IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Chatbot Based Self-diagnosis for Disease Prediction using Machine Learning

R.V.S.Lalitha[1][0000-0002-1808-9969], **Surekha Samsani**[2], **Dommu La Selene**[3] , **Yeluri Sai Rohith**[4]

[1]Aditya College of Engineering & Technology, Surampalem, India
[2] University College of Engineering Kakinada (A), JNTUK
[3]DXC - Bangalore
[4]TCS - Hyderabad
rvslalitha@gmail.coom

DOI: 10.48047/ijfans/v11/i12/476

Abstract
Healthcare sector is shifting focus on improving services by adopting new technological approaches. This enriches the functional characteristics of diagnostic data. With the increase in need of Medical services and lack of availability of the resources, Healthcare Chatbot is an attempt to assist common people with primary health care by reducing the burden on medical frontline workers. The objective of this research work is to design a 24/7 available chatbot that answers common medical queries, predicts diseases based on the symptoms and radiology images provided, aids with medication/ precautionary measures that are to be followed. A chatbot can provide a customized one on one interaction through text or voice interface and gives reply using Artificial Intelligence. It responds differently to messages containing certain keywords and uses Machine Learning to adapt their responses to fit the situation. The healthcare chatbot handles a large number of requested queries at a time making it reliable to use. The chatbot responds to medical queries only to the best of its knowledge database.

Keywords: RASA, Chatbot, CNN, Regression, Radiology Images, COVID19

Introduction

In India, with such a huge population it is very difficult to provide proper medical services and vaccination to the citizens which now has three categories viz., symptomatic, asymptomatic and non-infected. It is chaos in analyzing diseases like COVID-19, flu, influenza and pneumonia which are all similar to an extent. Many apps were released to help people resolve their queries but nothing proved to solve these common queries. A chatbot that can answer common queries of diseases and predicts disease with user given symptoms and when radiology images are uploaded, provides diagnosis/precautions to be taken as per the disease, gives info of the availability of nearby related hospitals too.

Literature Survey

The authors S. Shaikhet al., performed prediction on COVID-19 outbreak using Linear, Polynomial Regression models approach of Machine Learning.[1] The authors R. Kumari et al., presented a study on the developed prediction models for cases in India caused by COVID-19 using multiple linear regression, autocorrelation and auto regression [2] .The authors R. Lee et al., did analysis on whether a person might have diabetes or not by comparing 5 prediction algorithms which are Artificial Neural Networks, Logistic Regression, KNearest Neighbors, Decision Tree and Random Forest algorithm [3]. The author H. Q. Yu proposed an ecosystem using ML and NLP for discease prediction [4]. The authors R. B.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Mathewet al., proposed a recommendation system using chatbot for medical diagnosis using NLP and ML [5]. The authors M. Rohiniet al. analyzed COVID-19 using machine learning algorithmK-Nearest Neighbors (KNN) [6]. The authors S. Grampurohit et al., worked on classifier systems using ML classifiers [7]. The authors Jain, R et al., studied CNN models and compared for COVID-19 impact using chest X-ray images [8]. The authors Madaan, V et al., proposed image classification algorithm based on X-ray images called XCOVNet for early COVID-19 detection using CNN model [9]. The authors Umer, M et al., did research to predict COVID-19 with images of heart and Keras ImageDataGenerator [10]. The authors Singh et al., proposed an automated analysis of chest CT images for classifying whether COVID infected or not using Convolution Neural Networks (CNN) by tuning with multi-objective differential evolution MODE [11]. The authors Das et al., developed a deep learning model using X-ray images[12]. The authors, Puttagunta et al., carried out research on development of DLA for disease prediction using medical imaging applications [13]. The authors Abbas et al., suggested a deep CNN approach named as DeTraC, for the classification of COVID-19 chest X-ray images from various syndrome cases14]. The authors S. Bahri et al., worked on LSTM neural network for predicting the next COVID-19 recovered cases in USA, India and Italy for seven days ahead[15]. The authors Muhammadet al., used Supervised Learnig models and ANN for COVID-19 infection prediction on Epidemiology dataset.[16]

Query based Self Diagnosing Chatbot for COVID19

The objective of the Research is to predict the diseases from the given symptoms. A chatbot is developed to answer common medical queries. Along with, prediction of diseases based on the Symptoms and Radiology images provided. This aids with medication/facilities and precautionary measures that are to be followed are also articulated.

