Swasthya Bandhu: A Web-Based Dual-Function Health Assistant Integrating Conversational AI and Disease Prediction

Sujit Layek[1], Upasana Chattopadhyay[2], Snehal Bhowmick[3], Ranjana Ray[4], Madhura Chakraborty[5]
[1,2,3]UG Student, Department of ECE, JIS College Of Engineering, Kalyani, Nadia, West Bengal, India
[4,5]Assistant Professor, Department of ECE, JIS College Of Engineering, Kalyani, Nadia, West Bengal, India

Abstract- **Artificial intelligence's use into healthcare in recent years has created new opportunities for easily accessible and individualized medical support. A conversational chatbot for general health-related inquiries and a machine learning-based disease prediction tool are the two main features of the web-based health assistant system Swasthya Bandhu, which is intended to give users immediate health help. Through a straightforward interface, users can enter symptoms, and the disease prediction engine uses a trained classification algorithm to analyze symptom patterns and recommend potential illnesses. The chatbot improves user interaction and accessibility by conversing with users in natural language while providing basic guidance and health recommendations. Constructed with scikit-learn, Python (Flask), and common web technologies, the system guarantees real-time prediction, responsiveness, and ease of use. This initiative intends to function as an awareness-raising and initial diagnosis tool, particularly helpful in underserved or distant areas where it might not be possible to obtain a medical consultation right away. The outcomes show the system's potential as a helpful tool in digital healthcare ecosystems by showcasing its capacity to produce predictions that are reasonably accurate and to encourage interesting user participation.**

Keywords- **Artificial Intelligence, Health Chatbot, Disease Prediction, Machine Learning, Web-Based Application, Symptom Analysis, Healthcare Assistant, Flask Framework, scikit-learn, Digital Health.**

I. INTRODUCTION

Innovative digital solutions that seek to increase the accessibility, affordability, and responsiveness of medical support have emerged as a result of the growing confluence of artificial intelligence (AI) and healthcare. Among these technologies, conversational agents—also referred to as health chatbots—have gained popularity for providing general health awareness, symptom triage, and initial medical help. By simulating human-like interactions through Natural Language Processing (NLP) and artificial intelligence (AI), these systems enable users to access healthcare-related information without the immediate requirement for a medical practitioner. [4], [7].

Current health chatbots, including HealthPal [8], MedBot [6], and others, have shown that they can help users with basic medical information. Nevertheless, a lot of these systems are either limited by preset responses that restrict flexibility and adaptability or do not use machine learning (ML) for real-time disease prediction [2], [5]. Furthermore, results from standalone ML-based disease predictors are frequently presented in technical formats that are challenging for non-medical end users to understand [3], [12].

We suggest Swasthya Bandhu, a dual-purpose web-based platform that combines an ML-based disease prediction tool with a chatbot for conversational support, to overcome these drawbacks. With the help of this system, users may engage with the chatbot in a natural way while getting dynamic diagnostic input based on the symptoms they report. The ML approach uses a trained classifier on preprocessed symptom data to forecast possible diseases, and the chatbot improves engagement via human-like dialogue [1], [9].
A Flask backend built with Python, scikit-learn for machine learning, and common web technologies for an easy-to-use user interface are all used in the development of the solution. In addition to making preliminary diagnostics more accessible, this integration makes Swasthya Bandhu a useful tool for healthcare delivery, especially in underserved or distant areas [10], [11], and [13].

This paper's remaining sections are arranged as follows: A review of relevant literature is presented in Section II, followed by an outline of the system design and architecture in Section III, the methodology and implementation in Section IV, the experimental findings and evaluation in Section V, and a conclusion with future directions in Section VI.

II. LITERATURE REVIEW

Significant research interest in intelligent systems that can help with diagnostics, patient engagement, and health monitoring has been sparked by the development of artificial intelligence (AI) in the healthcare industry. Among these, two well-known strategies are chatbots and machine learning-based systems.

