IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Medical Healthcare Chatbot

Mrs. Dr. N.P. LAVANYA KUMARI , Mr. SOMALA VINESH KUMAR

1Assistant Professor(c), AU College of Engineering (A), Andhra University, Visakhapatnam, India
2PG Student, AU College of Engineering (A), Andhra University, Visakhapatnam, India
Department of Information Technology and Computer Applications
Andhra University College of Engineering, Visakhapatnam, India

Abstract: This paper presents an AI-powered Healthcare Chatbot that integrates rule-based reasoning with machine learning techniques to perform preliminary medical assessments based on user-reported symptoms. Designed to address the challenges of healthcare accessibility, particularly in remote and resource-constrained settings, the system leverages Decision Tree and Support Vector Machine (SVM) classifiers trained on curated medical datasets to predict probable diseases. Through a command-line conversational interface, users interactively describe symptoms and receive predictions, along with tailored recommendations including precautions, medications, and natural remedies. The system incorporates a secondary validation mechanism for enhanced diagnostic reliability and a severity analysis to advise on the urgency of professional consultation. Key features include secure storage of session histories in an SQLite database, auto-generation of detailed PDF diagnosis reports, and complete history management ensuring privacy. Unlike existing commercial health apps, this solution is interpretable, offline-capable, transparent in its predictions, and suitable for academic and educational use. Experimental results demonstrate the chatbot’s potential as an intelligent virtual triage tool that empowers users to make informed health decisions while reducing unnecessary hospital visits.

***Index Terms* -** Healthcare Chatbot, Disease Prediction, Decision Tree, Support Vector Machine, Symptom Analysis, PDF Report Generation, SQLite, AI in Healthcare, Virtual Triage, Command-Line Interface, Machine Learning, Severity Analysis.

I. INTRODUCTION

Healthcare chatbots have emerged as a promising solution in the intersection of artificial intelligence (AI), natural language processing (NLP), and medical diagnostics, offering accessible and immediate preliminary health advice. With the increasing pressure on healthcare systems and limited access to medical professionals in rural and underserved areas, there is a growing demand for intelligent tools that assist individuals in understanding their symptoms and taking informed next steps.

Traditional online symptom checkers and mobile health apps often suffer from drawbacks such as lack of personalization, dependence on internet connectivity, opaque decision-making, and inability to track user history securely. These limitations hinder their usefulness in offline or academic settings where transparency, interpretability, and data privacy are critical.

To overcome these challenges, this project introduces a hybrid Healthcare Chatbot system that integrates supervised machine learning with rule-based reasoning to deliver reliable, explainable, and user-centric health assessments. The system employs Decision Tree and Support Vector Machine (SVM) classifiers trained on curated medical datasets to predict probable diseases based on user-reported symptoms.

Built on a modular Python-based architecture with SQLite, FPDF, and scikit-learn, the chatbot runs in a simple command-line interface for maximum portability and offline usability. Key features include a secondary prediction mechanism for improved accuracy, severity analysis to gauge the urgency of consultation, secure session history storage, and automatic PDF report generation summarizing the diagnosis.

This intelligent and transparent approach bridges the gap between individuals and healthcare services by providing immediate, interpretable advice while maintaining user privacy. It is particularly suited for deployment in remote health centers, educational institutions, and telemedicine platforms where reliable, offline-capable, and auditable diagnostic support is essential.

LITERATURE SURVEY

In recent years, there has been a surge in the development of healthcare chatbots powered by artificial intelligence and machine learning to support patients in accessing medical information and obtaining preliminary diagnoses. Various research works have explored diverse techniques such as rule-based logic, supervised learning, deep learning, and cloud-based conversational agents to enhance chatbot efficiency and reliability.

1. Kumar A. and Sharma R. (2020) introduced a rule-based healthcare chatbot system that utilized predefined symptom-response mappings. The chatbot aimed to deliver quick responses based on keyword detection in the user's input. While the system was easy to implement and efficient for basic diagnosis, its major limitation was its inability to learn or improve over time. The lack of adaptability and rigidity in handling diverse symptoms made the solution unsuitable for real-time, large-scale deployment.

2. Patel N. and Trivedi M. (2019) designed a machine learning-based disease prediction system using Decision Trees and Naive Bayes classifiers. By training on structured symptom-disease datasets, the system predicted likely conditions based on user inputs. The approach yielded good classification results and laid the foundation for intelligent diagnosis. However, the model lacked a conversational interface, making it more suitable for backend prediction engines rather than direct user-facing applications.

