Health Consultant Bot: Primary Health Care Monitoring Chatbot for Disease Prediction

Asad Ur Rahman 1, Madiha Liaqat1, Ali Javeed1, Farman Hassan1

1 University of Engineering and Technology Taxila, Punjab Pakistan

* Correspondence: Asad Ur Rehman, Email ID: asadasif357@gmail.com.


Received | Feb 11, 2022; Revised | Feb 24, 2022 Accepted | Feb 27, 2022; Published | Feb 28, 2022.

This research paper presents a disease prediction chatbot that is intelligent enough to communicate with patients to predict their disease by detecting their symptoms through natural language processing. This system allows the user to describe their medical health condition in natural language, and by processing their natural language-based statement, our system detects the symptoms, predicts the disease, and provides basic precautions as well as a brief introduction about the disease. We have used IBM Watson Assistant to build this system. Watson assistant provides several machine learning algorithms to process user statements and symptoms extraction. In our system, symptoms were mapped by considering the community data which resulted in a predicted disease. Our system provides the relevant information about the predicted disease from the system’s database. In an experimental evaluation, we carried out a study having 156 subjects, who interact with the system in a daily use scenario. Results show the effectiveness and accuracy of our system to support the patient in taking good care of their health.

Keywords: Chatbot, Disease Prediction, Health monitoring, Healthcare

INTRODUCTION

In recent years, we have observed a substantial growth of interest in the use of eHealth services, which is a well-defined practice of healthcare services using electronic processes and communications [1]. Emerging technologies are revolutionizing healthcare services due to rapid advancements in technology-driven innovations [2]. People mostly depend upon health tracking devices and personalized medicines. Advancements in technologies have made it possible to design interactive models for patients. The novel coronavirus (Covid-19) has drastically shifted the people to use eHealth due to the risk of spreading the Covid-19. The patient’s needs can be met effectively and quickly by using eHealth services; eHealth facilitates the timely diagnosis and consistent monitoring of the disease. These rapid advancements in healthcare have significant advantages for society, which improve the average life expectancy.
and quality of life [3]. The digital evolution by introducing eHealth has enabled a speedy diagnosis process, assist in searching the nearest hospitals and physicians to treat the patients timely. Since 2010, the USA has significantly increased the health budget by 138% that is more than 30 billion dollars. It is expected that until 2025, eHealth will be generating an annual revenue of 350-450 billion dollars. Rapid evolution in health technologies will lead to reducing the massive cost of healthcare systems. There is a potential possibility to design computer-aided tools for telemedicine or to digitize the health data. We also designed a novel HealthConsultantBot, which integrates an intelligent virtual assistant.

Health maintenance is a set of activities that people practice on their own to maintain healthy life and wellbeing. These activities maintain health by making the immunity system strong enough to fight against disease [4]. Artificial Intelligence is a type of machine learning technique, the machine learns from experience datasets and predicts better results on testing datasets [5]. A chatbot is a computer program that interacts with end-users by using natural language processing (NLP). Chatbot technology started in the 1960s with the intention to check if automatic robots can fool the end-users by acting like real humans [6]. Chatbot integrates computational algorithms and NLP models to do an informal chat between humans and computers using the NLP [7]. NLP is a computerized approach that focuses on the automatic analysis of a language structure. NLP is being used in developing advanced technologies such as speech recognition and machine translation [8].

Even with the technological advancement, health care centres still completely rely on health care staff to carry out initial interviews and patient intake [9, 10]. It is observed that often there is a high workload and limited resources at the primary health care services results in a long wait for a patient before being advised by a specialist doctor [11]. Initially, health care staff has the responsibility to get complete details of every patient and know the symptoms of patients to treat patients accordingly. These individuals have a certain level of proficiency, and they have a major role in referring a patient to a medical expert. However, there are certain limitations of this traditional system of check-ups, such as medical staff that gets details from the patients can misunderstand the patient’s disease and may refer the patient to an irrelevant doctor. Therefore, staff needs to attend to more patients in a short period, this will increase the risk of not getting the valuable details.