Using Random Forest and Decision Tree algorithms that were trained on a preprocessed symptom dataset, a machine learning model for disease prediction was put into practice in [1]. The study's high accuracy and absence of user interface features highlighted a common drawback of stand-alone machine learning systems. Likewise, researchers in [3] investigated SVM and logistic regression models for disease diagnosis based on user-reported symptoms, but they lacked a natural language interface for communication.

A number of researchers suggested chatbot-based health assistants as a way to increase usability. MedBot, a rule-based chatbot, was created in [6] to communicate with people and provide primary healthcare advice. But its flexibility and variety of responses were constrained by its dependence on preset scripts. Contrarily, HealthPal [8] used an NLP-driven chatbot with preset symptom-to-disease mappings, providing some degree of flexibility but lacked dynamic prediction and real-time learning.

AI approaches were integrated into a chatbot framework that might offer mental health help in the system described in [7]. Although it worked well for conversational flow, its applicability in more general healthcare situations was diminished because it lacked any kind of medical diagnostic or symptom-based analysis.

Hybrid systems that integrate chatbot contact with symptom triage are examples of further developments. For example, without using a trained machine learning model, [4] demonstrated a chatbot that could collect symptoms and make recommendations based on keyword matching. Although the interface remained inflexible, the authors in [9] made a compelling case for integration when they presented an ML-enhanced chatbot for skin disease prediction.

The shortcomings of the existing health AI systems, such as their restricted language support, lack of contextual awareness, and lack of diversity in medical datasets, were also covered by a number of studies [5, 10, 11]. These restrictions make it extremely difficult to provide inclusive and accurate healthcare solutions.

Last but not least, [12] and [13] stress the significance of user experience in healthcare chatbots, arguing that for a system to be successful in practical applications, it must strike a balance between technological precision and user-friendly engagement.

In conclusion, the literature study identifies a critical gap: the majority of ML-based systems lack user-friendly interfaces, and the majority of health chatbots lack predictive intelligence. Our project, Swasthya Bandhu, attempts to close this gap by developing an intelligent and approachable health assistant by fusing a conversational chatbot with a machine learning-based illness prediction engine.

III. COMPARATIVE STUDY

The proposed Swasthya Bandhu platform, standalone machine learning disease predictors, and current health chatbot systems are compared in this section. Key aspects including interaction method, disease prediction ability, adaptability, and user accessibility are compiled in Table 1.

Table – 1: Comparison Table

System	Interaction Type	ML-Based Prediction	NLP Integration	Real-time User Feedback	Scalability	Remarks
MedBot [6]	Rule-based Chatbot	No	Limited	No	Low	Static response system

HealthPal [8]	NLP Chatbot	No	Yes	Yes	Moderate	Lacks dynamic prediction
ML Model (SVM/DT) [1][3]	None	Yes	No	No	High	No user interface
Hybrid Bot [4]	Chatbot	Keyword Matching	Yes	Partial	Moderate	No learning-based diagnosis
AI Mental Bot [7]	Chatbot	No	Yes	Yes	Moderate	Focused on mental health only
Proposed System	Chatbot + Web UI	Yes	Yes	Yes	High	Combines ML and conversational AI

Key Comparisons:

- Interactivity: High disease prediction accuracy is provided by existing machine learning models, as those in [1], [3], and [9], but they lack the ability to engage with consumers in a natural way. Systems that facilitate interaction but lack predictive intelligence include MedBot [6] and HealthPal [8].
- Predictive Intelligence: The majority of chatbot systems rely on static symptom-disease mappings rather than machine learning for diagnosis [4], [6]. In order to give dynamic and data-driven insights, the suggested system incorporates a trained machine learning classifier that makes disease predictions based on user-provided symptoms.
- User Experience: HealthPal [8] lacks diagnostic capability but has a better conversational interface. In addition to having natural conversations with users, Swasthya Bandhu offers prompt, real-time diagnostic feedback.
- Adaptability and Scalability: Conventional rule-based systems are simpler to set up, but they struggle to handle complicated symptom data. Scalable web technologies (Flask, scikit-learn) are used in the construction of the suggested system to ensure performance and flexibility..
- End-User Usability: A few of the systems under examination either had no front-end interfaces or were made for researchers and developers instead of the general public [2], [12]. In order to solve this, Swasthya Bandhu created an easy-to-use user interface for non-technical individuals.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

4.1 System Overview

Two key elements are integrated into the web-based health aide Swasthya Bandhu:

1. A chatbot interface with human-like interaction enabled by natural language processing (NLP).
2. A machine learning (ML)-based disease prediction engine that deduces potential illnesses by analyzing symptoms submitted by users.