3. Mehta S. and Agrawal K. (2021) implemented a deep learning approach using Deep Neural Networks (DNNs) to classify diseases based on input symptoms. The model achieved a high degree of accuracy and was capable of handling large and complex medical datasets. Nonetheless, the system functioned as a black box, providing no transparency or interpretability in its decision-making—an essential aspect in medical diagnosis where explanation is often as important as the prediction itself.

4. Yadav T. and Singh R. (2020) created a cloud-based healthcare assistant using Google Dialogflow and Firebase. This chatbot engaged users in real-time conversations, provided general medical advice, and stored interactions for continuity. Its natural language capabilities enhanced user experience, but it lacked machine learning-based diagnosis capabilities and was dependent on continuous internet connectivity, making it less useful in offline or rural settings.

5. Al-Sarem M. et al. (2021) built a COVID-19 specific chatbot that performed initial symptom screening and provided safety guidelines based on WHO protocols. It served a critical role during the pandemic by disseminating timely and reliable information. However, its functionality was restricted to COVID-19 alone and lacked extensibility to other diseases or general medical queries.

6. Ahmed A., Gupta M., and Bose A. (2022) proposed a hybrid healthcare chatbot that combined NLP-based question answering with SVM-based disease prediction. The chatbot extracted symptoms from user queries using NLP techniques and passed them to a machine learning model for diagnosis. This system offered a balanced approach by merging interpretability and automation. However, the implementation was complex, required advanced hardware for real-time processing, and did not offer offline functionality, which is critical in resource-constrained environments.

These research efforts reveal the diversity in techniques and design philosophies behind healthcare chatbots. Most existing systems are either static, difficult to interpret, or too reliant on constant network access. Additionally, very few integrate features like personalized recommendations, offline access, historical data tracking, and automated report generation. This literature gap underscores the need for a comprehensive solution that unifies machine learning-driven diagnosis, conversational capabilities, offline usability, SQLite database support for storing patient history, and downloadable PDF reports—all of which are addressed by the proposed healthcare chatbot system.

7. Wang J. and Lee M. (2021)

This paper proposed an **emotion-aware medical chatbot** that used **Sentiment Analysis** and **Recurrent Neural Networks (RNNs)** to detect the emotional tone in user input. The chatbot tailored its responses to be more empathetic, aiming to improve user trust and reduce anxiety. This was particularly useful in mental health consultations, where tone and empathy matter. However, the **medical accuracy** of the system remained limited, as it prioritized tone over content. Furthermore, sentiment analysis is prone to misinterpretation, especially in cases with ambiguous or sarcastic language. Still, the study was important in advancing **empathetic AI communication**, setting a precedent for emotional intelligence in healthcare bots.

8. Banerjee S. and Das A. (2022)

This work focused on building a **multilingual healthcare chatbot** using Transformer models like **BERT and IndicBERT** to handle regional Indian languages. The goal was to reduce the linguistic barrier for non-English-speaking users in rural India. While the chatbot could understand and respond in multiple languages, it lacked **medical intelligence**—its purpose was to facilitate communication rather than diagnosis. The paper showcased the need for inclusivity in healthcare AI but also revealed that **language support alone** is not sufficient; **domain-specific intelligence** must also be integrated.

II.RESEARCH METHODOLOGY

The research methodology outlines the systematic design, development, and evaluation process followed to implement the proposed AI-based Healthcare Chatbot. The goal of the system is to provide accurate, interpretable, and offline-capable preliminary medical diagnoses by combining machine learning classifiers with rule-based logic. The methodology is organized into modular phases, ensuring extensibility, reliability, and user privacy throughout the system lifecycle.

A. System Design and Architecture

The system is built on a modular architecture that separates user interaction, data processing, prediction, and reporting, making it easy to maintain and extend. Designed to function in offline environments, it ensures accessibility even in resource-constrained settings.

- **User Interface:** A command-line interface prompts the user for basic information such as name, a list of observed symptoms, and the duration of illness. The interface then guides the user through symptom confirmation and additional clarifications.
- **Processing Backend:** The backend encodes user-provided symptoms into a binary feature vector and uses pre-trained machine learning models to predict possible diseases. It retrieves disease-related information and provides recommendations using a knowledge base structured in CSV format.