Existing methods [12-23] have explored various Chatbots for disease prediction. Prakhar et al. [12] designed a chatbot for the prediction of diseases; the chatbot asked questions from users and it compute the probability of the disease then the chatbot asked questions about the specific disease. Three machine learning algorithms such as KNN, support vector machine (SVM), and naïve handles were employed for classification purposes, the SVM was used for the complex classification tasks. Rohit et al. [13] introduced a medical chatbot based on NLP and machine learning; the target audience was the individuals who used to visit hospitals for a scheduled check-up to know their medical condition. According to the authors the NLP-based bot is a great alternative for the patient in conducting daily check-ups. Nudtaporn et al. [14] designed an automated medical chatbot that was intended to work in collaboration with medical health care staff; the chatbot collects medical data of specific patients from an application named DoctorMe. The root source of this data was the patient’s medical consultant. The patient must have to visit the health care center to start interacting with this chatbot.
In [6], a trained chatbot was designed for medical assistance. The user interacts with the chatbot and sends his/her symptoms for the diagnosis of a certain disease. This Chatbot provides the details of the disease and possible medication used to cure the disease according to the entered symptoms by the end-user. This Chatbot [6] was artificially trained to suggest multiple available treatments. It has integrated JSON code to describe the medicine dosage categories with multiple age options and this module is named age-based medicine dosage. According to WHO reports, cancer is the second most effective disease in the world. In 2018, a total of 18 million cancer cases were reported all over the world [15]. Belfin R V et al. [16] introduced a graph-based chatbot for cancer patients, cancer patient shares their detail with this chatbot. Chatbot gets the cancer symptoms from the input string and then applies some filters on these tokens to make it understandable for the machine. The system picks the intents and entities from the user’s input string using machine learning, this system then uses a natural language tool kit (NLTK) to make it more understandable. The lemmatization process is also used to get the root words of input string tokens. UNIBOT [17] provides a web-based chatbot to answer university students’ queries. The main theme of this bot is to develop a run-time artificial intelligence (AI) bot which learns from users' input and answers the queries according to it. In this model, the user’s string passes from pre-processing, and then after connecting with the database, the system answers the user’s question by question mapping. Regular expressions in SQL queries were used to find the mapping in the database. In [18], a detailed survey was conducted to show the results of human behavior towards using chatbots in the marketing field. The survey was conducted on 60 respondents to know their experience about using chatbots in the marketing environment. Research shows that chatbot usage is very beneficial for many users for fast and efficient response.

As chatbots have presented themself as human-like agents but not the same as a human so there is always a factor of misbelief. Jitendra Purohit et al. [19] introduced the idea of a human resource (HR) chatbot for jobs to interview the candidates. The chatbot gets the CV from the applicant, which is then analyzed using NLP. Artificial intelligence makes the chatbot intelligent enough to ask relevant questions based on the candidate’s previous response. This chatbot stores the responses of candidates in the database so that they can be viewed by HR later. This chatbot solved the time limitation issue of interviewing many applicants.

MANDY [20] – is a medical chatbot, which assists the medical staff by automizing the patient intake. This chatbot interacts with a patient by dialogue, it asks questions from the patient and makes a history report of the patients for doctors. This system provides an interface for patients, a diagnostic unit, and an interface for doctors. MANDY’s comprises of three implementation modules such as chatbot application, web application, and e-health management system. This system combines data-driven natural language processing capability with knowledge-driven diagnostic capability. MANDY’s analysis engine uses word2vec for natural language processing which is Google’s word embedded algorithm. Word2vec maps the words into a vector space to get the word’s semantic similarity [20, 21]. The goal of the research is to overcome the problems faced by patients due to the high workload and limited resources at the primary care services. The proposed system is a chatbot that uses artificial intelligence to understand the user’s health condition by getting his/her input statement(s) using natural language. This system predicts the disease based on the symptoms observed from users’ input.
The proposed system has three major benefits, first, it reduces the workload on the medical staff, second, it provides personalized input service to the users, and third, the major one is the patients feel privacy to express their medical condition on a bot. Moreover, the research shows that patients have seem to be more honest and truthful while facing a robot instead of a human health care provider [22]. So, the proposed system will collect trustworthy information that will be a prime factor to achieve the desired output performance. The major contributions of our work are as under.

- The proposed system works on the diagnosis of several diseases present in the dataset.
- The proposed system is based on dialogues and conversations for each disease.
- The proposed system gets input in natural language.
- The proposed system gathers true details from patients.

The remaining paper is organized as follows; section II provides the details of the proposed working mechanism. Section III has detailed architecture of the HealthConsultantBot while section IV presents the experimental results and discussion. Finally, we conclude our work in section V.

**PROPOSED METHODOLOGY**

The main objective of the proposed HealthConsultantBot is to detect the disease based on the symptoms provided to the system. The proposed HealthConsultantBot comprises three modules such as design interface, gateway, and dialogue interface. In the initial step, the user interacts with the chatbot through a web platform. We integrated our chatbot with the famous social media platform, Facebook messenger to target the largest number of audiences around the globe. Webhook gateways are being used to route the traffic from the user interface to the cloud server. Once the user’s input gets routed to the backend, all the machine learning algorithm executes at IBM Watson Cloud. Proposed system logic is well defined in the following series of steps, Figure 1 shows the actual processing of the proposed chatbot.