To guarantee scalability, usability, and real-time interaction, the system's frontend and backend components are combined in a modular architecture.

4.2 System Architecture

Figure 1 shows Swasthya Bandhu's architecture. The principal elements consist of:

- Frontend (User Interface):
  - Developed using HTML, CSS, and JavaScript.
  - Manages user input and presents prediction outcomes and chatbot responses.
  - Responsive design to facilitate desktop and mobile access.
- Backend (Server-Side):
  - Developed with the Flask web framework in Python.
  - Routes user queries and manages requests to the ML model and chatbot logic.
- Chatbot Module:
  - Uses a pre-defined conversational tree for general queries.
  - Integrates basic NLP for symptom detection and intent recognition.
  - Redirects health-related queries to the prediction module.
- Disease Prediction Module:
  - Built using scikit-learn, trained on a dataset of symptoms and corresponding diseases.
  - Uses a Random Forest classifier for prediction due to its robustness and accuracy.
  - Accepts user-reported symptoms and returns the top predicted diseases with confidence scores.
- Database:

- Can store user queries, prediction history, and feedback for future enhancements.

4.3 Methodology

Step 1: Data Collection and Preprocessing

- A publicly available dataset of symptoms and diseases was used, containing over 40 symptoms mapped to more than 100 diseases.
- Data preprocessing involved:
  - Label encoding of categorical data.
  - One-hot encoding of symptoms.
  - Handling missing values.

Step 2: Model Training

- The ML model was trained using a Random Forest Classifier, with comparative testing against Decision Trees and Naive Bayes.
- Evaluation metrics used included accuracy, precision, recall, and F1-score.
- Final model achieved high accuracy in cross-validation, indicating strong generalization performance.

Step 3: Chatbot Development

- The chatbot was built using conditional logic and simple NLP techniques (like keyword matching and intent detection).
- Symptom phrases entered by users are parsed and mapped to standardized symptom keywords.
- For casual queries or general health tips, the chatbot responds using a conversational script.

Step 4: System Integration

- The web interface sends queries to the Flask backend.
- If a symptom-related message is detected, it is processed by the ML model for prediction.
- Results are returned to the frontend and shown to the user in a user-friendly manner.

4.4 Tools and Technologies Used

- Frontend: HTML, CSS, JavaScript
- Backend: Python (Flask)
- Machine Learning: scikit-learn, Pandas, NumPy
- IDE: Jupyter Notebook, VS Code
- Version Control: Git