Data Gathering

The database is created by scraping the data from the web. First, selected diseases information are scraped from nhp.gov.in, cdc.gov, seattlechildrens.org, nhsinform.scot and medicinenet.com websites. About 8000 diseases information is created and a Selenium script which searches the data from Google and scrapes the data available in the box that appears like this(Fig.1).

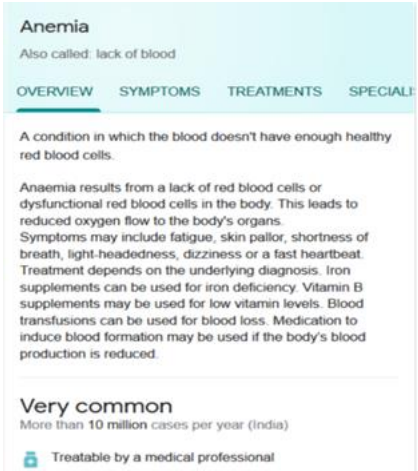

Fig.1.Search for Anemia from Google using Selenium script

The Selenium script will scrape the data from these and stores it in a Json format and data cleaning is performed which can be used for further data processing (Fig.2 and Fig.3). Chest X-ray images and pneumonia are collected from Kaggle.com and used for ML analysis.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

```
    # opening url in google
    driver.get("https://www.google.com/search?q="+disease.lower()+" symptoms")
    #result_dict[disease]=dict()
    time.sleep(2)
# SYMPTOMS
    otherName=driver.find_elements_by_xpath(othername_div)
    extraName=driver.find_elements_by_xpath(extraname_div)
    diagnoseType=driver.find_element_by_xpath(diagnose_div)
    symptomDesc=driver.find_element_by_xpath(symptom_desc_div)
    symtpoms=driver.find_element_by_xpath(symtpoms_div)
    #result_dict[disease]["otherName"]=otherName[0].text
    ## creating dict with name in google
```

Fig.2. Scraping of Symptoms

```
"Anemia": {
    "extraName": "Also called: lack of blood",
    "diagnoseType": "Requires a medical diagnosis",
    "symptomDescription": "Symptoms may include fatigue, skin pallor, shortness
    "symptoms": {
        "Whole body": "dizziness, fatigue, light-headedness, or malaise",
        "Heart": "fast heart rate or palpitations",
        "Also common": "brittle nails, headache, pallor, shortness of breath, or
    },
    "occurance": "Very common",
```

Fig.3. JSON data of symptoms

Preparing Dataset

After gathering all the data we obtained 491 diseases with 890 unique symptoms. For, building the dataset we created a dataset such that it will take the diseases and perform combinations of symptoms of that particular disease and append it to the dataset. If a particular disease has N symptoms, $2^{(N)}$-1 rows are obtained in the dataset. So in the dataset there will be a total of 891 columns as we have 890 diseases and one column will be of the disease name and if a particular symptom is present, it is marked as '1' in that column, otherwise '0'. In this way, the entire dataset is created(Fig.4)

