

MBBS - Med-Bay Bot System: A Chatbot for Disease Prediction and Treatment Recommendation

¹Rajat Gupta, ²Audarya Uttarwar, ³Pranjal Bhagade, ⁴Kalyani Magdum, ⁵Dr. Bharti P. Vasgi

¹⁻⁴Student, Department of Information Technology, Sinhgad College of Engineering, Pune, Maharashtra, India

⁵Assistant Professor, Department of Information Technology, Sinhgad College of Engineering Pune, Maharashtra, India

Abstract-One of the most important goals of Artificial Intelligence is the interaction of machines with humans in simplified natural language in such a way that even the non technically adept individuals can communicate with the system. In recent years, the interaction between a bot and a human, via the exchange of text messages, has gained popularity due to the evident ease of use. Many websites and apps are using these “chatbot” systems in order to reduce human workload be it the customer support bots on websites like makemytrip, swiggy or individual assistants like Google Assistant. Chatbots are known for being great at many things such as immediate response, large amounts of data availability, and most important is a reduced need of human supervision. The field of medicine is no stranger to this concept either, with chatbots such as Curable - Pain management, we observe that the arrival of artificial intelligence in the field of medicine is a welcome one.

Keywords- Disease prediction, Chatbot, Machine Learning, Nature Language Processing

I. INTRODUCTION

Billions of people around the world lack basic health care. Healthcare is a common and important topic in politics. Not only is it important in politics, but it is important to God. Healthcare is a basic civil right of every person living in a society. There are various quotes available on the internet but one that everyone will understand is “It is health that is real wealth and not pieces of gold and silver” (Mahatma Gandhi). According to an article published by WHO in 2017, half the world lacks access to essential health services and around 100 million are still pushed into extreme poverty because of health expenses [1]. In the proposed system a medical chatbot is built using as many of the free services and technologies as possible, so as to keep the chatbot free to use for everyone.

Nowadays Artificial Intelligence and machine learning based services are being used on a large scale. Virtual

assistants and chatbots are some of those services. Chatbots provide many services that are desirable for humans such as information retrieval in a natural language format for the layperson to understand easily, simple communication, round the clock support. In addition to these from the business perspective, chatbots are easier to maintain, require minimal supervision, increase customer satisfaction with their quick response. Also, Dialogue systems or conversational systems can support a wide range of applications in business enterprises, education, government, healthcare, and entertainment. In the medical sector, chatbots are used extensively for drug info, scheduling appointments with doctors, etc. Although it may never replace real doctors, chatbots can be used in the domain of drug prediction in order to provide immediate medical query resolution or for help with minor symptoms and save a visit to the hospital.

II. LITERATURE REVIEW

A. Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning [2]

This paper explains a medical chatbot which can be used to replace the conventional method of disease diagnosis and treatment recommendation. People consider it as the most reliable means to check their health status. Chatbot is an alternative to this conventional method of visiting a hospital and making an appointment with a doctor to get a diagnosis. This intends to apply the concepts of natural language processing and machine learning to create a chatbot application. People can interact with the chatbot just like they do with another human and through a series of queries, chatbot will identify the symptoms of the user and thereby predict the disease and recommend treatment.

Executing this framework can help people avoid the time-consuming method of visiting hospitals by using this free of cost application. The chat with users is completely personal and this helps users to be more open with their health matters and paves way for chatbots to efficiently identify the disease.

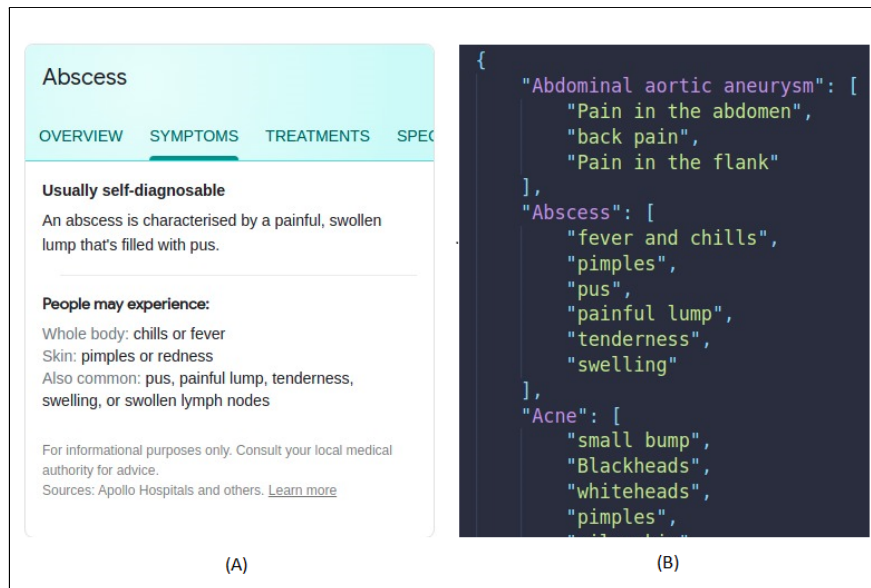


Figure 1. (A) Google Infobox for the disease Abscess. (B) Screenshot of the json file create

