Healthcare Chatbot Using Natural Language Processing

G.V.V. Sai Roshan, V. Koushik Raj, D. Mohan Satyendra, S. Sai Devesh, D. Lokesh, K. Swarna

Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India.

Abstract:- To get off to a good start in life, healthcare access is crucial. But seeing a doctor while you're unwell is a major pain. The way to go is to create a chatbot specifically for use in healthcare. Using Natural Language Processing (NLP), this AI component can identify ailments and administer basic medical treatment. Reduced healthcare costs and improved access to medical information are the two main aims of a healthcare chatbot. Chatbots have the potential to become patients' go-to resources for medical information, answering questions about symptoms and providing recommendations for treatment. The goal of developing a healthcare chatbot is to provide consumers with accurate information about any probable condition. Multiple languages for text or voice assistance are available to users of the technology. By analyzing user symptoms, bots may provide medical diagnoses, dietary recommendations, and doctor referrals. The public will then acquire self-defense skills and pay more attention to their health. An engine capable of providing relevant, human-like replies may be built using methods using AI and NLP, such NLTK for Python, to analyze speech and provide intelligent responses. Automated conversational software is known as a chatbot.

Keywords*: NLTK, NLP, HER, AI, Pre-processing.*

1. Introduction

Health care is of paramount importance in today's environment. These days, most individuals are worried about their jobs, their reputations on the job, and their online addiction. Their welfare is not a concern to them. Going to the doctor for minor issues could end up being a bad idea. So, to help people understand their condition and make an informed decision about seeking medical attention, we suggest developing an AI-driven healthcare chatbot system. Both the patients' health and their knowledge of their illness are improved. The health-related information that users receive might be accurate.

The healthcare industry is currently one of the most talked-about topics in the globe. As people adopt new ways of living, they are increasing the dangers to themselves. Therefore, it is crucial to identify and treat illnesses early on. The Indian healthcare system has a lot of challenges, one of the most significant being meeting the requirements of rural populations. This means the Indian government has its work cut out for them.

The majority of India's population (68.8%) lives in rural regions, where access to healthcare is limited and sickness death rates are high (census 2011). It could be quite a journey for people living in rural parts of several Indian states to reach bigger cities for medical treatment. In the healthcare industry, there is a severe lack of qualified medical professionals. Meeting face-to-face with medical experts is time-consuming and expensive.

Affordability and the possibility of communication issues with qualified specialists are additional considerations for rural areas. As a result, many people avoid necessary medical care, putting themselves and others at danger of potentially fatal illnesses and accidents. Natural language processing (NLP) is used in smart healthcare to interpret textual information and is linked to human-computer interaction communication. Data pertaining to clinical texts and other form of literature are two possible classifications for the text data. Electronic health record (EHR) systems primarily contain clinical text, which consists of unstructured text data such as medical notes, electronic prescriptions, and diagnostic findings. It is sourced from a wide variety of healthcare facilities. A variety of healthcare contexts include publications containing text for evidence-based reference and population screening surveys. This type of text is referred to as "other text data." From human-robot interaction in rehabilitative treatment to patient-provider communication in clinical investigation, communication is an integral part of every smart healthcare scenario. Machine translation and rehabilitation robot user interfaces are two technologies that bring even more improvement to these settings.

2. Literature survey

Simon Hoermann study lends credence to the viability and efficacy of online psychological state treatment conducted via text-based synchronous chat for individual clients. As a kind of online therapy for mental health issues, real-time text chats are gaining popularity. This evaluation primarily aims to analyze several synchronous web-based chat solutions.

Live text chat is a feature of many existing systems; yet, these systems are not without their drawbacks. For instance, patients who have to wait for a consultant to acknowledge their request do not receive a prompt response. In order to measure telephonic or online chat activity, several methods may apply quantity charges. It is important to consider in order to determine the future analytical research's cost-effectiveness in clinical settings using such technologies [1].

Saurav Kumar Mishra the chatbot would allow a patient to communicate with a virtual doctor, says Saurav Kumar Mishra. The chatbot's role is to simulate that of a doctor. The development of this chatbot made use of pattern matching regarding NLP and. Python is the name of the language that was used to create it. Twenty percent of chatbot responses were either ambiguous or wrong, according to the study, while eighty percent were spot on. Basic medical care, awareness-raising, and training might all benefit from this software, according to the results of the research and chatbot survey [2].