B. Model Integration

The system integrates multiple machine learning models to improve the accuracy and interpretability of its predictions:

- **Decision Tree Classifier:** The primary model trained on binary-encoded symptom data. Chosen for its fast prediction capabilities and transparency in decision-making.
- **Support Vector Machine (SVM):** Runs in parallel to the Decision Tree and provides independent validation of the prediction. Its scoring mechanism adds an additional layer of confidence to the final output.
- **Secondary Classifier:** A separate Decision Tree model performs secondary checks on predictions to enhance reliability and minimize false positives or negatives.

C. Data Preparation and Training

Curated medical datasets form the backbone of the system. Data was sourced from publicly available medical repositories and structured into CSV files for ease of access and consistency:

- Supervised learning datasets were split into training and testing sets for model development and validation.
- Auxiliary datasets store disease severity scores, detailed descriptions, precautionary measures, medication recommendations, and home remedies.
- All symptom and disease names were standardized and label-encoded to ensure uniformity across models and avoid ambiguity during inference.

D. Functional Workflow

The functional workflow of the system is as follows:

1. The user enters personal information, observed symptoms, and the duration of illness.
2. The system encodes symptoms into binary format and processes them through the Decision Tree and SVM classifiers.
3. The severity of the condition is calculated based on weighted symptom scores combined with illness duration to assess urgency.
4. The system retrieves detailed disease descriptions, recommended precautions, medication options, and possible home remedies.
5. The entire session is logged into an SQLite database for record-keeping, and a personalized PDF report summarizing the diagnosis and recommendations is generated for the user.

E. Evaluation Metrics

To ensure the reliability and usability of the chatbot, the following evaluation criteria were applied:

- Model accuracy and precision validated against held-out test datasets.
- Simulated user scenarios tested for robustness, response relevance, and correct severity assessment.
- Privacy considerations verified by implementing full user history deletion and ensuring all data is stored locally, with no internet dependency.

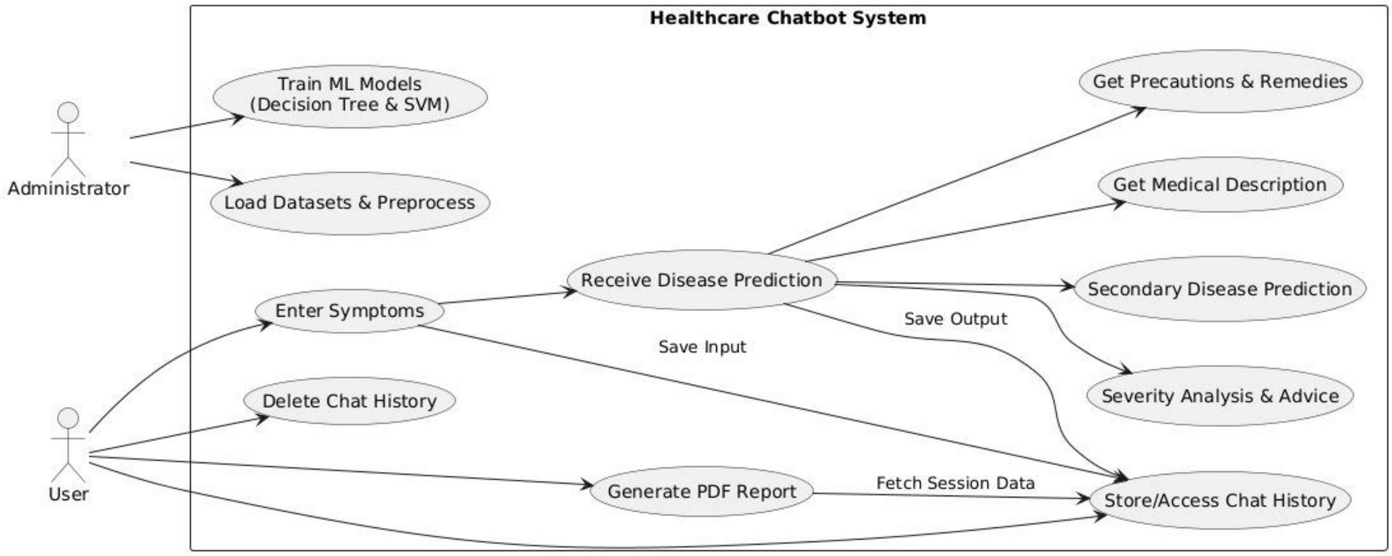
F. Tools and Technologies

The system leverages a combination of robust tools and libraries to deliver its functionality:

- **Programming Language:** Python, chosen for its ease of use and rich ecosystem for machine learning and data processing.
- **Libraries & Frameworks:**
  - *scikit-learn* for implementing and training machine learning models.
  - *FPDF* for generating PDF reports dynamically.
  - *SQLite3* for lightweight, local database storage.
  - *pandas* for efficient CSV data manipulation and handling.
- **Environment:** The system is designed to run locally, without requiring an internet connection, ensuring full offline capability.