**Primary health care monitoring chatbot for disease prediction**

We designed HealthConsultantBot based on client-server architecture to keep the end-user interaction separate from the actual implementation of the system. The aim of current study was to maintain the high-level useability of our chatbot for disease prediction.

**Dialog Interface**

We adopted the Facebook messenger interface, which is the most famous and easily accessible throughout the globe. Facebook is the largest social media platform in the world [23]. Our system is integrated with a Facebook page, which acts as the chatbot for Facebook users to interact. The user needs to login into his Facebook account and visits our page (Doctor Examine). The user just types a message to start a conversation with our system. This is the simplest way we found to interact with the largest audience in the world. The use of Facebook messenger enabled us to not implement a separate interface from scratch and to release our system to already enabled users. In this way, we have avoided the time and cost of end-user training; our system meets the highest standards of useability. We have built an intent named chitchat, which enabled the chatbot to interact with end-user other than the disease prediction. This will help the users to be more friendly and comfortable with the system. The dialogue interface has two sub-modules such as data acquisition and conclusive diagnosis response. The details are given in the subsequent sections.
Data Acquisition

Data acquisition is the process of collecting data. The users provide the medical data to the chatbot with the use of natural language. The collected data is then passed to the server-side (Watson assistant) through the gateway by employing webhooks to compute the probability of the disease.

Data Conclusive diagnosis response

Next, the conclusive response is displayed to the user as the output of his medical health statement. After finalizing the diagnosis, the patient receives a response containing the basic information about the disease, the patient is suffering from, and basic precautions to take until taking advice from a specialist medical consultant. At the last step a list of available medical consultants is being shown on output with respect to the predicted disease.

Gateway

The second module of the HealthConsultantBot is the gateway, which acts as the connection between the client and server-side. This is implemented using Restful web services that provide high-level APIs to allow the client to interact with the Watson chatbot. The connection between IBM Watson and Facebook messenger is made possible through the webhooks techniques. In this module, the client-side receives the message through push mode and webhooks distribute the communication through HTTPS access with an SSL certificate to Watson assistant.

Watson Side Implementation

Server-side implementation for HealthConsultantBot is performed on Watson assistant. We used NLP to understand the user’s statement about his health. Data acquisition is being done through Facebook messenger and webhooks are being used to transmit that data to Watson assistant for further processing. This data is tokenized and lemmatized to make it understandable for artificial intelligence algorithms. Intents are classified using SVM, with some pre-training by IBM Watson Assistant, entities use a fuzzy matching algorithm. We have already implemented a set of entities and intents with respect to each possible disease. After symptoms extraction from user input Watson assistant map the symptoms to the Intents and entity, and we get a list of hypothesis diseases with respect to the probability of each disease. A hypothesis disease with maximum probability is then being carried out as an assumption that the patient may have this certain disease. Then several questions are being asked to the user and record the response is yes and no to confirm the patient’s disease. We have stored each disease introduction and certain precautions in our Watson assistant. The final response contains the predicted disease with a brief introduction and precaution. Numerous steps such as tokenization, intent and entity mapping, symptoms to cause mapping, prioritization of hypothesis list, confirmation questions, and medical consultant database are performed in Watson assistant, and details are given in subsequent section.

Tokenization

The patient enters his medical condition in a natural language. The system applies string tokenization and Lemmatization to make the input understandable for machine learning algorithms.

Intent and Empty Mapping

The system then maps the user input on predefine intents and keywords taken from the tokenization process is being mapped on predefined entities. There is a predefined threshold value to get the mapping results of the user’s input.
Symptoms to Cause Mapping

Based on intent and entity mapping, the system identifies the basic user’s symptoms. At this stage, the system maps the user’s symptoms to the prebuild cause library and comes up with the hypothesis of root disease the patient is suffering from. Based on confidence received from the symptom, the system assigns a value to each hypothesis from 0-1 for mapping.

Prioritization of hypothesis list

In the next step, the system prioritizes the hypothesis list retrieved from the previous step and then considers the highest numerical value as the best match for the patient’s expected disease.

Confirm Questions

Based on the assumption built from the previous step, the system asks a set of questions from the patient to confirm the result. The user provides the natural language response which then gets stored in the backend and then the system analyses the context to approach the result. In case the patient response for the confirmation question does not match with the expected response to confirm a disease then the system asks to clarify the medical condition and controls get switched back to the data acquisition section.

