International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

A Hybrid Retrieval-Augmented Generation Framework for Intelligent Symptom Monitoring and Disease Guidance

ULLAGANTI SHREEYAA MADHURY[1], Dr.M.RAMESH[2]
[1]Student, Department of Computer Science and Engineering, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Professor & HOD, Department of CSE, Ellenki College of Engineering and Technology, Patelguda, Patancharu, Hyderabad, Telangana.
[2]Email: marpuramesh223@gmail.com

***Abstract*—** Early identification of diseases based on patient-reported symptoms can improve timely medical guidance and reduce the burden on healthcare services. This paper presents SymptoTrackAI, an intelligent symptom monitoring framework that integrates Retrieval-Augmented Generation (RAG) with FAISS-based similarity search to provide context-aware disease suggestions from user-entered symptoms. The proposed system utilizes a symptom–disease knowledge base and employs the Sequence-NQ transformer to retrieve semantically relevant medical information before generating an appropriate response. A secure web application is developed with modules for user registration, authentication, AI-assisted symptom analysis, and historical symptom log management. User interactions and predicted outcomes are stored to support continuous health monitoring and future analysis. The retrieval mechanism enables the chatbot to identify diseases by matching symptom descriptions with medically relevant records, thereby improving the relevance of generated recommendations. Experimental observations demonstrate that the hybrid retrieval framework consistently delivers accurate and context-sensitive responses for diverse symptom queries while maintaining efficient response time. The proposed solution offers a practical, scalable, and user-friendly approach for preliminary symptom assessment, enabling individuals to obtain timely health-related guidance and maintain a digital record of symptom history for informed healthcare decision-making.

***Keywords*—** Retrieval-Augmented Generation (RAG), FAISS similarity search, Sequence-NQ transformer, symptom monitoring, disease prediction, AI chatbot, semantic retrieval, healthcare informatics, symptom analysis, medical decision support, conversational artificial intelligence, health monitoring system.

INTRODUCTION

The increasing adoption of digital healthcare technologies has transformed the way individuals seek preliminary medical guidance for common health concerns. Many people rely on online resources to understand symptoms before consulting healthcare professionals. However, conventional symptom-checking methods often return generic or irrelevant information because they fail to consider the semantic relationship between user queries and medical knowledge. As a result, there is a growing demand for intelligent healthcare systems capable of providing more accurate, context-aware, and personalized symptom assessment while maintaining an accessible and user-friendly interface [2]–[7].

Recent developments in artificial intelligence and natural language processing have enabled conversational healthcare applications to retrieve and interpret medical information more effectively. Retrieval-Augmented Generation (RAG) has emerged as a promising approach by combining information retrieval with language generation, allowing responses to be grounded in trusted knowledge sources rather than relying solely on pretrained model knowledge. When integrated with FAISS similarity search, the retrieval process becomes significantly faster and more efficient by identifying semantically similar symptom descriptions from a large medical dataset. Furthermore, the Sequence-NQ transformer enhances semantic understanding by generating meaningful representations of user queries and

medical records, thereby improving the relevance and quality of the generated responses [18].

This work presents SymptoTrackAI, a web-based intelligent symptom monitoring system that combines Retrieval-Augmented Generation (RAG), FAISS similarity search, and the Sequence-NQ transformer to assist users in obtaining preliminary disease-related guidance based on their reported symptoms. The proposed framework includes secure user registration, authentication, AI-powered symptom analysis, and historical health log management for continuous symptom monitoring. User queries are matched against a symptom–disease knowledge repository to retrieve the most relevant medical information before generating context-aware recommendations. In addition, the system maintains previous consultation records to support long-term health tracking and future analysis. The proposed solution provides a practical and scalable approach for AI-assisted symptom assessment, helping users obtain timely health guidance while supporting informed healthcare decision-making [20].