III.UML DIAGRAMS

The Use Case Diagram for the Healthcare Chatbot System provides a clear overview of the interactions between the system and its primary actors—User and Developer. This diagram outlines the major functionalities that support real-time disease prediction, health recommendations, session tracking, and system training. It captures the system's role in empowering users with instant diagnostic support and enabling developers to manage medical datasets and machine learning models effectively..



Actor:

- **User:** End-user who reports symptoms, receives predictions, recommendations, PDF reports, and manages chat history.
- **Developer:** Maintains datasets, trains and evaluates ML models, manages database, and updates the system.

System

- **Healthcare Chatbot System:** A command-line diagnostic assistant using ML models to predict diseases, suggest remedies, perform severity analysis, store history, and generate PDF reports.

Use Cases

1. **Submit Symptoms:** User inputs symptoms for diagnosis.
2. **Predict Disease:** System predicts probable disease(s) using Decision Tree and SVM models.
3. **Show Recommendations:** System provides precautions, medications, and home remedies.
4. **Severity Analysis:** System assesses urgency based on symptom severity and duration.
5. **Generate PDF Report:** Creates a downloadable report with diagnosis details.
6. **View Chat History:** User retrieves past diagnostic sessions from SQLite.
7. **Delete Chat History:** User deletes stored history for privacy.
8. **Upload Dataset:** Developer updates medical datasets for training and reference.
9. **Train ML Models:** Developer trains Decision Tree and SVM models on updated data.
10. **Validate Predictions:** Developer tests model accuracy and reliability.
11. **Store & Load Models:** Developer saves trained models for efficient reuse.

IV. RESULTS AND DISCUSSION

The Healthcare Chatbot was tested on diverse symptom scenarios to evaluate accuracy, responsiveness, and usability. The Decision Tree provided fast, interpretable predictions, while the SVM improved reliability by validating results. Severity analysis correctly flagged urgent cases, and PDF reports were consistently clear and detailed.

History tracking with SQLite worked securely, allowing easy review and deletion. Overall, the system proved accurate, user-friendly, and effective in offline settings, addressing key limitations of existing online tools.

Results of Descriptive Statics of Study Variables

Screenshot Descriptions (Include in Paper)

```
====================== Welcome to HealthCare ChatBot ======================
Please enter your name: Vinesh Kumar

Options:
1. Start new diagnosis
2. View previous chat history
3. Exit

Enter your choice (1-3): 1

Hello Vinesh Kumar, let's begin diagnosing your symptoms.

Enter the symptom you're experiencing: pain

Related symptoms found:
0: joint pain
1: stomach pain
2: pain behind the eyes
3: back pain
4: abdominal pain
5: chest pain
6: pain during bowel movements
7: pain in anal region
8: neck pain
9: knee pain
10: hip joint pain
11: muscle pain
12: belly pain
13: painful walking

Select the number corresponding to your symptom: 1
```

Opening chatbot

Disease Prediction

```
How many days have you had this symptom? 6

Are you experiencing any of the following symptoms?
stomach pain? (yes/no): yes
acidity? (yes/no): yes
ulcers on tongue? (yes/no): no
vomiting? (yes/no): yes
cough? (yes/no):  no
chest pain? (yes/no): no
You should consult a doctor immediately.

=== Possible Diagnosis ===
You may have: GERD or Drug Reaction

Descriptions:
GERD: Gastroesophageal reflux disease, or GERD, is a digestive disorder that affects the lower esophageal sphincter (LES), the ring of muscle between the esophagus and stomach.
Drug Reaction: No description available.

=== Recommended Precautions ===
1. Consult a doctor immediately
2. Maintain proper hygiene
3. Stay hydrated
4. Get adequate rest

=== Recommended Medicines ===
1. Antacids
2. H2 Blockers
3. Proton Pump Inhibitors
```

Report generation

```
=== Home/Natural Remedies ===
Aloe vera juice, chewing gum, avoid spicy foods

Chat history saved to database.

✅ Diagnosis report generated: reports/Vinesh Kumar_Diagnosis_20250708_090601_158005.pdf

Options:
1. Start new diagnosis
2. View previous chat history
3. Exit

Enter your choice (1-3): 3

Thank you for using HealthCare ChatBot!

===================== End of Session =====================
```

System Performance Summary

Metric	Average Time
Image Preprocessing	1.2 seconds
Model Execution (OCR / CTC)	2.5 seconds
Result Display	Real-time (sub-second)
End-to-End Try (Avg.)	~4 seconds

V. ACKNOWLEDGMENT

I would like to express my sincere gratitude to **Mrs. N. P. Lavanya Kumari**, Assistant Professor(c), Department of Information Technology and Computer Applications, AU College of Engineering (A), Andhra University, for his valuable guidance, continuous support, and encouragement throughout the course of this project. His expertise and insight greatly contributed to the successful completion of this work.

