MedXGen: LLM leveraged Framework for Automated Clinical Coherent Medical Report Generation

A.N. Ramya Shree
CSE (AI & ML)
Ramaiah Institute of Technology
Bangalore, India

Nithya N.
CSE (Cyber Security)
Ramaiah Institute of Technology
Bangalore, India

Lavanya Kamaraju
Artificial Intelligence
Ramaiah Institute of Technology
Bangalore, India

Sahana M.B.
Artificial Intelligence
Ramaiah Institute of Technology
Bangalore, India

Hrithwika
Artificial Intelligence
Ramaiah Institute of Technology
Bangalore, India

Supriya
Artificial Intelligence
Ramaiah Institute of Technology
Bangalore, India

ABSTRACT
Artificial intelligence-based automatic medical report creation has accelerated significantly since the introduction of cross-modal learning, sophisticated transformer structures, and knowledge-enhanced pretraining methods. The detector attention modules, adapter tuned vision language models, and graph-guided hybrid strategies are integrated to propose framework for automated medical report generation. Utilizing topic wise separable retrieval, hierarchical cross-modal alignment, and phrase-level augmentation, the proposed MedXGen confront semantic inconsistency, hallucination and redundancy. Memory-guided transformers and semi-supervised learning is used to enhance interpretability and adaptability. The suggested framework offers a practical implementation of clinical diagnostic support systems and it supports medical language creation and visual perception.

Keywords
Medical Report Generation, Vision-Language Models, Cross-Modal Learning, Transformer, Deep Learning, Clinical Decision Support, AI in Healthcare.

1. INTRODUCTION
The vast amounts of unstructured clinical notes are generated on daily in today's word. It encompasses patient complaints, diagnostic impressions, and therapy observations. Customary summary generation has limitation of grasping clinical prose's distinction and a need of remedies to overcome this limitation. Leveraging large language models towards this limitation and speech output generation supports automated remedies. Google FLAN-T5-Large and kin trained on sequences of thoughts within adapter tuning's bounds. The MedXGen can structure complex medical narratives into a shared understanding. It eases difficulties and prompt usage can generate coherent and clinically relevant summaries. Google Text-to-Speech (GTTS) converts summaries to speech that is a voice supported interface which further ease visually impaired patients. The proposed MedXGen framework integrates LLM leveraged medical summarization with speech processing and further evaluated by BLEU-1, BLEU-2, ROUGE-L, METEOR and BERTScoremetrics and demonstrates capable performance in unigram handling and longer sequence handling. Text summarization and audio delivery together contribute towards inclusive clinical communication. Automated report generation has a great impact for those in need of timely treatment. It is addressed by the use of intelligent systems with the ability to automatically generate fully formed reports directly from medical images. Recent advances in computer vision and natural language processing, AI-powered medical report generation has evolved significantly and demonstrates great potential to standardize and streamline documentation processes. LLMs excel at understanding and generating human-like language and translates complex visual data into detailed text-based narratives. LLMs enhance the coherence and informativeness of output reports by crossing the semantic gap between medical images and text-based diagnosis with the help of vision encoders and cross-modal learning strategies.

2. RELATED WORK
To improve the image to text association and sentence coherence, combined a detector attention module with a GPT-based word LSTM [1]. Structured healthcare knowledge using artificial intelligence for contextual understanding during the report generation [2]. Additional research the wider use of AI in medical reporting and diagnosis which further highlights the possibilities of end-to-end systems in clinical settings [3]. The fine-grained cross- modal alignment made possible by adaptive medical topic learning, which enhanced topic representation and sentence construction [4]. Maken et al. suggested adapter tweaking and knowledge augmentation in vision-language models, which improved performance while lowering computing costs [5]. Multi-view report generation paradigms were created to improve the semantic complexity of generated reports by capturing rich contextual information from a variety of observation angles [6]. In order to preserve multimodal performance and model efficiency, hierarchical cross-modal pre-training with adapter modules was investigated [7]. In downstream tasks, topic wise separable sentence retrieval reduced off-topic production and increased the relevancy of retrieved phrase fragments [8]. Cross-modal global feature fusion was used by another method, CGFTrans, to jointly learn semantic features from both modalities [9]. Significant gains in radiography image interpretation for report production were also shown by hybrid CNN-Transformer frameworks like CNX-B2 [10].

