Proceedings of CONF-CDS 2026 Symposium: Machine Learning and Neural Network Applications in Engineering
DOI: 10.54254/2755-2721/2026.AST37152

A Review of Chest Imaging Report Generation Enhanced by Knowledge Graphs and Graph Neural Networks

Bingyu Song

School of Computer Science and Engineering, Changchun University of Technology, Changchun, China
736001310@qq.com

Abstract. Chest imaging report generation aims to automatically generate radiology reports from images such as chest X-rays and is an important task at the intersection of medical image analysis and natural language generation. In recent years, multimodal large language models have improved the fluency and organization of generated reports, but they still face limitations related to insufficient medical knowledge, inadequate modeling of anatomical relationships, and hallucination. This article reviews studies on chest imaging report generation, the integration of knowledge graphs and large language models, applications of graph neural networks in medical imaging, and the evaluation of multimodal foundation models for radiology. Existing studies indicate that knowledge graphs can provide structured medical concepts, relationships between diseases and anatomical sites, and traceable external knowledge, whereas graph neural networks are suited to relational propagation, selection, and compression in graph space.

Keywords: knowledge graph, graph neural network, large language model, multimodal fusion, artificial intelligence in radiology

1. Introduction

Chest X-ray is one of the most widely used medical imaging examinations in clinical practice, and the imaging report is the primary textual medium through which radiologists communicate diagnostic information to clinicians. Automated chest imaging report generation seeks to generate the Findings and Impression sections directly from images, reduce the reporting workload of radiologists, and improve report standardization in primary care, telemedicine, and routine imaging workflows. The task requires fluent text as well as accurate descriptions of lesion location, negative findings, severity, and clinical uncertainty.

Recent research has gradually shifted from traditional encoder-decoder models to multimodal large language models. Li and Zhang used ResNet50 and GPT-2 for chest X-ray report generation, showing that deep visual features and a text decoder can establish a basic image-text mapping [1]. A study noted that general-purpose LLMs must be adapted to radiology reports through in-domain sample induction and coarse-to-fine decoding [2]. In addition, RRG must generate fluent text and requires factual evaluation aligned with radiologists' judgments [3]. These studies have expanded the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AST37152

language organization capabilities of report-generation models, but medical knowledge constraints, anatomical relation modeling, and hallucination control remain insufficiently addressed.

This article uses a literature review method to examine CXR report generation, KG-LLM integration, GNN applications in medical imaging, and evaluation frameworks, with a focus on how KGs and GNNs can provide LLMs with medical knowledge that is selectable, compressible, and traceable. This review can help researchers distinguish among knowledge retrieval, graph-based relation filtering, and generation control in KG-GNN-LLM system design and provides a basis for selecting metrics for factuality control and clinical utility evaluation.

2. Evolution of CXR report generation techniques

Early chest imaging report generation generally followed an encoder-decoder framework, extracting chest X-ray features before generating a report. Raminedi et al. explored a multimodal report-generation architecture using a vision Transformer and GPT-2 [4], while Alqahtani et al. proposed the CNX-B2 model with a CNN encoder and Transformer decoder [5]. Research then shifted toward fine-grained vision-language alignment and patient context. Yin et al. noted that global image-text alignment can miss important disease descriptions and therefore introduced into DKA-RG a disease knowledge graph comprising 20 common chest diseases organized by anatomical location [6]. Liu et al. used current multiview images, prior images, prior reports, and examination indications to demonstrate the value of patient prior knowledge for report coherence and clinical accuracy [7]. Noh et al. improved local lesion descriptions through whole-image lesion prediction, 11 anatomical regions of interest (ROIs), lesion cross-attention, and contrastive learning [8]. Wang et al. incorporated images, reports, and RadGraph triplets into hierarchical contrastive learning, whereas Li et al. combined a medical knowledge graph, a Mamba/SSM encoder, and reinforcement learning rewards [9, 10]. Existing research has expanded from image-text matching to the joint modeling of diseases, anatomical sites, temporal relationships, and knowledge constraints.

Multimodal LLMs have further improved report organization. Building on MiniGPT-4, Liu et al. introduced relevant-instance retrieval, contrastive semantic ranking, and coarse-to-fine decoding to adapt an LLM to the style of radiology reports [2]. In CXR-LLaVA, Lee et al. connected an image encoder to an LLM, whereas Zambrano Chaves et al. trained a chest X-ray vision encoder, connected it to a language model through a lightweight adapter, and proposed CheXprompt for factuality evaluation [3, 11]. However, the error analysis by Lee et al. also showed that large models may still omit critical findings, such as pneumothorax and catheters, and generate hallucinations [11]. Therefore, LLMs improve language generation but still require external medical knowledge and explicit relational modeling.

