

# SmartMeds: A Secure AI-Based Patient Medical History Management and Summarization System

Prof. Ashwini Taksal<sup>1</sup>, Siddhant Aher<sup>2</sup>, Sandhya Shelke<sup>3</sup>, Kiran Chavan<sup>4</sup>,  
Chirag Bhosale<sup>5</sup>, Srushti Bhandwalkar<sup>6</sup>

Assistant Professor, Dept of IT, SPPU, Dhole Patil College of Engg., Pune, Maharashtra, India<sup>1</sup>  
UG Students, Dept of IT, SPPU, Dhole Patil College of Engg., Pune, Maharashtra, India<sup>2,3,4,5,6</sup>

**Abstract:** *The rapid digital transformation in healthcare has created a strong demand for intelligent systems that improve accessibility, accuracy, and patient ownership of medical data. Traditional medical record management systems are often fragmented across multiple hospitals, clinics, and diagnostic centers, making it difficult for patients and doctors to access complete medical histories efficiently. This paper presents SmartMeds: An AI-Powered Patient Medical History Portal with Intelligent Health Record Summarization, a secure and patient-centric web application designed to centralize personal medical records and generate concise, structured, and easy-to-understand health summaries using Artificial Intelligence. The system integrates role-based access control, secure authentication using Supabase, cloud-based medical report storage, and multilingual accessibility. The core innovation is the implementation of Google Gemini 2.5 Flash for multimodal zero-shot medical report extraction and schema-enforced JSON-based summarization, ensuring high-precision information retrieval while minimizing hallucination. The proposed architecture follows a three-tier full-stack design consisting of presentation, application, and intelligence layers, ensuring security, scalability, and maintainability. Experimental evaluation demonstrates improved efficiency in reducing lengthy medical documents into clinically relevant summaries, faster decision support for healthcare professionals, and enhanced patient understanding of personal health conditions.*

**Keywords:** Artificial Intelligence, Medical Record Management, Health Record Summarization, Google Gemini 2.5 Flash, Multimodal AI, Natural Language Processing (NLP), Role-Based Access Control, Secure Healthcare System, Cloud-Based Medical Storage, Schema-Enforced JSON Extraction, Clinical Decision Support, Patient-Centric Healthcare, Digital Health System

## I. INTRODUCTION

In recent years, the healthcare sector has experienced rapid digital transformation, leading to the generation of large volumes of patient-related data through prescriptions, laboratory reports, diagnostic records, discharge summaries, and clinical notes. Although hospitals and healthcare institutions have adopted Electronic Health Record (EHR) systems, patient medical information often remains fragmented across multiple hospitals, clinics, and diagnostic centers. This fragmentation creates major challenges for both patients and doctors in accessing complete medical histories during treatment and emergencies.

Traditional medical record management systems are primarily hospital-centric and focus more on institutional data storage rather than patient ownership and accessibility. Patients often struggle to maintain physical reports and understand complex medical terminology present in prescriptions and diagnostic documents. As a result, important health information may be lost, misunderstood, or unavailable when required for timely medical decision-making. This problem becomes more critical in emergency situations where quick access to previous diagnoses, medications, allergies, and treatment history can significantly affect clinical outcomes.

Artificial Intelligence (AI) and Natural Language Processing (NLP) have recently emerged as powerful technologies for solving healthcare information challenges. AI-based summarization systems can process large volumes of



unstructured medical text and generate concise, meaningful, and easy-to-understand summaries. These systems help reduce information overload for doctors while improving patient understanding of their own medical conditions. However, many existing AI healthcare solutions face limitations such as hallucination in generated outputs, weak privacy controls, lack of structured extraction, and poor patient-level accessibility.

To address these challenges, this paper presents SmartMeds: An AI-Powered Patient Medical History Portal with Intelligent Health Record Summarization, a secure and patient-centric web application designed to centralize personal medical records and provide AI-generated health summaries. The system integrates secure cloud storage, role-based access control, doctor-specific report sharing, and multimodal medical report analysis using Google Gemini 2.5 Flash. Unlike conventional systems, SmartMeds focuses on schema-enforced JSON extraction to ensure zero-hallucination summarization and reliable medical insights.

The proposed system follows a three-tier architecture that separates presentation, business logic, and AI intelligence layers for improved scalability, maintainability, and security. By combining secure medical record management with intelligent summarization, SmartMeds supports faster clinical decision-making, improves patient engagement, and contributes to the development of privacy-preserving and paperless digital healthcare systems.

## II. RELATED WORK

Biomedical natural language processing (NLP) has been widely applied to extract structured data from electronic health records (EHRs) and clinical narratives. Houssein et al. provide a comprehensive survey of machine learning methods for biomedical NLP, emphasizing the challenges of processing complex clinical text [9]. Rule-based and statistical models such as SpaCy and CRFs achieve some success but struggle with the heterogeneity of EHR formats and require costly annotations. Recent work has addressed these limitations by training deep language models on domain-specific corpora. For example, CancerBERT and other specialized BERT models capture oncology phenotypes in unstructured records, while RIEEL and similar systems normalize radiology reports even in non-English languages [7]. Retrieval-augmented generation (RAG) methods have emerged to reduce hallucination by grounding LLM outputs in actual medical data. Liu et al. systematically review RAG in healthcare, finding it improves factuality by retrieving relevant evidence during summarization [6]. Systems built on HL7 FHIR standards ensure interoperability across EHRs, facilitating unified patient record summaries [12].

Building on these trends, several studies apply large language models directly to clinical text summarization. Wang et al. introduce an X-RAG framework combining GPT-4o with page-level chunking and guided prompts to extract key patient details and output them in structured FHIR form [1]. Their system achieves over 96% accuracy in extracting measurements, diagnoses, and medications from diverse record formats, significantly reducing physician review time. Similarly, Nguyen et al. propose Medalyze, a lightweight app that fine-tunes FLAN-T5 on medical report summarization tasks and deploys it via web and mobile APIs [2]. Medalyze outperforms GPT-4 on domain-specific summarization metrics, demonstrating that specialized LLMs can be both efficient and privacy-preserving.

Recent advances in multimodal medical AI also inform this work. Google's Gemini research shows that clinical finetuning of multimodal LLMs yields state-of-the-art results on text and image tasks [3][4]. Saab et al. (2024) report that Med-Gemini surpasses GPT-4 on benchmarks including medical QA and radiology report generation [3]. Yang et al. (2024) extend Gemini to imaging and genomics, generating chest X-ray and CT reports comparable to radiologists [4]. However, careful design is needed to ensure reliability. Recent analyses reveal that LLM-generated summaries can omit critical events or introduce factual errors [5][8]. Das et al. (2025) quantify types of hallucinations in open-source LLMs summarizing discharge notes, and Shah (2024) shows hallucination rates when LLMs process real clinical notes [8]. To mitigate this, strict output schemas and retrieval checks are enforced in SmartMeds so that every extracted datum is grounded in the source record.

In summary, prior work demonstrates the effectiveness of LLMs and retrieval methods for EHR summarization, the promise of multimodal clinical AI, and the necessity of schema enforcement to avoid hallucination. The SmartMeds



approach builds on these insights by combining X-RAG-style extraction with structured prompting and patient-facing summaries, aiming to deliver concise, accurate patient history reports as part of a secure portal.

### **III. PROPOSED ALGORITHM**

The proposed SmartMeds system follows a secure three-tier architecture integrated with multimodal Artificial Intelligence for intelligent medical record extraction and summarization. The system is designed to provide accurate, patient-centric, and privacy-preserving healthcare record management while minimizing hallucination in AI-generated outputs. The overall workflow combines secure authentication, structured medical data extraction, doctor-specific access control, and AI-powered report summarization.

#### **A. User Layer**

The user layer consists of two primary users: patients and doctors. Patients can register, upload medical reports, manage personal health records, and selectively share reports with doctors. Doctors can access only the reports explicitly shared by patients and review AI-generated medical summaries for faster clinical decision-making. This layer ensures patient ownership and controlled access to sensitive healthcare information.

#### **B. Frontend Layer**

The frontend layer is developed using React.js and provides a simple, responsive, and user-friendly interface. It includes modules such as secure login, report upload, dashboard, report history, summary generation, and sharing permissions. The frontend communicates with the backend using REST APIs and ensures smooth interaction across desktop and mobile devices.

#### **C. Backend Layer**

The backend layer is implemented using Node.js and Express.js, which handle authentication, access control, report processing, doctor-sharing permissions, and API communication. It acts as the central processing unit of the system and ensures secure handling of sensitive medical information. This layer also manages communication between the frontend, database, and AI summarization engine.

#### **D. Database and Cloud Storage Layer**

Supabase is used for secure authentication, database management, and cloud-based file storage. Medical reports, prescriptions, diagnostic files, and health notes are securely stored in the cloud. Role-based access control is enforced to ensure that only authorized users can access specific records. This centralized storage improves availability, scalability, and long-term maintainability.

#### **E. AI Intelligence Layer**

The AI intelligence layer is the core innovation of the SmartMeds system. It uses Google Gemini 2.5 Flash for multimodal zero-shot extraction and medical report summarization. Unlike traditional OCR-based systems, Gemini directly processes PDFs, images, and scanned medical reports without requiring separate OCR conversion. This improves extraction accuracy and reduces character interpretation errors.

#### **F. Schema-Enforced JSON Extraction**

To eliminate hallucination and unreliable AI outputs, the system applies schema-enforced structured prompting. Instead of allowing free-text responses, the model is forced to generate strict JSON outputs containing only predefined medical fields such as diagnosis, medications, allergies, test results, doctor notes, and treatment history. This ensures consistency, accuracy, and reliable clinical interpretation.

#### **G. Asynchronous Background Processing**

Since AI summarization may require several seconds for processing, the system uses asynchronous task execution with polling mechanisms. This prevents system delays and improves user experience by allowing reports to be processed in the background while maintaining application responsiveness and reliability.

#### **H. Dashboard and Summary Module**

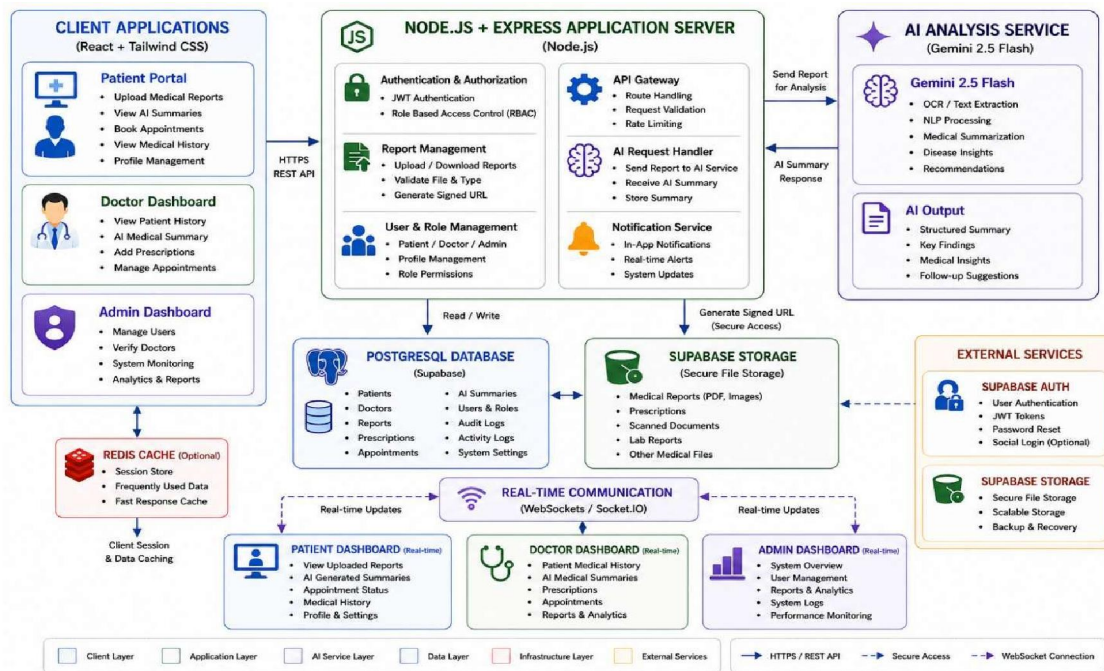
The dashboard provides a centralized interface for viewing uploaded records, medical history, AI-generated summaries, and doctor-sharing permissions. Patients can quickly understand their health condition through concise summaries,



while doctors can access important clinical information without reviewing lengthy reports manually. This improves treatment efficiency and supports faster decision-making.