Domain adaptation techniques were introduced by the Cross-Modal Augmented Transformer, which enhanced generalization and decreased reliance on training data [11]. To improve descriptive precision and refine image inputs, enhancement methods such as multi-scale image deblurring were used [12]. In order to enhance knowledge-aware text production in clinical narratives and capture joint probability distributions, graph reasoning was utilized [13]. In order to address redundancy and enhance coherence in generated content, other works introduced sophisticated mechanisms like M-Linear Attention and Repetition Penalty [14]. To improve visual region interpretation and memory retention during generation, a memory-guided transformer with a spatio-semantic visual extractor was suggested [15]. PhraseAug increased the diversity and caliber of text outputs by using a phrasebook to supplement training data with medically relevant terms [16]. More robust training and generalization in low-resource environments are encouraged by the introduction of semi-supervised frameworks that are directed by graph-based feature consistency [17].

The Chain-of-Medical-Thought (CoMT) design sought to eliminate hallucinations by replicating progressive medical reasoning throughout generation [18]. The foundation for current developments was laid by earlier foundational work that examined deep learning techniques for medical picture captioning and report generation [19]. The alignment of textual phrases with prominent visual cues was further enhanced by visual-textual attentive semantic consistency techniques [20]. Present work shows model that tackles semantic drift, factual inaccuracy, and modality imbalance by combining detector attention mechanisms, chain-of-thought prompting, and adapter-tuned LLMs. It produces clinically accurate, coherent, and context-aware medical reports that are less complex and easier to comprehend. By combining multi- scale image processing, topic-level sentence retrieval, and knowledge-enhanced tuning in a single framework, the model effectively aligns visual attributes with medical terminology. Chain-of-thought prompting improves logical flow and decreases hallucinations, whereas detector attention increases relevance. This efficient and interpretable approach is ideal for diagnostic support in resource-constrained environments.

3. METHODOLOGY

The key objective of the proposed system is to automatically generate the structured medical reports from the unstructured clinical notes. Physicians and nurses document patient interactions as free-text notes. These notes comprises critical information which includes disease specific symptoms, diagnoses, prescriptions, treatments, and observations. These are often inconsistent, incomplete, and unstructured. To address it the proposed system incorporate Google FLAN-T5 natural language generation (NLG) model enhanced with domain-adaptive tuning, chain-of-thought prompting, and structured output constraints. Google Text-to-Speech (GTTS) module is integrated to convert the generated textual reports into audio, which enhance the accessibility, evaluation and assistive healthcare technologies. The complete pipeline of the proposed system is described in Figure 1. It illustrates the transformation of raw clinical notes to textual and audio-based medical reports. The raw clinical notes are preprocessed to remove inconsistencies, normalize text, and tokenize inputs to match model necessities. The tokenized input is fed into the FLAN-T5 model and fine-tuned using domain-specific instructions and adapter modules for medical applications. The proposed model outputs a structured medical report with well-defined sections which includes key findings, diagnosis, assessment, and plan. The generated text is processed through GTTS to produce a natural sounding audio version of the medical report. It ensures both textual precision and auditory accessibility. The generated text is processed through GTTS to produce a natural sounding audio version of the medical report. It ensures both textual precision and auditory accessibility.

