

## SYMPTOM DETECTION CHATBOT USING NLP

Aryan Shishodia<sup>1</sup>, Shivam<sup>2</sup>, Tanish Sandhu<sup>3</sup>, Dr. Garima Srivastava<sup>4</sup>

<sup>1,2,3,4</sup>Netaji Subhas University of Technology, Delhi, India.

aryan.k.shishodia@gmail.com, shivam.22823@gmail.com, tanishsandhu10@gmail.com,  
garima.srivastava@nsut.ac.in

DOI: <https://www.doi.org/10.58257/IJPREMS37680>

### ABSTRACT

Healthcare plays a crucial role in starting and maintaining a good quality of life. However, consulting a doctor can often be challenging, especially when health issues arise unexpectedly. This research proposes the development of a healthcare chatbot leveraging Natural Language Processing (NLP), a branch of Artificial Intelligence, to assist in diagnosing diseases and offering preliminary information before a formal medical consultation.

The primary objective of this chatbot is to enhance access to medical knowledge and reduce healthcare expenses. By acting as a virtual medical assistant, the chatbot serves as a resource for patients, providing insights into their health conditions and empowering them to make informed decisions about their well-being.

For maximum utility, the chatbot must be capable of diagnosing a wide range of diseases and delivering accurate, relevant information. The system supports both text and voice-based interactions, enabling users to communicate in their preferred language. It analyzes the symptoms provided by the user to identify potential health conditions, recommends appropriate specialists, and offers dietary suggestions tailored to the diagnosed condition. Ultimately, this system aims to promote better health awareness and ensure users take appropriate preventive measures to safeguard their well-being.

### 1. INTRODUCTION

In today's fast-paced world, healthcare plays a vital role in our lives. However, many people, engrossed in their work, whether at home or in offices, and often consumed by the Internet, tend to neglect their health. This lack of attention may lead them to avoid visiting hospitals for minor health concerns, which could eventually escalate into more serious issues. A healthcare chatbot is designed to address such situations by responding to user queries in real-time. It analyzes user inputs, extracts key information, and provides relevant answers. Significant keywords are identified from the user's query, and the system matches them to generate appropriate responses. If a direct match is found, a precise answer is provided; otherwise, related suggestions are displayed.

The chatbot also aids in diagnosing potential diseases based on the symptoms shared by the user. Additionally, it provides information about relevant doctors for the identified health condition. By offering quick and accessible assistance, the chatbot helps reduce both the time and cost associated with healthcare, making it a practical solution for those unable to consult doctors immediately.

### 2. LITERATURE SURVEY

In conventional healthcare practices, individuals must often visit hospitals or consult doctors in person for routine check-ups or even minor health concerns. This approach can be both time-intensive and inconvenient. Moreover, many people lack comprehensive knowledge about the symptoms and treatments of various diseases, making professional medical advice essential. While telephonic consultations may seem like a viable alternative, they are typically impractical for addressing such needs effectively. To overcome these challenges, the integration of advanced technology, such as medical chatbots, offers a promising solution. These chatbots, powered by robust learning algorithms and exceptional problem-solving capabilities, have demonstrated their usefulness in assisting individuals with minor health issues. By utilizing natural language processing (NLP), medical chatbots can provide quick and accurate responses to health-related queries. Additionally, their user-friendly interfaces, often equipped with voice-to-text and text-to-voice functionalities, enable even inexperienced users to easily seek guidance regarding their health concerns [1]. The integration of TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity in chatbot development, as described in the paper [2] offers a promising approach to enhance the accuracy and relevance of chatbot responses. This method is particularly valuable in the context of medical chatbots, which have gained prominence as potential solutions to various healthcare challenges [2]. The application of Inverse Document Frequency (IDF) and cosine similarity enhances the chatbot's ability to ef-

ficiently navigate large-scale medical databases, ensuring the prompt retrieval of relevant information. These methodologies also contribute to improved contextual understanding by evaluating the frequency and significance of terms within user queries. This enables the chatbot to interpret the context more accurately, facilitating the generation of precise and customized responses tailored to the user's specific concerns [3]. This paper highlights that tokenization is