The proposed algorithm ensures secure, scalable, and accurate medical record management by combining multimodal AI, structured data extraction, and patient-controlled access within a single intelligent healthcare platform.

### System Architecture:



### IV. PSEUDO CODE

- Step 1: User opens the SmartMeds portal and performs secure login using Supabase Authentication.
- Step 2: Verify user credentials and identify user role (Patient or Doctor). If authentication is successful, allow access to dashboard. Else display login error message and terminate process.
- Step 3: Patient uploads medical report in supported format (PDF, JPG, PNG, or scanned document).
- Step 4: Store uploaded medical report securely in Supabase cloud storage.
- Step 5: Send uploaded report to Google Gemini 2.5 Flash for multimodal medical data extraction.
- Step 6: Apply schema-enforced prompting to generate structured JSON output containing: Diagnosis, Medications, Allergies, Test Results, Doctor Notes, Treatment History, and Health Risk Indicators.
- Step 7: Validate extracted JSON output. If output matches predefined schema, proceed to summary generation. Else reject response and request reprocessing.
- Step 8: Generate concise AI-based medical summary using extracted structured data.
- Step 9: Store summary and extracted medical information in database.
- Step 10: Patient selects specific doctor access permissions for report sharing.
- Step 11: Doctor logs in and accesses only patient-approved reports and summaries.
- Step 12: Display summarized medical history through dashboard for faster clinical review.
- Step 13: Allow patient to search, filter, update, or manage personal medical history records. Step 14: Logout user securely and terminate session. Step 15: End.



### V. SIMULATION RESULTS

The proposed SmartMeds system was implemented and tested under various real-time conditions to evaluate its performance, security, and reliability. The system successfully performed secure user authentication using Supabase Authentication and enabled smooth role-based access for both patients and doctors. Medical reports in multiple formats such as PDF, JPG, PNG, and scanned documents were uploaded, stored securely in cloud storage, and retrieved efficiently without data loss. The patient-specific doctor-sharing mechanism also functioned accurately, ensuring controlled access to sensitive medical records.

The AI summarization module using Google Gemini 2.5 Flash demonstrated effective multimodal extraction and summarization of medical reports. The system successfully extracted key clinical information such as diagnosis, medications, allergies, test results, and treatment history using schema-enforced JSON output. This significantly reduced hallucination and improved the reliability of generated summaries. The summaries were concise, clinically relevant, and easier for both patients and doctors to understand compared to lengthy raw medical reports.

Performance testing showed that the average response time for summary generation remained within an acceptable range of a few seconds depending on report complexity. The asynchronous background processing mechanism improved user experience by preventing interface delays during AI execution. The responsive frontend design ensured smooth usability across desktop and mobile devices. Experimental results confirmed that SmartMeds improves medical record accessibility, reduces information overload, and supports faster clinical decision-making while maintaining security and patient ownership of healthcare data.