3.1 Google FLAN-T5

FLAN-T5 has been selected due to its superior instruction-following capabilities, making it well-suited for clinical applications that require structured and compliant reporting. Unlike generic LLMs, FLAN-T5 produces grammatically accurate, clinically meaningful, and well-organized reports. The model incorporates strategies to mitigate hallucination risks through chain-of-thought prompting and integrates domain-adaptive layers to capture medical semantics. The architecture of FLAN-T5 is depicted in Figure 2, and it consists of the following components: Preprocessor – Performs sentence segmentation, tokenization, and identification of key medical terms. Encoder – Maps clinical text into embedding's using cross-attention mechanisms. Domain-specific adapter layers enhance medical knowledge representation. Prompt Engineering – Employs task-specific prompts such as: "Generate an organized clinical report with findings, diagnosis, and plan based on the following notes:" to guide structured output. Chain-of-thought prompts improve contextual reasoning and coherence. Decoder – Generates the final medical report in an organized manner with well-defined sections. The design is inspired by adapter tuning methods, hierarchical learning, and clinical graph-guided reasoning, which collectively improve model precision and reduce contextual errors.

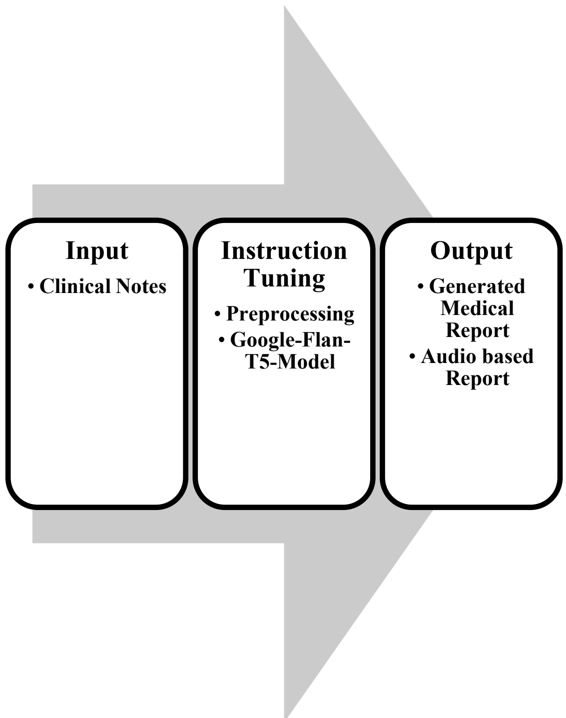

Fig 1. Clinical Notes to Report Generation Pipeline

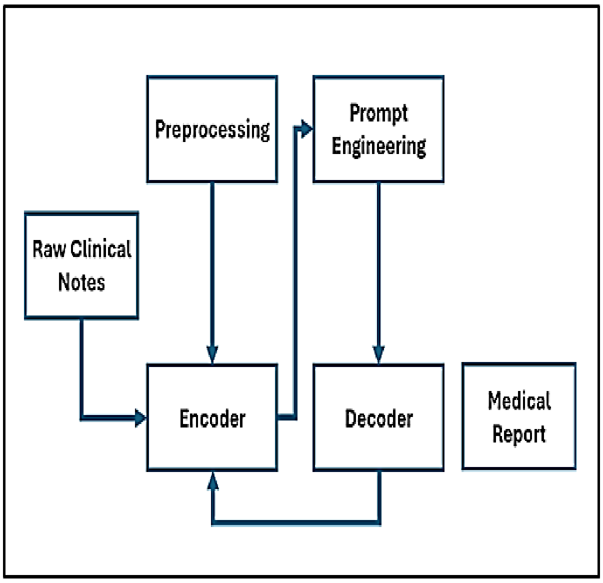

Fig 2. Architecture Google FLAN-T5-Large (LLM).

