A MULTIMODAL AI FRAMEWORK FOR MEDICAL IMAGE DIAGNOSIS USING GOOGLE GEMINI

Project Reference No.: *49S_BE_0767*

College : BGS College of Engineering and Technology, Bengaluru
Branch : Department of Artificial Intelligence and Data Science
Guide(s) : Dr. Chandini A G
Student(s) : Mr. Darshan N A
Mr. Abhishek H G
Mr. Nikhil N C
Mr. Niranjan Gowda K S

Keywords:

Artificial Intelligence, Deep Learning, CNN, Google GenAI, Multimodal AI, Medical Imaging.

Introduction:

Artificial Intelligence (AI) has emerged as one of the most transformative technologies in modern healthcare, enabling automated diagnosis, decision support, and improved patient outcomes. The rapid growth of medical data, including imaging and patient records, has created a need for intelligent systems capable of analyzing complex information efficiently. This project focuses on developing a multimodal AI-based diagnostic assistant that integrates image analysis and textual reasoning to detect multiple diseases.

The system is capable of processing MRI scans, X-ray images, dermoscopic images, and fundus images along with structured patient information such as age, symptoms, and medical history. By combining these different data modalities, the system provides more accurate and context-aware diagnostic results.

Unlike traditional machine learning models that require extensive training datasets, this system leverages Google GenAI (Gemini), which enables zero-shot reasoning and interpretation. This allows the model to analyze unseen data without retraining, making it highly scalable and adaptable. The system also generates confidence scores and explanations, which improve transparency and trust among healthcare professionals. A web-based interface is developed using modern technologies to ensure accessibility and ease of use. The platform allows users to upload images and receive real-time diagnostic outputs.

This approach significantly reduces the time required for diagnosis and helps in early detection of diseases. It is particularly beneficial in rural and underdeveloped areas where access to medical specialists is limited. The integration of AI with healthcare systems can improve efficiency, reduce human errors, and enhance clinical decision-making. Overall, this project demonstrates the potential of multimodal AI systems in transforming the future of healthcare diagnostics.

Objectives:

1. To design and develop an AI-based diagnostic system capable of detecting multiple diseases with high accuracy and efficiency.
2. To integrate deep learning and generative AI techniques for providing accurate predictions along with meaningful explanations.
3. To analyze medical images and patient data in real time for faster and more reliable decision-making.
4. To support multimodal inputs such as medical images, clinical reports, and patient history for comprehensive diagnosis.
5. To develop a user-friendly web interface for seamless interaction between users and the system.
6. To improve accessibility to healthcare services, especially in remote and underserved areas.
7. To reduce dependency on manual diagnosis and minimize human errors while ensuring consistent results.
8. To ensure interpretability of AI outputs so that healthcare professionals can understand and trust the system.
9. To design a scalable system capable of handling large volumes of data and multiple users efficiently.
10. To build a reliable and high-performance system that enhances clinical decision-making and improves patient outcomes.

Methodology:

1. The proposed system follows a structured and systematic workflow for multi-disease diagnosis using multimodal AI. Initially, medical images such as MRI, X-ray, dermoscopic, and fundus images are collected along with patient metadata including age, gender, and symptoms. The input data undergoes preprocessing, where images are resized, normalized, and validated for quality to ensure accurate analysis. After preprocessing, a structured prompt is generated by combining the image data with patient information.
2. The system utilizes Google GenAI (Gemini) to perform multimodal analysis, where both visual and textual inputs are processed simultaneously. Gemini generates a diagnosis, confidence score, key observations, and a detailed explanation based on the provided inputs. The backend system, developed using FastAPI or Flask, handles API requests, processes responses, and ensures proper formatting of outputs.
3. The frontend, built using React and TypeScript, provides an interactive interface for users to upload medical images and input patient details. The system displays results in real time, including diagnostic predictions and visual explanations. Error handling and validation mechanisms are implemented to ensure reliability and robustness.

Finally, the entire system is deployed locally or on cloud platforms to enable real-time accessibility. This methodology ensures an efficient, scalable, and explainable AI-based diagnostic system for multiple diseases.