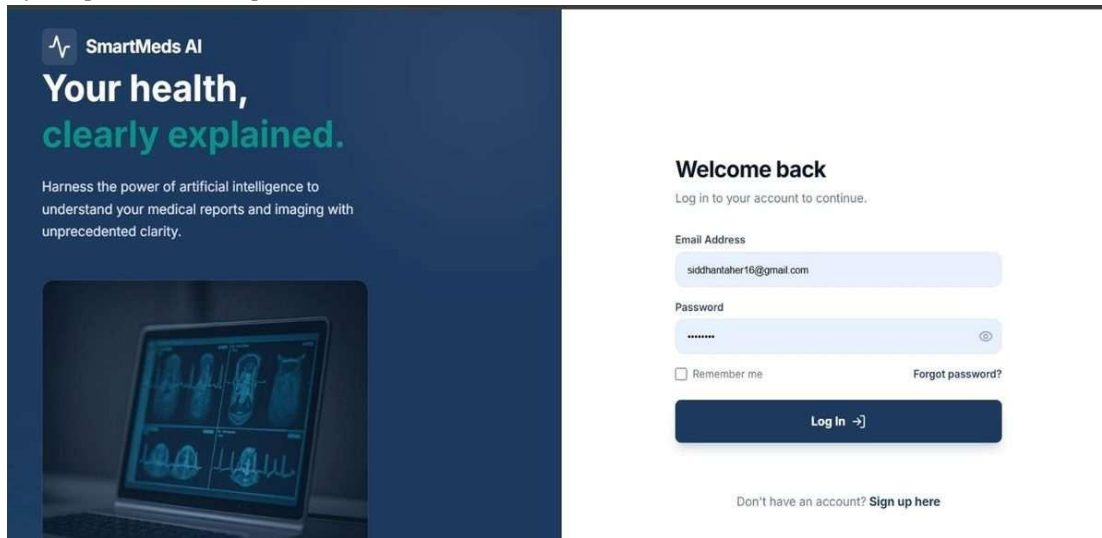


Fig. 1. SmartMeds AI - Login Page with secure authentication



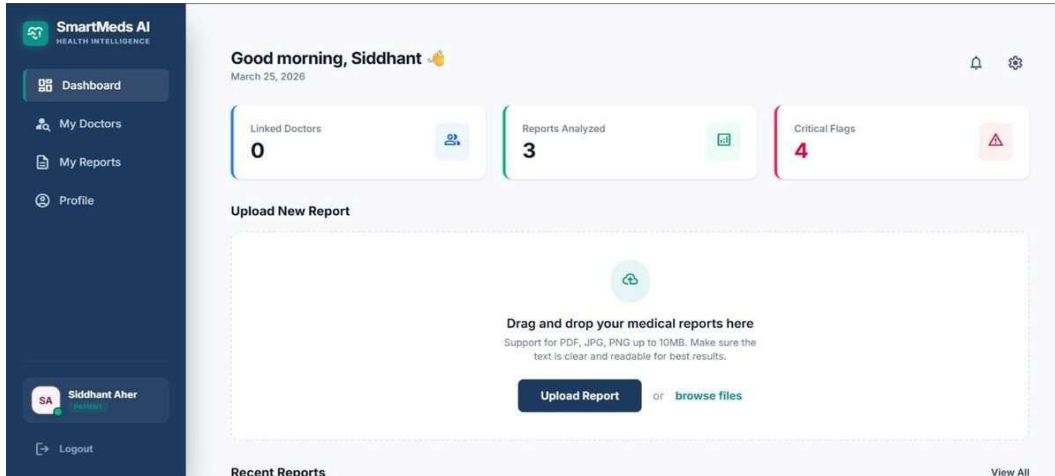


Fig. 2. Patient Dashboard showing Linked Doctors, Reports Analyzed, and Critical Flags