3. Three approaches to KG+LLM integration

Based on the existing literature, the integration of KGs and LLMs can be broadly categorized into three approaches: using KGs to enhance LLMs, using LLMs to construct KGs, and end-to-end integration of KGs and LLMs. All three seek to alleviate missing knowledge, unstable reasoning, and hallucinations in medical tasks, but they differ in emphasis.

3.1. KG-enhanced LLMs

The central idea of KG-enhanced LLMs is to use structured external knowledge to compensate for limitations in the models' implicit knowledge. Soman et al. proposed KG-RAG, which uses the

DOI: 10.54254/2755-2721/2026.AST37152

SPOKE biomedical knowledge graph to provide token-optimized knowledge context for Llama-2-13b, GPT-3.5-Turbo, and GPT-4. By using a minimal graph schema and embedding-based pruning, the method reduced token consumption by more than 50% and improved Llama-2 performance on an MCQ dataset by 71% [12]. Liu et al. noted that iterative retrieval can improve accuracy, although question ambiguity, fixed external datasets, and inadequate retrieval evaluation continue to limit RAG systems [13].

In chest imaging report generation, KGs can provide disease definitions, anatomical locations, abnormal findings, and diagnostic associations, but simply stacking triplets is insufficient. Tian et al. noted that directly inputting KG triplets may introduce irrelevant context and noise [14]. Wang et al. therefore retrieved similar reports during inference, extracted RadGraph triplets, and constructed an image-conditioned dynamic knowledge graph. Li et al. used a medical knowledge graph module with 78 entities and injected knowledge vectors into the visual encoding stage through attention and gating mechanisms [9, 10]. These results shift the focus of KG enhancement from "how much knowledge is introduced" to three more specific questions: whether retrieval is accurate, whether the representation is sufficiently compressed, and whether relational constraints agree with the current imaging evidence.

3.2. LLM-based KG construction

LLM-based KG construction uses the information extraction, summarization, and structured generation capabilities of language models to extract entities and relations from medical texts, patient records, reports, and image descriptions. Siragusa et al. used Llama 3.1 Instruct 8B to construct a knowledge graph based on MedPix 2.0 and applied it to a multimodal biomedical dataset and a RAG system [15]. Dehal et al. noted that sequence-to-sequence models such as T5 and BART can support structured output, summarization, and KG-to-text generation, indicating that LLMs can be enhanced by KGs and can also participate in KG construction [16].

For chest imaging report generation, LLM-based KG construction can extract "disease-anatomical site-imaging finding" triplets from historical reports, identify positive, negative, and uncertain statements in the Findings and Impression sections, and convert disease progression into a temporal graph structure. This approach can reduce the manual effort required to construct specialized knowledge graphs, but the extraction results remain susceptible to hallucinations, differences in reporting style, and inconsistent entity boundaries. Medical rules, expert review, or automated consistency checks therefore remain necessary for correction.

3.3. End-to-end integration of KGs and LLMs

End-to-end KG-LLM integration places graph structures directly into representation learning, reasoning and planning, or generation control, rather than using retrieval results only as textual prompts. Xu et al. noted that biomedical data include tables, text, images, and time series and are shaped by data schemas, coding systems, and quality standards. LLMs still suffer from knowledge gaps, ambiguous reasoning, and hallucinations. In such tasks, unifying KGs and LLMs can serve as a representation-level constraint, improve the organization of medical information, and reduce reasoning drift caused by unconstrained generation.

In RRG, end-to-end integration can bring image regions, anatomical sites, disease entities, clinical indications, prior reports, and generated text into a unified framework: the visual encoder extracts regional features, the KG provides disease-anatomical site relations, the GNN propagates information over a subgraph, and the LLM generates the report from the fused representation. In

DOI: 10.54254/2755-2721/2026.AST37152

theory, such end-to-end designs can improve multi-hop reasoning and interpretability, but training complexity, annotation costs, and difficulties in unifying cross-modal encoders currently limit their use. In the near term, they are more suitable for parallel validation alongside modular approaches.