According to theory of **Divya Madhu**, AI has the ability to identify possible treatments and diagnose diseases using symptoms alone. In order to catch any problems before they do harm to the body, it is important to have regular physical checkups. Several constraints, including government regulations and the costs of research and execution, impede the effective application of customized medicine; however, these aspects are not addressed in the study [3].

Hameed Ullah Kazi explains an AIML-powered chatbot that medical students can use. The free and open-source Chatter bean served as the foundation for the bot's development. The AIML-powered chatbot may be configured to convert user requests into relevant SQL queries. After collecting 97 question samples, they were categorized according to their kind. We divided the following categories into subgroups based on the number of questions in each. The queries were built upon the database and computer language. Here is the p-ISSN for this query: 2395-0072. documenting the results, whether they were created or acquired. After selecting a database answer using the random() function, this code utilizes the text() method to accept input and trims() any extra punctuation. The purpose of the chatbot is to entertain and reassure users on symptoms, treatments, and, eventually, cures. A key component of AI, natural language processing (NLP) was utilized in the development of this chatbot. To make sure our chatbot gives accurate and appropriate responses, we might apply the same strategy [4].

The article asserts that the proliferation of smartphones is a direct outcome of the exponential growth in mobile phone use brought about by technological advancements. In this case, the patient will receive the necessary answer after the mobile app has collected user information. This response not only offers therapy and therapeutic support, but it also allows for the early detection of a specific illness. The main goal was to create a system that would make real-time data transmission and reception easier. After receiving this data in real-time, the web server encrypts it and processes it further. Intelligent wireless interactive healthcare system development and implementation is outlined.

Cancer is usually discovered much later in life by most study participants. Cancer develops when abnormal cells keep reproducing and spreading. No one ever gets a second chance at a healthy, longer life when they have cancer.

The health crisis of depression is rapidly becoming the most pressing issue of our day. Findings from this study point to the importance of communication in promoting psychological health. If the patient attempts to confide in someone at the right time and nobody is around, the problem is only partially remedied. Because of this, the chatbot is unveiled. The chatbot can facilitate communication between those in distress by utilizing natural language processing (NLP). To aid machines in comprehending human speech, this scenario makes use of one branch of AI known as natural language processing (NLP). Processing of natural languages (NLP) aids in text comprehension, while artificial intelligence (AI) enables the chatbot to assess the conversation. The bots' ability to successfully acquire and retain vast amounts of cancer-related data from the internet allows them to provide patients with accurate and valuable information tailored to their specific needs.

3. Methodology

We are using three algorithms to develop a health care chatbot system.

1. N-gram Algorithm.

2. TF-IDF (Term frequency-inverse data frequency).

3. Cosine similarity algorithm.

N-gram Algorithm: N-grams are designed to make it easier for computers to understand a word in the material. One way to organize text is into N-grams, which are really just a continuous string of n-items. The N-items symbol indicates the possibility of having two, three, etc. things. So, it's an endless string of particulars. Being able to guess the next word in a phrase was helpful. There will be letters, words, and sentences. We may refer to them as bigrams when n=2 and trigrams when n=3. A different number for "n" could be used depending on the statement.

Term frequency (tf): It is common practice to exclude all stop words while training a model to comprehend the text. Another method is to use TF-IDF to rank the terms in order of importance. By dividing the entire word count by the word frequency, we may find the word's occurrence in a text. Every paper has the same set of sentences.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

Wi = tf * idf

The abbreviation tf stands for "inverse document frequency," which is the overall frequency of a word in a phrase. We produce the resultant vector by applying the aforementioned method to the user input and finding the weight of each sentence. The question database's pre-processing and phrase weighting are used to build a comparable vector.

Inverse Data Frequency (idf): As a percentage of total number of papers using the word in the multiple document log. Every text in the corpus has its own weight, which is determined by the inverse data frequency.

$$idf(w) = log(\frac{N}{df_t})$$

Cosine Similarity Algorithm: By calculating an angle's cosine that separates two non-zero vectors in an inner product space, cosine similarity finds how similar they are. In information (data) mining, this method is also employed to measure cluster cohesiveness.

To find the cosine of similarity, divide A by the product of the squares of B.

However, in an n-dimensional space, the sum of the cosine distances of two vectors cannot equal the gap between them. On the other side, distance stands for linked words.