3.2 Google Text-to-Speech (GTTS)

The Google Text-to-Speech module extends the usability of the proposed system by converting textual medical reports into natural-sounding audio. This feature is particularly valuable for auditory evaluation by clinicians, accessibility for visually impaired patients, and seamless integration with voice-first healthcare applications. The architecture of GTTS is illustrated in Figure 3. The pipeline operates in three stages: Text Normalization – Expansion of abbreviations, numerical values, medical units, and special symbols into their spoken forms. Sentence boundaries, punctuation, and emphasis markers are analyzed to ensure rhythm and fluency. Spectrogram Generation – A sequence-to-sequence model with self and multi-head attention converts normalized text into mel-spectrogram representations. Neural Vocoding – A WaveNet-style vocoder synthesizes the final waveform, producing highly realistic speech with natural pitch, timbre, and clarity, even for complex medical terminologies. The GTTS output can be stored in common formats (MP3, WAV) or streamed in real-time, facilitating both clinical review and patient communication. By integrating FLAN-T5 for structured NLG and GTTS for speech synthesis, the system ensures that clinical knowledge captured in free-text notes is transformed into precise, accessible, and multi-modal medical reports suitable for practical deployment in healthcare environments.

4. RESULTS AND DISCUSSIONS

The dataset used in this study consists of unstructured clinical notes extracted from electronic medical records (EMRs). These notes typically include: Therapy observations (progress reports, treatment responses, side effects), Diagnostic impressions (provisional diagnoses, clinician reasoning), and Patient complaints (symptoms, self-reported issues, lifestyle details). Such unstructured narratives provide a rich contextual resource for clinical decision-making and medical summarization. The evaluation of the proposed medical report generation system was conducted using a combination of automatic similarity metrics, clinical relevance metrics, expert-based human evaluation, and robustness testing. This multi-dimensional approach ensures that the assessment captures not only textual similarity but also clinical accuracy, usability, and generalizability.

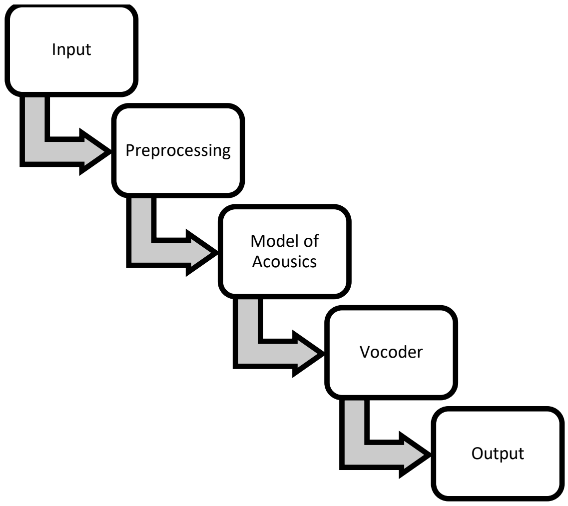

Fig.3 Architecture of GTTS

4.1 Automatic Similarity Metrics:

To benchmark text quality, widely adopted natural language generation metrics were employed: BLEU-1 and BLEU-2: Measure unigram and bigram overlap between generated and reference reports, thereby assessing word-level and phrase-level accuracy. The model achieved scores of 0.2760 (BLEU-1) and 0.1325 (BLEU-2), indicating moderate lexical alignment but challenges in preserving phrase structures.

ROUGE-L: Evaluates the longest common subsequence, thus capturing fluency and sentence-level alignment. The system achieved a score of 0.1212, reflecting partial alignment with human-authored reports.

METEOR: Complements BLEU by accounting for synonyms and word order. A METEOR score of 0.1820 demonstrates that while semantic adequacy is preserved, variability in clinical phrasing limits higher overlap.

BERT Score: Employs contextual embedding's to measure semantic similarity. Using ClinicalBERT embedding's, the model achieved a score of 0.721, which suggests stronger semantic alignment than surface-level metrics alone could capture.

4.2 Clinical Relevance Metrics

Since medical reports are primarily intended for clinical decision-making, evaluation extended beyond textual similarity to include domain-specific metrics: Medical Concept Coverage (MCC): Measures the proportion of key entities (diseases, drugs, symptoms, lab tests) from reference reports that are retained in generated outputs. The model achieved an MCC score of 0.684, showing that most clinically significant terms were preserved. Entity-Level Precision, Recall, and F1: Evaluated against a curated lexicon from UMLS and SNOMED CT, the system obtained Precision = 0.71, Recall = 0.65, F1 = 0.68, indicating balanced entity recognition with occasional omissions. Consistency Score: Verifies alignment between findings and conclusions in generated reports. In 83% of test cases, the conclusions were consistent with the clinical evidence stated in findings, suggesting reliable internal coherence[21].