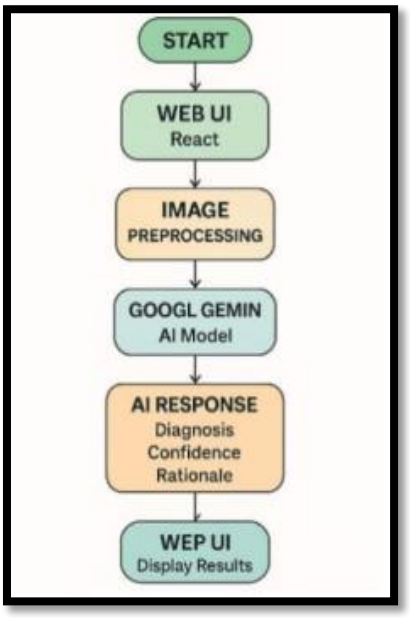

Figure 1: Architecture Overflow

Result and Conclusion:

1. The system successfully delivers real-time diagnostic outputs with a high level of accuracy and reliability. It is capable of processing multiple types of medical data efficiently, enabling faster analysis compared to traditional methods. One of the key strengths of the system is its ability to generate confidence scores alongside each prediction, helping users understand the certainty of the results. In addition, it provides detailed explanations, making the outcomes more transparent and easier to interpret.
2. The integration of Generative AI significantly enhances the interpretability of the model by converting complex outputs into user-friendly insights. This improves usability, especially for healthcare professionals who may not have deep technical expertise in AI. The web-based platform ensures easy accessibility, allowing users to interact with the system from different locations without requiring specialized software.
3. Furthermore, the system demonstrates strong performance in handling multimodal data, combining inputs such as medical images, text, and clinical records. This capability highlights the effectiveness of multimodal AI in improving diagnostic accuracy and providing a more comprehensive analysis. The results clearly indicate that the system can support early detection and reduce the chances of human error.
4. Overall, the project proves that AI-based diagnostic systems can play a crucial role in modern healthcare. It enhances decision-making, improves efficiency, and supports clinicians in delivering better patient care. The findings confirm that integrating AI into healthcare workflows has the potential to significantly transform and optimize diagnostic processes.