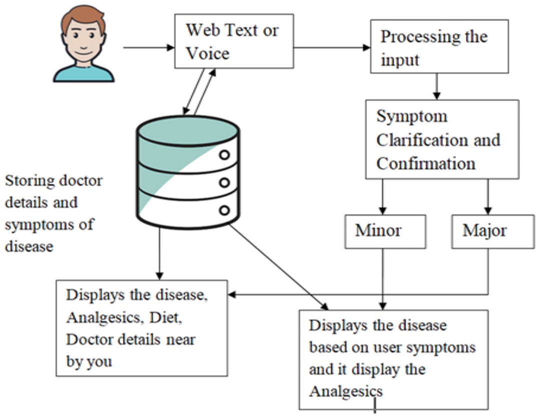
Web Text or Voice
Processing the input
Symptom Clarification and Confirmation
Storing doctor details and symptoms of disease
Minor
Major
Displays the disease, Analgesics, Diet, Doctor details near by you
Displays the disease based on user symptoms and it display the Analgesics

Fig1: System Architecture.

Fig1 shows that users may start a pleasant conversation with the chatbot, which is recorded in the database for further usage. To better understand the user's symptoms, the chatbot will ask them a series of questions. We shall find out how bad the disease is. The severity of the sickness will determine the chatbot's reaction. If the situation is severe enough, the user will be given recommendations for pain relievers, healing foods, and the contact information for a doctor in the area who may offer further treatment. The chatbot's UI may have a friendly, approachable tone as well. Chatbots can save you a trip to the ER that isn't required.

The first step is for the chatbot in Fig2 to take user input and run it through certain algorithms. With the user's input, the bot will run the algorithms. Algorithms and a database of symptoms will work together to understand the input. The chatbot will ask a set of inquiries to the user designed to confirm and clarify their symptoms. The disease will be categorized as either mild or severe. No matter how dire the situation, the chatbot will respond. If the situation is serious, the user will obtain the doctor's contact information, so they may seek pain medication, dietary guidance, and follow-up care.

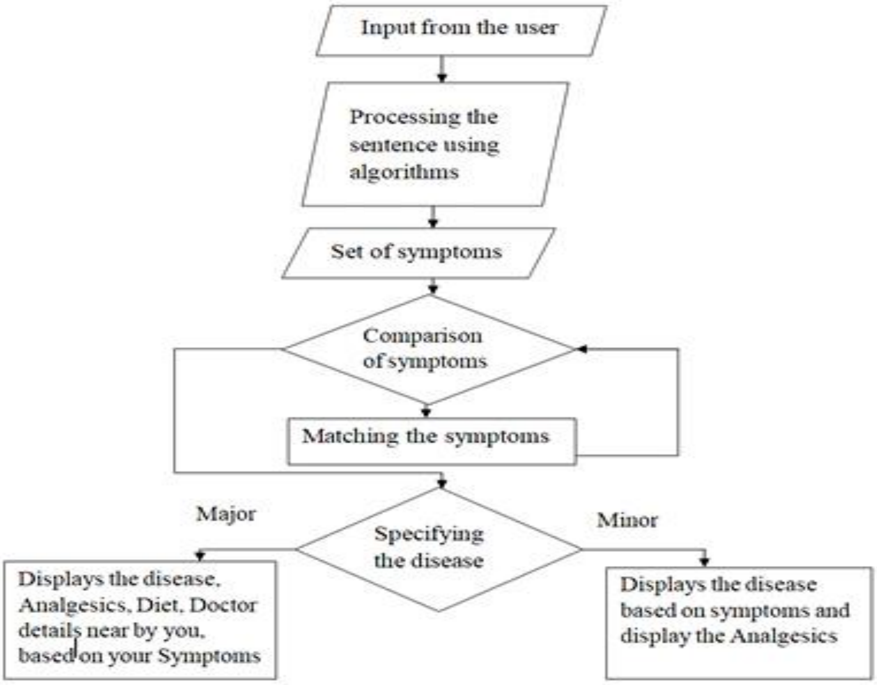
Input from the user
Processing the sentence using algorithms
Set of symptoms
Comparison of symptoms
Matching the symptoms
Major
Specifying the disease
Minor
Displays the disease, Analgesics, Diet, Doctor details near by you, based on your Symptoms
Displays the disease based on symptoms and display the Analgesics

Fig 2: Dataflow of Chatbot.