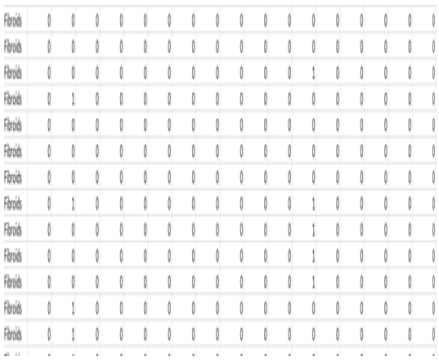

Fig.4. Dataset for symptoms

Fig.5. Architecture of chatbot

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

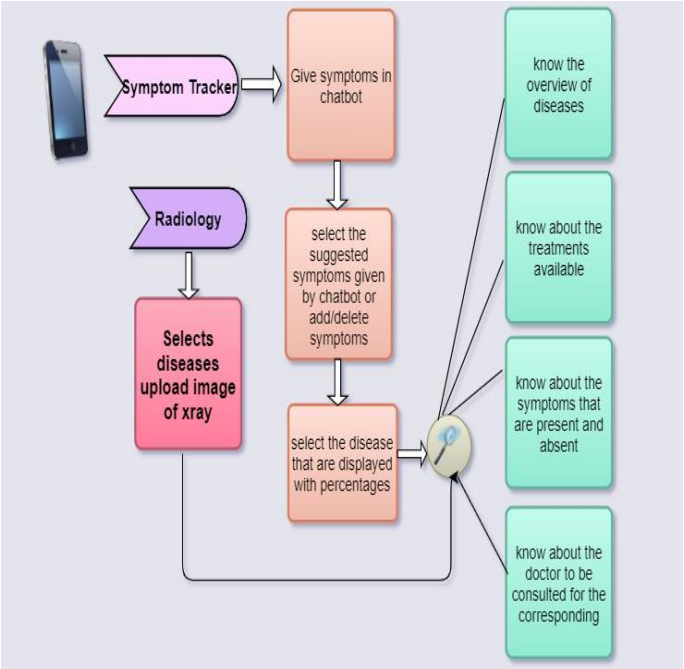

Fig.6. Workflow of Chatbot

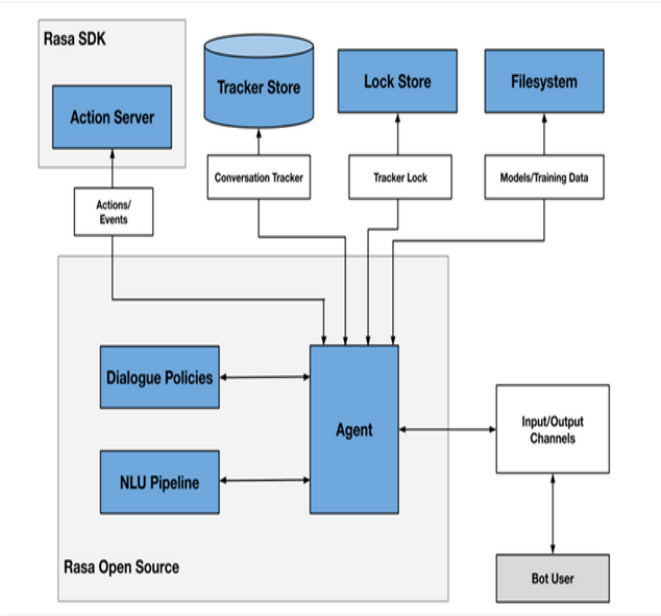

Fig.7. RASA architecture

Fig.5 and Fig.6 illustrates architecture and workflow of chatbot. Flutter is a cross platform application where we can develop android, iOS, web apps on a single code base. We used the REST API service available in the flutter module to connect to our chatbot through the rasa REST channel which is deployed in the Heroku server. Users can directly create an account and login using the credentials to access the services provided by the application. No data of user chat is stored on the databases or in the native applications. As the

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

architecture best chooses our problem, RASA is chosen over other alternatives(Fig.7). We can connect to rasa using the rest api service pro-vided by rasa.

Performance Evaluation Measures

In the Disease prediction through Symptoms, user will be able to get the prediction of a disease if the probability of that particular disease meets a threshold value of 55%. Accuracy obtained using Radiology images is 98.33% and Pneumonia is 91.18%. Based on the symptoms given, the user can get the probability of the occurrence of the disease, description, treatment. There is a feasibility to upload an X-ray image of lungs, for which the bot prompts the message as per the prediction of either the presence of Covid-19 or Pneumonia. NLU manages classification, extraction and response so as to decide the next action to be occurred.

