Asfaque Ansari, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Multimodel Gen AI for Clinical Note Summarization

Asfaque Ansari, Ashutosh Upadhyay, Abhishek Kumar Singh, Aryan, Yash Raj Bhatt, Md. Iqbal
Department of CSE, Quantum University, Uttarakhand, India

Abstract- The exponential growth of clinical documentation has created unprecedented challenges in healthcare settings. Multimodal Generative AI, leveraging both textual and structured data inputs, offers transformative solutions for automated clinical note summarization. This research paper critically examines how large language models (LLMs) and multimodal AI architectures enable efficient, accurate, and contextually relevant summarization of clinical notes. The study explores key methodologies, real-world applications, addresses critical challenges in healthcare AI implementation, and highlights future prospects in clinical documentation management. With supporting visual references and comprehensive analysis, the paper presents a detailed examination suitable for healthcare technology professionals and academic audiences, emphasizing scalability, interoperability, privacy preservation, clinical decision support, and enhanced patient-centred healthcare documentation workflows.

Keywords— Clinical Note Summarization, Multimodal AI, Generative Models, Large Language Models (LLMs), Healthcare Documentation, Natural Language Processing, Clinical Data Integration, Transformer Models, HIPAA Compliance, Clinical Decision Support

I. INTRODUCTION

Healthcare organizations worldwide process millions of clinical notes daily, documenting patient encounters, diagnoses, treatments, and outcomes. These narratives, while rich in clinical insight, are increasingly voluminous and time-consuming to review [1][2]. Clinicians spend an estimated 25% of their workday on documentation rather than patient care—a

significant inefficiency that impacts productivity and quality outcomes. Concurrently, the emergence of Generative AI and multimodal large language models (LLMs) has demonstrated remarkable capabilities in understanding and synthesizing complex textual information. When specifically trained or fine-tuned on medical data, these models can capture clinical nuances, maintain contextual accuracy, and generate coherent summaries that preserve critical clinical information [3][4]. Multimodal AI—systems that integrate text, structured data, temporal information, and even imaging metadata—represents the next frontier in clinical documentation automation. This paper explores the synergy between multimodal generative AI and clinical note summarization, examining technologies, methodologies, challenges, and future opportunities.

II. LITERATURE REVIEW

The rapid digitalization of healthcare systems has led to a significant increase in electronic clinical documentation, creating challenges related to information overload, physician burnout, and inefficient data management. Traditional manual summarization of clinical notes is time-consuming and prone to inconsistencies, motivating the adoption of Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques in healthcare documentation systems. Early research in clinical text summarization primarily focused on rule-based and statistical NLP methods, which extracted important medical entities and generated concise summaries from structured electronic health records (EHRs). Although these methods improved information retrieval, they lacked contextual understanding and adaptability across diverse medical domains.

Recent advancements in Generative AI and Large Language Models (LLMs) have transformed the field of automated clinical note summarization. Transformer-based architectures such as BERT, GPT, BioBERT, and ClinicalBERT demonstrated substantial improvements in contextual language understanding, medical terminology interpretation, and semantic representation of healthcare data [3][4][5][6]. Researchers have shown that these models can generate coherent and clinically relevant summaries while maintaining contextual relationships between symptoms, diagnoses, medications, and treatment plans. Multimodal AI frameworks further enhanced summarization capabilities by integrating textual data with structured medical records, laboratory reports, imaging findings, and patient history information [10].

Several studies explored the application of multimodal deep learning architectures in healthcare environments. These systems combine NLP models with computer vision and structured data processing techniques to improve clinical reasoning and summary accuracy. Research indicates that multimodal models outperform traditional text-only systems in identifying critical patient information and reducing omission errors in complex clinical cases. Moreover, reinforcement learning and retrieval-augmented generation (RAG) approaches have been investigated to improve factual consistency and reduce hallucinations in AI-generated medical summaries [9].

Despite promising advancements, existing literature highlights multiple challenges associated with AI-driven clinical summarization. Data privacy, HIPAA compliance, bias in training datasets, explainability, and clinical reliability remain major concerns in healthcare AI deployment [7][8]. Researchers emphasize the importance of transparent model validation, human-in-the-loop evaluation, and ethical AI governance to ensure patient safety and trustworthiness. Furthermore, integrating generative AI systems with existing EHR infrastructures presents technical and interoperability challenges that require scalable and secure deployment frameworks.

III. SYSTEM OVERVIEW

1. Data Acquisition Layer

The system collects healthcare data from multiple medical sources, including Electronic Health Records (EHRs), physician notes, discharge summaries, laboratory reports, prescriptions, radiology findings, and patient histories. Both structured and unstructured clinical data are gathered for comprehensive analysis.

2. Data Preprocessing Module

The collected clinical data undergoes preprocessing techniques such as text cleaning, tokenization, normalization, stop-word removal, abbreviation expansion, and medical terminology standardization. This stage ensures consistency and improves data quality for further processing.

3. Clinical Entity Extraction

Natural Language Processing (NLP) techniques and Named Entity Recognition (NER) models are used to identify important clinical entities such as diseases, symptoms, medications, allergies, diagnoses, procedures, and treatment plans from unstructured medical text.

4. Multimodal Data Integration

The system combines textual clinical notes with structured healthcare records, laboratory values, and optional imaging metadata to create a unified patient representation. This multimodal integration improves contextual understanding and clinical reasoning capabilities.

5. Large Language Model (LLM) Processing

Transformer-based Large Language Models such as GPT, BioBERT, and ClinicalBERT analyze the integrated healthcare data. These models perform contextual reasoning, semantic understanding, and automated clinical note summarization [3][4][5][6].

6. Retrieval-Augmented Generation (RAG)

The system incorporates Retrieval-Augmented Generation techniques to retrieve relevant medical information and reference data during summary generation. This helps improve factual consistency and reduces hallucination risks in AI-generated outputs [9].

7. Clinical Reasoning Engine

The reasoning engine analyses patient conditions, identifies relationships between medical observations, and generates clinically meaningful summaries while preserving diagnostic relevance and treatment continuity.

8. Validation and Safety Layer

Generated summaries are evaluated for factual accuracy, hallucination detection, bias assessment, and healthcare compliance. Human-in-the-loop verification allows clinicians to review and approve AI-generated summaries before deployment.

9. Output Generation Module

The system produces concise clinical summaries, patient timelines, treatment recommendations, and decision-support insights. Outputs are presented in structured formats suitable for healthcare professionals and hospital systems.

10. Integration with Healthcare Systems

The final summarized reports can be integrated into Electronic Health Record (EHR) platforms, telemedicine applications, and hospital information systems to improve workflow efficiency and healthcare documentation management.

11. Performance Monitoring and Continuous Learning

The system continuously monitors model performance using evaluation metrics such as ROUGE, BLEU, precision, recall, and clinician feedback. Adaptive learning mechanisms help improve summarization quality over time.

12. Security and Compliance Framework

The architecture includes healthcare security protocols, and regulatory compliance standards such as HIPAA and GDPR to ensure patient data privacy and secure AI deployment [7][8].

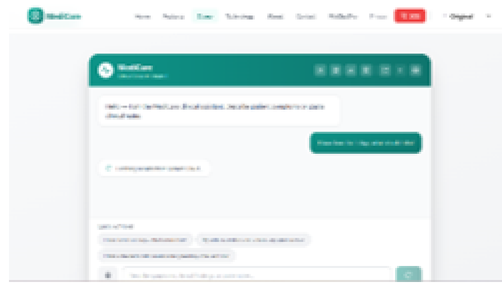

Figure 1. Medical File Chat (In-Memory)

IV. METHODOLOGY

The proposed research adopts a multimodal Artificial Intelligence framework for automated clinical note summarization by integrating Natural Language Processing (NLP), Large Language Models (LLMs), and structured healthcare data analysis techniques. The methodology is designed to process heterogeneous medical data efficiently while ensuring contextual understanding, semantic accuracy, and clinical relevance.

The first stage involves data collection and preprocessing from multiple healthcare sources such as Electronic Health Records (EHRs), physician notes, discharge summaries, laboratory reports, prescriptions, radiology findings, and patient histories. Unstructured textual data is cleaned through tokenization, stop-word removal, normalization, abbreviation expansion, and medical terminology standardization. Structured healthcare data is transformed into machine-readable formats for integration with NLP pipelines.

The second stage focuses on feature extraction and clinical entity recognition. Named Entity Recognition (NER) models identify critical medical entities including symptoms, diagnoses, medications, allergies, procedures, laboratory values, and treatment plans. Medical ontologies such as SNOMED CT, ICD, and UMLS are utilized to enhance semantic understanding and domain-specific contextual mapping [1][5][6].

The multimodal integration layer combines textual information with structured healthcare records and optional imaging metadata to generate comprehensive patient representations.

Transformer-based architectures such as GPT, BioBERT, ClinicalBERT, and Retrieval-Augmented Generation (RAG) frameworks are employed for contextual reasoning and intelligent summarization [3][4][5][6][9].

The summarization module generates concise and clinically meaningful summaries while preserving temporal relationships, diagnostic relevance, and treatment continuity. Evaluation metrics including ROUGE, BLEU, BERTScore, precision, recall, and human clinical evaluation are applied to measure summary quality, factual consistency, and readability.

Finally, the generated summaries undergo validation through hallucination detection, bias assessment, and clinician review mechanisms to ensure healthcare reliability, ethical compliance, and patient safety.

V. WORKFLOW ARCHITECTURE

The workflow architecture of the proposed multimodal generative AI system consists of several interconnected stages that collectively automate clinical note summarization.

