International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4400 - 4406
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0574

Intelligent Medical Assistance Framework Powered by Large Language Models and Retrieval-Augmented Generation

Mrs. Ashwini Kadam[1], Piyush Jethwa[2], Atharva Sonawane[3], Sanket Pawar[4], Pranav Palve[5], Vaishnavi Pawar[6]
[1]Assistant professor, Department of CSE, Yashoda Technical Campus, Satara, Maharashtra, India.
[2,3,4,5,6]UG, Department of CSE, Yashoda Technical Campus, Satara, Maharashtra, India.
Emails: *ashwinijadhavr@gmail.com[1], piyushjethwa727@gmail.com[2], atharvasonawane96@gmail.com[3], pawarsanket8100@gmail.com[4], pranavpalve20@gmail.com[5], vaishnavipawar2906@gmail.com[6]*

Abstract
Artificial Intelligence (AI) has significantly transformed modern healthcare systems by improving healthcare accessibility, intelligent diagnosis support, and patient interaction. Traditional healthcare systems mainly depend on manual consultation processes and static healthcare platforms, which often fail to provide personalized and real-time medical assistance. Recent progress in generative language technologies, retrieval-based intelligence, conversational systems, and multimodal AI has supported the creation of advanced healthcare platforms that can interpret patient inputs and provide relevant medical assistance based on contextual understanding. This review paper presents a detailed study of AI-based intelligent healthcare assistance systems focusing on conversational healthcare technologies, retrieval-supported reasoning systems, and multimodal healthcare analysis frameworks. The paper reviews existing research related to machine learning-based disease prediction systems, healthcare chatbot technologies, medical image analysis systems, and Retrieval-Augmented Generation architectures. The study also discusses the role of semantic retrieval systems, vector databases, and intelligent reasoning approaches in improving the reliability and contextual understanding of healthcare responses. Furthermore, the paper analyses major challenges associated with AI-powered healthcare systems including misinformation generation, lack of real-time medical understanding, multimodal integration complexity, healthcare data privacy concerns, and infrastructure dependency. Future research directions such as explainable AI, multilingual healthcare systems, cloud-based healthcare platforms, and real-time healthcare analytics are also explored.
Keywords: *Artificial Intelligence (AI); Healthcare Assistance Systems; Retrieval-Augmented Generation (RAG); Conversational AI; Large Language Models (LLMs); Multimodal Learning, Deep Learning.*

1. Introduction
Healthcare is one of the most essential sectors affecting human well-being and social development. Access to accurate medical information, timely diagnosis, and healthcare guidance plays an important role in improving patient outcomes and reducing healthcare-related risks. However, healthcare systems across many regions continue to face major challenges such as shortage of healthcare professionals, delayed consultation services, increasing patient load, and limited medical accessibility. These challenges become more severe in rural and underdeveloped areas where healthcare infrastructure and medical facilities are limited. Traditional healthcare information systems mainly provide static medical information and depend heavily on manual consultation processes. Although these systems help in healthcare record management and basic medical support, they lack intelligent reasoning capabilities and personalized patient interaction. Modern healthcare environments require intelligent systems capable of providing automated healthcare guidance, contextual reasoning, and interactive communication support. The advancement of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) has significantly transformed healthcare technologies. AI-based healthcare systems can analyze healthcare information, identify disease patterns, support diagnosis processes, and provide intelligent

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4400 - 4406
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0574