Symptoms prediction ML Modeling

After creating the dataset, this is tested on different models like multinomial Naïve Bayes, Random Forest, KNN, Logistic Regression, Support Vector Machine and Decision Tree. When a user utters a symptoms query then all the related symptoms pop up and the user can select the symptoms he is suffering with and can skip the remaining symptoms. All the selected symptoms go to one list and remaining symptoms go to the removed list and those questions will not be asked again. In this way we can gather all the symptoms the user is suffering with and can predict the probability of occurrence of a particular disease using Logistic regression. Fig.7 represents architecture and Fig.14 shows accuracies obtained. We got a 91.02% Cross Validation accuracy for Logistic Regression.

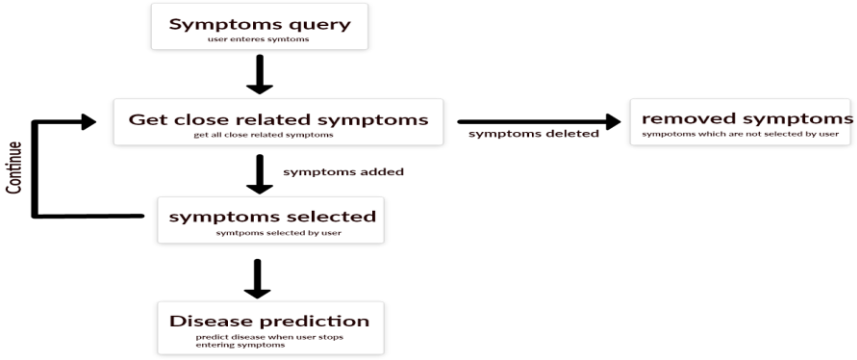

Fig.7. Various phases for disease prediction

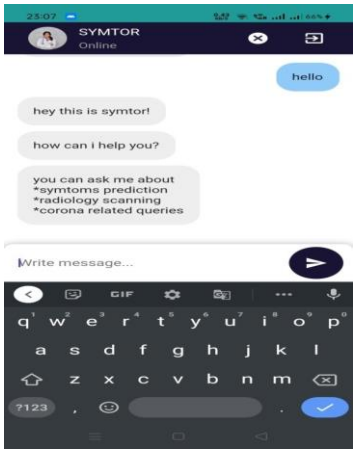

Fig.8. Symptom Analysis

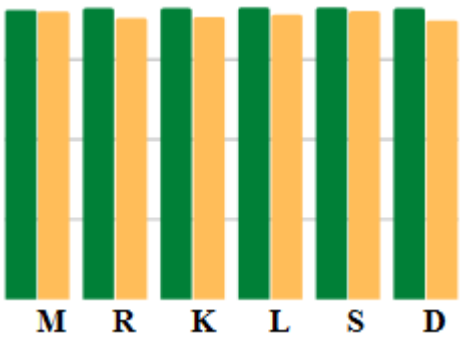

Fig.9. Accuracies of ML models.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

CNN can be used to handle variance in images – translation, rotation, illumination, size. Also, captures the spatial features from an image and handles overfitting(Fig.10)

