Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 1s (2026) | Pages: 224-229

Review of Specialist Recommendation System Based on User-Reported Symptoms

Yashica Mayaramani[1], Anupama Phakatkar[2], Rutuja Uplenchwar[3], Dhanvantari Pawar[4], Pornima Sadanshiv[5]

[1,2,3,4,5]Department of Computer Engineering Department, SCTR's PICT, Pune, 411043, Maharashtra, India.

[1]yashica928@gmail.com, [2]agphakatkar@pict.edu, [3]rutujauplenchwar@gmail.com, [4]dppawar174@gmail.com, [5]sadanshivsudhir794@gmail.com

Peer Review Information

Type: *Article*
Received: *04 April 2026*
Revised: *30 April 2026*
Accepted: *05 June 2026*
Published: *26 June 2026*

Abstract

Finding the right medical specialist according to symptoms is one of the main problems that healthcare patients face. This can lead to delayed diagnoses, increased healthcare costs, and compromised health outcomes. A variety of dig-ital solutions have been developed over the years, such as rule-based symptom checkers, machine learning models, and assistants driven by large language mod-els(LLMs). This paper reviews these approaches used within the healthcare sector for the identification of medical specialty. Machine learning models, large language models, and retrieval-augmented generation frameworks applied for patient interaction, symptom analysis, and specialist recommendation have been explored. In addition, the paper discusses challenges related to ethical considera-tions and data privacy. Finally, it outlines emerging research directions with the aim of improving the robustness, reliability, and ethical deployment of AI-driven healthcare assistants.

Keywords: Machine Learning; Large Language Models; Retrieval-Augmented Generation; Medical Specialist Consultation; Clinical Decision Support; Healthcare Artificial Intelligence.

How to Cite This Article

Mayaramani, Y., Phakatkar, A., Uplenchwar, R., Pawar, D., & Sadanshiv, P. (2026). Review of specialist recommendation system based on user-reported symptoms. *Open Access International Journal of Science and Engineering*, 9(1s), 224–229.

Introduction

The complexity of the healthcare system often presents significant challenges for patients seeking medical care. One of the critical obstacles is to identify the right medi-cal specialist based on the symptoms reported by the patients. Misidentification at this stage can lead to delayed diagnosis and increased healthcare costs [1]. This ultimately makes the treatment inefficient. To address this issue, there has been an increasing focus on the development of digital health technologies that empower patients with preliminary health assessments and specialist guidance [4].

Over the years, many solutions have emerged to assist in symptom assessment and specialist recommendation, ranging from rule-based symptom checkers to sophisti-cated machine learning models. Recently, the use of large language models (LLMs) in the healthcare sector has revolutionized conversational AI by enabling more natural, context-aware, and flexible patient interactions [16]. These LLM-powered assistants are capable of interpreting even complex symptom descriptions and asking rele-vant follow-up questions, thus guiding patients to the most appropriate healthcare providers.

This research work is intended to provide an extensive review of different approaches used in healthcare systems and chatbots. The evolution from traditional methods to modern approaches is covered in this paper. Furthermore, the paper dis-cusses significant issues such as maintaining accuracy of the diagnosis and privacy issues within AI healthcare systems.

By synthesizing current advances and identifying existing gaps, this work also outlines future research directions that could enhance the robustness, reliability, and ethical deployment of AI-powered healthcare assistants. The ultimate goal is to con-tribute to the development of intelligent agents that effectively support patients in navigating healthcare pathways, optimizing resource use, and improving clinical outcomes.

Related Work

Proper classification of medical specialties according to the symptoms reported by the patient is critical to efficiently navigate the healthcare system, but this is a challeng-ing task due to the complexity involved in describing the symptoms and the medical knowledge required. The main area of focus in [1] is improving the automation of medical specialty recommendation to help patients in selecting the right medical spe-cialist without requiring advanced medical expertise. The problem identified by the study carried out within the framework of China's health system is that patients tend to sign up for incorrect specialties because they lack medical knowledge dur-ing the hospital signing-in process. To solve this, a hybrid neural network model has been proposed that integrates convolutional neural networks (CNNs) with a Bidirec-tional Encoder Representations from Transformers (BERT) framework enhanced by an attention mechanism. This combination leverages BERT's powerful contextual lan-guage understanding and ability of CNNs to capture local feature patterns in symptom descriptions.