healthcare recommendations. Conversational AI technologies have become particularly important because they allow users to interact naturally with healthcare systems using human language. These systems improve healthcare accessibility by enabling patients to ask healthcare-related questions and receive automated medical guidance. Large Language Models (LLMs) have further improved the ability of AI systems to process complex healthcare conversations and generate human-like responses. However, standalone language models may generate inaccurate or misleading healthcare information due to limited domain-specific medical understanding and lack of real-time healthcare knowledge. In healthcare applications, misinformation generation can create serious risks because medical recommendations directly affect patient safety and treatment decisions. To overcome these limitations, Retrieval-Augmented Generation (RAG) frameworks have emerged as effective solutions for improving response reliability in healthcare systems. RAG combines external healthcare knowledge retrieval with intelligent language generation models. Before generating responses, the system retrieves relevant healthcare information from external knowledge sources, improving factual consistency and reducing misinformation generation. Vector databases and semantic retrieval systems are commonly used for efficient healthcare information retrieval. Research studies have highlighted the growing importance of RAG systems in healthcare applications because they improve response grounding and reliability. Modern healthcare research is increasingly focusing on multimodal intelligent systems for improving medical analysis and patient support. These systems can process multiple forms of healthcare information including textual symptom descriptions, voice recordings, and medical images. Deep learning models such as Convolutional Neural Networks (CNNs) are widely used for disease detection and healthcare image interpretation tasks. Multimodal learning improves healthcare accessibility and enables flexible methods of patient interaction. Modern intelligent healthcare systems are increasingly integrated with cloud computing technologies, wearable healthcare devices, and smart healthcare monitoring systems. Such integration enables scalable healthcare assistance, real-time healthcare analytics, and continuous patient monitoring. These technologies have the potential to reduce healthcare workload, improve patient engagement, and support healthcare professionals in decision-making processes. Despite significant technological advancements, several challenges still affect the development and implementation of intelligent healthcare systems. These include healthcare data privacy concerns, limited availability of healthcare datasets, computational complexity, infrastructure dependency, explainability limitations, and ethical concerns related to automated healthcare guidance. Therefore, continuous research is necessary to improve the transparency, scalability, reliability, and practical usability of healthcare AI systems. This review paper presents a comprehensive study of intelligent healthcare assistance systems using Artificial Intelligence technologies including Large Language Models, Retrieval-Augmented Generation, conversational AI, multimodal learning, and semantic healthcare retrieval systems. The paper reviews existing research works, analyzes major healthcare AI approaches, discusses current challenges and limitations, and explores future opportunities in intelligent healthcare systems.

2. Related Work

The advancement of Artificial Intelligence technologies has significantly influenced the development of intelligent healthcare systems. Researchers have proposed multiple approaches for improving disease diagnosis, healthcare analytics, patient interaction, and medical support systems using Machine Learning, Deep Learning, Natural Language Processing, and intelligent conversational models. This section discusses important research contributions related to AI-based healthcare assistance systems.

2.1. Esteva et al. (2017)

Esteva et al. [1] explored the use of deep neural network architectures for identifying skin-related diseases from healthcare images. Their study demonstrated that convolutional neural networks can effectively learn visual healthcare patterns and improve disease classification accuracy. Although the model achieved strong performance in image analysis, the system was mainly limited to visual diagnosis tasks and did not support conversational healthcare guidance or contextual patient interaction.

2.2. Rajkomar et al. (2019)

Rajkomar et al. [2] investigated the role of machine learning techniques in clinical prediction systems and electronic healthcare data analysis. Their research highlighted the ability of machine learning algorithms to analyze patient records and support healthcare decision-making processes. However, the proposed approaches mainly focused on structured healthcare information and lacked interactive communication features for patient assistance.

2.3. Brown et al. (2020)

Brown et al. [3] introduced advanced Large Language Models capable of performing multiple natural language understanding tasks using few-shot learning techniques. Their work improved conversational AI systems by enabling human-like response generation and contextual understanding. Despite these advancements, the models still faced limitations related to misinformation generation and lack of specialized medical reasoning capability.

2.4. Lewis et al. (2020)

Lewis et al. [4] proposed the Retrieval-Augmented Generation framework, which combines external information retrieval mechanisms with language generation models. Their approach improved response grounding by retrieving relevant information before generating responses. The framework reduced misinformation generation and improved contextual understanding, although healthcare-specific integration and real-time medical adaptation were not extensively addressed.

2.5. Chen et al. (2020)

Chen et al. [5] studied the impact of Artificial Intelligence technologies in healthcare applications including diagnosis support systems, healthcare automation, and intelligent patient monitoring. Their work discussed both the advantages and implementation challenges associated with healthcare AI systems. The research also emphasized ethical concerns, transparency requirements, and healthcare data privacy considerations.