Medical Consultant Database

The system contains a database having the medical consultant’s contact information. The system provides a list of nearby specialized consultants’ contact information based on the predicted disease and patient residential information.

EXPERIMENTAL RESULTS AND DISCUSSION

Dataset

Up to our best knowledge, no real-time dataset with real-time information is available that relates to the actual patient's symptoms. Else, due to privacy rules and regulations real medical data may not be available publicly, therefore we used the method mentioned in [22] to deal with this limitation. We used an unofficial dataset available on the platform of Kaggle [23] for testing purposes. We used 13 diseases and assumed P number of patients are going to use this system by providing their medical
conditions. We extracted the symptoms from their medical condition statements and follow the
probability of each symptom falling on an already stored intent of a specific disease. To check the
accuracy of our system, we used N number of observations for each disease with respect to the P
number of patients.

**Evaluation Metrics**

The purpose of this experiment is to evaluate the performance of the proposed
HealthConsultantBot on thirteen diseases such as peptic ulcer, migraine, hypertension,
gastroenteritis, fungal infection, fever, drug reaction, diabetes, chronic cholestasis, cervical
spondylosis, bronchial asthma, allergy, and aids. For this purpose, we designed a client-server
architecture based HealthConsultantBot for the prediction of diseases. In the initial step, the
user interacts with the proposed HealthConsultantBot by using Facebook messenger. The
HealthConsultantBot takes input from the users and assumes the probability of disease based
on the input provided by the users. We stored questions in Watson assistant based on
symptoms of diseases. The HealthConsultantBot then asks questions of the assumed disease
that are stored in the Watson assistant. From the results reported in TABLE 1, we can observe
that our system performs worst on gastroenteritis and achieved an accuracy of 88.88%,
precision of 85.71%, recall of 100%, and F1-score of 92.30%.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Accuracy%</th>
<th>Precision%</th>
<th>Recall%</th>
<th>F1-Score%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptic ulcer</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90.90</td>
</tr>
<tr>
<td>Migraine</td>
<td>91.66</td>
<td>90</td>
<td>100</td>
<td>94.73</td>
</tr>
<tr>
<td>Hypertension</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Gastroenteritis</td>
<td>88.88</td>
<td>85.71</td>
<td>100</td>
<td>92.30</td>
</tr>
<tr>
<td>Fungal infection</td>
<td>91.66</td>
<td>100</td>
<td>88.88</td>
<td>88.88</td>
</tr>
<tr>
<td>Fever</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Drug reaction</td>
<td>90</td>
<td>100</td>
<td>85.71</td>
<td>92.30</td>
</tr>
<tr>
<td>Diabetes</td>
<td>93.33</td>
<td>91.66</td>
<td>100</td>
<td>95.65</td>
</tr>
<tr>
<td>Chronic cholestasis</td>
<td>92.85</td>
<td>100</td>
<td>90.90</td>
<td>95.23</td>
</tr>
<tr>
<td>Cervical spondylosis</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Bronchial asthma</td>
<td>93.3</td>
<td>100</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Allergy</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The proposed system performs second-best on diabetes and achieved an accuracy of
93.33%, precision of 91.66%, recall of 100%, and F1-score of 95.65% while it performs best
and achieved an accuracy of 100%, precision of 100%, recall of 100%, and F1-score of 100%
for four diseases such as peptic ulcer, hypertension, fever, and allergy, respectively. The
detailed results of each disease in terms of accuracy, precision, recall, and F1-score are given
in TABLE 2. Experimental results demonstrate that the proposed HealthConsultantBot
assumes the correct disease and is reliable to be used for quick access. From the results
reported in Figure 2, we can observe that our system performs well and achieved an accuracy
of 94.83%, precision of 97.24%, recall of 95.49%, and F1-score of 96.36%.