The model has been trained and tested on a large dataset containing more than 40,000 patient symptoms from eight different medical departments, including Otorhi-nolaryngology and Pediatrics. The experimental results demonstrated an accuracy of more than 93.5%, which outperforms traditional methods of classification and shows strong generalization capabilities. This model indicates the success and efficiency of advanced neural network architectures for symptom-based specialty classification, offering a promising solution for automated doctor referral systems.

Large language models (LLMs) have greatly influenced the medical field, specif-ically in disease diagnosis. The use of advanced LLMs such as GPT-4, ChatGPT, GPT-3.5, and LLaMA in medical diagnostics has been explored in [2]. Among these, GPT-4 is recognized for its popularity and widespread adoption due to its diverse uses, improved accuracy, and efficiency. It has been emphasized that LLMs utilize medical data from various sources, including general medical databases, specialized docu-ments, medical imaging, and genomic information to support diagnostic reasoning. These models have been applied across a wide range of healthcare domains, addressing chronic diseases, respiratory diseases, cancer, and even rare conditions, highlighting their adaptability and broad utility in healthcare.

Recent developments in AI-driven diagnostic systems focus on imitating the sequential decision-making process of medical professionals during patient consul-tations. This has been presented in [3] by addressing the challenge of integrating symptom inquiry with disease diagnosis in a cohesive framework. The authors for-mulate the diagnostic process as a sequential interaction in which a system asks targeted symptom-related questions before the final diagnosis. To improve efficiency and accuracy, an adaptive mechanism has been proposed that aligns reinforcement learning-based symptom inquiry with classification-based disease diagnosis, using dis-tribution entropy to optimize the interaction between these components. A valuable contribution of this work is the creation of a comprehensive new dataset derived from the MedlinePlus knowledge base, featuring a wide range of diseases and extensive symptom and examination data, which addresses the limitations of previous datasets. Experimental results demonstrate that this integrated approach outperforms sev-eral state-of-the-art methods across multiple datasets by achieving higher diagnostic accuracy along with minimizing the number of required symptom inquiries. This work highlights that the adaptive AI systems capable of dynamic questioning and precise diagnosis are important, which can significantly improve symptom-based healthcare assistants and specialist recommendation tools.

A major improvement brought by the integration of machine learning technology with electronic medical records has been described in [4]. This system provides a per-sonalized approach to collect data about patients, medical documentation, as well as decision support for doctors. In contrast to traditional systems driven by symptoms, the SmartTriage uses the EMRs for bidirectional communication that allows the sys-tem it to condition its symptom inquiries and predictions on the patient's medical history

The system also provides processing for the patient's free text entries to identify the main complaint based on a fine-tuned transformer model, as well as a set of focused questions based on a random forest model to gather the detailed symptomatology.

This patient-specific data obtained is then used to accurately predict detailed ICD-10-CM codes, in addition to the right medication, laboratory, and imaging orders, with the clinical decision support (CDS) outputs reintegrated into the EMR for seamless clinician use.

The machine learning components of SmartTriage have been extensively trained on vast datasets comprising over 25 million primary care encounters and 1 million patient visit reasons in the form of free-text narratives. These datasets support the design of models such as LSTM for patient history representation, transformers for complaint extraction, random forests for sequencing of questions, and feed-forward networks for CDS predictions. This integrative approach indicates the potential of combining patient history, symptom inquiry, and clinical decision support in a personalized, data-driven manner, thus setting a benchmark for future healthcare assistants.

A network-integrated medical chatbot framework has been introduced in [5] with the purpose of enhancing communication between patients and virtual healthcare assistants. The objectives of the research include leveraging artificial intelligence (AI) and natural language processing (NLP) technology, using Google's Dialog Flow tool to build an auto-response chatbot that is capable of interacting with users in the language preferred by the user.

