIDEr-D)	↓ Clinical entity-relation accuracy (RadGraph F1)	METransformer R2GenGPT
Contrastive alignment	↑ Entity-relation accuracy	↓ N-gram precision	CXR-RePaiR-2/ Select BioViL-T

- Models incorporating multiple optimization strategies demonstrated more balanced performances
- 4 of 14 studies included human evaluation
  - AI reports were rated equal or better than human-written reports in 40–60% of cases
  - AI reports had higher rates of omission and errors, especially for complex findings

Conclusion

- Although AI VLMs show potential for automated radiology report generation, they remain limited by a lack of multimodal training datasets, unstandardized evaluation metrics, and errors
- If limitations are addressed and VLMs are properly integrated with human oversight, VLMs could serve as assistive tools that enhance efficiency and patient care

Key Takeaways

- AI radiology report generation has advanced rapidly with VLMs
- VLM-generated reports (2020-2024) are promising but require radiologist oversight for safe clinical use, especially in complex cases
- No single optimization strategy dominates → hybrid approaches show the most promise
- Current limited dataset diversity restricts generalizability and model performance
- Lack of standardized evaluation metrics remains a major barrier to clinical adoption

References

1. Tran C, Pal B, Stirrat T, Shi V, Umair M. Recent advances in artificial intelligence for radiology report generation: a brief review. *BJR Artif Intell*. 2026;3(1):ubag003. doi:10.1093/bjrai/ubag003
2. Hong AS, Levin D, Parker L, Rao VM, Ross-Degnan D, Wharam JF. Trends in diagnostic imaging utilization among Medicare and commercially insured adults from 2003 through 2016. *Radiology*. 2020;294(2):342-350. doi:10.1148/radiol.2019191116
3. Bhargavan M, Sunshine JH, Schepps B. Too few radiologists? *AJR Am J Roentgenol*. 2002;178(5):1075-1082. doi:10.2214/ajr.178.5.1781075
4. Wang Z, Liu L, Wang L, Zhou L. METransformer: Radiology report generation by transformer with multiple learnable expert tokens. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR)*. 2023:11558-11567.
5. Chen Z, Song Y, Chang TH, Wan X. Generating radiology reports via memory-driven transformer. In: Webber B, Cohn T, He Y, Liu Y, eds. *Proc 2020 Conf Empir Methods Nat Lang Process (EMNLP)*. Association for Computational Linguistics; 2020:1439-1449. doi:10.18653/v1/2020.emnlp-main.112
6. Endo M, Krishnan R, Krishna V, Ng AY, Rajpurkar P. Retrieval-based chest X-ray report generation using a pre-trained contrastive language-image model. In: *Proc Mach Learn Health*. PMLR; 2021:209-219.
7. Bannur S, Hyland S, Liu Q, et al. Learning to exploit temporal structure for biomedical vision-language processing. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR)*. 2023:15016-15027.