I. RELATED WORK

S. Divya, V. Indumathi, S. Ishwarya, M. Priyasankari, and S. K. Devi (2018) proposed "A Self-Diagnosis Medical Chatbot Using Artificial Intelligence" [5]. The authors developed an intelligent conversational system that assists users in identifying possible health conditions based on the symptoms they describe. The chatbot employed artificial intelligence techniques to interpret user queries and generate suitable disease-related suggestions through an interactive interface. The system was designed to provide preliminary medical guidance before professional consultation, making healthcare information more accessible. Experimental observations indicated that the chatbot effectively supported symptom-based interactions while reducing the time required to obtain basic health advice. This work demonstrated the practical value of AI-driven conversational systems in improving patient engagement and enhancing the accessibility of healthcare services. [5]

S. S. Kumar, S. T. Ahmed, A. S. Fathima, M. Nishabai, and S. Sophia (2024) presented "Medical ChatBot Assistance for Primary Clinical Guidance Using Machine Learning Techniques" [7]. The researchers introduced a machine learning-based chatbot intended to provide initial clinical guidance by analyzing user-reported symptoms. Their framework incorporated natural language processing and predictive learning methods to improve the understanding of patient queries and recommend appropriate medical advice. The study highlighted that intelligent conversational systems can reduce the workload of healthcare professionals by offering immediate responses to common medical concerns. Experimental evaluation confirmed that the proposed chatbot generated reliable recommendations while maintaining efficient interaction with users. The research demonstrated the growing role of machine learning in supporting primary healthcare services. [7]

I.-C. Hsu and J.-D. Yu (2022) presented "A Medical Chatbot Using Machine Learning and Natural Language Understanding" [9]. The proposed work focused on integrating machine learning algorithms with natural language understanding techniques to improve communication between patients and healthcare systems. The chatbot was capable of interpreting user input, identifying medical intent, and producing meaningful responses based on learned knowledge. By combining language processing with predictive modeling, the system achieved improved accuracy in understanding health-related conversations compared with conventional rule-based chatbots. The study emphasized that intelligent language understanding enhances user experience and supports more effective digital healthcare communication. [9]

Babu and S. B. Boddu (2024) introduced "BERT-Based Medical Chatbot: Enhancing Healthcare Communication Through Natural Language Understanding" [10]. The authors proposed a healthcare chatbot built upon the Bidirectional Encoder Representations from Transformers (BERT) model to improve the interpretation of medical queries. The contextual language representation provided by BERT enabled the chatbot to understand complex patient inputs and generate more meaningful responses. The study demonstrated that transformer-based language models significantly improve conversational quality by capturing semantic relationships within user queries. Experimental findings showed that the proposed approach enhanced communication efficiency and supported more accurate healthcare information delivery in conversational environments. [10]

S. D. Bhavani Peri, S. Santhanalakshmi, and R. Radha (2024) proposed "Chatbot to Chat with Medical Books Using Retrieval-Augmented Generation Model" [12]. The authors developed a Retrieval-Augmented Generation (RAG)-based conversational system that retrieves relevant information from medical literature before generating responses. Instead of relying solely on pretrained language models, the framework

combined document retrieval with text generation to produce responses supported by medical knowledge sources. This approach improved the factual consistency and reliability of chatbot recommendations while minimizing inaccurate or fabricated information. The research demonstrated that integrating retrieval mechanisms with generative models can substantially improve the quality of AI-assisted healthcare conversations and provide users with more trustworthy medical guidance. [12]

II. DATASET DETAILS

The proposed SymptoTrackAI framework utilizes a symptom–disease dataset containing a collection of medical symptoms, associated disease names, and corresponding health advice for training the intelligent chatbot. The dataset acts as the primary knowledge source that enables the system to retrieve relevant medical information based on symptoms entered by users. Each record represents a combination of symptom descriptions and the corresponding disease, allowing the retrieval model to learn meaningful associations between user queries and medical conditions. Before the knowledge base is incorporated into the system, the dataset is organized and indexed to eliminate redundant records and improve retrieval efficiency. This preprocessing stage ensures consistent representation of symptom information and enhances the overall quality of semantic search.

