International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1958-1963
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0286

AI-Chatbot for Disease Prediction

Dr.R. G Suresh Kumar[1], Mr. Karthikeyan K[2], Mr. Kille Venkatsan S[3], Mr. Madhivadhanan M[4], Mr.Deepak Kumar[5]

[1]Head of the department, Dept of CSE, Rajiv Gandhi College of Engineering and Technology, Puducherry, India.

[2,3,4,5] UG Scholar, Dept of CSE, Rajiv Gandhi College of Engineering and Technology, Puducherry, India.

Emails: *sureshkumar_rg@rgcet.edu.in[1], kskarthik8220@gmail.com[2], killevenkatasan@gmail.com[3], pricemathi2004@gmail.com[4], dkgtpro@gmail.com[5]*

Abstract

This AI-based medical Chabot, particularly in disease prediction, highlighting the limitations of earlier Chabot's that rely on Long Short-Term Memory (LSTM) networks, such as narrow disease coverage and reduced accuracy. Recurrent neural networks (RNNs) are proposed as a promising alternative due to their enhanced capabilities in processing complex patient queries and providing accurate health recommendations. A systematic review emphasizes the need for medical Chabot's to integrate extensive disease databases and advanced natural language processing (NLP) techniques to improve interaction quality. Despite NLP's potential, current systems face challenges like language asymmetry, leading to response inaccuracies. Additionally, emerging trends such as hybrid AI models and modular design approaches are fostering greater adaptability in Chabot development. This survey identifies key challenges, including data dependency, scalability, and the necessity for improved accuracy in real-world applications, suggesting that a shift to RNN-based architectures and broader training data could significantly enhance the practical utility of medical Chabot's, ultimately improving patient outcomes and healthcare support reliability.

Keywords: *AI, LSTMs, RNNs, NLP.*

1. Introduction

AI-based medical Chabot's are emerging as integral tools for healthcare by aiding in patient assessment, early disease detection, and personalized health support. Earlier Chabot systems for disease prediction are often limited by their reliance on specific models, such as LSTM networks, which, while effective for some sequence-based tasks, struggle with expanding disease coverage and maintaining high accuracy across a broader set of conditions. The proposed solution in this survey is to examine the role of RNNs, which have shown promise in handling sequential data more efficiently and could improve Chabot accuracy and applicability across various medical conditions. [1]

1.1.Current Challenges and Limitations in Disease Prediction Chabot's

The limitations of current Chabot technologies, such as LSTM-based models, in terms of narrow disease scope and prediction accuracy. Specific problems, such as scalability and language asymmetry that limit the effectiveness of traditional Chabot's.

1.2.Advancements in AI Models for Chabot's

An advanced AI models, particularly RNNs, as promising alternatives to existing LSTM models. This RNNs are better suited for handling sequential data and interpreting complex patient queries for a broader disease spectrum.

1.3.Role of NLP in Enhancing Chabot Interaction

The importance of natural language processing (NLP) in medical Chabot's for interpreting patient queries accurately and responding meaningfully. The major challenges in NLP such as medical jargon and language diversity, which affect response accuracy.

1.4.Research Goals and Scope of the Survey

The main focus of the survey, including evaluating RNNs and hybrid models in improving disease prediction. Highlight the survey's goal to suggest

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1958-1963
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0286

more effective Chabot design strategies for enhanced reliability in healthcare

1.5.Research Questions

This paper seeks to answer the following questions

- RQ1. What are the primary methodologies and approaches used in AI-based medical Chabot's?
- RQ2. How do these models manage the complexity of sequential data, and what are their limitations?
- RQ3. What opportunities exist to enhance disease prediction and scalability in Chabot applications?

2. Literature Overview

2.1.Current Landscape of AI-Based Medical Chabot's

AI Chabot's in Healthcare: Existing AI Chabot, which focuses on infectious disease prediction, are typically built on LSTM models and have shown utility in Managing patient queries and symptom analysis, especially during the COVID-19 pandemic. These models operate on symptom-based interaction, relying on limited datasets and specific disease focus, which restricts their versatility and scalability for broader healthcare applications Natural Language Processing (NLP) for Chabot Interactions: NLP is critical to Chabot functionality, enabling natural, human-like interactions. However, studies reveal that NLP models are limited by their ability to interpret complex medical language and varied user expressions. This issue, known as language asymmetry, results in reduced accuracy in Chabot responses, especially for nuanced or ambiguous medical inquiries

2.2.Key Methodological Approaches in Medical Chabot's

The studies surveyed indicate a range of methodological approaches, including:

LSTM Networks: LSTMs, commonly used for their capability to process sequential data, are foundational in current Chabot models but face scalability and accuracy limitations when applied to datasets beyond a few specific diseases. Additionally, LSTM networks, while effective in capturing short-term dependencies, often struggle with learning long-term dependencies when applied to complex, multi-step dialogues, which can lead to degraded chatbot performance in sustaining context over extended interactions.