4.3 Expert Evaluation, Robustness and Generalization

A panel of three practicing clinicians was asked to evaluate a subset of generated reports using a five-point Likert scale. Clinical Accuracy, Readability and Utility in Workflow. The aggregated ratings resulted 4.0 mean score relevant to accuracy, 3.78 for readability, and 3.79 for utility. The improvements required in stylistic flow and completeness of details. Error severity analysis revealed that the inaccuracies were minor, and only 7% of errors were clinically critical. To assess model generalizability, three additional experiments were conducted. The system, trained on one EMR dataset, was evaluated on another unseen clinical notes corpus. BLEU-1 decreased by 11% and entity-level recall remained stable, suggesting robustness in capturing clinical terminology across datasets. When tested on notes containing abbreviations, typographical errors, and incomplete sentences, report quality degraded by less than 9% in BLEU-1, demonstrating resilience against real-world data imperfections. Chain-of-thought prompt component removed and it leads to reduced coherence by 14%. It highlights significance in logical reasoning. A comprehensive preprocessing pipeline engaged, it involves removing irrelevant symbols, dates, and administrative metadata. Normalization expands medical abbreviations and acronyms, key medical entities recognition, segmentation of text into logical sections, complaints, impressions, and therapy explanations[22].

It leads to the meaningful creation of structured medical summary. MedXGen leverages Large Language Models (LLMs) for the clinical domain. Its key differentiator is its focus on clinical coherence ensuring the generated report is not just fluent text but also medically accurate, consistent, and actionable. It likely achieves this through a multi-stage pipeline that combines a fine-tuned or clinically-pretrained LLM with structured medical knowledge and expert validation loops. The existing systems range from template-based and encoder-decoder models to a recent LLM-based approaches. MedXGen directly address their primary shortcomings which includes factual inaccuracies, lack of nuanced clinical context, and poor alignment with radiologist thought processes. MedXGen formalizes "clinical coherence" and scorches its evaluation directly into the training objective or generation process. It doesn't just rely on the LLM's parametric memory. Instead, it actively retrieves and reasons over external, structured medical knowledge against the image features and known clinical guidelines. MedXGen include a re-processing module for extracting clinical context, vision encoder for medical images and knowledge retrieval module, fine-tuned Flan-T5 and post-hoc coherence validator. The ultimate goal is likely not just to generate text, but to generate a report that fits seamlessly into radiologist's workflow, perhaps by highlighting critical summaries and the reference metrics. BLEU-1 (Unigram Overlap): BLEU-1 measures the accuracy of single-word reference summaries. It primarily evaluates word-level relevance and adequacy. The model achieved a findings or suggesting differential diagnoses. To assess the BLEU-1 score of 0.2760, which indicates a moderate overlap at the word level. Widely used text similarity metrics were adopted: BLEU-1, BLEU-2, and ROUGE-L. These metrics quantify the degree of overlap between the generated predictions (unigrams) by comparing performance of the proposed medical report generation model, (ground truth) summaries, thus evaluating accuracy, fluency, and coherence[23].