The processed dataset is integrated with a Retrieval-Augmented Generation **(**RAG**)** framework and FAISS similarity search to support intelligent symptom analysis. The Sequence-NQ transformer converts both user queries and medical records into semantic embeddings, enabling the retrieval engine to identify the most relevant symptom–disease relationships from the knowledge repository. Instead of directly predicting diseases through conventional classification, the system retrieves contextually similar medical records and generates responses based on the retrieved information. This hybrid retrieval mechanism improves response relevance, minimizes incorrect recommendations, and produces medically meaningful guidance for a wide variety of symptom combinations. The indexed dataset therefore serves as the foundation for accurate semantic retrieval and context-aware response generation.

The same dataset also supports real-time interaction through the developed web-based application. After user authentication, individuals can submit symptom descriptions to the chatbot, which retrieves the most relevant disease information and provides appropriate health guidance within a short response time. In addition, every user query and generated recommendation is stored in a database to maintain a history of symptom records for future reference and health monitoring. This continuous logging mechanism enables users to review previous consultations while supporting long-term health analysis. The combination of a structured symptom–disease dataset, semantic retrieval, and conversational AI provides a scalable and reliable framework for intelligent symptom monitoring and preliminary disease guidance.

III. PROPOSED METHODOLOGY

The proposed SymptoTrackAI framework introduces an intelligent healthcare assistance system that combines Retrieval-Augmented Generation (RAG), FAISS similarity search, and the Sequence-NQ transformer to provide preliminary disease guidance based on symptoms entered by users. The primary objective of the framework is to deliver context-aware and reliable medical recommendations by retrieving relevant information from a symptom–disease knowledge repository instead of relying solely on predefined responses. A web-based application integrates secure user authentication, AI-powered symptom analysis, retrieval-based response generation, and historical health record management into a single platform, enabling users to interact with the chatbot efficiently while maintaining their consultation history.

The methodology begins with preparing a structured symptom–disease dataset that contains symptom descriptions, corresponding diseases, and associated medical guidance. Before integrating the dataset into the chatbot, the knowledge base is organized and indexed to improve retrieval speed and eliminate duplicate or inconsistent records. The processed medical information is then converted into semantic vector representations using the Sequence-NQ transformer, allowing both the stored medical records and incoming user queries to be represented within the same embedding space. This semantic representation enables the system to recognize the contextual meaning of symptoms rather than depending only on keyword matching, thereby improving the relevance of retrieved medical information.

After the knowledge base is prepared, the semantic embeddings are indexed using the FAISS similarity search algorithm to support efficient retrieval from a large collection of symptom records. When a user submits symptoms through the chatbot interface, the input query is transformed into an embedding using the same transformer model and compared with the indexed dataset to identify the most semantically relevant medical records. The retrieved information is then supplied to the Retrieval-Augmented

Generation (RAG) framework, which generates a context-aware response by combining retrieved medical knowledge with natural language generation. This hybrid retrieval mechanism improves the quality of chatbot responses, reduces irrelevant recommendations, and provides users with meaningful preliminary health guidance based on their reported symptoms.

The developed web application further enhances usability by providing modules for user registration, login, AI-assisted symptom analysis, and access to previous consultation records. Every interaction between the user and the chatbot is securely stored in a database, enabling long-term symptom tracking and future health analysis. Users can review their previous symptom history and generated recommendations whenever required, supporting continuous health monitoring. By combining semantic retrieval, transformer-based language understanding, efficient similarity search, and conversational AI within a unified framework, the proposed methodology offers a scalable and practical solution for intelligent symptom monitoring and preliminary disease assessment while improving the accessibility of digital healthcare services.