The chatbot system identifies diseases based on patient-entered symptom keywords and offers personalized medical services that can be used by the patients according to their needs. Beyond symptom detection, the system has the capability to man-age medication by allowing patients to record their medication schedules and receive timely reminders of adherence to medication. In addition, the framework supports patient education by providing accessible information about diseases and encouraging preventive health measures. This comprehensive approach highlights how integrating AI chatbots into healthcare workflows can enhance patient engagement, adherence, and awareness, ultimately contributing to improved health outcomes.

Accurate capture of the nuances of patient complaints, such as severity, onset, and duration of symptoms, has been a challenge in clinical conversational AI systems. This shortcoming has been addressed in [6] by proposing a two-stage approach to identify detailed characterizations of symptom entities expressed by patients during clinical interactions.

The proposed method utilizes Word2Vec and BERT embeddings tailored for clini-cal text to effectively encode patient symptom descriptions.The model reformulates the problem as a multi-label classification issue to capture the primary symptoms as well as their contextual characterizations (e.g., the severity of a headache). The encoded representations are further processed with Linear Discriminant Analysis (LDA) to perform classification, significantly improving the model performance.

Experimental evaluations show a 40-50% increase in accuracy compared to exist-ing models, highlighting the importance of incorporating contextual embeddings and sophisticated classification techniques for more precise symptom characterization. This advancement is critical for improving symptom extraction modules in healthcare chatbots, thus improving the quality of patient interaction and subsequent specialist recommendation.

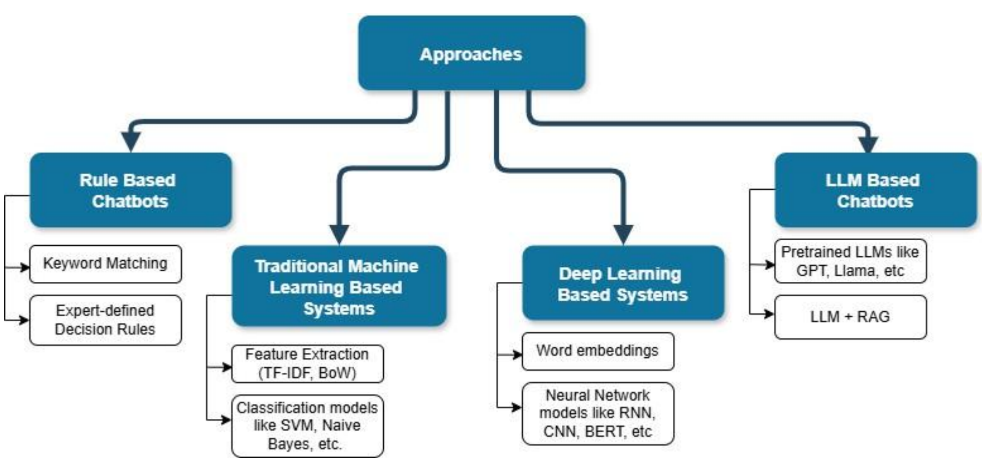

***Fig. 1**. Categorization of the methodologies for analyzing healthcare symptoms*

Methodologies Studied

Rule Based Chatbots

This is the most basic and oldest approach. Early chatbot systems in healthcare relied on rule-based algorithms, which use predefined decision trees or expert-written rules to match symptoms to conditions or specialties. The chatbot's responses are based on a set of predefined rules, keywords, or if-then logic. This approach has been highlighted in [5] where symptom to disease mapping is done using database keyword lookups.

These chatbots are easily interpretable and easy to build for simple tasks. But they are limited in their scalability as they cannot handle complex or ambiguous input and are not context-aware.

Traditional Machine Learning Based Systems

Prior to the dominance of LLMs, the core methodological approach for symptom-to-diagnosis or symptom-to-specialist mapping relied heavily on traditional Machine Learning (ML) and Deep Learning (DL) architectures. These models focus pri-marily on supervised classification, where patient-reported features (symptoms) are mapped to a defined output class (the suggested specialist, e.g., 'Cardiologist' or 'Gastroenterologist').