B. 'K-Bot' Knowledge Enabled Personalized Healthcare Chatbot [3]

In the present scenario when most of the things are just a click away, people consider visiting hospitals as the most effective, reliable and convenient way for their routine check-ups or disease diagnosis. The proposed system or approach focuses on creating an alternate using Decision Tree Algorithm where people can interact with chatbot and it will identify other symptoms and predict the disease along with the confidence level and thus recommending a specialized doctor. Using the above framework can help people save their time and money as well.

C. Self- Diagnosis Medical Chat-Bot Using Artificial Intelligence [4]

This Paper explains Medical chatbots were built to reduce medical costs and improve access to medical knowledge. The recommended notion is to develop a medical chatbot that can adopt AI to analyze the ailment and produce necessary information concerning the condition were discussing with a doctor. Text Diagnosis Bot enables sufferers to join in analyses of their medicinal matters and resent a personalized analysis report with reference to the symptoms. Therefore, people have opinions about their well-being and individual stability.

D. Healthcare Chatbot using Natural Language Processing [5]

This paper gives an idea to create a healthcare chatbot using Natural Language Processing technique. It is the part of Artificial Intelligence that can diagnose the disease and provide basic information. The user can achieve the benefit of a healthcare chatbot only when it can diagnose all kinds of disease and provide necessary information. The system provides text or voice assistance, that means the user can use his own convenient language, Bot will provide which type of

disease based on the user symptoms, and provides a doctor and also provides food suggestions that means which type of food you have to take. Thus, people will have an idea about their health and have the right protection. Chatbots are programs that work on Machine Learning (ML) as well as Artificial Intelligence (AI) Natural Language Processing (NLP) techniques such as NLTK for Python can be applied to analyses speech, and intelligent responses can be found by designing an engine to provide appropriate human like responses.

III. DATASET

The previously available datasets are either not freely available on the internet or they are restricted to some specific diseases.

For the disease prediction model to work, we need a few things in the dataset:

1. List of Diseases.
2. List of symptoms for each disease.
3. Age range in which the disease can affect.
4. Gender for which the disease is prevalent.

There were a few datasets available which had disease and their symptoms but they did not have the age range or the gender for which the gender is prevalent. Also, the number of diseases was very low. Hence the dataset for diseases and their symptoms along with their age range and gender prevalence was created by scraping the web using a few python scripts and through some manual searching.

A. List of Diseases

First step is to get a list of available diseases. For this we referred to a website called <https://www.nhsinform.scot>. We scraped the website using BeautifulSoup4 library and collected around 315 disease names. Then we manually searched the internet for more diseases and collected around 234 more diseases. So, finally after merging and removing

redundant diseases we are left with around 506 total diseases.

B. List of symptoms for each disease

Next step is to collect symptoms for each of the 506 diseases. For this we wrote a script that will do a google search for that disease and scrape the symptoms from the google infobox. The symptoms are stored in a list corresponding to the disease. So the dataset is in json format where the key is the disease name and the value is the list of symptoms for that disease.

As we can see in the above image (A), symptoms are picked up using a python script and then after cleaning this data it is saved in a json file as shown in the image (B) on the right.

During the cleaning process, first we separate symptoms on the basis of ','(comma) and 'or' keywords. Then all the symptoms are saved into a single list and saved into a dictionary with key being the disease name and values being the list of symptoms. Once the data is collected for all the possible diseases, it is saved in the JSON format.

Using the above process we were able to create a dataset having 400+ diseases with almost 1300+ unique symptoms.

C. Age range in which the disease can affect & Gender for which the disease is prevalent

Finally after getting the dataset for symptoms we manually created the age and gender dataset for each disease. This dataset was created by referring various medical websites. The format of the dataset thus created is as shown below:

```
1 {
2   "Acute cholecystitis": {
3     "age_group": [
4       "65-150"
5     ],
6     "gender": [
7       "MALE",
8       "FEMALE",
9       "PRIORITYM"
10    ]
11  },
```

Figure 2. Age and Gender Dataset

Finally, a csv file is created from the above json files using a python script with it's first row containing all the symptoms. In each successive row

IV. PROPOSED SOLUTION

A. Problem Statement

To develop a user friendly Chatbot which can be used for disease prediction based on provided symptoms, obtain disease information and any medicinal drug related information.