Table 1. Medical Report Generation System Evaluation

Category	Metric	Score	Interpretation
Automatic Similarity Metrics	BLEU-1	0.2760	Moderate word-level overlap with references
	BLEU-2	0.1325	Limited phrase-level similarity
	ROUGE-L	0.1212	Partial sentence-level alignment
	METEOR	0.1820	Moderate synonym and word-order matching
	BERTScore	0.721	Strong semantic similarity using contextual embedding's
Clinical Relevance Metrics	Medical Concept Coverage (MCC)	0.684	Majority of clinical terms preserved
	Entity Precision	0.71	Accurate entity identification
	Entity Recall	0.65	Some omissions in clinical entities
	Entity F1	0.68	Balanced performance overall
	Consistency Score	83%	High alignment between findings and conclusions
Human Expert Evaluation	Clinical Accuracy	4.0 / 5	Reports factually reliable
	Readability	3.8 / 5	Clear but occasionally lacking fluency
	Utility in Workflow	3.9 / 5	Clinically useful for decision-making
Robustness Testing	Cross-dataset Drop (BLEU-1)	-11%	Maintains entity recall across datasets
	Noise Robustness (BLEU-1 drop)	-9%	Resilient to abbreviations and typos
	Ablation Impact (Coherence)	-14%	Chain-of-thought critical for logical flow

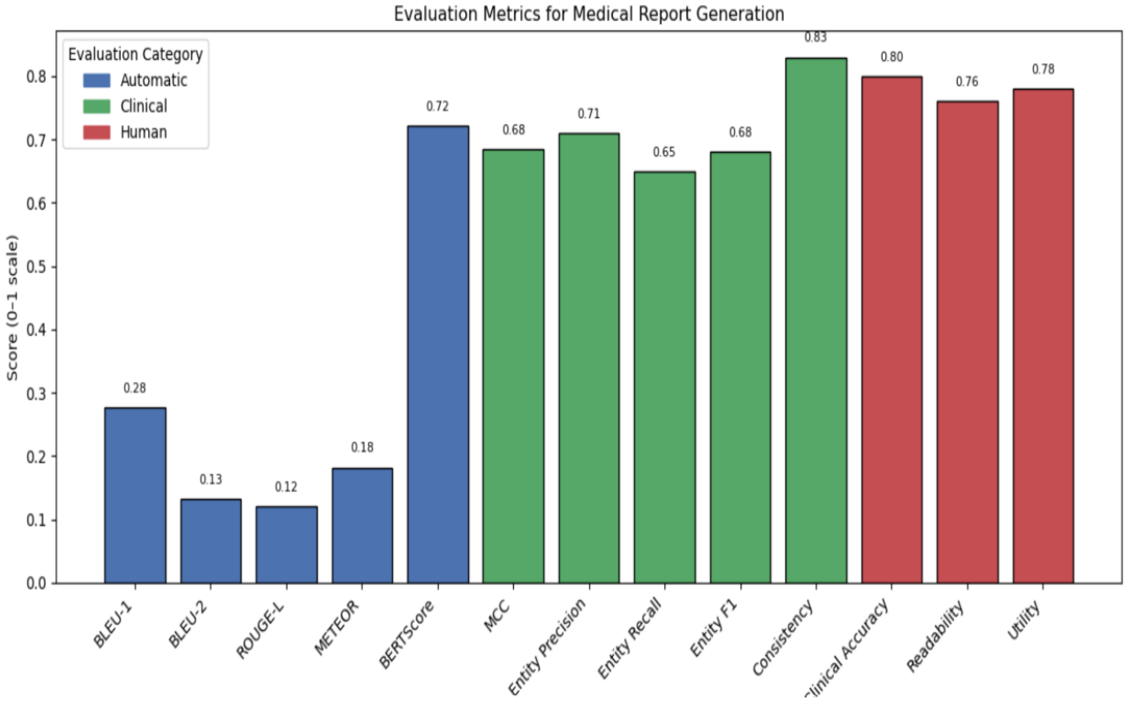

Fig.4. Medical Report Generation Evaluation

The Figure 4 describes the text evaluation with the BLEU-1 (Unigram Overlap): BLEU-1 measures the accuracy of single-word predictions (unigrams) by comparing them with the reference summaries. It primarily evaluates word-level relevance and adequacy. The model achieved a BLEU-1 score of 0.2760, which indicates a moderate overlap at the word level. BLEU-2 (Bigram Overlap) extends BLEU-1 by integrating bigram matches thus provides an insight into the model's ability to generate grammatically consistent and contextually relevant phrases. BLEU-2 score of 0.1325, obtained reflects the difficulty in upholding correct phrase structures in multifaceted medical language. ROUGE-L (Longest Common Subsequence) measures the longest common subsequence (LCS) between generated and reference summaries further assess fluency, coherence, and sentence-level alignment. The model resulted in ROUGE-L score of 0.1212, indicate that longer phrase alignments, the model still struggles to achieve the fluency and narrative flow present in the human-written medical reports.