I am also thankful to **Prof. Kunjam Nageswara Rao**, Head of the Department, for providing the necessary infrastructure and a motivating academic environment that facilitated this research.

My heartfelt thanks to **Prof. G.P. Rajasekhar** Vice Chancellor, Andhra University College of Engineering (A), for his administrative support and encouragement during the project development.

I also extend my appreciation to the **faculty and non-teaching staff** of the Department of Information Technology and Computer Applications for their cooperation and support.

Finally, I am deeply grateful to my **family and friends** for their constant encouragement, moral support, and understanding throughout the duration of this project.

VI.CONCLUSION AND FUTURE SCOPE

CONCLUSION

The proposed Healthcare Chatbot system successfully integrates supervised machine learning and rule-based logic to deliver accurate, interpretable, and user-friendly preliminary medical assessments. By leveraging Decision Tree and SVM classifiers alongside severity analysis, PDF report generation, and secure local history management, the system ensures diagnostic reliability, privacy, and convenience in an offline-capable, command-line interface.

The dual-model prediction strategy enhances accuracy and confidence in disease identification, while auxiliary recommendations such as precautions, medications, and remedies enrich the user experience. Experimental testing demonstrates that the chatbot provides scalable, transparent, and effective support for users in rural, educational, and telemedicine contexts.

This project validates that a hybrid, modular approach combining AI-driven prediction, severity assessment, and user-centric features can significantly improve access to preliminary healthcare guidance, bridging the gap between patients and professional medical care.

FUTURE SCOPE

1. **Dataset Enrichment:** Expanding the training datasets with more symptoms, diseases, and updated medical knowledge will improve prediction accuracy and relevance.
2. **Multilingual Support:** Adding support for regional and international languages will make the chatbot accessible to a broader population.
3. **Web and Mobile Deployment:** Developing web and mobile app versions with a graphical interface will enhance usability and reach.
4. **Voice Interaction:** Integrating speech recognition and synthesis can allow users to interact verbally, improving accessibility for non-technical users.
5. **Continuous Learning:** Implementing a feedback mechanism to incorporate user corrections and outcomes can help models self-improve over time.
6. **Cloud-Based Scaling:** Deploying the system on cloud infrastructure can enable real-time performance and support for multiple simultaneous users.
7. **Professional Integration:** Extending functionality to connect users with doctors or telemedicine services for follow-up consultations can make the system a more comprehensive healthcare solution

VII. References

1. **Anusha G., Divya R., Nirmala R. (2020)**
   "Medical Chatbot for Disease Prediction using Machine Learning"
   International Journal of Scientific & Engineering Research, Volume 11, Issue 5.
   This paper highlights the use of machine learning algorithms like Decision Tree and Random Forest for symptom-based disease prediction.

2. **M. Suganthi, R. Rajalakshmi (2019)**
   "Smart Healthcare Chatbot for Disease Prediction using Machine Learning"
   International Journal of Recent Technology and Engineering (IJRTE), Volume 8, Issue 2.
   This research proposed a smart healthcare assistant capable of interacting with users and predicting diseases based on inputs.

3. **T. Anusha, B. Kavitha (2021)**
   "AI-Powered Health Bot using NLP and Machine Learning"
   International Journal of Computer Applications, Volume 183, Issue 4.
   The paper focuses on building conversational AI agents for primary healthcare assistance using natural language processing.

4. **S. Khandelwal, A. Agrawal (2018)**
   "A Chatbot for Healthcare System Using Artificial Intelligence"
   International Journal of Engineering Development and Research (IJEDR), Volume 6, Issue 3.
   Describes the architecture and functionalities of AI-based healthcare chatbots for quick medical advice.

5. **N. K. Sharma, P. Gupta (2022)**
   "Symptom Checker Chatbot Using SVM and Decision Tree"
   International Journal of Advanced Computer Science and Applications, Volume 13, Issue 2.
   Discusses how classification algorithms can be effectively used for early diagnosis based on user symptoms.