2.3.Recurrent Neural Networks (RNNs)

RNNs, though less commonly used in medical Chabot's, demonstrate potential for handling complex data sequences and accommodating diverse disease datasets due to their internal memory. This makes them a promising choice for expanding the scope and accuracy of disease predictions in medical Chabot's. RNNs can also capture temporal dependencies in patient data, which is critical in medical applications where disease symptoms and progression are time-dependent. This characteristic allows RNN-based models to analyse patterns over time, enhancing the model's ability to predict disease onset and progression with greater accuracy, especially in cases where symptoms may evolve gradually or recur.[2]

2.4.Hybrid Deep Learning Models

Table 1 Comparison of AI Models in Medical Chabot's

Model	Common Use	PROS	CONS
LSTM	Disease prediction in limited conditions	Good for handling sequential data	Limited scalability for multiple diseases, lower accuracy for diverse datasets accuracy for diverse datasets
RNN	Broader disease prediction	Improved sequential data	Complex implementation and higher resource requirements
CNN	Image recognition and classification	Efficient feature extraction from images.	Requires large datasets and computational power.

An emerging approach combines multiple AI models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), to improve disease prediction accuracy. By leveraging CNN's ability to extract complex features and RNN's proficiency with sequential data, hybrid models can process a wider range of patient inputs and symptoms. This multi-layered approach enhances the Chabot's adaptability to diverse medical cases. Such architectures are also more resilient to noise in user inputs, providing more accurate and context-aware health predictions. (Table 1) (Figure 1)

2.5.Chart View

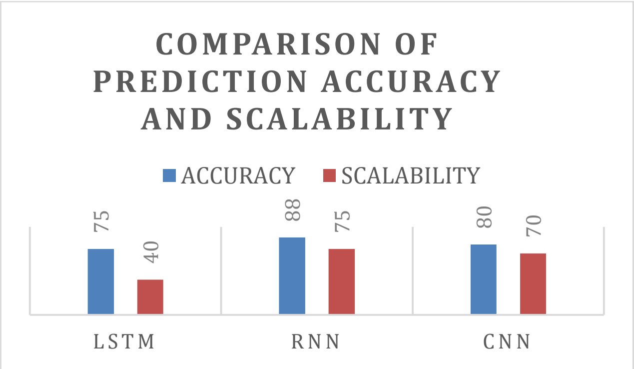

Figure 1 Comparison of Prediction Accuracy and Scalability of LSTM, RNN and CNN

3. Methodologies and Approaches

3.1.Data Collection and Preparation

- **Data Sources:** The common sources used for Chabot training datasets, such as electronic health records (EHRs), symptom databases, and medical literature from databases like PubMed and Medline. [3]
- **Data Quality and Pre-processing:** The challenges in data quality, including handling missing data, standardizing medical terminology, and removing biases. Pre-processing steps like data cleaning, tokenization, and stemming/lemmatization can enhance the model's ability to understand and interpret user inputs more accurately.

3.2.Model Selection and Training Techniques

3.2.1. AI Model Selection

- **Long Short-Term Memory (LSTM):** The use of LSTMs in handling sequence-based data, such as symptom progression, while highlighting their limitations in managing multi-disease datasets and in longer response generation. [4]
- **Recurrent Neural Networks (RNNs):** RNNs has the advantage in sequential data processing, explaining how their ability to retain contextual information allows for a more comprehensive analysis of patient symptoms across a broader range of conditions.
- **Hybrid Models:** This approaches combining RNNs with other architectures, like Convolutional Neural Networks (CNNs) or Transformers, to enhance pattern recognition and contextual understanding in symptom-based Chabot's. [5]

3.3.Training Techniques

- **Supervised vs. Unsupervised Learning:** That the role of both learning methods in disease prediction Chabot's, noting that supervised learning is typically used for labelled medical datasets while unsupervised learning can assist in discovering symptom patterns in unstructured patient conversations.
- **Transfer Learning:** The pre-trained models on large medical datasets can be fine-tuned for disease prediction tasks, which reduces the need for extensive labelled data and accelerates training times.
- **Reinforcement Learning (RL):** It explore the use of RL to continuously improve Chabot responses based on patient feedback, allowing Chabot's to learn over time and provide more accurate predictions. [6]

3.4.Natural Language Processing (NLP) for Symptom Analysis

- **Feature Extraction:** NLP techniques like Named Entity Recognition (NER) and Part-of-Speech (POS) tagging identify relevant medical terms and symptoms in patient inputs, enabling Chabot's to extract and interpret essential features accurately.
- **Language Models:** Language models like BERT or GPT have been adapted to process and interpret medical queries, improving the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1958-1963
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0286

Chabot's understanding of medical terminology, context, and patient intent.