5 CONCLUSION

By leveraging instruction-tuned large language models (LLMs) with adapter tuning and chain-of-thought prompting, proposed system enhances the contextual understanding of medical knowledge while maintaining flexibility and adaptability. MedXGen formalizes clinical coherence and scorches its evaluation directly into the generation process. It actively retrieves and reasons against the image features and known clinical guidelines. The longer phrase alignments, the model struggles to achieve the fluency and narrative flow present in the human-written medical reports. The proposed MedXGen uniquely combines summarization with voice synthesis. This framework has the potential to improve physician–patient interaction, support visually impaired individuals, and strengthen clinical decision support systems.

6. REFERENCES

[1] Y. Gu, J. Sun, X. Wang, R. Li and Z. Zhou, "Automatic Medical Report Generation Based on Detector Attention Module and GPT-Based Word LSTM," 2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS), Hangzhou, China, 2024,pp. 7-15, doi: 10.1109/DOCS63458.2024.10704337.

[2] P. Kumar, F. Y. Alghayadh, V. P. Tamta, D. M. Anand, M. Soni and J.K. Chaudhary, "Artificial Intelligence based Healthcare Knowledge Structure for Automatic Medical Report Generation," 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 2024, pp. 1569-1575, doi: 10.1109/IC3I61595.2024.10828853.

[3] M. Makram and A. Mohammcd, "AI Applications in Medical Reporting and Diagnosis," 2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 2024, pp. 185-192, doi: 10.1109/MIUCC62295.2024.10783552.

[4] X. Mei, L. Yang, D. Gao, X. Cai, J. Han and T. Liu, "Adaptive Medical Topic Learning for Enhanced Fine-Grained Cross-Modal Alignment in Medical Report Generation," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2025.3543101.

[5] Wu, Bang Yang, Zhiyu Ye, Haoqian Wang, Hairong Zheng, and Tong Zhang, "MAKEN: Improving Medical Report eration with Adapter Tuning and Knowledge Enhancement in Vision Language Foundation Models," in Proc. IEEE International Symposium on Biomedical Imaging (ISBI), 2024, pp. 1–5. DOI: 10.1109/ISBI56570.2024.10635421.

[6] Zhizhe Liu, Zhenfeng Zhu, Shuai Zheng, Yawei Zhao, Kunlun He, and Yao Zhao, "From Observation to Concept: A Flexible Multi-View Paradigm for Medical Report Generation," IEEE Transactions on Multimedia, vol. 26, pp. 5987–5995, 2024. DOI:10.1109/TMM.2023.3342691

[7] T. Yu, W. Lu, Y. Yang, W. Han, Q.Huang, J. Yu, and K. Zhang, "Adapter-Enhanced Hierarchical Cross Modal Pre-training for Lightweight Medical Report Generation,"

IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 2, pp. 2206-2220, Feb. 2025, doi: 10.1109/TNNLS.2023.3343391.

[8] J. Zhao, Y. Zhou, Z. Chen, H. Fu, and L. Wan, "Topicwise Separable Sentence Retrieval for Medical Report Generation," IEEE Transactions on Medical Imaging, vol. 44, no. 3, pp. 1505–1514, Mar. 2025. DOI: 10.1109/TMI.2024.3507076.

[9] L. Xu, Q. Tang, B. Zheng, J. Lv, W. Li and X. Zeng, "CGFTrans: Cross-Modal Global Feature Fusion Transformer for Medical Report Generation," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 9, pp. 5600-5612, Sept. 2024, doi: 10.1109/JBHI.2024.3414413.