4. Role and representative architectures of GNNs in knowledge graph-enhanced chest imaging report generation

4.1. Value of relational modeling and supporting evidence

The value of GNNs lies in explicitly incorporating relationships among medical objects into report generation. Chest imaging reports involve relationships among anatomical sites, diseases and imaging findings, image regions and report entities, and prior and current examinations. CNNs and ViTs are effective at extracting visual features, while LLMs are effective at organizing text, but neither typically provides stable, explicit relational reasoning. Introducing a GNN between the KG and the LLM can therefore encode sparse and potentially noisy graph knowledge into relational representations that better reflect the current imaging context, thereby providing structured constraints for report generation.

The results of Tian et al. directly support this role. They used a GNN to encode a knowledge graph, selected relevant node embeddings through cross-modal pooling, and connected the graph and text domains with a domain projector. With the LLM frozen, the model outperformed the baseline by 13.5% [14]. The ablation results of Wang et al. showed that after introducing an image-conditioned dynamic knowledge graph, entity coverage on MIMIC-CXR increased from 32.7% to 76.5%, while the CE F1 score increased from 0.441 to 0.492 [9]. These results indicate that the role of the GNN in the system is to select disease, anatomical site, and attribute relations according to the current image, reducing the amount of irrelevant knowledge passed to the generation model.

4.2. Representative architectures and application pathways

Fine-grained image-text alignment enhanced by a disease knowledge graph is the first pathway. In addition to global image-text alignment, the DKA-RG model of Yin et al. introduced a disease knowledge graph comprising 20 common chest diseases organized by anatomical location, allowing the model to learn correspondences between disease findings and report descriptions [6]. Wang et al. further incorporated RadGraph triplets and an image-conditioned dynamic knowledge graph into contrastive learning [9]. Such methods refine whole-image matching into correspondences among diseases, anatomical sites, and findings, making them suitable for cases in which lesion descriptions tend to be overly general or omitted. If a GNN is added, relations within the graph can also be propagated during alignment, allowing these medical constraints to enter report generation.

Knowledge retrieval, graph neural prompting, and LLM generation form the second pathway. Soman et al. used KG-RAG to retrieve and compress biomedical knowledge graph context, while Tian et al. used a GNN to encode a knowledge graph, selected relevant node representations through cross-modal pooling, and projected them into LLM prompts [12, 14]. For chest imaging report generation, suspected abnormalities and anatomical regions can first be identified, followed by retrieval of a chest disease subgraph. A GNN then filters the relations before they are provided to a multimodal LLM. The value of this combination lies in removing relations unrelated to the current image before passing the compressed subgraph to the LLM, thereby making noise easier to control while retaining the basis for generation.

DOI: 10.54254/2755-2721/2026.AST37152

Coordination between multimodal foundation models and clinical evaluation forms the third pathway. In CXR-LLaVA, Lee et al. connected an image encoder to an LLM for chest X-ray interpretation and report generation. Zambrano Chaves et al. trained a chest X-ray vision encoder and proposed CheXprompt for factuality evaluation, while Tanno et al. involved radiologists in comparing AI-generated and human reports [3, 11, 15]. These studies place generation back into a clinical evaluation setting to assess whether reports are consistent with imaging evidence and physicians' review requirements. This pathway does not itself center on GNNs; KG-GNN components can be integrated as knowledge-enhancement modules, and their effects still need to be validated jointly through factuality metrics and physician evaluation.

5. Evaluation frameworks, benchmarks, and discussion

Existing studies generally assess text similarity, clinical factuality, and human evaluation [1], using automatic metrics on public chest imaging datasets [16].

Several clear gaps remain in current evaluation. Dataset partitioning and preprocessing lack unified standards, and traditional NLG metrics have difficulty identifying medical factual errors. For KG-GNN-LLM systems, specialized methods have not yet been established to evaluate knowledge retrieval, graph relation filtering, and reasoning paths, and the correspondence of generated results with image regions and knowledge nodes remains unclear. The expert evaluation by Tanno et al. showed that automatic reports are more suitable as drafts, error alerts, and tools for clinical review [15]. Future studies should standardize chest disease-anatomy knowledge graphs, evaluate KG retrieval, GNN filtering, and LLM generation separately, increase cross-institution, cross-device, cross-population, and long-tail abnormality testing, and retain radiologist evaluation.