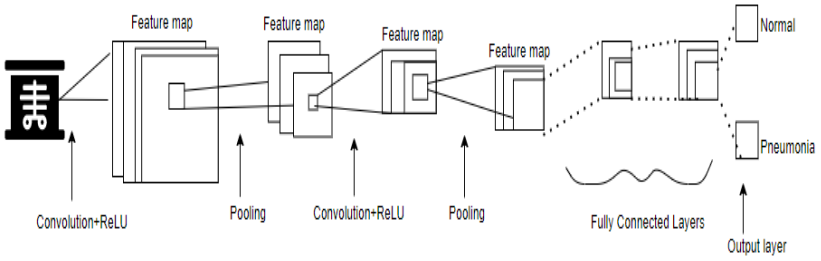

Fig.8. CNN model for Pneumonia

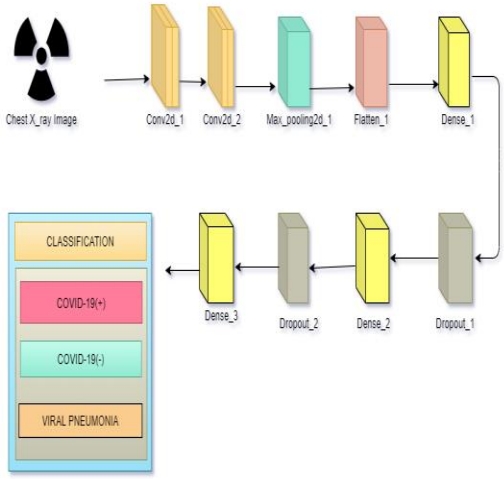

Fig.10. CNN model for classification

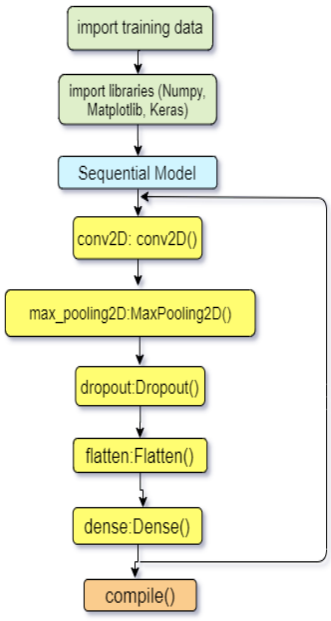

Fig.11.Covid-19 Detection – CNN Model

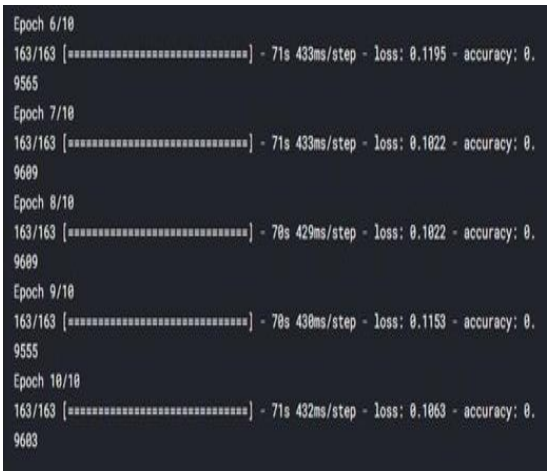

Fig.12. Accuracies observer over 10 epochs.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Results and Discussion

The sample outputs for various queries are presented through (Fig.13-20)