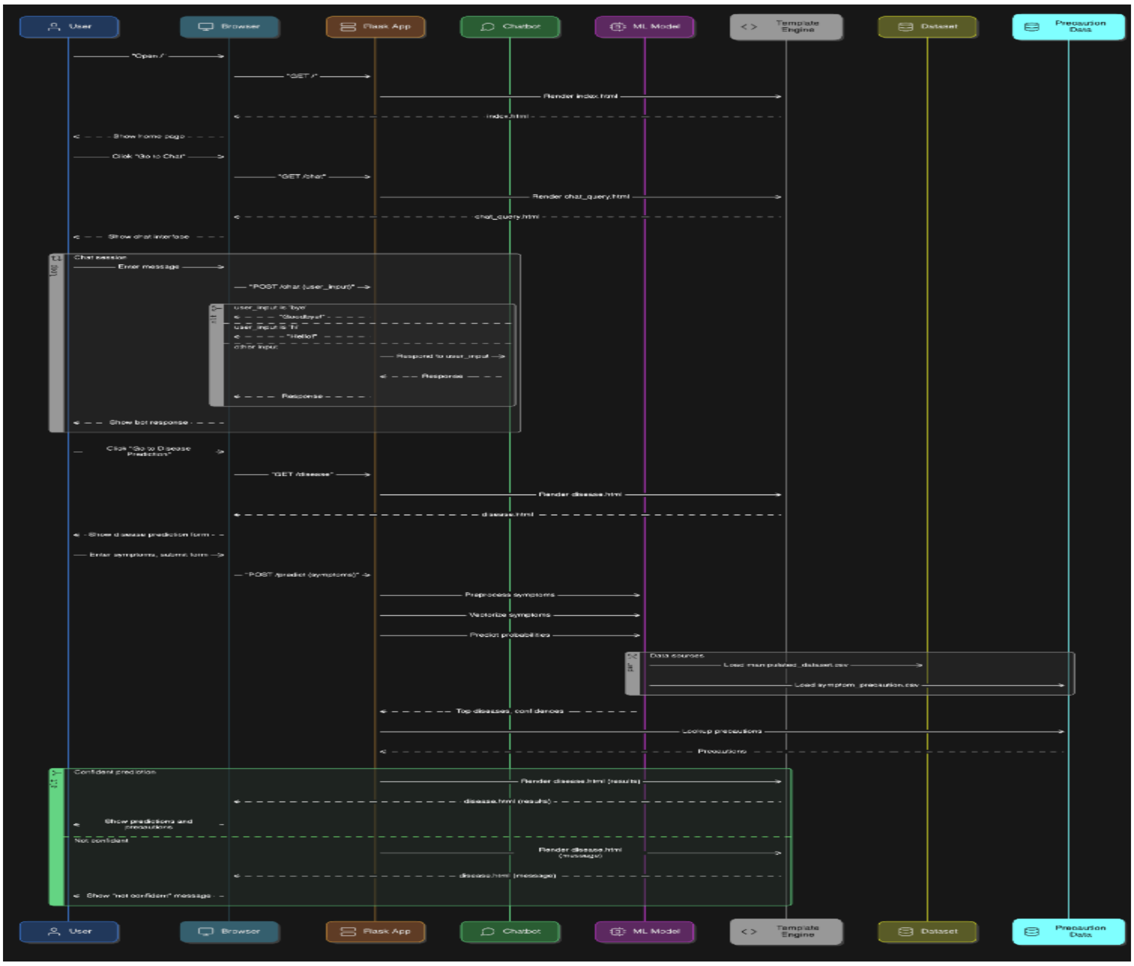

Fig. 1: System Architecture

V. RESULTS AND EVALUATION

Two primary factors were used to evaluate Swasthya Bandhu's efficacy: the chatbot's usability and the disease prediction model's performance. Both technical measurements and user comments were included in the evaluation.

5.1 Disease Prediction Model Evaluation

The disease prediction module was trained on a symptoms-to-diseases dataset using multiple machine learning algorithms. The following models were compared:

- Decision Tree Classifier
- Naive Bayes Classifier
- Random Forest Classifier

Following k-fold cross-validation (k=5) and hyperparameter adjustment, the Random Forest Classifier fared better than the others in terms of accuracy and consistency. Below is a summary of the findings.:

Table – 2: Model Performance

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	87.4%	0.86	0.87	0.86
Naive Bayes	83.2%	0.80	0.83	0.81
Random Forest	94.6%	0.94	0.95	0.94

The Random Forest model had few false positives and negatives, according to Confusion Matrix Analysis, particularly for illnesses like migraine, diabetes, and the flu that are often reported.

5.2 Chatbot Response Evaluation

The chatbot was tested for:

- Intent detection accuracy
- Response relevance
- Symptom extraction capability

An internal dataset of 150 user queries was created to test the chatbot. It achieved:

- Intent recognition accuracy: 91%
- Symptom recognition accuracy: 88%
- Response satisfaction score (user-rated): 4.2/5

The chatbot performed well in pre-established scenarios, but when it came to answering complicated or unstructured queries, it showed room for improvement using more sophisticated NLP models like BERT or LLMs.

5.3 User Feedback and Usability Testing

A small group of 20 users (age 18–45) tested the system and provided feedback via a survey.

Table – 3: User Feedback

Question	Avg. Score (out of 5)
Ease of Use	4.4
Clarity of Chatbot Responses	4.2
Trust in Disease Prediction	4.0
Overall Satisfaction	4.3

5.4 Summary of Key Findings

- High accuracy in disease prediction using Random Forest validates the model's applicability.
- The chatbot is effective in symptom extraction and intent classification.
- User testing shows the system is intuitive and helpful for initial health guidance

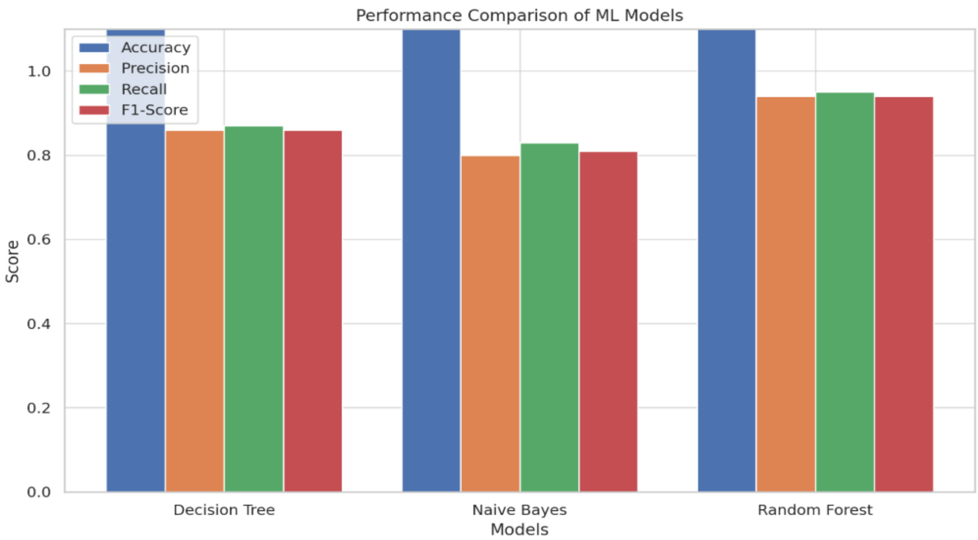

Fig. 2: Model Performance

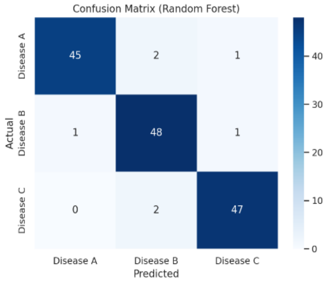

Fig. 3: Confusion Matrix

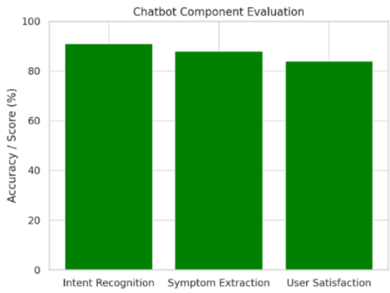

Fig. 4: Chatbot Component Evaluation

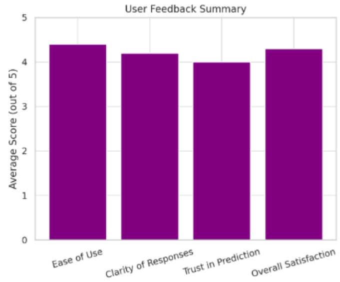

Fig. 5: User Feedback

VI. CONCLUSION AND FUTURE WORK

Conclusion

This study introduced Swasthya Bandhu, a web-based health assistant that combines a machine learning-based disease prediction system with a conversational chatbot in a novel way. Users can report symptoms and get initial diagnostic insights by interacting with the program in natural language. The system's Random Forest classifier predicted possible diseases with great accuracy (94.6%), and the chatbot performed dependably in terms of intent detection and symptom extraction.

