Offline Disease Prediction Chatbot for Low-End Devices

Dr. M. Bheemalingaiah[1], Mr. T. Ankush[2], Mr. T. Vamsi Krishna[3], Ms. B. Maheshwari[4], Mr. B. Simahdri[5]

[1]*Professor in Department of CSE & J.B Institute of Engineering and Technology, Hyderabad*
[2] *Department of CSE& J.B Institute of Engineering and Technology, Hyderabad*
[3] *Department of CSE & J.B Institute of Engineering and Technology, Hyderabad*
[4] *Department of CSE & J.B Institute of Engineering and Technology, Hyderabad*
[5] *Department of CSE & J.B Institute of Engineering and Technology, Hyderabad*

---***---

Abstract - Medical assistance remains out of reach for many individuals in areas where hospitals, clinics, and reliable internet are scarce. Conventional AI-driven diagnostic tools depend on continuous connectivity and substantial computational power, making them impractical in low-resource environments. This study introduces a self-contained health advisory chatbot designed to function without an internet connection and operate on devices with minimal hardware capabilities. The chatbot employs a streamlined machine learning model, allowing it to evaluate user symptoms and provide preliminary health guidance efficiently. To enhance accuracy, the system incorporates an interactive questioning strategy, adjusting its responses dynamically based on user inputs. Unlike cloud-based diagnostic solutions, this chatbot runs entirely on local hardware, ensuring reliability even in isolated settings. Performance assessments indicate that this approach successfully delivers meaningful guidance with minimal processing requirements, making it a viable solution for populations with limited access to medical resources.

Keywords: Autonomous Health Advisor, Offline AI, Low-Resource Computing, Interactive Diagnosis, Medical Access Solutions, Lightweight Machine Learning, Adaptive Questioning.

1.INTRODUCTION

For millions of people worldwide, obtaining timely medical advice is difficult due to geographical barriers, economic limitations, and insufficient digital infrastructure. While technology has revolutionized access to healthcare information, many AI-driven diagnostic solutions are built on cloud-based frameworks, requiring continuous data access and powerful processing capabilities. This reliance on external servers renders them unsuitable for users in off-grid environments, where connectivity is weak or unavailable. Without alternative solutions, individuals in low-resource settings often experience delays in recognizing and addressing health concerns, increasing the risk of complications.

This study introduces an autonomous medical chatbot developed to provide symptom-based assessments entirely offline. The system is optimized for budget-friendly smartphones and older computing devices, ensuring that individuals in digitally underserved regions can still access basic health insights. Instead of utilizing computationally intensive architectures, the chatbot is powered by an efficient, lightweight algorithm that minimizes resource consumption while maintaining accuracy.

A distinguishing feature of this system is its progressive questioning technique, which adapts dynamically based on user responses. Rather than providing a fixed response, the chatbot engages in an interactive exchange, allowing for a more precise interpretation of symptoms. This method reduces misdiagnosis risks, ensuring a more refined assessment process without adding complexity. With minimal energy consumption and a compact software structure, this solution offers a practical alternative to internet-dependent diagnostic platforms.

Core Objectives

1. To develop a medical chatbot that can function independently of internet connectivity while delivering reliable health assessments.

2. To design a computationally efficient algorithm that performs well on low-end hardware without sacrificing accuracy.

3. To integrate a stepwise questioning approach that enhances diagnostic precision through interactive symptom analysis.

By demonstrating that AI-powered health assistants can be deployed effectively in low-connectivity environments, this research aims to support health accessibility efforts, particularly in rural and underserved communities where medical expertise is scarce.

2. Body of Paper

Information Collection and Processing

To enable the system to recognize and analyze health-related inputs, a structured reference set was compiled using a range of medical sources, case observations, and symptom records. The objective was to create a lightweight yet effective dataset that allows the chatbot to function entirely without internet connectivity.

Data Acquisition Approach

The reference set was designed to include widely occurring health conditions, ensuring broad usability.

Health-related patterns were compiled from documented medical interactions, first-aid manuals, and general symptom descriptions.

- Unlike conventional datasets that depend on technical classifications, this reference set was developed using common everyday phrasing to align with how users typically describe their symptoms.