This approach follows a pipeline based architecture consisting of feature engi-neering and classification steps. During feature engineering, raw symptom data are manually cleaned, normalized, and transformed to a structured, numerical feature vec-tors using techniques like TF-IDF, Bag-of-Words, etc. These feature vectors are then provided as input to classification models like SVM, Naive Bayes, etc.

Deep Learning Based Systems

Deep learning models overcome the reliance on manual feature engineering by enabling representation learning, where neural networks automatically discover complex pat-terns and hierarchical structures from raw input data. These models learn dense semantic embeddings that capture latent relationships among symptoms. RNNs, par-ticularly LSTMs, have been used to model the temporal progression of symptoms and patient history, where the order of events (e.g., fever followed by rash) is crucial for accurate specialist routing. They maintain an internal state, or "memory," that cap-tures context across time steps in the patient dialog. CNNs can also be adapted for sequential symptom data or for multimodal input fusion.

More recently, Transformer-based architectures, such as BERT and its clinical variants, have been employed to generate contextualized representations of symptom descriptions, significantly improving performance over earlier DL models. Nevertheless, traditional machine learning and early deep learning approaches remain limited in their ability to handle the scale, ambiguity, and conversational complexity inherent in natural language symptom reporting.

Large Language Model (LLM)-Based Chatbots (Generative AI)

This represents a major leap in chatbot technology. LLMs provide flexible human-like interactions, ask intelligent follow-up questions, and leverage vast medical corpora to recommend conditions and specialists with high contextual awareness.

This approach is highly conversational in nature, can handle complex and unex-pected questions, and provides more natural and flexible responses. However, it can be computationally expensive. And without RAG, it is prone to inaccuracies.

To overcome these limitations, recent systems integrate LLMs with Retrieval-Augmented Generation (RAG), which significantly improves accuracy, reliability, and domain grounding. In an LLM+RAG pipeline, the model does not rely solely on its parametric memory; instead, it retrieves the most relevant, up-to-date medi-cal guidelines, specialist profiles, and symptom–disease documents from an external knowledge base before generating a response. This ensures that the output remains fact-based, consistent with clinical knowledge, and less prone to hallucinations. RAG also enables customization of the system with institution-specific data (such as hospi-tal departments or available specialists), making recommendations more precise and context-aware. As a result, the LLM+RAG architecture balances the conversational strength of LLMs with the factual reliability of structured medical information, mak-ing it one of the most effective and scalable approaches for specialist recommendation systems.

Challenges

Chatbots in the healthcare sector have emerged as transformative tools for patient engagement and medical support, but their deployment faces significant challenges that need to be addressed. One primary issue is to ensure that the medical advice provided by the chatbots is accurate as well as reliable. Since the human health is complex in nature with various nuances of symptoms, there is a risk of misdiagnosis or inappropriate recommendations by the chatbot and the nuances of symptoms, which could adversely affect patient outcomes. This challenge is compounded by limitations inherent in training data and algorithmic biases, potentially leading to inequitable healthcare delivery.

