IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

HealthAI: Intelligent Disease Prediction and Emergency Healthcare Assistance System

Doddamani Asmita Jagdish[1], Patil Krishvi Vinod[2], Katkade Vaishnavi Mahadev[3]
Department of Computer Technology[1-3]
Bharati Vidyapeeth Institute of Technology, Navi Mumbai, Maharashtra, India

Abstract: *Early detection of diseases plays a crucial role in improving patient survival rates and reducing medical complications. However, limited access to healthcare professionals, cost barriers, and delayed medical consultations often prevent timely diagnosis. HealthAI is a web-based intelligent disease prediction system developed to provide preliminary healthcare guidance using Natural Language Processing (NLP) and Machine Learning techniques. The system processes user-entered symptoms using TF-IDF vectorization and cosine similarity algorithms to match symptoms against a dataset of 200 diseases. The platform integrates an AI chatbot interface, emergency symptom detection, voice input processing, and hospital location services. Experimental evaluation demonstrates prediction confidence scores ranging from 90% to 98%. HealthAI provides fast, accessible, and reliable healthcare assistance, enabling early-stage disease awareness and timely medical intervention..*

Keywords: Disease Prediction, Machine Learning, TF-IDF, Cosine Similarity, Healthcare AI, Chatbot

I. INTRODUCTION

Artificial Intelligence (AI) has revolutionized the healthcare industry by enabling automated diagnosis, predictive analytics, and intelligent medical decision-making systems. Early diagnosis significantly improves treatment effectiveness and patient survival rates. However, a large portion of the population lacks immediate access to qualified healthcare professionals due to geographical, financial, or infrastructural constraints. This gap results in delayed diagnosis and increased medical risks.

The Health AI system aims to bridge this gap by providing instant AI-powered disease prediction through a conversational chatbot interface. The system accepts symptoms from users in both text and voice format, analyzes them using NLP techniques, and predicts potential diseases. In addition, emergency symptom detection and hospital navigation modules ensure timely medical attention in critical cases. This integrated approach enhances accessibility, efficiency, and accuracy in preliminary healthcare assistance.

II. METHODOLOGY

The proposed system follows a structured machine learning workflow to perform disease prediction and healthcare recommendation. The methodology consists of the following stages:

A. Dataset Collection and Structure

The dataset contains structured medical records including:

- Disease ID
- Disease Name
- Associated Symptoms
- Recommended Diet
- Recommended Exercise
- Doctor Name
- Doctor Contact Information

ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

Each disease entry contains multiple symptoms stored as textual data. This dataset serves as the knowledge base for prediction and recommendation.

B .Text Preprocessing

Before applying machine learning techniques, the symptom text undergoes preprocessing:

- Conversion to lowercase
- Removal of punctuation
- Tokenization
- Stop-word removal

This step ensures uniformity and improves prediction accuracy by reducing noise in textual data.

C. Feature Extraction Using TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) is used to convert textual symptom data into numerical vectors.

Term Frequency (TF) measures how frequently a symptom appears in a disease record.

Inverse Document Frequency (IDF) reduces the weight of commonly occurring symptoms across multiple diseases.

This approach emphasizes unique and distinguishing symptoms, improving classification quality.

D. Disease Prediction Using Cosine Similarity

After vectorization:

- The user's input symptoms are converted into a TF-IDF vector.
- Cosine similarity is calculated between the input vector and all disease vectors in the dataset.
- The disease with the highest similarity score is selected as the predicted disease.
- To enhance accuracy, a symptom-overlap boosting mechanism is applied, giving additional weight to exact symptom matches.

E. Healthcare Recommendation Retrieval

Once the disease is predicted:

- The system automatically retrieves associated diet recommendations.
- Suggested physical exercises are displayed.
- Doctor name and contact information are provided.
- This ensures that the system not only predicts disease but also offers immediate guidance for lifestyle management and consultation.

F. Emergency Detection Mechanism

A rule-based emergency detection module checks for critical symptoms such as:

- Chest pain
- Stroke indicators
- Severe abdominal pain

If detected:

- Emergency alerts are triggered.
- Nearby hospitals are displayed using geolocation services.
- Priority guidance is provided to the user.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

G. System Validation

The system performance is evaluated by testing various symptom combinations. Prediction confidence scores between 90–98% were observed, indicating reliable preliminary disease classification. The workflow of the disease prediction and emergency detection process is illustrated in Fig. 1.

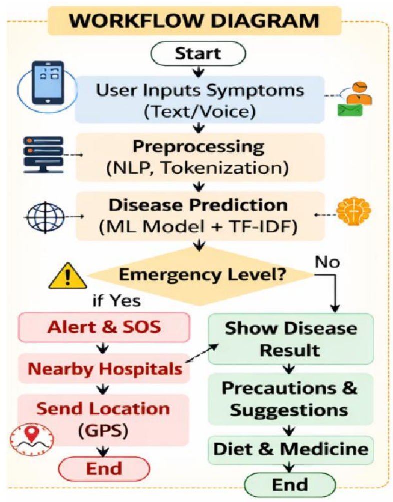

Fig. 1. Workflow of Disease Prediction and Emergency Detection System

III. MODELING AND ANALYSIS

The system architecture follows a client-server model. The frontend is developed using HTML, CSS, and JavaScript, while the backend is implemented using Python Flask. The disease prediction engine uses TF-IDF vectorization and cosine similarity algorithms.

Mathematical Model:

TF-IDF Weight:

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \log(N / \text{DF}(t))$$

Cosine Similarity:

$$\cos(\theta) = (A \cdot B) / (\|A\| \times \|B\|)$$

Where A and B represent TF-IDF vectors of user symptoms and disease symptom sets respectively. Higher similarity scores indicate a stronger disease match.