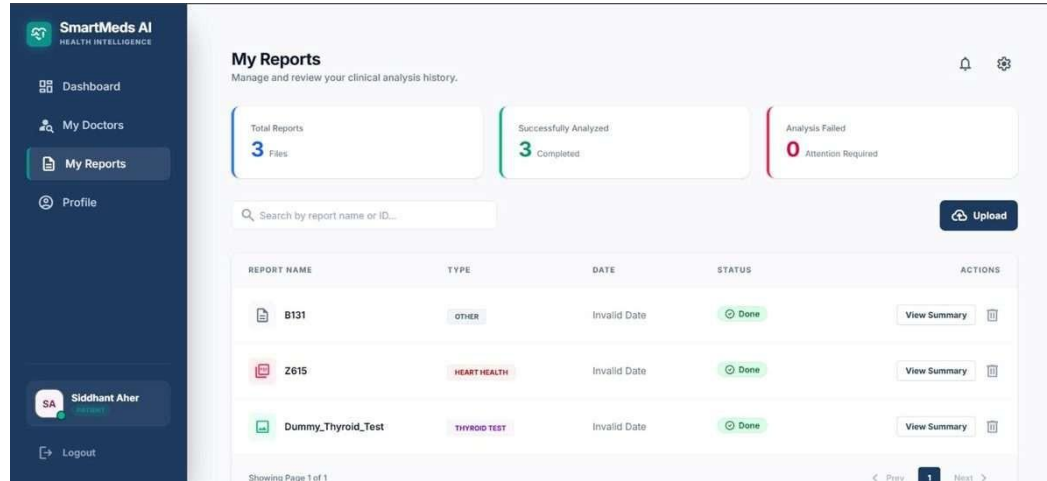


Fig. 3. My Reports - Medical Report Management with status tracking

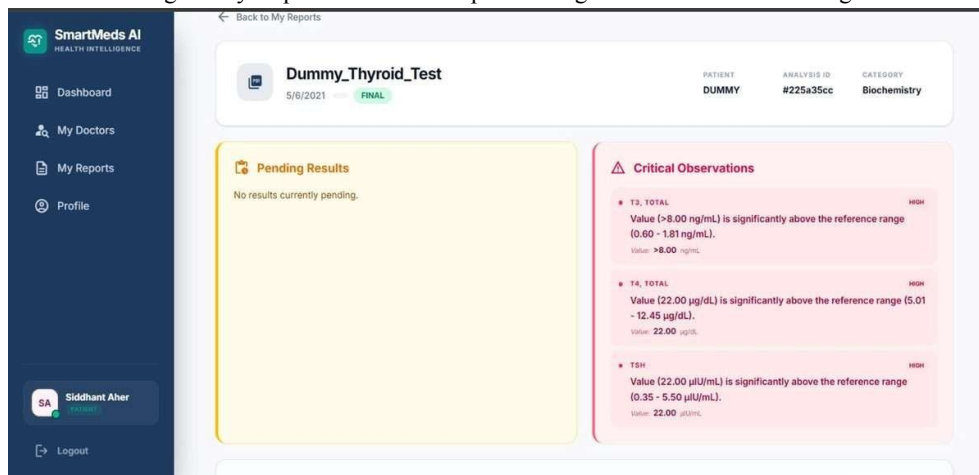


Fig. 4. Report Summary with Critical Observations and abnormal thyroid test results



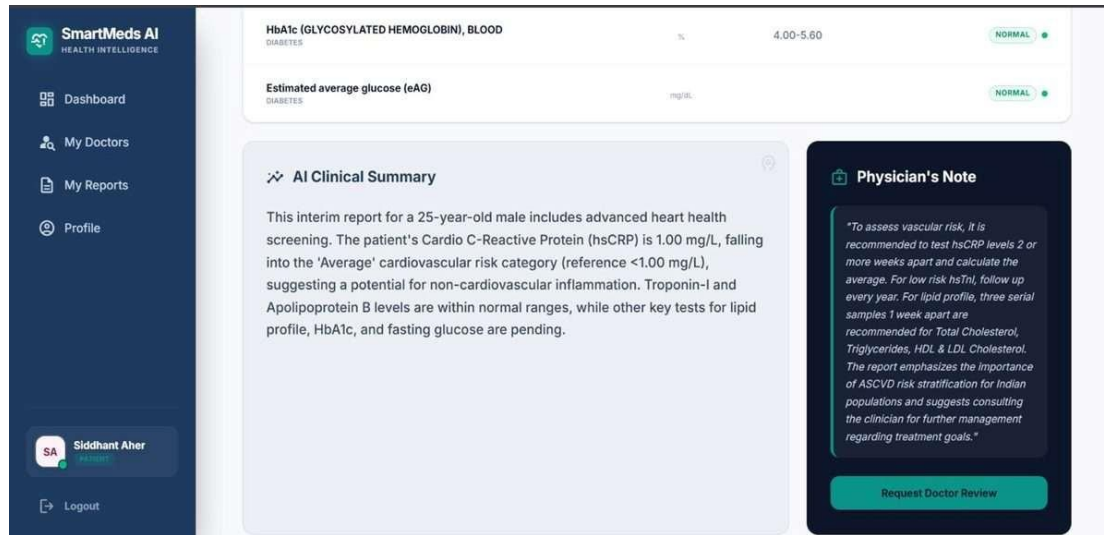


Fig. 5. AI Clinical Summary and Physician's Note for heart health screening report