- **Handling Ambiguity and Language Diversity:** The methods like multilingual embedding's and domain-specific language models to tackle language diversity and asymmetry.by explaining how these methods help interpret medical slang, regional expressions, or non-standard medical terminology.

3.5.Model Evaluation and Validation

- **Evaluation Metrics:** By identifying specific metrics relevant for medical Chabot performance evaluation, such as accuracy, precision, recall, F1 score, and AUROC (Area under the Receiver Operating Characteristic) curve. Explain why each is important for assessing Chabot reliability in predicting diseases. [7]

3.6.Validation Techniques

- **Cross-Validation:** k-fold cross-validation in improving model robustness, especially for limited medical datasets.
- **Real-world Simulation Testing**: It describes the importance of real-world scenarios where Chabot's are tested with diverse queries to assess their practical effectiveness.
- **A/B Testing and User Feedback:** It explain how feedback loops, including A/B testing with users and physicians, allow Chabot's to refine responses over time and adapt to different user demographics.

3.7.Implementation Frameworks and Tools

- **AI Frameworks:** Most commonly used frameworks, such as Tensor Flow, PyTorch, and Keras, for model development, training, and deployment. [8]
- **Data Integration and Management Tools:** The use of data management tools like Apache Hadoop and data pre-processing frameworks like NLTK and SpaCy for efficient handling of large medical datasets.
- **Deployment Considerations:** key factors in deploying medical Chabot's, such as compliance with healthcare regulations (e.g., HIPAA) and integration with electronic health record (EHR) systems to provide real-time patient support and health tracking.

4. Findings and Trends

4.1.Shift Toward Advanced Neural Networks

Preference for RNN over LSTM: An emerging trend is the shift toward RNNs for medical Chabot development. Unlike LSTMs, RNNs better support complex sequential data handling, enabling Chabot's to interpret and predict patient symptoms across a broader spectrum of diseases, thus enhancing predictive accuracy and expanding potential healthcare applications. Another advantage of using RNNs in medical chatbot development is their lower computational complexity compared to LSTMs. Since RNNs have simpler architectures without the added memory cell structures, they require fewer resources, making them more efficient for real-time symptom prediction and response generation. This efficiency is especially beneficial for applications that need to operate quickly on mobile or cloud-based platforms with limited processing power, enabling broader access and faster patient support. (Figure 2)

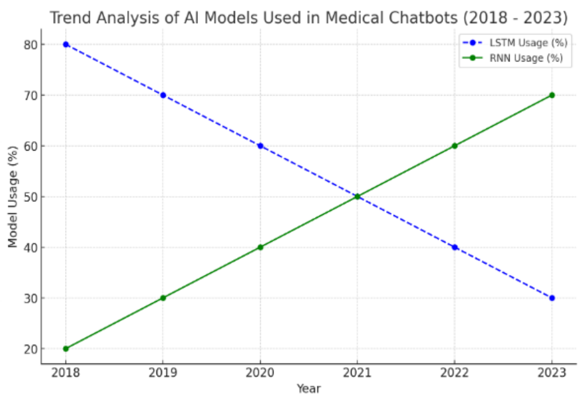

Figure 2 Graph on Trend Analysis of AI Models Used in Medical Chabot's (2018 - 2023)

This graph illustrates the trend in model usage over recent years, highlighting the increasing preference for RNNs and the corresponding decline in LSTM usage for AI-based medical Chabot's. [9]

4.2.Enhanced Dataset Integration

To improve disease coverage, recent studies emphasize the integration of comprehensive disease databases. This trend aligns with the proposed Chabot enhancement in the provided abstract, as it moves toward a system capable of analysing a broader set of symptoms and conditions, ultimately offering more personalized healthcare recommendations. [10]

4.3.NLP and Symptom-Based Responses

NLP advancements facilitate more accurate patient interactions by interpreting symptoms and providing tailored recommendations. However, limitations remain in processing complex medical terminology and symptom descriptions, particularly in cases involving rare or ambiguous diseases. (Table 2) [11]

Table 2 Findings and Trends in AI Chabot Development

Trend	Description	Impact on Chabot Design
Shift to RNN models	Use of RNNs over LSTMs for improved accuracy	Enhanced disease prediction scope
Use of multi-disease databases	Incorporation of extensive health data	Increased scalability and adaptability
Advanced NLP integration	Improved natural language processing	Enhanced user experience and accuracy

5. Challenges and Gaps

5.1.Limitations in Sequential Data Processing

LSTMs, though capable of processing sequential data, exhibit weaknesses in scaling for multiple diseases and adapting to varied patient queries. This limitation results in reduced accuracy for broader applications and restricts Chabot's to a narrower range of health conditions, as seen in single-disease models like those for COVID-19. [12]