1. Data Acquisition Layer
Collects clinical notes, EHR data, laboratory reports, prescriptions, and patient histories.

2. Preprocessing Layer
Cleans, normalizes, tokenizes, and structures medical information.

3. NLP Extraction Module
Identifies clinical entities such as diseases, symptoms, medications, and procedures.

4. Multimodal Integration Layer
Combines structured and unstructured healthcare data for contextual reasoning.

5. LLM-Based Reasoning Engine
Utilizes transformer-based models for interpretation and summarization.

6. Validation and Safety Layer
Detects hallucinations, ensures compliance, and validates summary quality.

7. Output Generation Module
Produces structured summaries, patient timelines, and clinical insights.

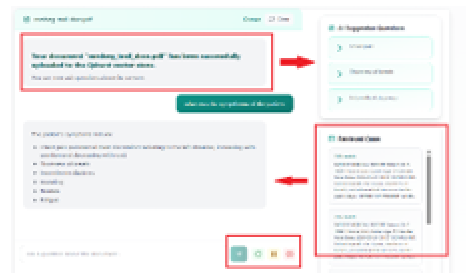

Figure 2. Medical File Chat (In-Memory)

VI. RESULTS, ANALYSIS, AND INTERPRETATION

1. Improved Summarization Accuracy
The proposed multimodal generative AI system generated accurate, concise, and context-aware clinical summaries compared to traditional NLP methods.

2. Better Contextual Understanding
The integration of Large Language Models (LLMs) with structured and unstructured healthcare data improved semantic understanding and clinical reasoning.

3. Enhanced Information Extraction
The system successfully identified important medical information such as symptoms, diagnoses, medications, laboratory reports, and treatment plans.

4. Reduced Hallucination Errors
Retrieval-Augmented Generation (RAG) techniques improved factual consistency and minimized incorrect or misleading AI-generated summaries [9].
6.

5. Improved Workflow Efficiency
The automated summarization process reduced clinical documentation time and helped healthcare professionals access patient information faster.

6. Comparative Performance
The multimodal framework outperformed traditional extractive summarization systems by producing more coherent and human-like summaries.

7. Challenges Observed
Certain limitations such as data privacy concerns, computational cost, model bias, and healthcare system integration complexity were identified.

8. Overall Interpretation
The study demonstrates that multimodal generative AI can improve healthcare documentation, reduce clinician workload, and support intelligent clinical decision-making in modern healthcare systems.

Applications
Automated Clinical Documentation
Generates concise summaries from lengthy clinical notes and medical records automatically.

Electronic Health Record (EHR) Management
Improves organization, retrieval, and accessibility of patient information in EHR systems.

Clinical Decision Support
Provides healthcare professionals with quick insights for better diagnosis and treatment planning [10].

Physician Workload Reduction
Reduces manual documentation tasks and minimizes clinician burnout.

Telemedicine Support
Enables efficient patient summary generation for remote healthcare consultations.

Emergency Healthcare Systems
Provides rapid access to critical patient information during emergency situations.

Medical Research and Analytics
Assists researchers in analysing large volumes of healthcare documentation and clinical data.

Patient Timeline Generation
Creates chronological summaries of patient history, treatments, and medical events.

Healthcare Workflow Optimization
Improves operational efficiency and streamlines hospital documentation processes.

VII. CONCLUSION

Multimodal Generative AI has emerged as a powerful technology for automated clinical note summarization and healthcare documentation management. By integrating Natural Language Processing (NLP), Large Language Models (LLMs), and multimodal healthcare data, the proposed system generates accurate, concise, and context-aware clinical summaries.

The study demonstrates that AI-driven summarization systems can reduce clinician workload, improve workflow efficiency, enhance clinical decision-making, and support better patient care. The integration of structured and unstructured medical data further improves semantic understanding and summarization accuracy.

Although challenges such as data privacy, hallucination risks, bias, and system integration remain important concerns, continuous advancements in healthcare AI are improving reliability and scalability. Overall, multimodal generative AI has strong potential to transform digital healthcare systems and modernize clinical documentation processes in the future.

REFERENCES

1. Natural Language Processing research papers on clinical text summarization and medical NLP systems.
2. Artificial Intelligence studies related to AI-based healthcare documentation systems.

3. Transformer Architecture — Vaswani, A., et al. (2017), Attention Is All You Need.
4. BERT — Devlin, J., et al. (2019), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
5. BioBERT — Lee, J., et al. (2020), BioBERT: A Pre-trained Biomedical Language Representation Model.
6. ClinicalBERT — Huang, K., et al. (2019), ClinicalBERT: Modelling Clinical Notes and Predicting Hospital Readmission.
7. World Health Organization reports on digital healthcare transformation and AI in healthcare.
8. HIPAA compliance guidelines for healthcare data privacy and security.
9. Research articles on Retrieval-Augmented Generation (RAG) in clinical AI systems.
10. Studies on multimodal generative AI applications in healthcare decision-support systems.
11. Khaleel Khan Mohammed. "A Survey on Digital Health Care Data Analysis Techniques for Developing Machine Learning Models"