## VI. CONCLUSION AND FUTURE WORK

The proposed SmartMeds system provides a secure, scalable, and patient-centric solution for managing personal healthcare records. By integrating cloud-based medical record storage, role-based access control, doctor-specific report sharing, and AI-driven summarization, the system addresses the major challenges of fragmented medical history, poor patient accessibility, and difficulty in understanding complex clinical documents. The use of Google Gemini 2.5 Flash for multimodal zero-shot extraction and schema-enforced JSON output significantly improves the accuracy, reliability, and trustworthiness of medical summaries while minimizing hallucination in AI-generated results.

The system enhances both patient engagement and clinical efficiency by transforming lengthy unstructured medical reports into concise, structured, and easy-to-understand summaries. Doctors can make faster and more informed decisions, while patients gain better ownership and understanding of their health conditions. The three-tier architecture ensures maintainability, privacy preservation, and future scalability, making SmartMeds suitable for modern digital healthcare environments.

In future work, the system can be extended into a full-scale mobile application for Android and iOS platforms to improve accessibility and real-time usage. Advanced predictive analytics and disease risk detection models can be integrated to support preventive healthcare and early diagnosis. Multilingual voice assistance, wearable device integration, and direct hospital API connectivity can further enhance usability and interoperability. The platform can also be expanded to support telemedicine services and personalized treatment recommendations, making SmartMeds a more comprehensive intelligent healthcare ecosystem.

## VII. ACKNOWLEDGMENT

The authors sincerely thank the faculty and staff of the Department of Information Technology, Dhole Patil College of Engineering, Pune, for their continuous guidance and support throughout this research work. Special thanks are extended to all contributors who supported the design, development, and testing of the SmartMeds system.

## REFERENCES

- [1]. J.-F. Wang et al., "An innovative X-RAG technique combined with GPT-4o for summarizing medical information from EHR and EMR to assist doctors in clinical decision-making effectively and efficiently," Health Inform. J., vol. 31, no. 3, 2025.



- [2]. V.-T. Nguyen et al., "Medalyze: Lightweight Medical Report Summarization Application Using FLAN-T5Large," arXiv:2505.17059, 2025.
- [3]. K. Saab et al., "Capabilities of Gemini Models in Medicine," arXiv:2404.18416, 2024.
- [4]. L. Yang et al., "Advancing Multimodal Medical Capabilities of Gemini," arXiv:2405.03162, 2024.
- [5]. A. B. Das, S. Ahmed, and S. K. Sakib, "Hallucinations and Key Information Extraction in Medical Texts: A Comprehensive Assessment of Open-Source Large Language Models," arXiv:2504.19061, 2025.
- [6]. S. Liu, A. B. McCoy, and A. Wright, "Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines," *J. Am. Med. Inf. Assoc.*, vol. 32, no. 4, pp. 605-615, 2025.
- [7]. Q. Xu et al., "Towards normalized clinical information extraction in Chinese radiology report with large language models," *Expert Syst. Appl.*, vol. 271, p. 126585, 2025.
- [8]. S. V. Shah, "Accuracy, consistency, and hallucination of large language models when analyzing unstructured clinical notes in electronic medical records," *JAMA Netw. Open*, vol. 7, no. 8, e2425953, 2024.
- [9]. E. H. Houssein, R. E. Mohamed, and A. A. Ali, "Machine Learning Techniques for Biomedical Natural Language Processing: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 140628-140653, 2021.
- [10]. L. Li et al., "A scoping review of using large language models (LLMs) to investigate electronic health records (EHRs)," arXiv:2405.03066, 2024.
- [11]. J. C. Mandel et al., "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records," *J. Am. Med. Inf. Assoc.*, vol. 23, no. 5, pp. 899-908, 2016.
- [12]. A. S. Tang et al., "Harnessing EHR data for health research," *Nat. Med.*, vol. 30, pp. 1847-1855, 2024.