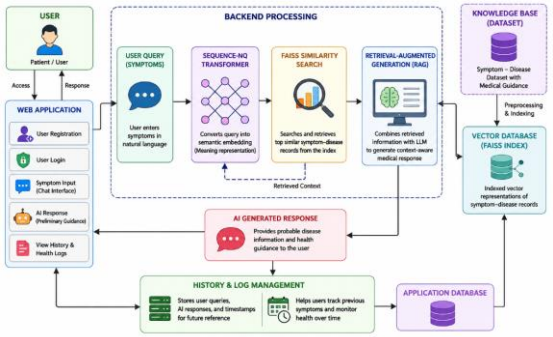

Figure 1: System Architecture of the Proposed SymptoTrackAI System

Figure 1 illustrates the architecture of the proposed SymptoTrackAI system for intelligent symptom monitoring and preliminary disease guidance. The process begins with user registration and login through a secure web application, followed by symptom submission using a chatbot interface. The entered symptoms are converted into semantic embeddings using the Sequence-NQ transformer. Simultaneously, a symptom–disease knowledge base is indexed using FAISS similarity search, enabling efficient retrieval of medically relevant records. The retrieved information is then processed by the Retrieval-Augmented Generation (RAG) framework to generate context-aware health recommendations. The generated response is displayed to the user through the chatbot interface. In addition, user queries, predicted advice, and timestamps are stored in the application database for future reference. The health log module allows users to review previous consultations, supporting continuous symptom monitoring and effective AI-assisted healthcare guidance.

V.RESULT AND DISCUSSION

The proposed SymptoTrackAI system was developed to provide intelligent symptom monitoring and preliminary disease guidance by integrating Retrieval-Augmented Generation (RAG), FAISS similarity search, and the Sequence-NQ transformer. The experimental study was conducted using a symptom–disease knowledge base containing symptom descriptions, associated diseases, and medical guidance. Before deployment, the dataset was organized, indexed, and converted into semantic embeddings to support efficient retrieval and context-aware response generation.

The developed framework was implemented as a web-based application that allows users to register, log in securely, submit symptoms through a conversational chatbot, and review previous consultation records. When a user enters symptoms, the Sequence-NQ transformer converts the query into semantic representations, while the FAISS retrieval engine identifies the most relevant medical records. The RAG framework then combines the retrieved information with language generation techniques to produce meaningful health recommendations.

Experimental observations demonstrate that the hybrid retrieval framework consistently retrieves contextually relevant symptom–disease information and generates accurate, natural-language responses for different symptom combinations. The integration of semantic retrieval with conversational AI improves response relevance while reducing unrelated recommendations. Furthermore, the health log module successfully records user interactions for future reference, enabling continuous symptom monitoring. These results indicate that the proposed framework provides an efficient, scalable, and reliable solution for AI-assisted symptom assessment and supports informed healthcare decision-making.

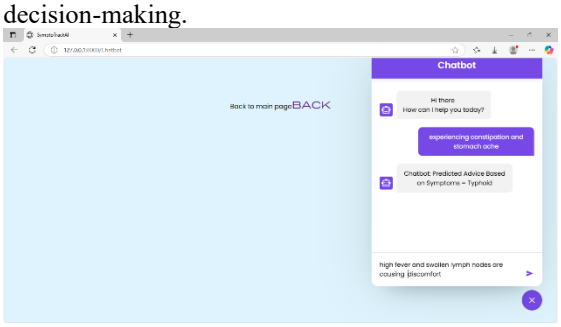

Figure: Chatbot-Based Symptom Analysis and Disease Prediction

The above figure illustrates the chatbot-based symptom analysis module of the proposed SymptoTrackAI system. After successful user authentication, the chatbot interface allows users to

enter their symptoms in natural language. In this example, the user provides symptoms such as constipation and stomach ache, followed by high fever and swollen lymph nodes. The system processes these inputs using the Sequence-NQ transformer, FAISS similarity search, and the Retrieval-Augmented Generation (RAG) framework to identify the most relevant symptom–disease relationship from the medical knowledge base. Based on the retrieved information, the chatbot generates a preliminary disease prediction and recommends Typhoid as the most probable condition. The response is displayed instantly within the chat window, enabling interactive communication between the user and the system. This module demonstrates the capability of the proposed framework to analyze symptom descriptions, retrieve contextually relevant medical knowledge, and provide timely AI-assisted health guidance through an intuitive conversational interface.