Table 1. *Existing Work*

Ref	Approach	Implementation & Results
[6]	Deep Learning	Uses BERT embeddings combined with LDA for classification. The approach is a two-stage method to detect entity characteristics (time, severity) in addition to the main symptom. The clinical dialogue dataset for this study was curated from text inputs of potential users of a healthcare bot and labeled using a semi-supervised technique. Result: The proposed BERT+LDA model achieved an accuracy of 64.2% and precision of 92.4%. This indicates 40–50% improvement in accuracy over state-of-the-art models.
[3]	Reinforcement Learning (RL) Hybrid	Formulates diagnosis as a sequential decision-making problem, using RL for symptom inquiring and classification for final diagnosis, aligned via distribution entropy. Dataset for this study is extracted from MedlinePlus using a systematic procedure. The final proposed dataset consists of 893 diseases. Result: Achieved 96.2% accuracy on the MedlinePlus dataset. This outperforms SOTA methods by achieving higher accuracy with fewer inquiring turns, efficiently managing the dialogue flow.
[9]	Machine Learning (Classification)	Compares SVM and Naïve Bayes algorithms for liver disease prediction using the Indian Liver Patient Dataset (ILPD). The dataset has 576 instances and 10 attributes like age, gender, etc. Result: SVM showed the best accuracy (79.66%), while Naïve Bayes had the minimum execution time (1670 ms). This mentions SVM as a better classifier for liver diseases.
[5]	Rule-Based / Intent-Matching	Uses Google Dialogflow for intent classification and symptom-to-disease mapping via database keyword lookups. Focuses on a database-driven diagnosis system, with added features like medicine reminders and health education. The dataset is obtained from Kaggle, containing data about diseases, symptoms, and precautions to be taken. Result: Tested with 10 diseases and passed all tests.
[16]	Large Language Models (LLMs)	Focuses on Domain-Specific LLMs (e.g., Clinical LLMs) trained on biomedical and clinical corpora (EHRs). Result: Discusses the superior performance of domain-specific LLMs; Med-PaLM 2 excels on USMLE. Advantage: Potential to enhance the entire patient journey (triage, diagnosis support, management) due to exceptional language comprehension.

Data privacy and security concerns are also important in medical AI applications, as chatbots handle sensitive patient information. Thus, proper security practices must be followed that comply with regulations such as HIPAA and GDPR. Vulnerabilities to data breaches or unauthorized access pose serious ethical and legal risks, demanding strict measures across system design and deployment stages.

Integration of chatbot systems with existing healthcare infrastructure, such as elec-tronic medical records (EMRs) and hospital networks is also complicated in nature. This is mainly because many healthcare systems use legacy software with limited oper-atability. Without proper integration, chatbot systems may operate in silos, reducing clinical utility and burdening healthcare providers instead of reducing their workloads. Finally, while AI chatbots can simulate empathetic interactions, there remain inherent limitations in their ability to replicate the human touch and emotional intelligence vital to effective healthcare communication, especially in mental health contexts.

Overall, overcoming these intertwined challenges will require multidisciplinary col-laboration among AI researchers, healthcare professionals, and policy makers to ensure AI chatbots are safe, effective, equitable, and trusted components of future healthcare ecosystems.

Future Scope

Significant advances in the future of healthcare chatbots can help transform the accessibility of medical care for patients. One major direction is the improved inte-gration with electronic health records (EHRs) and hospital information systems. Next-generation chatbots will be able to access and update patient records in real time, allowing personalized, context-aware conversations that leverage a patient's medical history to provide more accurate assessments and tailored specialist recommendations. The future will also see chatbots contributing greatly to preventive healthcare and chronic disease management through predictive analytics. By continuously mon-itoring patient data from various sources, chatbots can detect early warning signs, suggest lifestyle modifications, and provide timely reminders for medication adherence or screenings, reducing hospital admissions and healthcare costs.

Ensuring ethical AI practices, transparency, and fairness will be a critical theme moving forward. Research will increasingly focus on explainable AI methods, bias mitigation strategies, and data privacy protections to build trust among patients and clinicians. Regulatory frameworks will evolve to govern the safe deployment of AI chatbots in clinical environments.

Lastly, AI healthcare chatbots will shift from stand-alone tools to integrated com-ponents of broader telemedicine and virtual care ecosystems, supporting seamless transitions from digital assessment to human clinician intervention. This convergence will improve healthcare scalability, efficiency, and patient outcomes, marking a new era of AI-assisted medicine.

Conclusion

AI-powered healthcare chatbots have revolutionized the healthcare industry, offer-ing tremendous improvements in terms of interaction, accessibility, and efficiency. Using advances in natural language processing, machine learning, and large language models, these chatbots provide timely symptom assessment, personalized specialist recommendations, appointment scheduling, and ongoing patient support. Their abil-ity to operate 24/7 and manage routine administrative tasks reduces the burden on healthcare providers while improving patient satisfaction and health outcomes.