Google and Apple's Siri both rely significantly on natural language processing (NLP) for language detection. Two main parts of it are NLG and NLU, or natural language understanding and generation. It paves the way for machine to comprehend human-written and spoken directions. Because of its intricate structure and shape, understanding natural language requires more mental energy than creating it. It analyzes several aspects of the language and utilize NLU Natural Language Understanding to analyze and covert formless information into a format than the system can readily understand.

4. Feature extraction

NLP text pre-processing: Using natural language processing (NLP), businesses may provide customers with an excellent chat experience. In order to extract symptoms and relevant keywords from user-inputted text, pre-processing is necessary. The user's request has been satisfactorily addressed with the application of pre-processing techniques such as stemming, tokenization, TF-IDF [12], and cosine similarity.

1) **Tokenization:** Case folding is the process of automatically changing every user-entered text to lowercase. The user's initial, lowercase input is transformed into a string of words through the tokenization process. As a result, the text will be shown in its component words. Tokenization might be helpful when individual words need to be handled and evaluated. Also, the last bag of words is left after all punctuation is removed.

2) **Stop word removal:** Word bags created in the prior pre-processing stage are filtered for valuable keywords by removing stop words such as "a," "an," "the," etc. Because they consume so much space and time during pre-processing, stop words must be eliminated.

3) **Stemming:** Then, in order to get at their root form, all of the words in the word bag are iteratively stripped of their prefixes and suffixes. It is called stemming in natural language processing. Once tokens are manufactured, they are sent to the stemmer. In this instance, the system employs the Porter-Stemmer algorithm, which outperforms other stemmers.

4) **Sentence similarity:** Two assertions can be distinguished from one other by observing the degree to which their resulting vectors are identical. To determine how far apart the vectors are, the cosine similarity method is employed. A chatbot will respond with pertinent information if the angle of cosine between two phrase vectors is more than 0, and it will ask the user for relevant information if it is not.

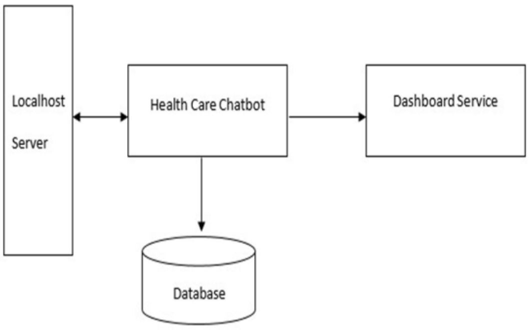

Fig 3: Architecture of the proposed system.

Classification Algorithms The system has compared the following classification algorithms for disease classification:

1) Random Forest Classifier [7]

2) K-Nearest Neighbors (KNN) [6]

3) Support Vector Machine (SVM) [11]

4) Decision Tree [8]

5) Multinomial Naive Bayes (MNB) [9]

The Random Forest Classifier groups all decision trees into clusters. Entering this place is like stepping into a jungle. People often believe that a forest with more trees would last longer. Random forests also make use of several decision trees [7]. To train the decision trees, we use a subset of the original training data. The result is an improvement in the Random Forest Classifier's accuracy and a decrease in overfitting.

KNN works by assuming that nearby items are connected. When classifying, the KNN approach looks for similarities between characteristics. The degree to which the new data point resembles the samples used for training determines its value. It is one of the earliest and most basic forms of categorization [8].

In n-dimensional space, the goal of the support vector machine method is to locate a hyperplane. A number of hyperplanes can be considered for any categorization issue. Maximizing the distance between data points of various classes is necessary to guarantee that the selected hyperplane has the largest margin [11]. Classifications are delineated by the hyperplanes.

An organizational tool for categorization known as a decision tree is shaped like a tree. Attributes are used to constantly segregate the training data [9].

That every feature is completely separate from every other feature is the foundation of both the MNB algorithm and the Bayes theorem. Based on this assumption, it sorts the given data [10].

5. Experimental results

In order to improve patient engagement and communication between practitioners and clients, healthcare chatbots that use natural language processing (NLP) have emerged as a viable option in the modern healthcare industry. Experiments testing the effectiveness of such systems have shown encouraging results, as shown in Fig. 1.