DISCUSSION

The proposed SymptoTrackAI framework combines Retrieval-Augmented Generation (RAG), FAISS similarity search, and the Sequence-NQ transformer to provide intelligent symptom monitoring and preliminary disease guidance. A structured symptom–disease knowledge base is prepared and indexed to support efficient semantic retrieval, enabling the system to understand the contextual meaning of user-entered symptoms instead of relying solely on keyword matching. The web-based application offers secure user registration, authentication, chatbot-based symptom analysis, and health history management within a unified platform.

Experimental observations indicate that the hybrid retrieval framework consistently retrieves medically relevant information and generates context-aware responses for diverse symptom queries. The integration of semantic embeddings with FAISS indexing improves retrieval speed, while the RAG framework enhances the quality and reliability of generated recommendations. In addition, the health log module successfully records previous consultations, allowing users to monitor their symptoms over time. The proposed system therefore provides an efficient, scalable, and user-friendly solution for AI-assisted symptom assessment, supporting timely health guidance and informed healthcare decision-making.

VI.CONCLUSION

This work presents SymptoTrackAI, an intelligent symptom monitoring system that integrates Retrieval-Augmented Generation (RAG), FAISS similarity search, and the Sequence-NQ transformer to provide preliminary disease guidance based on user-reported symptoms. The proposed framework combines semantic retrieval with conversational artificial intelligence to deliver context-aware and relevant health recommendations through a secure web-based application. The system includes user registration, authentication, chatbot-based symptom analysis, and historical consultation management, enabling continuous health monitoring and easy access to previous records. Experimental observations demonstrate that the hybrid retrieval approach effectively identifies relevant symptom–disease relationships and generates meaningful responses with improved retrieval efficiency and contextual understanding. The developed application offers a scalable, reliable, and user-friendly solution for AI-assisted symptom assessment, supporting timely healthcare guidance and informed decision-making. Future enhancements may include expanding the medical knowledge base, integrating multilingual support, incorporating multimodal medical information such as clinical reports and medical images, and deploying the system on cloud-based healthcare platforms to improve scalability, personalization, and real-time healthcare assistance.

REFERENCES

[1] World Health Organization (WHO) and International Bank for Reconstruction and Development/The World Bank, Tracking Universal Health Coverage: 2017 Global Monitoring Report. Geneva, Switzerland: World Health Organization, 2017.

[2] World Health Organization, “Health Workforce,” 2025. [Online]. Available: https://www.who.int/health-topics/health-workforce

[3] World Economic Forum, Patient-First Health With Generative AI: Reshaping the Care Experience, Jan. 2024. [Online]. Available: https://www3.weforum.org/docs/WEF_Patient_First_Health_with_Generative_AI_2024.pdf

[4] E. Itelman, G. Golovchiner, A. Barsheshet, R. Kornowski, and A. Erez, “Balancing innovation and professionalism: The emerging role of AI-powered chatbots in medical consultation,” Heart Rhythm, vol. 21, no. 10, pp. 2037–2039, Oct. 2024.

[5] S. Divya, V. Indumathi, S. Ishwarya, M. Priyasankari, and S. K. Devi, “A self-diagnosis medical chatbot using artificial intelligence,” Journal of Web Development and Web Designing, vol. 3, no. 1, pp. 1–7, 2018.

[6] S. H. Alqaidi, S. M. Albugami, W. S. Alzahrani, S. Badri, and A. Wali, “Network-integrated medical

chatbot for enhanced healthcare services," Telematics and Informatics Reports, vol. 15, Art. no. 100153, Sep. 2024.

[7] S. S. Kumar, S. T. Ahmed, A. S. Fathima, M. Nishabai, and S. Sophia, "Medical chatbot assistance for primary clinical guidance using machine learning techniques," Procedia Computer Science, vol. 233, pp. 279–287, Jan. 2024.