Although promising progress has been made, challenges related to accuracy, pri-vacy, ethics, and seamless integration with clinical workflows need to be addressed. Future innovations focused on the explainability of AI, multimodal interaction, and personalization will help healthcare chatbots reach new levels of functionality in healthcare delivery. The future of healthcare will see these healthcare assistants become an essential tool in delivering accessible and affordable patient-centric care.

By synthesizing symptom analysis with intelligent specialist referral, agentic healthcare assistants have the potential to transform the patient journey, making healthcare navigation more intuitive and effective. Continued interdisciplinary collab-oration and rigorous validation are essential to unlock the full benefits of AI chatbots and ensure their safe and ethical adoption in real-world clinical settings.

References

1. Mao, C., Zhu, Q., Chen, R. and Su, W., 2023. Automatic medical specialty classification based on patients' description of their symptoms. BMC Medical Informatics and Decision Making, 23(1), p.15.
2. Almubark, I., 2025. Exploring the impact of large language models on disease diagnosis. IEEE Access.
3. Yuan, H. and Yu, S., 2024. Efficient symptom inquiring and diagnosis via adaptive alignment of reinforcement learning and classification. Artificial Intelligence in Medicine, 148, p.102748.
4. Valmianski, I., Frost, N., Sood, N., Wang, Y., Liu, B., Zhu, J.J., Karumuri, S., Finn, I.M. and Zisook, D.S., 2021, November. SmartTriage: A system for person-alized patient data capture, documentation generation, and decision support. In Machine Learning for Health (pp. 75-96). PMLR.
5. Alqaidi, S.H., Albugami, S.M., Alzahrani, W.S., Badri, S. and Wali, A., 2024. Network-integrated medical chatbot for enhanced healthcare services. Telematics and Informatics Reports, 15, p.100153.
6. Saha, B., Lisboa, S. and Ghosh, S., 2020, July. Understanding patient complaint characteristics using contextual clinical BERT embeddings. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5847-5850). IEEE.
7. Divya, S., Indumathi, V., Ishwarya, S., Priyasankari, M. and Devi, S.K., 2018. A self-diagnosis medical chatbot using artificial intelligence. Journal of Web Development and Web Designing, 3(1), pp.1-7.
8. Chakraborty, C., Pal, S., Bhattacharya, M., Dash, S. and Lee, S.S., 2023. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Frontiers in Artificial Intelligence, 6, p.1237704.
9. Vijayarani, S. and Dhayanand, S., 2015. Liver disease prediction using SVM and Na¨ıve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), pp.816-820.
10. Alty, S.R., Millasseau, S.C., Chowienczyc, P.J. and Jakobsson, A., 2003, Decem-ber. Cardiovascular disease prediction using support vector machines. In 2003 46th Midwest Symposium on Circuits and Systems (Vol. 1, pp. 376-379). IEEE.
11. Arumugam, K., Naved, M., Shinde, P.P., Leiva-Chauca, O., Huaman-Osorio, A. and Gonzales-Yanac, T., 2023. Multiple disease prediction using Machine learning algorithms. Materials today: proceedings, 80, pp.3682-3685.
12. Ayanouz, S., Abdelhakim, B.A. and Benhmed, M., 2020, March. A smart chatbot architecture based NLP and machine learning for health care assistance. In Pro-ceedings of the 3rd international conference on networking, information systems & security (pp. 1-6).
13. Kevin, B., VIkin, B. and Nair, M., 2023. Building a Chatbot for healthcare using NLP. Authorea Preprints.
14. Prabanand, S.C., Akilan, S., Bhoopesh, S.J., Ananda Kumar, A. and Saswin, V., 2024. AI HEALTHCARE CHATBOT SYSTEM USING LLM-LAMA3.
15. Cascella, M., Montomoli, J., Bellini, V. and Bignami, E., 2023. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. Journal of medical systems, 47(1), p.33.
16. Yang, R., Tan, T.F., Lu, W., Thirunavukarasu, A.J., Ting, D.S.W. and Liu, N., 2023. Large language models in health care: Development, applications, and challenges. Health Care Science, 2(4), pp.255-263.