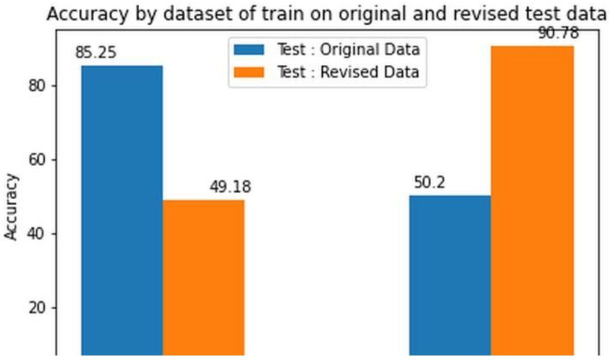

Fig 4: Accuracy by dataset of train on original and revised test data.

Investigating healthcare chatbot’s natural language processing (NLP) capabilities, the study analyzed and responded to user inquiries, supplied pertinent data, and made personalized suggestions. Impressive comprehension of natural language inputs was shown by the chatbot's accurate answers to various medical inquiries. The chatbot might deliver context-appropriate replies by mining user chats for data using natural language processing (NLP) methods. This is of the utmost importance in the healthcare industry, where clear and concise communication is essential for both patients and healthcare providers. All of computer-based predictive model's trials were conducted using Python 3.8.8 and the Scikit-Learn Artificial Intelligence module as showed in the above Fig 4. We compared five different classification algorithms, all of which are described in the suggested approach, to get the findings for the illness classification technique.

6. Conclusion

One form of AI is the chatbot. system that allows computers and humans to communicate through language. This program's goal is to provide the user with accurate findings and prompt responses from the bot. It has been concluded that chatbots are easy to use and accessible to anybody who can type in their native language. The ability to provide personalized symptoms is a key feature of a diagnosis-backed chatbot. Trial results show that natural language processing (NLP) chatbots might change healthcare like no other industry before. The solutions' efficacy in enhancing healthcare communication and accessibility is demonstrated by high user contentment, rapid response times, and great user query understanding performance. The integration of natural language processing (NLP) into healthcare chatbots has the potential to revolutionize patient-centered therapy and information sharing as technology progresses. Various natural language processing (NLP) tasks focused on text and speech are assessed in order to identify the most up-to-date methods for addressing these issues. We demonstrate the power natural language processing's possibilities, (NLP) in smart healthcare by developing applications that use approach to natural language processing in a range of smart healthcare contexts, such as clinical practice, hospital administration, personal care, public health, and medication research. You may find a comprehensive list of examples of smart healthcare applications and the associated natural language processing technologies in Table II. We go on to talk about how NLP-driven smart healthcare is essential for two particular medical issues: mental health and the COVID-19 epidemic.

Refrences

[1] K. Oh, D. Lee, B. Ko, and H. Choi, "A Chatbot for Psychiatric Counselling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375. Doi: 10.1109/MDM.2017.64.

[2] Du Preez, S.J. & Lall, Manoj & Sinha, S. (2009). An intelligent web-based voice chatbot. 386 - 391.10.1109/EURCON.2009.5167660 Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING. 5. 158-161.2017.

[3] Bayu Setiadi, Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modelling", Intelligent Systems Modelling.

[4] Simulation (ISMS) 2016 7th International Conference on, pp. 72-77, 2016.

[5] Dahiya, Menal. (2017). A Tool of Conversation.

[6] C.P. Shabari Ram, V. Srinath, C.S. Indhuja, Vidhya (2017). Ratwatte: Chatbot Application Using Expert System, International Journal of Advanced Research in Computer Science and Software Engineering,2.

[7] Tin Kam Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 1995, pp. 278-282 vol.1, doi: 10.1109/ICDAR.1995.598994.

[8] L. Jiang, Z. Cai, D. Wang and S. Jiang, "Survey of Improving K Nearest-Neighbor for Classification," Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007), Haikou, China, 2007, pp. 679-683, doi: 10.1109/FSKD.2007.552.

[9] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, May-June 1991, doi: 10.1109/21.97458.

[10] Kaviani, Pouria and Dhotre, Sunita, “Short Survey on Naive Bayes Algorithm,” International Journal of Advance Research in Computer Science and Management 04, 2017.

[11] Mrs. Rashmi Dharwadkar, Dr.Mrs. Neeta A. Deshpande "A Medical ChatBot". International Journal of Computer Trends and Technology (IJCTT) V60(1):41-45 June 2018. ISSN:2231-2803.

[12] Qaiser Shahzad, Ali Ramsha, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” International Journal of Computer Applications, 2018.