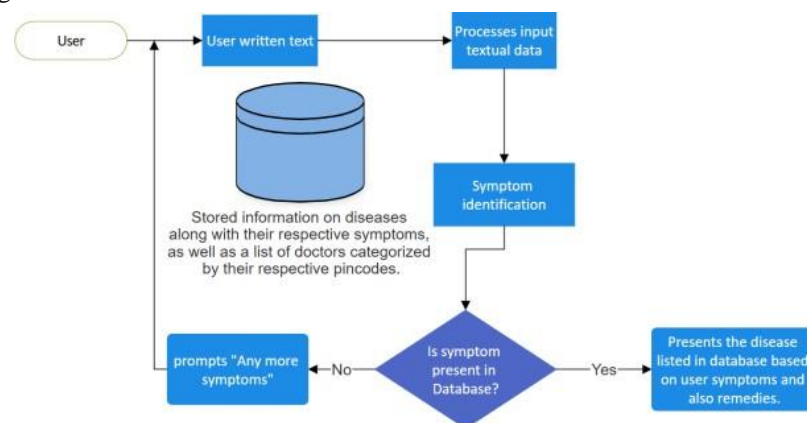
vital for medical chatbots, as it ensures accurate interpretation and processing of complex user inputs. However, it presents unique challenges, such as handling intricate medical terminology, abbreviations, and specialized language like drug names, medical conditions, and anatomical references, which often deviate from standard linguistic norms. Proper tokenization is essential for identifying and preserving critical medical concepts, delivering reliable information, ensuring patient safety, and enabling effective communication in healthcare applications [1]. Virtual assistants have proven highly beneficial for patients and various medical tasks. Chatbots like Florence, Molly, and Ada are particularly effective, using advanced algorithms to provide medical assistance anytime, anywhere, thus supporting human health. Designed with foundational artificial intelligence principles, these chatbots enable seamless dialogue with users, eliminating the need for physical hospital visits by offering assistance through apps, text messages, and instant messaging. While chatbots are already enhancing processes in industries such as automotive, robotics, and retail, their adoption in healthcare is accelerating, where they assist patients in performing numerous tasks efficiently [4]. The paper describes the development of an AIML-powered chatbot designed for medical students, built on the free and open-source Chatter bean framework. This chatbot can transform user queries into SQL requests, ensuring accurate database interactions. Using 97 categorized question samples, queries were structured around database and programming logic. By leveraging NLP, a core AI component, the chatbot delivers responses that entertain and reassure users about symptoms, treatments, and cures. Similar methods could enhance the accuracy and reliability of other chatbots [5].

#### A. Existing System

Simon Hoermann [6] discusses the available evidence supporting the feasibility and effectiveness of online mental health interventions that use synchronous text-based chat. Such written exchanges are gaining popularity as web-based methods for delivering mental health services. This review evaluates various synchronous web-based chat systems. While many current systems offer live text-based chats, they often come with limitations, such as delays in patient responses, requiring users to wait extended periods for expert acknowledgment. Additionally, some services may charge fees for live chats or phone consultations. The challenge remains in determining whether these technologies can be cost-effective in clinical practice, which warrants further research.

#### B. Proposed System

Our suggested solution allows users to communicate with a bot via voice or text for queries to respond to the queries, the system makes use of an expert system. Additionally, the user can see the available physicians for that specific illness. Numerous individuals can get online counselling sessions with this technology. The chatbot's data is kept in the database as a pattern template. The bot will recommend foods that are appropriate for your condition and deliver analgesics.



**Fig. 1. System Architecture**

### 3. SYSTEM ARCHITECTURE

#### A. Tokenization

Tokenization involves breaking down sentences into individual words to facilitate processing. This process splits text at predefined character points, removing punctuation. It ensures that each word is separated for analysis, leading to the next stages of text processing.

#### B. Red-Ex

Redundant expressions are eliminated from sentences to emphasize significant keywords. This process removes repetitive or non-essential terms, such as 'a,' 'an,' or 'the,' which do not contribute to the core meaning. It helps reduce computational complexity and enhances processing efficiency.

### C. Explanatory Variable Extraction Based on N-gram and TF-IDF

Explanatory variables extraction involves simplifying data by prioritizing variables based on their importance to the document. This process improves the efficiency and accuracy of analysis by identifying significant keywords and their frequencies within the document.

### ALGORITHMS

Three algorithms are employed in the development of the health-care chatbot system:

1. N-gram Algorithm
2. TF-IDF (Term Frequency-Inverse Document Frequency)
3. Cosine Similarity Algorithm

### D. N-gram Algorithm

N-grams assist machines in understanding words within context, enhancing their ability to predict upcoming words in a sequence. An N-gram is a sequence of  $n$  contiguous items from a sample of text. The 'n' can vary, for instance, with  $n = 2$ , it's called a bigram, and with  $n = 3$ , it's called a trigram. This method helps to predict future words based on sentence structure.

### E. TF-IDF

#### E.1. Term Frequency (TF)

This metric reflects the frequency of a word in a document. It is calculated by dividing the number of times a word appears by the total number of words in the document. Each document has its own term frequency.

$$TF(t) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

#### E.2. Inverse Document Frequency (IDF)

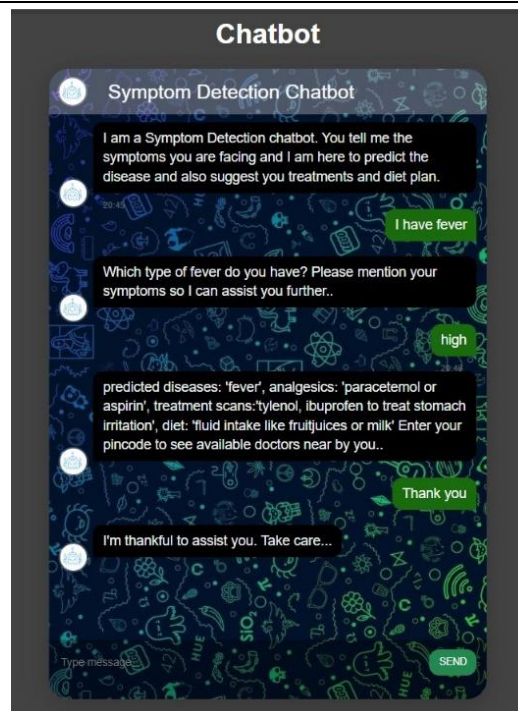
IDF measures the importance of a word across all documents by calculating the logarithm of the total number of documents divided by the number of documents containing the word. It helps to weigh unique words across the entire dataset.

$$IDF(t) = \log \frac{N}{n_t}$$

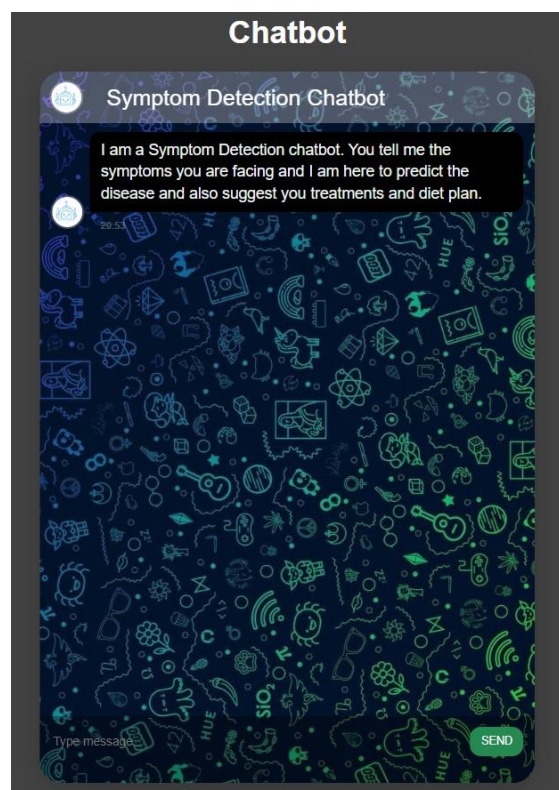
#### Algorithm 1. TF-IDF Calculation

- 1: **procedure** TF-IDF( $D, t$ )
- 2:  $N \leftarrow |D|$
- 3: **for all**  $d \in D$  **do**
- Total terms in  $d$
- 5:  $df(t) \leftarrow |\{d \in D : t \in d\}|$
- 6:  $idf(t) \leftarrow \log \frac{N}{df(t)}$
- 7: **for all**  $d \in D$  **do**
- $tfidf(t, d) \leftarrow tf(t, d) \times idf(t)$
- 9: **return**  $\{tfidf(t, d) : d \in D\}$