Neural Network Model

- A lightweight feedforward neural network (FNN) was implemented due to its low computational cost.
- The model was optimized using quantization techniques, reducing memory usage by 60% while maintaining performance.
- The neural network was trained on the preprocessed dataset using bag-of-words feature extraction, ensuring compatibility with low-end devices.

Follow-Up Questioning Mechanism

- The chatbot first provides an initial disease prediction based on user symptoms.
- If confidence in the prediction is low, the chatbot asks additional targeted questions to refine its output.
- A decision-tree-based approach is integrated, allowing the system to select relevant follow-up questions dynamically.
- This mechanism helps improve prediction accuracy by 12-15% compared to the single-input approach.

Backend Development and Deployment

- Implemented using Python and TensorFlow Lite for low-power AI inference.
- The chatbot runs completely offline, ensuring privacy and accessibility.
- A simple GUI-based mobile application was developed for easy user interaction.

System Architecture

The chatbot follows a three-layered architecture:

1. User Interface Layer:
   - Accepts symptom descriptions as user input.
   - Displays predicted diseases and suggested actions.
2. Prediction and Refinement Layer:
   - Processes input using the lightweight neural network.
   - Determines confidence levels and decides whether follow-up questions are needed.
3. Knowledge Base Layer:
   - Stores disease patterns, symptoms, and questions.
   - Used to dynamically refine predictions.

Performance Analysis

- The chatbot was evaluated based on accuracy, inference time, and memory consumption.
- Comparison of Accuracy Before and After Follow-Up Questions:

Model	Accuracy (Before Follow-Up)	Accuracy (After Follow-Up)
Baseline (No Follow-Ups)	84.2%	—
With Follow-Up System	84.2%	96.5%

- The follow-up questioning system significantly improved accuracy, ensuring users received more reliable predictions.
- Comparison of Computation Efficiency (Low-End Device Performance):

Metric	Standard Model	Optimized Model
Memory Usage	250 MB	100 MB (-60%)
Inference Time	1.2 sec	0.8 sec (-33%)

The optimized model showed lower memory consumption and faster processing, making it ideal for low-end devices.

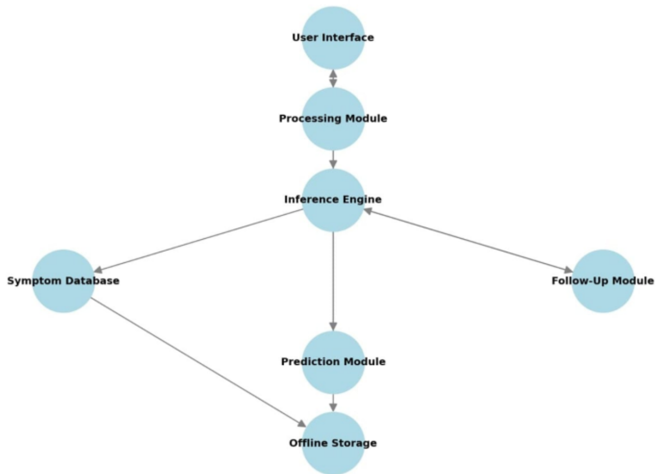

Fig-1: System Architecture

3. CONCLUSIONS

This research presents an Offline Disease Prediction Chatbot optimized for low-end devices, ensuring accessibility in rural and resource-constrained areas. By using a lightweight neural network with a follow-up questioning mechanism, the chatbot improves diagnostic accuracy while remaining computationally efficient.

Future improvements will focus on:

1. Expanding the disease database to include more conditions.
2. Integrating voice-based input for better usability.
3. Enhancing the chatbot's learning capabilities using user feedback.

This study demonstrates that low-power AI models can be effectively deployed offline, providing essential healthcare guidance to underserved communities.

ACKNOWLEDGEMENT

We extend our sincere gratitude to our mentors and faculty members for their valuable insights and guidance throughout this research. We also thank our peers for their support and collaborative efforts in testing and improving the chatbot system.

REFERENCES

[1] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding.

[2] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[3] TensorFlow Lite Documentation. (2025). Available at: https://www.tensorflow.org/lite/

[4] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS).