International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI HEALTH ASSISTANT USING MACHINE LEARNING AND DATA SCIENCE

Raman Singh Raghuwanshi[1] Ahmedun Nisha[2], Vaishnavi Mishra[3],
Shruti Rathore[4],Mulayam Singh Baghel[5] Prof.Ashish Gawande[6]
B.Tech scholar[1,2,3.4,5], Assistant Professor[6]
Department of Artificial Intelligence & Data Science
Shri Balaji Institute of Technology & Management, Betul, RGPV University, M.P., India

ABSTRACT:

The paper presents an AI Health Assistant that leverages Machine Learning and Data Science to provide personalized, conversational health consultations. It integrates advanced Natural Language Processing for symptom understanding across multiple languages (English, Hindi, Hinglish) and offers real-time voice-based and chat-based assistance. The system utilizes dynamic machine learning Ensembling models like Random Forest Classifier for continuous learning and improved diagnostic accuracy, aiming to make initial health guidance accessible, empathetic, and efficient for users and medical professionals alike [1] [2].

Keywords: Natural Language Processing, Voice Assistant, Machine Learning, Data Science

1.INTRODUCTION

In an era demanding accessible healthcare, the AI Health Assistant emerges as a pivotal solution, integrating Machine Learning and Data Science to redefine preliminary health consultations. This system offers a Personalized & Conversational Interface,

Enabling users to articulate health concerns in natural language through both chat and voice. Crucially, it understands multiple languages, including English, Hindi, and Hinglish, significantly broadening its accessibility [1]. For immediate, hands-free assistance, users can engage in Real-Time Health Consultation via live voice sessions with an AI Health Expert, providing direct and personal health advice without typing [2]. This study proposes a Natural Language Processing (NLP) for flexible understanding of conversational language, ensuring natural and empathetic interactions [3]. This system changes Static Knowledge to Dynamic Learning, utilizing Machine Learning models, including Random Forest Classifier and Ensemble learning, which continuously train on new data to enhance accuracy and expand knowledge over time [4].

Beyond reactive responses, the assistant delivers Wellness Tips, such as advice on sleep or hydration, fostering a holistic approach to well-being.The Centralized Health UI offers a clean, intuitive interface, designed to feel like a personal wellness companion. Furthermore, features like the "Mindset Check" provide Inspirational Mindset Checking through voice-based AI, offering mental health relaxation and ensuring users feel connected during stress or anxiety [5].

2 .LITERATURE REVIEW

Early AI health support used expert systems and decision trees for basic symptom assessment [1], but lacked flexibility for complex medical dialogues and limited user expression [2]. Ensemble methods like Random Forest and Ensembling techniques enhanced diagnostic accuracy. Model is Trained on vast medical Public datasets, Voice-enabled interfaces, exemplified by "Jarvis Health," further revolutionized this domain for intuitive symptom checking [3].

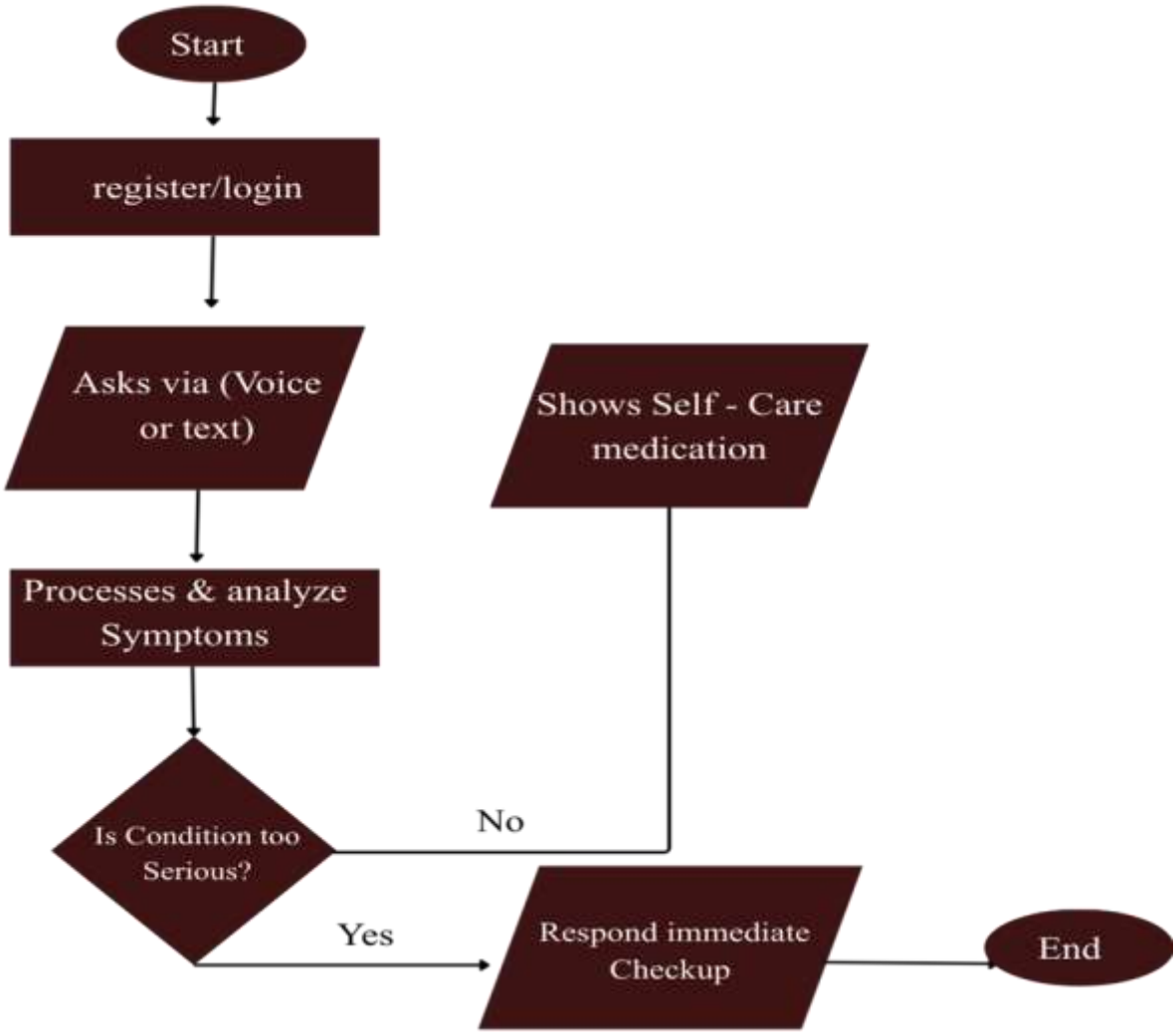
Recent literature emphasizes real-time conversational AI and explainable AI (XAI) to improve user experience and clinical utility. Multilingual support and fuzzy logic matching are crucial for accurate interpretation of diverse user inputs [4] [5] [6]. These models offered dynamic learning capabilities, crucial for evolving medical understanding. The future of virtual assistants in healthcare is redefined by their ability to offer proactive wellness tips, mental health support, and seamless integration into daily life, transforming them into personal wellness companions [7] .

This research underscores the need for accurate, empathetic, accessible, and continuously learning systems [8] . Furthermore, the integration of XAI into conversational AI systems, including chatbots, is crucial for developing next-generation assistants that can not only provide accurate information but also explain their rationale in an understandable manner, fostering greater user confidence and engagement [9] [10]

3. METHODOLOGY

The proposed study develops interpretable machine learning model and CNN for an AI Health Assistant and Image Visualization. The workflow encompasses data acquisition, preprocessing, natural language processing, model training, and an interactive user interface.

FLOWCHART



A. Dataset Description

The dataset used in this system consists of structured medical data sourced from (WHO) publicly available symptom-disease. It contains symptom-to-disease mappings across 41 disease categories, together covering over 4,800 labeled records with 132 symptom features as input attributes and disease name as target variable.

B. Data Preprocessing

Data was cleaned and transformed through:

1. **Missing Value Handling :-** Null and empty symptom columns were dropped entirely.
2. **Encoding :-** Applied Label Encoder to convert Disease names to numerical class labels
3. **Feature Consistency Check:-** All 132 symptom columns hold only binary values (0 or 1)
4. **Splitting :-** Dataset was divided 70% for training and 30% for testing to evaluate model performance on unseen data.

C. Model Development

Three models were trained and compared:

- **Single Decision Tree**: The Ground-state model for initial symptom-to-disease mapping.
- **Random Forest:** Improved accuracy and robustness through ensemble voting.
- **Random Forest with Bagging:** Delivers predictability training each tree on random data subsets. [2]

After ensemble tuning, Random Forest achieved ~85% accuracy, outperforming the single Decision Tree.

D. Evaluation Metrics

Model performance was assessed using Accuracy, Precision, Recall, and F1-score. Random Forest showed the highest accuracy, confirming its strong multi-class disease classification over a single Decision Tree.

E. Risk Categorization

Predicted diseases were classified into three actionable care levels for user guidance:

- **Urgent Care (immediate medical attention required)**
- **Monitor Symptoms (watch and manage at home)**
- **Self Care (manageable with basic precautions)**

This provides effective Care and personalization of patient's health..

F. System Interface and Interaction

A Flask and React based conversational interface allows users to describe symptoms through text or live voice input in English, Hindi, or Hinglish. The system guides users through a natural dialogue, starting with symptom collection, moving into NLP extraction and fuzzy matching, before delivering a predicted disease, precaution advice, and care-level recommendation all accompanied by empathy quotes and optional voice output, making the experience feel personal rather than clinical. Through DOM Manipulation, as the Python backend generates results, JavaScript dynamically injects these into the chat window without refreshing the page, maintaining the flow of a real-time conversation.

G. Discussion and Critical Analysis

Random Forest outperformed single tree models by capturing complex symptom co-occurrence patterns across 41 disease categories. Key prediction drivers were symptom combinations, language normalization accuracy, and fuzzy mapping precision. The system enhances usability through voice interaction, Hindi and English support, and risk-based care categorization

4. RESULT/OUTPUT:

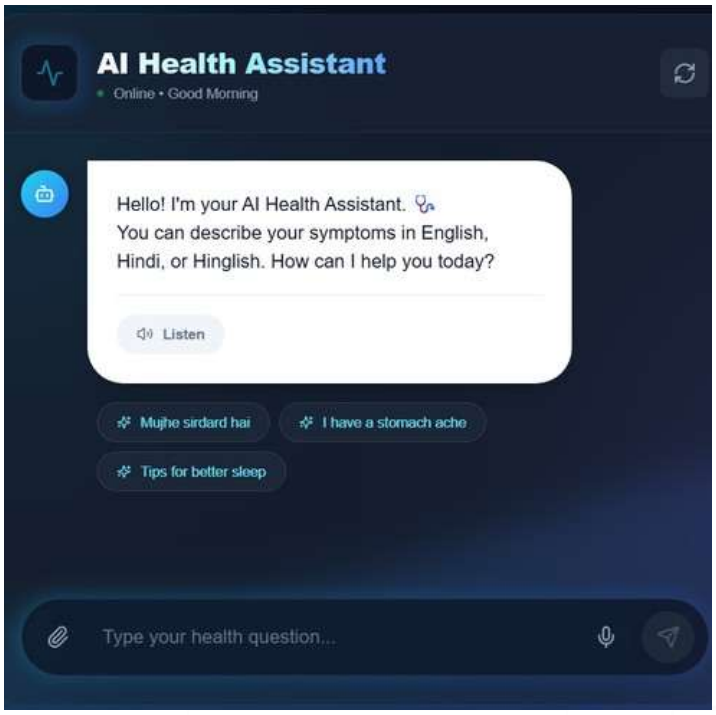


Fig 1: Chat based Result interaction

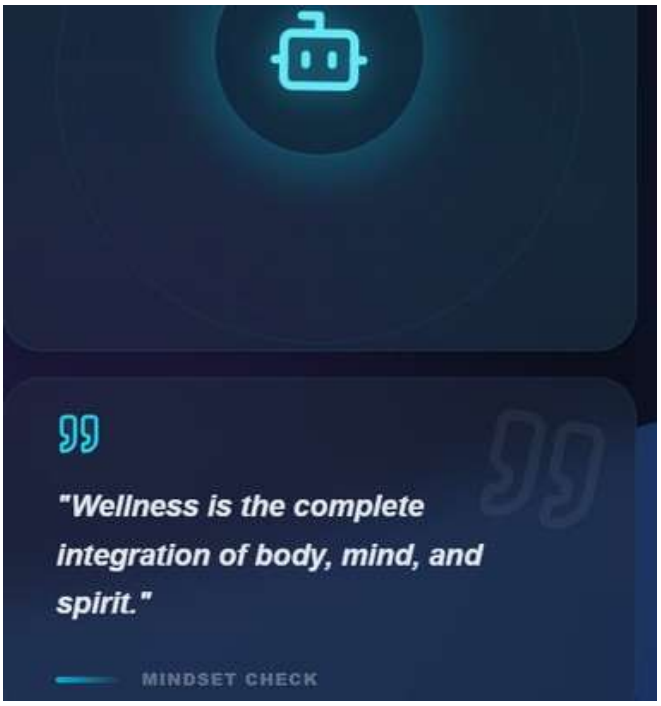


Fig 2 : Mindset Check Panel

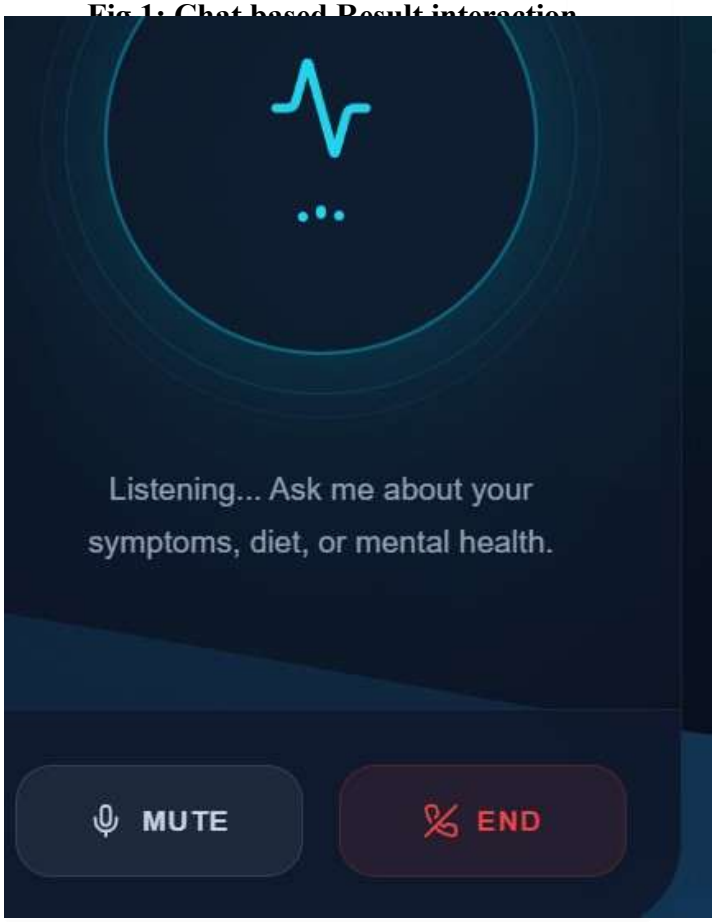


Fig 3: LIVE Voice Assistant

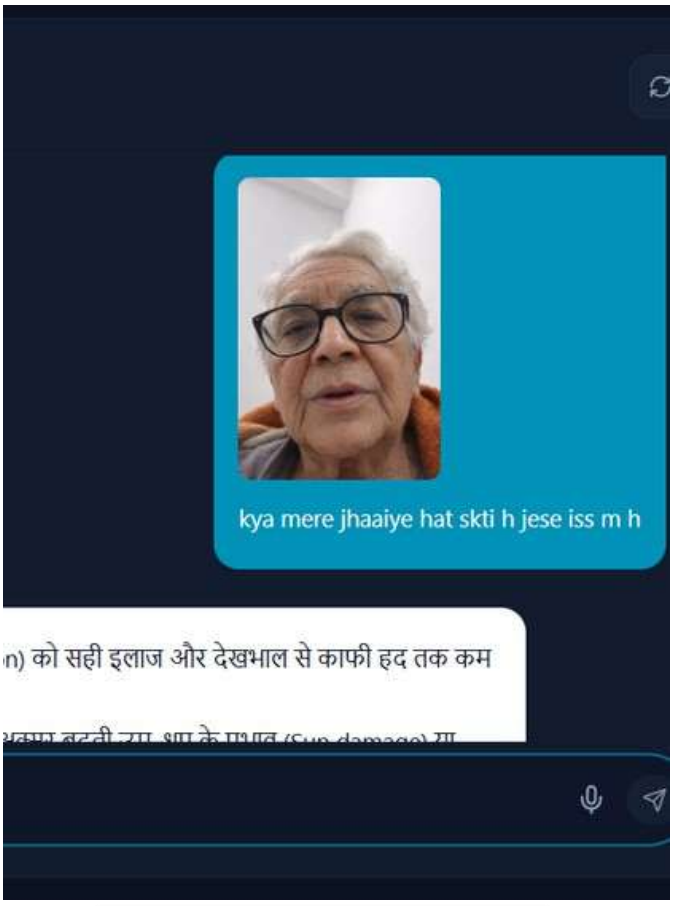


Fig 4: Image - Based Interaction Result

5. RESEARCH GAP:

Most models act as black boxes, offering symptom predictions without explaining underlying health risk factors or providing multimodal consultation support. This research bridges gaps by introducing an interpretable, voice-enabled health consultation system that offers severity-based health risk segmentation (Self-Care, Monitoring Required, Urgent Care) and empathetic wellness insights, enhancing both usability and decision support for medical professionals and patients. Existing studies mainly focus on symptom prediction accuracy using models like Decision Trees, and basic NLP techniques, but they lack personalized interaction and practical healthcare [1][3].

6. CONCLUSION:

The proposed system integrates data preprocessing, feature extraction, and predictive modeling through algorithms such as Random Forest Classifier and Ensemble Learning. The system not only predicts whether a user's symptoms indicate a health concern but also provides severity-based risk classification categorizing users into Self-Care, Monitoring Required, and Urgent Care risk segments. This enables healthcare professionals to prioritize critical cases and allocate resources efficiently. The implementation of a voice-enabled interactive interface allows users to consult the AI assistant in both text and voice formats, enhancing accessibility and user engagement transparency. Moreover, the inclusion of empathetic wellness tips and pre-visit patient summarization helps medical professionals understand critical symptoms influencing patient health. The multimodal interaction design, powered by NLP and speech processing, creates a human-like consultation experience.[5][6]

7. FUTURE SCOPE:

For future research, deep learning-based architectures, transformer models for advanced NLP, and real-time **IoT** health data streaming can be explored to further enhance prediction accuracy and personalization. Integration with **Electronic Health Record (EHR)** systems and deployment on cloud platforms with **Telemedicine APIs** would make the solution more adaptive for large-scale healthcare operations.

REFERENCES:

1) Mahmoud Nasr, Md. Milon Islam, Shady Shehata, Fakhri Karray, Yuri Quintana, “Smart Healthcare in the Age of AI, 2021, Publisher: IEEE Access , DOI: 10.1109/ACCESS.2021.3118960
2) Kambam Sai Charan, Buda Dhanushya, Velayudham, Bolla Manoj Sai, “Intelligent Symptom Conversation 2025, Publisher: IEEE - DOI: 10.1109/ICFT66708.2025.11336630
3) Varisha Khanam, “Jarvis Health: An AI-Based Voice-Enabled Symptom Checker Chatbot” 2025, Publisher : IJERT , DOI : https://doi.org/10.5281/zenodo.18106489
4) Emre Sezgin, “Redefining Virtual Assistants in Health Care : The Future With Large Language Models” 2024, Publisher : Journal of Medical Internet Research (JMIR) DOI: 10.2196/53225
5) Marrina Barreda, David Canterro – Prieto, Daniel Coca ..etc, Transforming healthcare with chatbots: Uses and applications , Publisher: PubMed Central (PMC) doi: 10.1177/20552076251319174
6) Sharad Rakshit Wani, Shahid Hussain Attri, Sachin Setia, "A Review on Disease Predictions” 2024, Publisher: Springer (Lecture Notes in Networks and Systems) , , DOI: 10.1007/978-981-99-9037-5_26
7) Shaswat Chakraborty, IEEE Access “Medical Chatbot” , 2023 - DOI: 10.1109/ACCESS.2022.3227208
8) Rijul Gupta, Dhanshree R Gunjawate, for "Voice Disorder Recognition,” 2024, Publisher: BMJ Open (British Medical Journal), DOI: 10.1136/bmjopen-2023-076998
9) Prasannakumar P, Paul J, Vani V,“ "AI-Driven Symptom Analysis” 2025, Publisher: Springer (Smart I nnovation, Systems and Technologies), DOI: 10.1007/978-981-97-8355-7_25
10) Hiba Malik Mohssen,“Mental Healthcare Chatbot Based on a Fine-Tuned GPT-2,” 2025, Publisher: Springer Nature (Lecture Notes in Networks and Systems), DOI: 10.1007/978-981-96-3102-5_5