2.6. Ramesh et al. (2020)

Ramesh et al. [6] analyzed opportunities and limitations associated with Artificial Intelligence applications in healthcare environments. Their study highlighted the importance of intelligent healthcare systems in improving healthcare accessibility and reducing operational workload. However, the work also identified issues related to system explainability, practical implementation complexity, and healthcare infrastructure dependency.

2.7. WHO Digital Healthcare Reports (2021)

The World Health Organization [7] emphasized the growing importance of digital healthcare technologies and intelligent healthcare systems for improving healthcare accessibility and patient management services. The reports highlighted the need for scalable and technology-driven healthcare solutions capable of supporting healthcare delivery in both urban and rural environments.

2.8. Devlin et al. (2019)

Devlin et al. [8] introduced transformer-based language representation models that significantly improved contextual understanding in Natural Language Processing tasks. Their research enhanced language understanding capabilities in conversational systems and healthcare text analysis applications. However, applying such models effectively in healthcare environments requires domain-specific training and high computational resources. From the reviewed literature, it can be observed that existing healthcare systems mainly focus on specific functionalities such as healthcare prediction, disease classification, healthcare analytics, or conversational interaction. Limited research has focused on integrating conversational intelligence, Retrieval Augmented Generation, multimodal healthcare analysis, and intelligent reasoning into a unified healthcare assistance framework. Therefore, integrated healthcare systems combining these technologies have strong potential to improve the reliability, accessibility, scalability, and effectiveness of future healthcare assistance platforms.

3. Analysis of Existing Approches

The rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) has significantly transformed intelligent healthcare assistance systems. Several approaches have been proposed for disease

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4400 - 4406
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0574

prediction, healthcare monitoring, conversational medical support, and healthcare analytics. This section discusses the major technologies and approaches used in modern healthcare assistance systems.

3.1. Machine Learning Based Healthcare Systems

Machine Learning techniques are widely used in healthcare systems for disease prediction, healthcare analytics, patient monitoring, and clinical decision support. Algorithms such as Decision Trees, Random Forests, Support Vector Machines, and Logistic Regression are commonly applied to healthcare datasets for identifying disease patterns and generating healthcare predictions. These systems improve healthcare analysis accuracy but generally depend on structured datasets and predefined training information [9].

Advantages:

- Good prediction accuracy
- Effective healthcare data analysis
- Useful for disease prediction and healthcare monitoring

Limitations:

- Limited conversational capability
- Dependence on structured healthcare datasets
- Lack of real-time contextual understanding

3.2. Deep Learning Based Healthcare Systems

Deep Learning technologies have significantly improved medical image analysis and intelligent disease detection systems. Convolutional Neural Networks (CNNs) are widely used for identifying abnormalities in medical images such as X-rays, MRI scans, CT scans, and dermatological images. Deep learning systems improve disease classification accuracy and automate healthcare image analysis. However, these systems require large healthcare datasets and high computational resources.

Advantages:

- High disease detection accuracy
- Automated feature extraction capability
- Effective medical image analysis

Limitations:

- Large dataset requirements
- High computational complexity
- Limited explain ability of predictions

3.3. Conversational AI Based Healthcare Systems

Conversational healthcare systems enable patients to interact naturally with healthcare platforms using human language. AI-powered healthcare chatbots provide healthcare guidance, answer patient queries, and improve healthcare accessibility. These systems enhance user engagement and support basic healthcare interaction. However, many conversational systems still generate static or context-limited responses.

Advantages:

- Natural language communication
- Improved patient interaction.
- Better healthcare accessibility

Limitations:

- Limited reasoning capability
- Response inconsistency in complex healthcare scenarios
- Lack of real-time medical understanding

3.4. Large Language Model (LLM) Based Systems

Large Language Models have significantly improved conversational healthcare systems by enabling advanced natural language understanding and response generation. These models can understand healthcare-related questions and generate human-like responses. LLMs improve user experience and accessibility in healthcare applications. However, standalone language models may produce inaccurate or misleading healthcare information due to hallucination problems and lack of domain-specific healthcare grounding.

Advantages:

- Human-like conversational support
- Better contextual understanding
- Improved natural language interaction

Limitations:

- Hallucination issues
- Lack of real-time healthcare knowledge
- Dependence on pre-trained information

3.5. Retrieval-Augmented Generation (RAG) Systems

Retrieval-Augmented Generation improves the reliability of healthcare conversational systems by combining external healthcare information retrieval with language generation models. Before generating responses, RAG systems retrieve relevant healthcare knowledge from vector databases and medical document repositories. This reduces misinformation generation and improves healthcare response accuracy.

Advantages:

- Improved response reliability
- Reduced misinformation generation
- Better contextual healthcare recommendations

Limitations:

- Dependence on knowledge base quality
- Increased system complexity
- Requirement of efficient retrieval mechanisms

3.6. Multimodal Healthcare Systems

Multimodal healthcare systems combine text processing, voice interaction, and medical image analysis to provide comprehensive healthcare assistance. These systems improve healthcare accessibility and allow users to interact using different communication methods. Deep learning models are widely used for medical image interpretation and disease detection.

Advantages:

- Supports multiple input formats
- Improved healthcare analysis capability
- Flexible patient interaction

Limitations:

- Complex system integration
- High computational requirements
- Sensitive to image and data quality

The analysis of existing healthcare approaches reveals that most systems mainly focus on specific functionalities such as disease prediction, healthcare analytics, chatbot interaction, or medical image analysis. Limited systems successfully integrate conversational AI, Retrieval-Augmented Generation, multimodal learning, and intelligent healthcare reasoning into a unified healthcare assistance platform. Therefore, intelligent systems integrating these technologies have strong potential to improve the reliability, scalability, accessibility, and effectiveness of future healthcare assistance systems.

4. Challenges & Limitations

Artificial Intelligence-based healthcare systems have demonstrated significant advancements in healthcare support, disease prediction, and intelligent patient interaction. However, several technical and practical limitations still affect the reliability, scalability, and implementation of intelligent healthcare assistance systems in real-world healthcare environments.

4.1. Hallucination in Large Language Models

One of the major challenges in Large Language Models is hallucination, where AI systems generate incorrect or misleading healthcare information. Since healthcare recommendations directly affect patient safety, inaccurate medical responses can create serious risks. Although Retrieval-Augmented Generation reduces misinformation, achieving complete healthcare reliability remains difficult [10].

4.2. Lack of Real-Time Healthcare Understanding

Many healthcare AI systems mainly depend on static datasets and pre-trained information. Therefore, they may fail to provide recommendations according to changing healthcare conditions, updated medical guidelines, or real-time patient situations. Real-time healthcare understanding remains a major challenge for intelligent healthcare systems.

4.3. Multimodal Analysis Challenges

Poor-quality images can reduce disease detection accuracy. In addition, integrating text, image, voice, and healthcare sensor data into a unified intelligent framework remains technically challenging.

4.4. Privacy and Security Issues

Healthcare assistance systems may collect sensitive patient information including medical history, healthcare reports, and personal healthcare data. Secure storage and protection against unauthorized access are essential for maintaining patient trust and healthcare system reliability. Continuous advancements in Artificial Intelligence, Retrieval-Augmented Generation, multimodal learning, and real-time healthcare analytics are creating new opportunities for improving intelligent healthcare assistance systems. Addressing these limitations can lead to more reliable, scalable, and patient-friendly healthcare technologies.

5. Future Scope

The rapid development of Artificial Intelligence (AI),

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 06 June 2026
Page No: 4400 - 4406
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0574

Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and intelligent healthcare technologies has created significant opportunities for improving healthcare assistance systems. Future healthcare systems can become more intelligent, scalable, accessible, and reliable through continuous research and technological advancements.

5.1. Advanced Agentic AI Systems

Future healthcare assistance systems can utilize autonomous AI agents capable of multi-step reasoning, intelligent decision-making, and healthcare task automation. These systems can combine intelligent information retrieval, healthcare resources, and adaptive reasoning techniques to deliver more accurate and personalized medical guidance.