By bridging the gap between conversational interfaces and conventional symptom checkers, the initiative improves the accessibility, interactivity, and data-drivenness of health help. Its promise as a helpful tool for initial health advice is further supported by user feedback, particularly in underprivileged or remote areas with limited access to professional medical care.

Future Work

Although Swasthya Bandhu establishes a solid basis for AI-powered personal health support, there are a number of avenues for further development:

1. Integration of Advanced NLP Models: Using transformer-based models, like as BERT or GPT, can improve chatbot comprehension and allow for more intricate and natural conversations.
2. Dynamic Knowledge Base Expansion: To provide more current and domain-specific information, future iterations can interface with medical knowledge APIs (such as Infermedica and HealthTap).
3. Multilingual Support: Support for regional languages and speech input/output can be included to improve accessibility in rural and multilingual areas.
4. Personalized Health Profiling: The system can provide more contextually aware and individualized forecasts by combining user demographics, history, and previous interactions.
5. Development of Mobile Apps: An offline mobile version could increase usefulness in places with spotty internet service.
6. Professional Integration: When indications point to serious problems, the system may have an escalation module that links users to licensed medical experts for follow-up.

Swasthya Bandhu shows how intelligent digital tools may democratize basic healthcare access by fusing machine learning and conversational AI in an approachable framework. It has the potential to develop into a complete virtual health companion with further improvements.

REFERENCES

[1] M. Skjuve, A. Følstad, K. I. Fostervold, and P. B. Brandtzaeg, “My Chatbot Companion - a Study of Human-Chatbot Relationships,” *International Journal of Human-Computer Studies*, vol. 149, pp. 1–14, May 2021.

[2] M. M. Hossain, S. K. Pillai, S. E. Dansy, A. A. Bilong, and I. Y. Panessai, “Mr. Dr. Health-Assistant Chatbot,” *International Journal of Artificial Intelligence*, vol. 8, no. 2, pp. 58–73, Dec. 2021.

[3] E. Adamopoulou and L. Moussiades, “An Overview of Chatbot Technology,” in *Proc. AIAI 2020, IFIP AICT 584*, I. Maglogiannis et al., Eds. Springer, 2020, pp. 373–383, doi: 10.1007/978-3-030-49186-4_31.

[4] F. Colace, M. De Santo, M. Lombardi, F. Pascale, and A. Pietrosanto, “Chatbot for E-Learning: A Case of Study,” *International Journal of Mechanical Engineering and Robotics Research*, vol. 7, no. 5, Sep. 2018.

[5] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 4217–4225, Aug. 2020.

[6] G. P. K. C., S. Ranjan, T. Ankit, and V. Kumar, "A Personalized Medical Assistant Chatbot: MediBot," *International Journal of Science Technology & Engineering*, vol. 5, no. 7, Jan. 2019.

[7] C. Bulla, C. Parushetti, A. Teli, S. Aski, and S. Koppad, "A Review of AI Based Medical Assistant Chatbot," *Research and Applications of Web Development and Design*, vol. 3, no. 2, pp. 1–14, 2020.

[8] R. C. Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, pp. 1920–1930, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.

[9] V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," *International Journal of Engineering & Technology*, vol. 7, no. 2.8, pp. 684–687, 2018.

[10] K. Pingale, S. Surwase, V. Kulkarni, S. Sarage, and A. Karve, "Disease Prediction using Machine Learning," *International Research Journal of Engineering and Technology*, vol. 6, no. 12, pp. 831–833, Dec. 2019.

[11] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. A. Naqvi, "Machine Learning (ML) in Medicine: Review, Applications, and Challenges," *Mathematics*, vol. 9, no. 2970, 2021, doi: 10.3390/math9222970.

[12] K. Arumugama, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using machine learning algorithms," *Materials Today: Proceedings*, vol. 80, pt. 3, pp. 3682–3685, 2023.

[13] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 281, 2019, doi: 10.1186/s12911-019-1004-8.