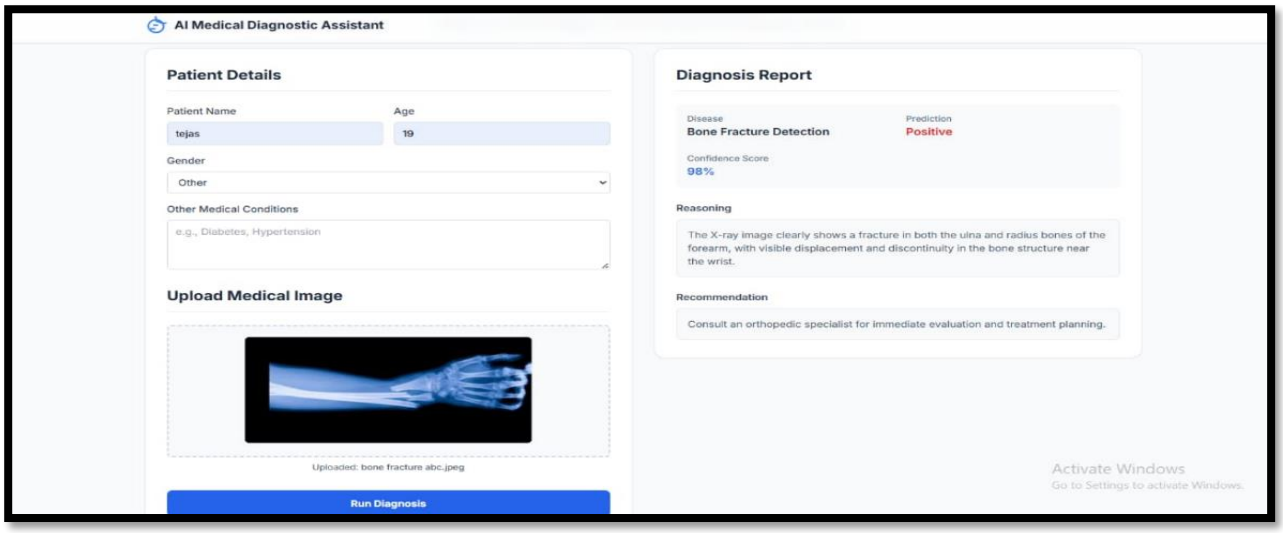

Figure 2: Bone Fracture Detection using AI Medical Diagnostic System

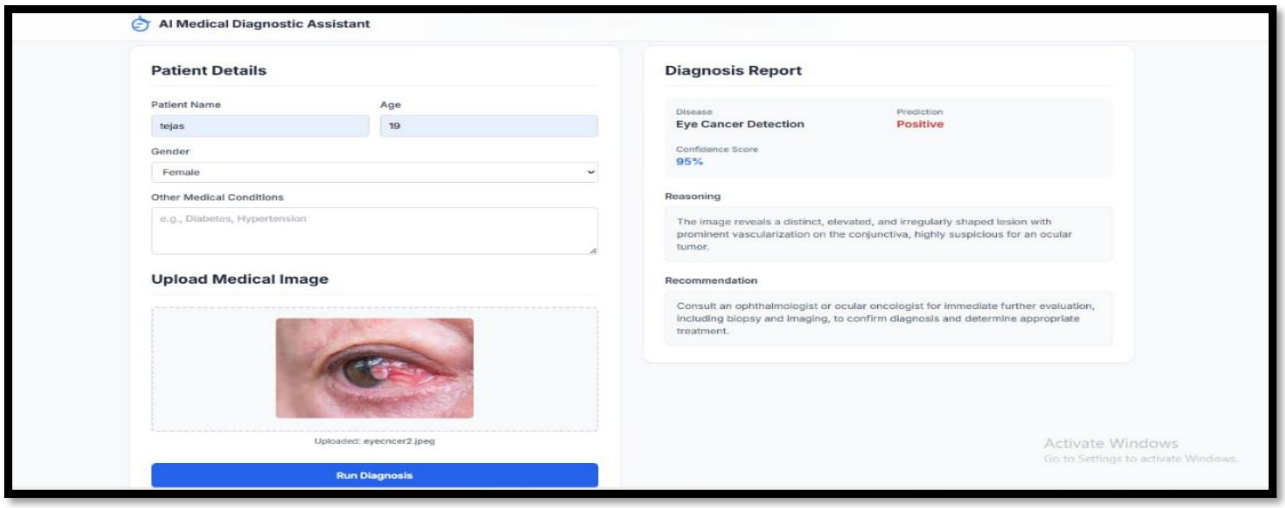

Figure 3: Eye Cancer Detection using AI Medical Diagnostic System

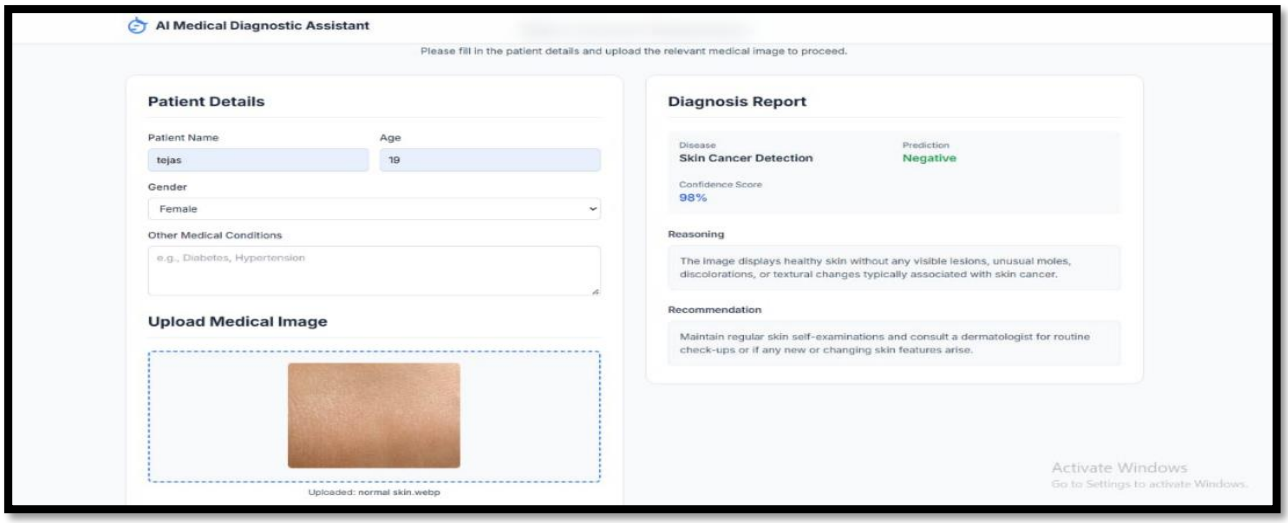

Figure 4: Skin Cancer Detection using AI Medical Diagnostic System

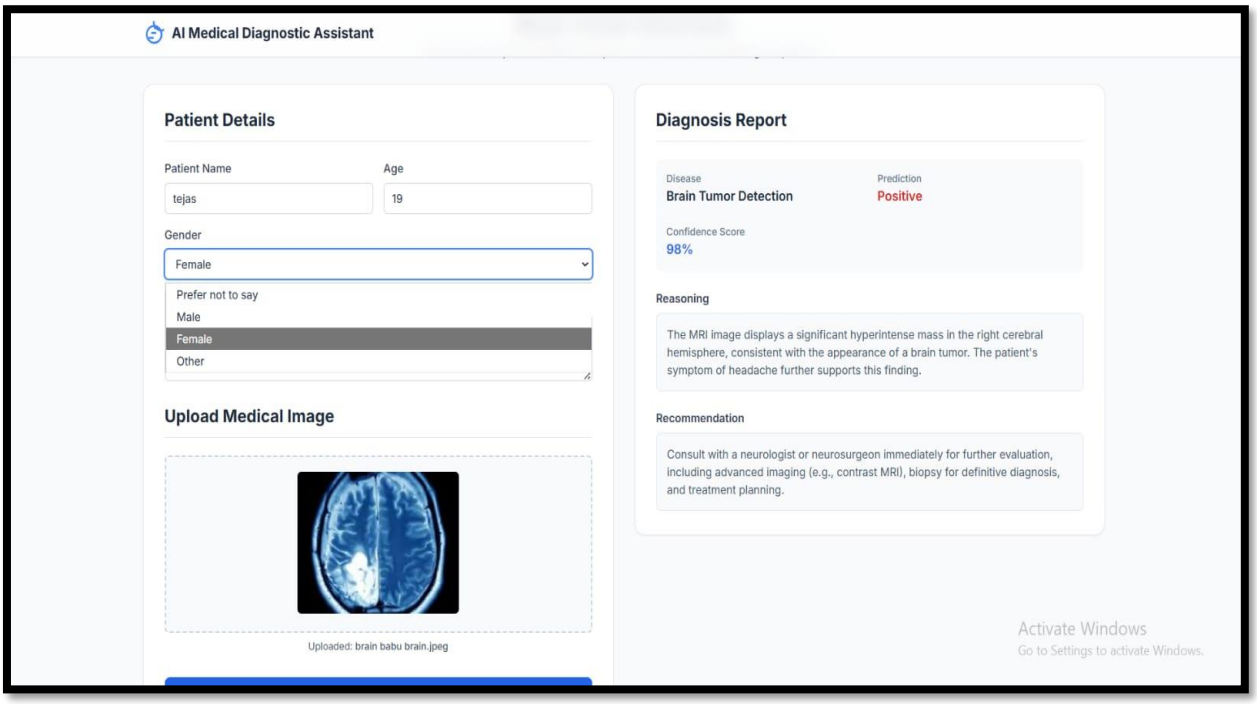

Figure 5: Brain Tumor Detection using AI Medical Diagnostic System

Project Outcome & Industry Relevance:

1. The proposed multimodal AI diagnostic system successfully demonstrates the application of advanced artificial intelligence techniques in medical image analysis and disease detection. By integrating image-based inputs with patient data, the system provides accurate, real-time diagnostic results along with confidence scores and detailed explanations. The use of Google GenAI (Gemini) enhances the system's ability to perform zero-shot reasoning, enabling it to analyze previously unseen medical data without the need for extensive retraining.
2. The project holds significant relevance in the healthcare industry, particularly in improving diagnostic efficiency and accessibility. It can assist healthcare professionals in early disease detection, reducing diagnostic time and minimizing human errors. The ability to handle multimodal inputs makes the system highly effective in providing comprehensive and context-aware analysis.
3. Furthermore, the system can be integrated into telemedicine platforms, making healthcare services more accessible in rural and underserved areas. The web-based interface ensures ease of use and scalability, allowing multiple users to access the system simultaneously. Overall, the project contributes to the advancement of AI-driven healthcare solutions and demonstrates strong potential for real-world clinical applications.

Working Model vs. Simulation/Study:

1. This project is developed as a **working software-based model** that performs real-time medical diagnosis using multimodal AI. It is not a theoretical study but a fully functional system implemented using Google GenAI (Gemini), integrated with a backend framework such as FastAPI/Flask and a frontend built using React.

2. The system allows users to upload medical images and input patient details, after which it generates diagnostic predictions, confidence scores, and explanations in real time. The use of API-based integration with Gemini enables dynamic and intelligent processing of both image and textual data.
3. Although the system does not involve a physical hardware model, it operates as a complete working prototype simulating real-world healthcare diagnostics. Therefore, it can be classified as a **working model based on simulation with real-time AI processing and deployment capabilities**.

Project Outcomes and Learnings:

1. The project resulted in the successful development of a multimodal AI-based diagnostic system capable of analyzing medical images and patient data for disease detection. It demonstrated the effectiveness of integrating deep learning and generative AI for real-time healthcare applications.
2. Through this project, we gained practical knowledge in multimodal AI systems, prompt engineering, and the use of large language models such as Google Gemini. We also developed skills in backend development using FastAPI/Flask and frontend development using React.
3. The project enhanced our understanding of handling medical image data, preprocessing techniques, and integrating AI models into user-friendly applications. We also learned the importance of explainability and interpretability in AI-based healthcare systems.
4. Overall, the project improved our technical, analytical, and problem-solving skills, while providing valuable experience in building scalable and real-world AI solutions for the healthcare domain.

References:

1. L. Zhang et al., “Automated Brain Tumor Segmentation Using Hybrid CNN Models,” *IEEE Transactions on Medical Imaging*, 2023.
2. M. Patel and A. Soni, “Deep Learning for Fracture Detection Using X-ray Images,” *International Journal of Computer Applications*, 2022.
3. Y. Chen et al., “Multimodal Large Language Models in Medicine: Applications and Challenges,” *arXiv preprint arXiv:2305.xxxxx*, 2023.
4. Google DeepMind, “Gemini 1.5: Technical Report,” *Google Research*, 2024.
5. N. Esteva et al., “Dermatologist-level Classification of Skin Cancer with Deep Neural Networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

Future Scope:

The future scope of this project includes:

1. The system can be expanded to detect a wider range of diseases by incorporating larger and more diverse datasets, improving accuracy and applicability across multiple medical domains. Continuous updates with new medical data will further enhance model performance.

2. Integration with IoT-based healthcare devices such as wearable sensors can enable real-time patient monitoring and early detection of health issues, providing timely alerts to patients and healthcare professionals.
3. Deployment on cloud platforms and enhancement using advanced AI models can improve scalability, accessibility, and performance. Future improvements may also include multilingual support and personalized healthcare recommendations based on patient data.