B. Problem elaboration

Chatbots are used in different fields and used for various purposes in day-to-day life. Our chatbot is mainly for medical purposes. Disease prediction based on

provided symptoms, obtaining disease information and any medicinal drug related information that are the main features of this project.

C. Proposed Methodology

Chatbots provide many services that are desirable for humans. They interact with them in natural language(English). In this model users can input unstructured symptoms or select the symptoms suggested by the system, based on which, a list of probable diseases is provided back to the user. Further, the user can select any of the output diseases to get more information about its other symptoms, causes, diagnosis, possible treatment, etc. to help the user better understand the disease and current medical condition. The system also suggests other symptoms based on the ones that the user has input. If the user wants information about a particular drug then the system also gives that feature. It provides all the necessary information of that drug. In this way System totally used for medical purposes we named the project as MBBS (Med-Bay-Bot System) This MBBS project mainly contains the following features.

- Talk back service
- Symptoms suggestion and selection
- Disease Prediction
- Drug Information
- Disease Information

D. Architecture Diagram

The architecture diagram gives the abstract view of the system which involves the main components and entities of the system. The input to the system is the text entered by the user in natural language. The text goes to the text processing module, where it's intent is identified and an appropriate response from the list of responses of the particular intent is provided to the user in the form of chat conversation.

If the intent is recognised as a disease prediction intent then symptoms from the text are extracted, stored in a list and passed to the machine learning model for prediction. Now, more symptoms are required for the prediction for accurate results. So, the bot prompts the user with a list of symptoms repetitively which are relatable to the initial symptoms entered.

E. Proposed System

1) Talk back service

This feature collects primary information of the user like name, age, gender, etc. Also users can chat with the system like "What is your name?", "What is your work?" Users' information will be helpful to correctly diagnose the disease.

This is the most necessary and mandatory part to move further and for correct prediction of disease.

2) Symptoms suggestion and selection

Here the user first enters Symptoms they have. Then System suggests some topmost matching symptoms to user. If users have symptoms from the suggested list they can select among them otherwise they can select "Done. But I have more symptoms". Then again it

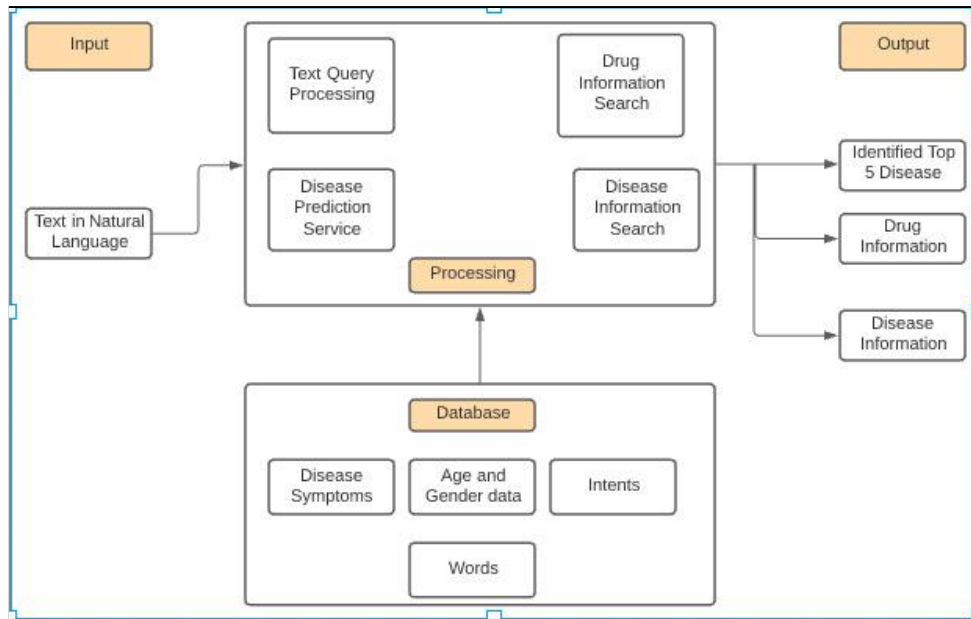


Figure 3 Architecture Diagram

suggests symptoms based on previous symptoms. Once all symptoms are covered user can select “Done! These are all the symptoms I have”.