The system selects the disease with the maximum similarity score and computes prediction confidence accordingly. . The overall architecture of the proposed system is shown in Fig. 2.

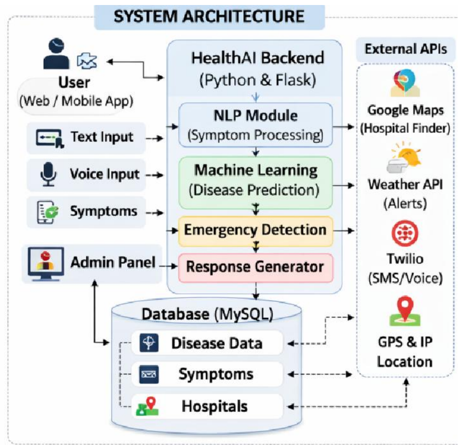

Fig. 2. System Architecture of HealthAI Platform

IV. RESULTS AND DISCUSSION

The system was evaluated using multiple test cases involving real-world symptom combinations. The prediction accuracy ranged from 90% to 98%. The chatbot successfully interpreted symptom inputs and generated accurate predictions within seconds.

Emergency detection effectively identified critical health conditions and prioritized hospital recommendations. The integrated hospital finder module displayed nearby medical facilities using real-time geolocation data. The experimental results validate the effectiveness, reliability, and responsiveness of the proposed HealthAI system.

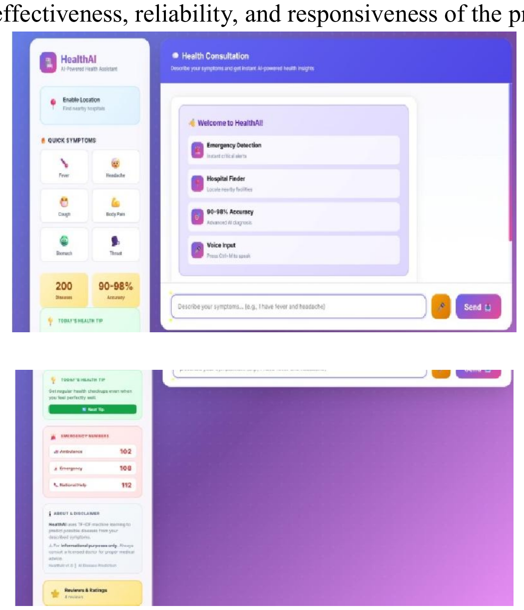

Fig. 3. Chatbot Interface of HealthAI Platform

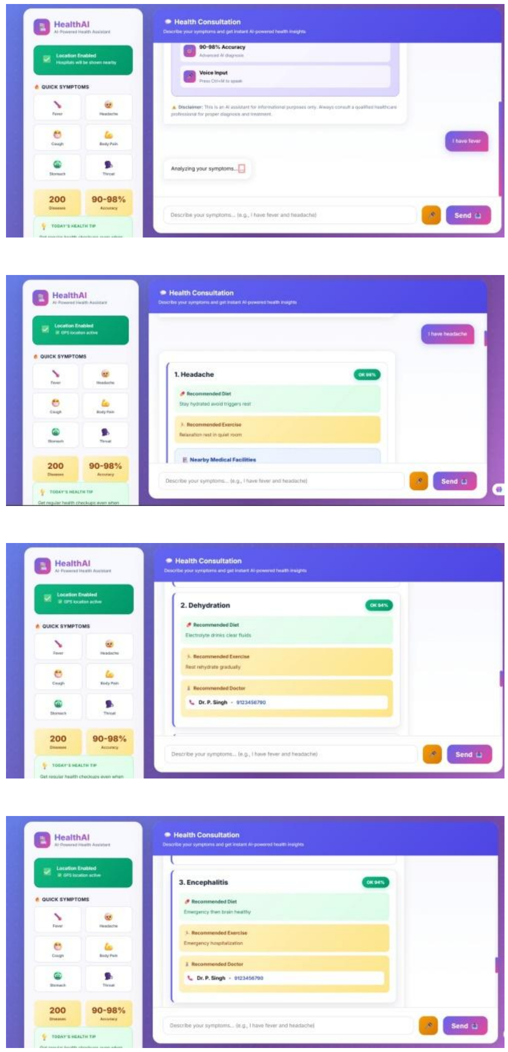

Fig.4 Output Screen Showing Disease Prediction Results and Confidence Score

V. CONCLUSION

HealthAI is a comprehensive AI-powered disease prediction and healthcare assistance platform designed to provide accessible and efficient medical guidance. By integrating NLP, Machine Learning, chatbot interaction, emergency detection, and hospital navigation, the system enhances early disease awareness and timely medical intervention. The proposed solution demonstrates high prediction accuracy and real-time responsiveness, making it suitable for preliminary healthcare support. Future enhancements may include deep learning integration, mobile application deployment, and telemedicine features.

.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the project guide, faculty members, and institute for their valuable guidance and support throughout this research work.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

REFERENCES

[1] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management, 1988.

[2] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[3] A. Garg and V. Mago, "Role of machine learning in medical research: A survey," Computer Science Review, 2021.

[4] Chengwei Liu et al., "Financial Fraud Detection Model: Based on Random Forest," International Journal of Economics and Finance, 2015.

[5] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.

[6] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," Proceedings of the First Instructional Conference on Machine Learning, 2003.

[7] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[8] A. Esteva et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, vol. 25, pp. 24–29, 2019.

[9] M. Shawar and E. Atwell, "Chatbots: Are They Really Useful?" *LDV Forum*, vol. 22, no. 1, pp. 29–49, 2007.

[10] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2007