5.2. Real-Time Healthcare Data Integration

Future intelligent healthcare systems can integrate real-time healthcare data from wearable devices, smart healthcare sensors, hospital databases, and healthcare monitoring systems. This integration can improve patient monitoring, disease prediction, and personalized healthcare guidance.

5.3. Multilingual Conversational Healthcare Systems

Future healthcare systems should support multiple regional languages and dialects to improve healthcare accessibility for users from different linguistic backgrounds. Multilingual conversational AI can allow patients to communicate naturally in their preferred language.

5.4. Enhanced Multimodal Learning

Future healthcare systems can integrate text, voice, medical images, and sensor data for more comprehensive healthcare analysis and decision support. Advanced multimodal learning techniques can improve disease detection accuracy and healthcare monitoring capabilities.

5.5. Integration with Hospital and Government Healthcare Systems

Future intelligent healthcare platforms can integrate government healthcare schemes, hospital databases, healthcare guidelines, and medical policies to provide comprehensive patient support and healthcare awareness.

5.6. Cloud-Based Scalable Healthcare Platforms

Cloud computing technologies can be used to develop scalable and efficient healthcare systems capable of supporting large numbers of users and healthcare datasets. Cloud infrastructure can improve system availability, processing efficiency, and healthcare accessibility.

5.7. Improved Medical Image Analysis Systems

Future healthcare systems can utilize advanced deep learning architectures, transfer learning techniques, and larger healthcare datasets to improve disease detection accuracy and medical image interpretation capabilities across multiple healthcare domains. The integration of conversational AI, Retrieval-Augmented Generation, multimodal learning, cloud computing, and intelligent healthcare analytics has strong potential to transform modern healthcare systems. Future advancements in these technologies can contribute significantly to the development of smarter, more reliable, and patient-centered healthcare assistance systems.

Conclusion

Artificial Intelligence-based healthcare assistance systems are significantly transforming modern healthcare environments through intelligent, interactive, and context-aware healthcare support. Technologies such as Large Language Models, Retrieval-Augmented Generation, conversational AI, deep learning, and multimodal healthcare analysis have improved healthcare accessibility, disease detection, patient interaction, and intelligent medical support systems. This review paper studied various AI-driven healthcare technologies including machine learning-based healthcare systems, deep learning-based disease detection systems, conversational healthcare platforms, Retrieval-Augmented Generation frameworks, and multimodal healthcare assistance systems. The paper also analyzed major challenges associated with healthcare AI systems including hallucination in language models, lack of real-time healthcare understanding, healthcare data limitations, privacy concerns, multimodal integration complexity, and infrastructure dependency. The study concludes that integrating Artificial Intelligence technologies with retrieval-supported reasoning systems, real-time healthcare analytics, and multimodal learning frameworks can significantly improve the reliability and effectiveness of future healthcare assistance systems. Continuous advancements in AI technologies, cloud computing, intelligent healthcare analytics, and explainable AI are expected to create more scalable, patient-friendly, and trustworthy healthcare systems in the future.

References

[1]. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[2]. D. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," The New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

[3]. T. B. Brown et al., "Language Models are Few-Shot Learners," in Proceedings of NeurIPS, 2020.

[4]. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Proceedings of NeurIPS, 2020.

[5]. H. Chen, B. Yang, J. Pei, and J. Liu, "Artificial Intelligence in Healthcare: Applications, Risks, and Ethical Considerations," IEEE Access, vol. 8, pp. 184–197, 2020.

[6]. S. Ramesh, C. Kambhampati, J. Monson, and P. Drew, "Artificial Intelligence in Medicine," Annals of The Royal College of Surgeons of England, vol. 86, no. 5, pp. 334–338, 2020.

[7]. World Health Organization (WHO), "Digital Health and Artificial Intelligence in Healthcare," WHO Technical Report, 2021.

[8]. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, 2019.

[9]. E. Topol, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY, USA: Basic Books, 2019.

[10]. F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial Intelligence in Healthcare: Past, Present and Future," Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017.