In this way the system can collect all symptoms from the user. e.g. In below pictures user entered experiencing fever (Low-grade fever). Then the system again suggests more symptoms related to that. This chain will continue until the user selects “Done! These are all the symptoms I have”.

3) Disease Prediction

Once all symptoms are entered, the system can suggest up to top 5 diseases. e.g. user enters symptoms like low grade fever, muscle pain, feeling ill, loose motion, coughing, etc. This system suggests Norovirus and Chest Infection as the top most occurring diseases.

Diseases with a likeness percentage are calculated using the formula.

$$\frac{n((\text{Final symptoms selected by the user}) \cap (\text{Total Symptoms for that disease}))}{n(\text{Total Symptoms for that Disease})}$$

4) Drug Information

This feature gives us Drug information. If a user wants to know about some medicines, what they are used for, their contents, benefits, and side effects, they can enter the name of that medicine and the system will suggest some related drug names. A user can select an option which they want. e.g. If a user enters Crocin, then the system can suggest related options like Crocin Advance, Crocin Pain Relief, Crocin 240 DS suspension mixed fruit, Crocin baby drop, Crocin 1000mg etc. so a user can select any one for more information.

F. Concepts

1) Logistic Regression

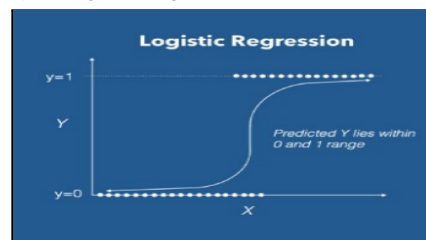


Figure 4 Logistic Regression

Logistic Regression is a Machine Learning algorithm which is used for classification problems. It uses a complex cost function, defined as the Sigmoid Function. The hypothesis of logistic regression tends to limit the cost function between 0 and 1.

When using linear regression we use the formula of the hypothesis:

$$h\theta(x) = \beta_0 + \beta_1 x$$

For logistic regression we modify it a little bit:

$$\sigma(Z) = \sigma(\beta_0 + \beta_1 x)$$

Now our hypothesis will give values between 0 and 1.

$$h\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Figure 5 Sigmoid Function

2) Neural Networks

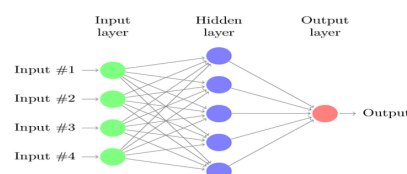


Figure 6 Neural Network

Artificial Neural Networks consist of three layers: a layer of input units is connected to a layer of hidden units, which is connected to a layer of output units.

- Input layer: The activity of the input units represents the raw information that is fed into the network.
- Hidden layer: The activity of each hidden unit is determined by the activities of the input units and weights on the connections between the input and hidden units.
- Output layer: The behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

V. RESULTS

The evaluation of the Logistic Regression ML model (89.82%) is done by comparing it with the KNN model (89.63%). The chatbot predictions are found to be 97.14% accurate. After Unit testing, the accuracy after reducing the symptoms was found to be 76.33% while with full symptoms it was found to be 88.00%

```
(env) rajat@lucifer:~/Documents/MedbayBot_heroku$ python wsgi.py
Testing chatbot predictions
Chatbot Testing Accuracy: 97.14285714285714%
.Prediction on Reduced Symptoms Accuracy: 76.33333333333333%
.Predictions Accuracy: 88.0%

-----
Ran 3 tests in 193.419s

OK
```

Figure 7. Accuracy

The System's performance is also evaluated against the well known web-md symptom checker (<https://symptoms.webmd.com/>) Figure (1) shows the result on the same set of symptoms from our system, while figure (2) shows the result from web-md.