[8] Prathwini and Prathyakshini, "PreScienceMED: Leveraging text classification for precise disease prediction," in Proc. IEEE Int. Conf. Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Mangalore, India, Oct. 2024, pp. 90–95, doi: 10.1109/DISCOVER62353.2024.10750775.

[9] I.-C. Hsu and J.-D. Yu, "A medical chatbot using machine learning and natural language understanding," Multimedia Tools and Applications, vol. 81, no. 17, pp. 23777–23799, Jul. 2022.

[10] A. Babu and S. B. Boddu, "BERT-based medical chatbot: Enhancing healthcare communication through natural language understanding," Exploratory Research in Clinical and Social Pharmacy, vol. 13, Art. no. 100419, Mar. 2024.

[11] S. Akilesh, S. A. Abdullah, R. Abinaya, S. Dhanushkodi, and R. Sekar, "A novel AI-based chatbot application for personalized medical diagnosis and review using large language models," in Proc. Int. Conf. Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, Nov. 2023, pp. 1–5, doi: 10.1109/RMKMATE59243.2023.10368616.

[12] S. D. Bhavani Peri, S. Santhanalakshmi, and R. Radha, "Chatbot to chat with medical books using retrieval-augmented generation model," in Proc. IEEE North Karnataka Subsection Flagship Int. Conf. (NKCon), Bagalkote, India, Sep. 2024, pp. 1–5, doi: 10.1109/NKCON62728.2024.10774900.

[13] A. Kelly, E. Noctor, and P. V. de Ven, "Design, architecture and safety evaluation of an AI chatbot for an educational approach to health promotion in chronic medical conditions," Procedia Computer Science, vol. 248, pp. 52–59, Jan. 2024.

[14] J. H. Lee, M.-S. Jeong, J.-U. Cho, H.-K. Jeon, J.-H. Park, K.-D. Shin, S.-J. Song, and Y.-G. Cheong, "Developing an ophthalmic chatbot system," in Proc. 15th Int. Conf. Ubiquitous Information Management and Communication (IMCOM), Jan. 2021, pp. 1–7.

[15] C. Li, C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon, and J. Gao, "LLaVA-Med: Training a large language-and-vision assistant for biomedicine in one day," in Advances in Neural Information Processing Systems (NeurIPS), 2024, pp. 28541–28564.

[16] Y. Guo and Z. Wan, "Performance evaluation of multimodal large language models (LLaVA and GPT-4-based ChatGPT) in medical image classification tasks," in Proc. IEEE 12th Int. Conf. Healthcare Informatics (ICHI), Jun. 2024, pp. 541–543.

[17] S. Gupta, M. Joshi, P. Handa, R. Yadav, and N. Goel, "Revolutionizing medical imaging: Unveiling ChatGPT's potential in diagnostics," in Proc. 2nd World Conf. Communication and Computing (WCONF), Jul. 2024, pp. 1–6, doi: 10.1109/WCONF61366.2024.10691986.

[18] Y. Natarajan, K. R. S. Preethaa, S. Pandian, S. U. Geetha, and S. Senthilkumar, "Enhancing medical information retrieval with a language model," in Proc. Int. Conf. Cognitive Robotics and Intelligent Systems (ICCROBINS), Coimbatore, India, Apr. 2024, pp. 437–441, doi: 10.1109/ICCROBINS60238.2024.10533903.

[19] P. I. Prayitno, R. P. Pujo Leksono, F. Chai, R. Aldy, and W. Budiharto, "Health chatbot using natural language processing for disease prediction and treatment," in Proc. 1st Int. Conf. Computer Science and Artificial Intelligence (ICCSAI), vol. 1, Oct. 2021, pp. 62–67.

[20] B. Ihnaini, Y. Huang, L. Li, J. Wei, and S. Qi, "Enhancing Chinese medical diagnostic chatbot through supervised fine-tuning of large language models," in Proc. 6th Int. Conf. Internet of Things, Automation and Artificial Intelligence (IoTAAI), Guangzhou, China, Jul. 2024, pp. 205–212, doi: 10.1109/IOTAAI62601.2024.10692873.