1. Define the procedure TF-IDF with inputs: a document set  $D$  and a term  $t$ .
2. Calculate  $N$ , the total number of documents in the document set  $D$ .
3. For each document  $d$  in  $D$ , calculate the term frequency ( $tf$ ) of  $t$ :
  - (a) Divide the frequency of  $t$  in  $d$  by the total number of terms in  $d$ .
4. Compute the document frequency ( $df$ ) of  $t$ , which is the count of documents in  $D$  that contain  $t$ .
5. Compute the inverse document frequency ( $idf$ ) using the formula:



## SCREENSHOTS



$$\text{idf}(t) = \log \frac{N}{\text{df}(t)}$$

5. For each document  $d$  in  $D$ , calculate the TF-IDF value for  $t$  as:

$$\text{tfidf}(t, d) = \text{tf}(t, d) \times \text{idf}(t)$$

6. Return the TF-IDF values for  $t$  across all documents in  $D$ .

## F. Cosine Similarity Algorithm

Cosine similarity calculates the cosine of the angle between two non-zero vectors in an inner product space, determining how similar they are. This method is also used to assess cohesion within clusters in data mining. The formula for Cosine similarity



#### 4. CONCLUSION

Cosine Similarity =

$$\frac{\vec{U} \cdot \vec{V}}{\|\vec{U}\| \|\vec{V}\|}$$

The chatbot serves as an effective tool for interaction between humans and machines. The system provides rapid, accurate responses without delay, offering personalized diagnostic infor-

Cosine distance measures how far apart two vectors are in an n-dimensional space, indicating the relationship between words.

mation based on symptoms. Its user-friendly interface makes it accessible to anyone who can type in their native language.

#### 5. FUTURE ENHANCEMENTS

The future is poised for messaging apps, as users are spending more time on these platforms compared to other applications. Implementing personalized medicine via chatbots could save lives and raise medical awareness. Regardless of geographical location, people can engage in medical conversations as long as they have access to a computer or smartphone with an internet connection. The chatbot's efficiency can be further enhanced by expanding the word database and incorporating more word combinations, enabling it to handle a broader range of diseases.

- Expanding the database to handle all types of diseases.
- Enhancing natural language processing to understand complex user inputs.
- Supporting multilingual interactions.

#### 6. REFERENCES

- [1] Dharwadkar, R., Deshpande.: A medical chatbot. International Journal of Computer Trends and Technology (IJCTT), vol. 60, no. 1, (2018).
- [2] Shrivastava, R., Pujahari, A., Singh, S.P., Bhowmik, T. (2023). Efficient Question Answering in Chatbot Using TF-IDF and Cosine Similarity. In: Garg, D., Kumar, N., Iqbal, R., Gupta, S. (eds) Innovations in Information and Communication Technologies. Algorithms for Intelligent Systems. Springer, Singapore.
- [3] Ahmed, M. Anouar Abdelhakim, Boudhir Ayanouz, Soufyane. (2021). An Intelligent Chatbot Using NLP and TF-IDF Algorithm for Text Understanding Applied to the Medical Field.
- [4] Bates, M.: Health Care Chatbots Are Here to Help. IEEE Pulse, vol. 10, no. 3, pp. 12-14 (2019).
- [5] Ramalingam, Jegadeesan Srinivas, Dava Nagappan, Umapathi Ganesan, Karthick Venkateswaran, Natesan. (2023). Section A-Research paper Personal Health-care Chatbot for Medical Suggestions Using Artificial Intelligence and Machine Learning Eur. 12. 6004-6012. 10.31838/ecb/2023.12.s3.670.
- [6] Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review. J Med Internet Res. 2017 Jul 21;19(8):e267. doi: 10.2196/jmir.7023. PMID: 28784594; PMCID: PMC5595406.