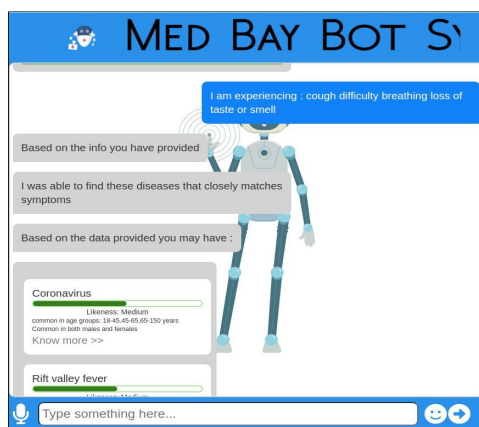


Figure 8. Med-Bay Bot Prediction

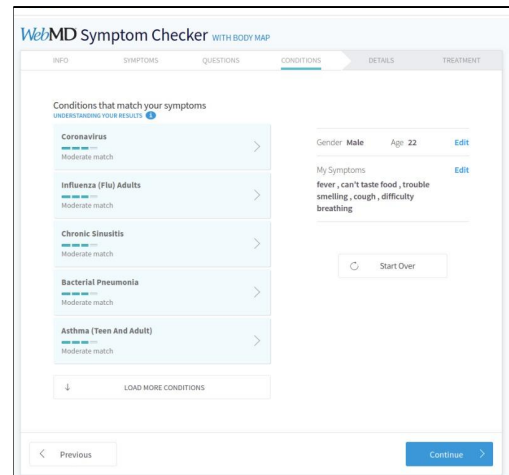


Figure 9. WebMD Prediction

Upon clicking on the know more section of a disease we were able to provide information related to a disease. Also we were able to display information about a drug.

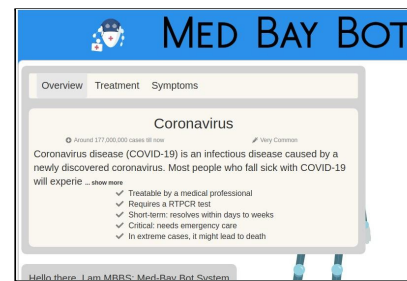


Figure 10. Disease Information

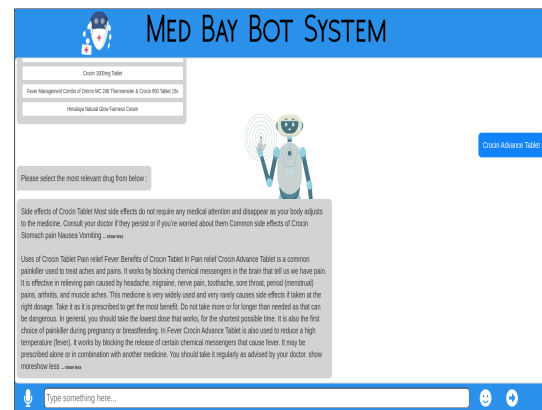


Figure 11. Drug Information

VI. FUTURE SCOPE

Our aim is to implement a full fledged interactive system which can later help the user in scheduling doctors appointments, finding relevant doctors upon detection of a disease, which are practicing near the user's location. An interactive speech to text and text to speech element can also be added for the differently abled individuals.

The chatbot can also suggest drugs related to the diagnosis it has made. A system can be implemented where the chatbot can remind the user about their scheduled drug dosage and scheduled doctor appointments.

VII. CONCLUSION

This project is an adequate solution for many services. The main key feature of this project is disease prediction. We can easily and precisely predict the disease on the basis of symptoms. So the user gets to know about the respective disease as early as possible and according to that user can consult a doctor. It is useful for other services like Disease information, Drug information. It gives information about particular disease and drugs.

The application is developed for obtaining a fast response from the bot which implies with minimal delay it provides the correct result to the user. It is found that chatbot is user friendly even for the naive user. Chatbot provides personalised diagnosis supported symptoms.

The system was able to suggest a diagnosis using a direct approach of the question and answering technique. The new development in artificial intelligence and the new way of thinking have the potential to entirely change the experience of customers. Especially in the field of medicine, a medical based chatbot offers a personalized analysis based upon symptoms.

REFERENCES

1. "World Bank and WHO: Half the world lacks access to essential health services, 100 million still pushed into extreme poverty because of health expenses" WHO, December 13, 2017, accessed June 12, 2021, <https://www.who.int/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses>
2. R. B. Mathew, S. Varghese, S. E. Joy and S. S. Alex, "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 851-856, doi: 10.1109/ICOEI.2019.8862707.
3. A. Ait-Mlouk and L. Jiang, "KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding Over Linked Data," in IEEE Access, vol. 8, pp. 149220-149230, 2020, doi: 10.1109/ACCESS.2020.3016142.
4. Habib, Fakihi & Shakil, Ghare & Iqbal, Shaikh & Sajid, Shaikh. (2021). Self-Diagnosis Medical Chatbot Using Artificial Intelligence. 10.1007/978-981-15-6707-0_57.
5. Papiya P. Mahajan, Rinku B. Wankhade, Anup A. Jawade, Pragati P. Dange, Aishwarya S. Bhoge, "Healthcare Chatbot using Natural Language Processing" International Research Journal of Engineering and Technology (IRJET), 2020, Volume 07, Issue 11, pp. 1715-1720