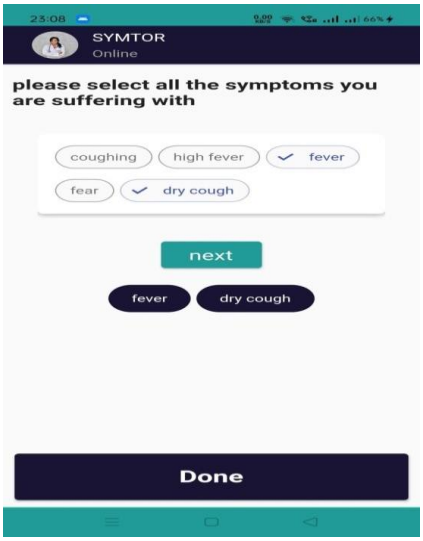

Fig.13Selection of symptoms

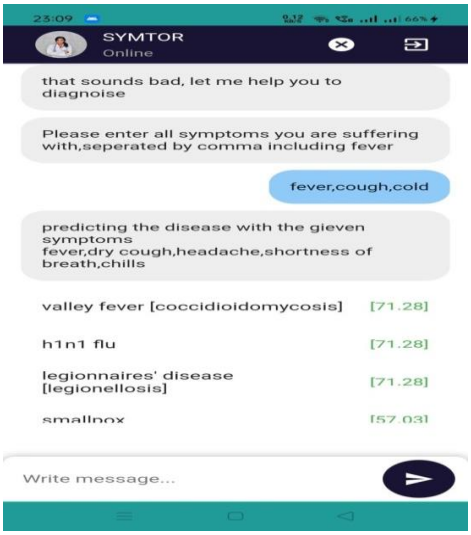

Fig.14 Disease prediction with >50%

Fig.15. About treatments

Fig.16 Specialists

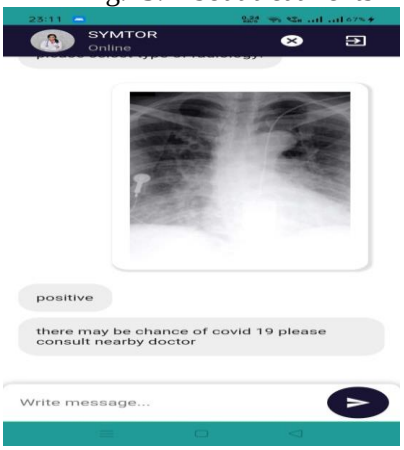

Fig.17. COVID 19 prediction

Fig.18. COVID cases in Andhra

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

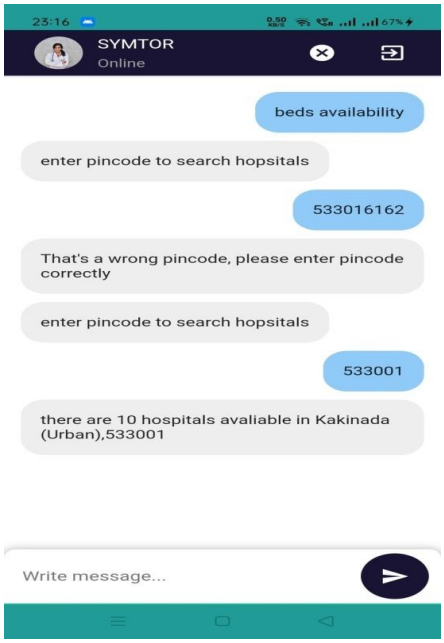

Fig.19. Hospitals

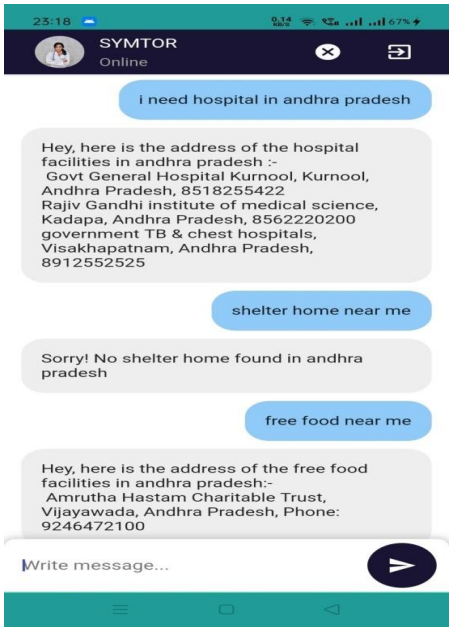

Fig.20. Hospitals and food in AP

Conclusion

Customized chatbots can reach out to a large audience on messaging apps and be more effective than humans. This kind of chatbot makes the lives of people easy by providing diagnosis remotely, especially in pandemic situations. A Medical chatbot has has broad scope where they can lend a helping hand to both patients and physicians. This kind of automation is being seen and is reliable to use with much advanced learning algorithms in the future.

References

1. D S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburgh, Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting,Confluence,2021.
2. R. Kumari et al., Analysis and predictions of spread, recovery, and death caused by COVID-19 in India, in Big Data Mining and Analytics.
3. R. Lee and C. Chitnis, Improving Health-Care Systems by Disease Prediction, CSCI, 2018.
4. H. Q. Yu, Experimental Disease Prediction Research on Combining Natural Language Processing and Machine Learning, ICCSNT, 2019.