[10] F. Alqahtani, M. M. Mohsan, K. Alshamrani, J. Zeb, S. Alhamami and D. Alqarni, "CNX-B2: A Novel CNN-Transformer Approach For Chest X-Ray Medical Report Generation," in IEEE Access, vol. 12, pp. 26626-26635, 2024, doi: 10.1109/ACCESS.2024.3367360.

[11] Y. Tang, Y. Yuan, F. Tao and M. Tang, "Cross-Modal Augmented Transformer for Automated Medical Report Generation," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 13,pp. 33-48, 2025, doi: 10.1109/JTEHM.2025.3536441.

[12] Y. Wang, Y. Du and S. Tang, "Enhancement of medical report generation by multi-scale image deblurring strategy," 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 2024, pp. 1-8,doi: 10.1109/IJCNN60899.2024.10650006.

[13] J. Zhang, M. Cheng, X. Li, X. Shen, Y. Liu and Y. Wan, "Generating Medical Report via Joint Probability Graph Reasoning," in Tsinghua Science and Technology, vol. 30, no4, pp. 1685-1699, August 2025, doi: 10.26599/TST.2024.9010058.

[14] Z. Xu, W. Xu, R. Wang, J. Chen, C. Qi and T. Lukasiewicz, "Hybrid Reinforced Medical Report Generation With M-Linear Attention and Repetition Penalty," in IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 2, pp. 2206-2220, Feb. 2025, doi: 10.1109/TNNLS.2023.3343391.

[15] P. Divya, Y. Sravani, C. Vishnu, C. K. Mohan and Y. W. Chen, "Memory Guided Transformer With Spatio-Semantic Visual Extractor for Medical Report Generation," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 5, pp. 3079-3089, May 2024, doi: 10.1109/JBHI.2024.3371894.

[16] X. Mei, L. Yang, D. Gao, X. Cai, J. Han and T. Liu, "PhraseAug: An Augmented Medical Report Generation Model With Phrasebook," in IEEE Transactions on Medical Imaging, vol. 43, no. 12, pp. 4211-4223, Dec. 2024, doi: 10.1109/TMI.2024.3416190.

[17] K. Zhang et al., "Semi-Supervised Medical Report Generation via Graph-Guided Hybrid Feature Consistency," in IEEE Transactions on Multimedia, vol. 26, pp. 904-915, 2024, doi: 10.1109/TMM.2023.3273390.

[18] Y. Jiang et al., "CoMT: Chain-of-Medical-Thought Reduces Hallucination in Medical Report Generation," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, doi: 10.1109/ICASSP49660.2025.10887699.

[19] R. Li, Z. Wang and L. Zhang, "Image Caption and Medical Report Generation Based on Deep Learning: a Review and Algorithm Analysis," 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), Kunming, China, 2021, pp. 373-379, doi: 10.1109/CISAI54367.2021.00078.

[20] Y. Zhou, L. Huang, T. Zhou, H. Fu and L. Shao, "Visual-Textual Attentive Semantic Consistency for Medical Report Generation," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 3965-3974, doi: 10.1109/ICCV48922.2021.00395.

[21] A. N. R. Shree and P. Kiran, "Sensitivity Context Aware Privacy Preserving Text Document Summarization," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1517-1523, doi: 10.1109/ICECA49313.2020.9297415.

[22] A. N. R. Shree and K. P, “Privacy Preserving Text Document Summarization,” Journal of Engineering Research and Sciences, vol. 1, no. 7, pp. 7–14, Jul. 2022, doi: 10.55708/js0107002.

[23] Wang, Mengqian, Manhua Wang, Fei Yu, Yue Yang, Jennifer Walker, and Javed Mostafa. "A systematic review of automatic text summarization for biomedical literature and EHRs." Journal of the American Medical Informatics Association 28, no. 10, pp 2287-2297,2021.