5.2.Data Dependency and Scalability Constraints

Current AI Chabot's heavily rely on disease-specific datasets, which, while ensuring accuracy within those constraints, limit applicability in broader, real-world healthcare environments. The absence of generalized datasets or multi-disease training data constrains their practical utility, reducing effectiveness for diverse health inquiries and generalized disease prediction. [13]

5.3.Language Processing Asymmetry

NLP remains a bottleneck, with language asymmetry affecting accurate interpretation of medical terms and context. This issue poses a barrier for Chabot's aiming to handle diverse patient expressions and complex symptom descriptions, impacting their response accuracy and user experience. [14]

Conclusion

This literature survey highlights significant advances in the development of AI-based medical Chabot's, with a growing emphasis on utilizing RNNs to improve sequential data processing and disease coverage. RNNs offer a promising solution for Chabot's seeking to expand their predictive capabilities and provide reliable healthcare information across a wide range of conditions. Future research should focus on integrating multi-disease datasets, advancing NLP techniques, and adopting modular frameworks to enhance Chabot versatility and reliability in healthcare. This approach has the potential to improve patient outcomes and position AI Chabot's as vital components of digital healthcare systems. [15]

References

[1]. J. Cahn, ''CHATBOT: Architecture, design, & development,'' Dept. Comput. Inf. Sci., Univ. Pennsylvania School Eng. Appl. Sci., Philadelphia, PA, USA, Tech. Rep. EAS499, 2017.

[2]. A. Kumar, P. K. Meena, D. Panda, and M. Sangeetha, ''Chatbot in Python,'' Int. Res. J. Eng.

[3]. S. A. Sheikh, ''Artificial intelligence based Chabot for human resource using deep learning,'' Ph.D. dissertation, Dept. Comput. Sci. Eng., Manipal Univ., Manipal, India, 2019.

[4]. M. M. Hossain, S. Krishna Pillai, S. E. Dansy, and A. A. Bilong, ''Mr. Dr. Health-assistant chatbot,'' Int. J. Artif. Intell., vol. 8, no. 2, pp. 58–73, Dec. 2021.

[5]. G. Battineni, N. Chintalapudi, and F. Amenta, ''AI chatbot design during an epidemic like the novel coronavirus,'' Healthcare, vol. 8, no. 2, pp. 1–8, Jun. 2020.

[6]. R. Dharwadkar and N. A. Deshpande, ''A

medical chatbot,'' Int. J. Comput. Trends Technol., vol. 60, no. 1, pp. 41–45, 2018.

[7]. D. Madhu, C. J. N. Jain, E. Sebastain, S. Shaji, and A. Ajayakumar, ''A novel approach for medical assistance using trained chatbot,'' in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 243–246.

[8]. F. Mehfooz, S. Jha, S. Singh, S. Saini, and N. Sharma, ''Medical chatbot for novel COVID-19,'' in ICT Analysis and Applications. Singapore: Springer, 2021, pp. 423–430.

[9]. M. Herriman, E. Meer, R. Rosin, V. Lee, V. Washington, and K. G. Volpp, ''Asked and answered: Building a chatbot to address COVID-19-related concerns,'' NEJM Catalyst Innov. Care Del., vol. 1, pp. 1–13, Jun. 2020.

[10]. S. Altay, A. S. Hacquin, C. Chevallier, and H. Mercier, ''Information delivered by a chatbot has a positive impact on COVID-19 vaccines attitudes and intentions,'' J. Exp. Psychol., Appl., vol. 27, pp. 1–11, Oct. 2021.

[11]. P. Amiri and E. Karahanna, ''Chatbot use cases in the COVID-19 public health response,'' J. Amer. Med. Inform. Assoc., vol. 29, no. 5, pp. 1000–1010, Apr. 2022.

[12]. M. Almalki and F. Azeez, ''Health chatbots for fighting COVID-19: A scoping review,'' Acta Inf. Medica, vol. 28, no. 4, p. 241, 2020.

[13]. P. Weber and T. Ludwig, ''(Non-) interacting with conversational agents: Perceptions and motivations of using chatbots and voice assistants,'' in Proc. Conf. Mensch Comput., vol. 1, Sep. 2020, pp. 321–331.

[14]. D. Shin, The perception of humanness in conversational journalism: An algorithmic information-processing perspective, NewMediaSoc.,vol.24, no. 12, pp. 26802704, Dec. 2022, doi: 10.1177/1461444821993801.

[15]. T.N.K.Hung,N.Q.K.Le,N.H.Le,L.VanTuan, T.P.Nguyen,C.Thi,and J.-H. Kang, An AI-based prediction model for drug-drug interactions in osteoporosis and Pagets diseases from SMILES, Mol. Informat., vol. 41, no. 6, Jun. 2022, Art. no. 2100264.