6. Conclusion

This review focuses on the role of knowledge graphs and graph neural networks in enhancing chest imaging report generation. The literature shows that models have evolved from encoder-decoder architectures to multimodal large language models, but gaps in medical knowledge, inadequate modeling of anatomical relationships, hallucinations, and insufficient clinical evaluation remain major bottlenecks. The theoretical contribution of this article is to place report generation, knowledge graph integration, and applications of graph neural networks in medical imaging within a common analytical framework, showing that relation-aware reasoning can compensate for the limitations of text-only generation. In practice, this framework can be used to identify key design considerations in factuality control, knowledge prompting, and clinically interpretable evaluation. Because the study relies on a literature review, its data sources, metrics, and scope depend on the available literature, and the generalizability of its conclusions requires further empirical support. Future work could construct standardized chest disease-anatomy knowledge graphs, evaluate the effectiveness of GNN-based knowledge filtering, and conduct validation with physician participation in real radiology workflows. Existing research suggests that the value of knowledge graphs and graph neural networks lies not in replacing generation models, but in providing three forms of structural support for report generation: disease relations, anatomical constraints, and evidence tracing.

References

[1] Li, J., & Zhang, Y. (2025). Automatic chest X-ray image caption for report generation. In *ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE '25)*.

DOI: 10.54254/2755-2721/2026.AST37152

https://doi.org/10.1145/3721201.3725429

[2] Liu, C., Tian, Y., Chen, W., Song, Y., & Zhang, Y. (2024). Bootstrapping large language models for radiology report generation. In *The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)*.

[3] Zambrano Chaves, J. M., et al. (2025). A clinically accessible small multimodal radiology model and evaluation metric for chest X-ray findings. *Nature Communications, 16*, 3108. https://doi.org/10.1038/s41467-025-58344-x

[4] Raminedi, S., Shridevi, S., & Won, D. (2024). Multi-modal transformer architecture for medical image analysis and automated report generation. *Scientific Reports, 14*, 19281. https://doi.org/10.1038/s41598-024-69981-5

[5] Alqahtani, F. F., Mohsan, M. M., Alshamrani, K., Zeb, J., Alhamami, S., & Alqarni, D. (2024). CNX-B2: A novel CNN-Transformer approach for chest X-ray medical report generation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3367360

[6] Yin, H., Wu, W., & Hao, Y. (2024). DKA-RG: Disease-knowledge-enhanced fine-grained image-text alignment for automatic radiology report generation. *Electronics, 13*, 3306. https://doi.org/10.3390/electronics13163306

[7] Liu, K., Ma, Z., Kang, X., Li, Y., Xie, K., Jiao, Z., & Miao, Q. (2025). Enhanced contrastive learning with multi-view longitudinal data for chest X-ray report generation. *arXiv:2502.20056*.

[8] Noh, W.-J., Pi, S.-W., & Lee, B.-D. (2026). Hybrid framework for lesion-aware, clinically coherent chest X-ray report generation using contrastive learning and large language models. *Scientific Reports, 16*, 4645. https://doi.org/10.1038/s41598-025-34799-2

[9] Wang, B., et al. (2026). Toward bias-resilient radiology report generation: Hierarchical contrastive learning and adaptive knowledge graph integration. *Knowledge-Based Systems, 340*, 115672. https://doi.org/10.1016/j.knosys.2026.115672

[10] Li, Z., Yang, D., Li, R., & Zhang, J. (2026). Knowledge-guided and reinforced selective state space model for radiology report generation. *Journal of Biomedical Informatics, 177*, 105011. https://doi.org/10.1016/j.jbi.2026.105011

[11] Lee, S., Youn, J., Kim, H., Kim, M., & Yoon, S. H. (2025). CXR-LLaVA: A multimodal large language model for interpreting chest X-ray images. *European Radiology, 35*, 4374–4386. https://doi.org/10.1007/s00330-024-11339-6

[12] Soman, K., et al. (2024). Biomedical knowledge graph-optimized prompt generation for large language models. *Bioinformatics, 40*(9), btae560. https://doi.org/10.1093/bioinformatics/btae560

[13] Liu, S., McCoy, A. B., & Wright, A. (2025). Improving large language model applications in biomedicine with retrieval-augmented generation: A systematic review, meta-analysis, and clinical development guidelines. *Journal of the American Medical Informatics Association, 32*(4), 605–615. https://doi.org/10.1093/jamia/ocaf008

[14] Tian, Y., Song, H., Wang, Z., Wang, H., Hu, Z., Wang, F., Chawla, N. V., & Xu, P. (2024). Graph neural prompting with large language models. In *The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)*.

[15] Tanno, R., et al. (2025). Collaboration between clinicians and vision-language models in radiology report generation. *Nature Medicine, 31*, 599–608. https://doi.org/10.1038/s41591-024-03302-1

[16] Savage, C. H., et al. (2025). Open-source large language models in radiology: A review and tutorial for practical research and clinical deployment. *Radiology, 314*(1), e241073. https://doi.org/10.1148/radiol.241073