**Table 2.** Details of the Dataset

<table>
<thead>
<tr>
<th>Disease</th>
<th>Test</th>
<th>Number of Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 2022</td>
<td>Vol 4</td>
<td>Issue 1</td>
</tr>
<tr>
<td>Observation</td>
<td>Symptoms</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>abdominal pain, internal itching, loss of appetite, the passage of gases, vomiting</td>
<td></td>
</tr>
<tr>
<td>Migraine</td>
<td>blurred and distorted vision, depression, headache, indigestion, stiff neck, irritability</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>chest pain, dizziness, headache, lack of concentration, loss of balance</td>
<td></td>
</tr>
<tr>
<td>Gastroenteritis</td>
<td>dehydration, diarrhea, sunken eyes, vomiting</td>
<td></td>
</tr>
<tr>
<td>Fungal infection</td>
<td>dyschromic patches, itching, nodal skin eruptions, skin rash</td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td>Dehydration, General weakness, Headache, Sweating</td>
<td></td>
</tr>
<tr>
<td>Drug Reaction</td>
<td>burning micturition, itching, skin rash, spotting urination, stomach pain</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>blurred and distorted vision, excessive hunger, fatigue, increased appetite, lethargy, obesity, restlessness, polyuria, weight loss</td>
<td></td>
</tr>
<tr>
<td>Cervical spondylosis</td>
<td>back pain, dizziness, loss of balance, neck pain, weakness in limbs</td>
<td></td>
</tr>
<tr>
<td>Chronic cholestasis</td>
<td>abdominal pain, itching</td>
<td></td>
</tr>
<tr>
<td>Bronchial Asthma</td>
<td>breathlessness, cough, fatigue, High fever, Mucoid Sputum</td>
<td></td>
</tr>
<tr>
<td>Allergy</td>
<td>chills, epiphora, fever, shivering, Sneezing</td>
<td></td>
</tr>
</tbody>
</table>
Next, we designed a confusion matrix to better analyze the performance of the proposed HealthConsultatinBot for the prediction of diseases. From TABLE 3, so, that we can observe that our system misclassified 3 healthy persons as patients while 5 patients as healthy persons. More specifically, the FP rate of the proposed system is 0.93% while the FN rate is 0.96%. These results illustrate that our system successfully predicted the disease and is reliable to be used for quick and accurate prediction of diseases.

Table 3. Confusion Matrix

<table>
<thead>
<tr>
<th>Predicted Class</th>
<th>Actual Class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>141</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>5</td>
<td>41</td>
<td></td>
</tr>
</tbody>
</table>

Performance comparison with other methods

This experiment is conducted to justify the superiority and performance of the proposed system against the existing method to detect various diseases. For this purpose, we compared the performance of the proposed systems based on accuracy and F1-score. From the results reported in TABLE 4, we can observe that Polignano, M. et al. [3] achieved an accuracy of 94.20% and F1-score of 94.2% while our system achieved an accuracy of 94.83%, precision of 97.24%, recall of 95.49%, and F1-score of 96.36%. We observe an accuracy gain of 0.63% and an F1-score gain of 2.16%. Experimental results and comparative analysis of the proposed system indicate that our HealthConsultantBot is accurate and reliable to be used for the prediction of diseases.
Table 4. Performance Evaluation of the proposed System.

<table>
<thead>
<tr>
<th>Author</th>
<th>Accuracy %</th>
<th>Precision %</th>
<th>Recall %</th>
<th>F1 Score %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polignano, M. et al. [26]</td>
<td>94.20</td>
<td>--</td>
<td>--</td>
<td>94.2</td>
</tr>
<tr>
<td>Proposed</td>
<td>94.83</td>
<td>97.24</td>
<td>95.49</td>
<td>96.36</td>
</tr>
</tbody>
</table>

CONCLUSION

In this research article, we presented HealthConsultantBot, a Facebook messenger-based conversational agent for the user looking to get assistance about their health condition. This conversational agent is designed in a modular fashion so that new features can easily be attached when needed. The conversation with this system is being carried out through the simple text-based interface that makes this system very easy to use for the users. The architecture of this system is divided into three main parts, User interface, Gateway, and server-side Watson implementation. The system understands the user-health condition presented in the natural language then predicts a disease based on the symptoms extracted from user input. This system also provides the precautions to fight against the certain predicted disease. In the future, we are focusing on performance improvement and the addition of new features in our system like patient profile management, food suggestion, and physical activity suggestion based on the user's health condition. We are looking to maintain the record of the user's health condition with time as if it's getting better or patients need to get assistance in some other way.

Acknowledgment. Acknowledgments are considered necessary.

Author's Contribution. All the authors of this article have contributed to its completion, author Asad Ur Rehman gave the idea for starting research on this topic, he proposed the methodology of this research, he gathered the input dataset and other useful material needed for research purpose, he also done work in coding to get results; author Madiha Liaqat supervised the overall progress of work from start to end, author Ali Javed completed the backend coding for getting required results, he analyzed the results as well; author Farman Hassan gave the idea to publish the article in the IJIST Journal, he provided the template and contributed in checking formatting and plagiarism of the article.

Conflict of interest. Authors claim that there exists no conflict of interest for publishing this manuscript in IJIST.

REFERENCES


[23] “Disease Symptom Prediction | Kaggle.”

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.