International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1599-1607 www.ijaem.net ISSN: 2395-5252

AI HealthCare

Sumit
Department of Information Technology, Maharaja Agrasen Institute of Technology, New Delhi – 110086

Submitted: 01-06-2021 Revised: 14-06-2021 Accepted: 16-06-2021

ABSTRACT: Healthcare bot is a technology that makes interaction between man and machine possible by using Artificial Intelligence with the support of dialog flow. Now a day people tend to seek knowledge or information from internet that concern with health through online healthcare services.
To lead a good life healthcare is very much important. But it is very difficult to obtain the consultation with the doctor in case of any health issues.
The basic aim of this system is to bridge the vocabulary gap between the doctors by giving self-diagnosis from the comfort of one's place. The proposed idea is to create a medical chatbot using Artificial Intelligence that can diagnose the disease and provide basic details about the disease before consulting a doctor. To reduce the healthcare costs and improve accessibility to medical knowledge the medical bot is built. Certain bots act as a medical reference books, which helps the patient know more about their disease and helps to improve their health. The user can achieve the real benefit of a bot only when it can diagnose all kind of disease and provide necessary information. Hence, people will have an idea about their health and have the right protection.

I. INTRODUCTION

The proposed idea is to create a medical chatbot using Artificial Intelligence that can diagnose the disease and provide basic details about the disease before consulting a doctor. To reduce the healthcare costs and improve accessibility to medical knowledge the medical chatbot is built. Certain chatbots acts as a medical reference books, which helps the patient know more about their disease and helps to improve their health. The user can achieve the real benefit of a chatbot only when it can diagnose all kind of disease and provide necessary information. A text-to-text diagnosis bot engages patients in conversation about their medical issues and provides a personalized diagnosis based on their symptoms. Hence, people will have an idea about their health and have the right protection.

In India the government doctor to patient ratio is 1:10189 and WHO recommends the ratio of 1:1000 which means India is in the deficit of 6,00,000 the doctors.
So, many of the people are not able to get required diagnosis
Due to this the rate of private hospitals are also increased which common people cannot afford. In some places the hospitals are far away from the person's home. Some people are not very comfortable with talking with doctor about their issues.
The proper time that a doctor should give to a patient is also much less which sometimes leads to criticalities.
Privacy and security are also a problem these days.

Healthcare chatbots will never substitute doctors. But they provide plenty of opportunities to facilitate their job or to improve their performance Acknowledging the improvements in information and communication technology. When talking about the health care industry, the possibilities for chatbots grow are relatively high. Although the adoption rate is not yet widespread. The number of healthcare chatbots is increasing. As health services become patient-centric, offering personalized and satisfactory experience is of utmost priority for healthcare providers**.**

It is being proved that in the unusual case of chatbots, we have a technology where we are not forced to choose between cost and efficacy; the chatbots replacing humans for certain functions definitely makes the process more effective while saving money for both in the short and long terms. A win-win situation which can only get better as the bots become more and more "learned" in their tasks.

These AI implementations in healthcare shows us how the chatbots are improving the state of healthcare in India and will be going very far with its use in a greater number of tasks. It will increase reliability and cost effectiveness to the current scenario of health which proves chatbots to be a boon to mankind
The Goal is to introduce Health Bot, a system designed to improve the eHealth paradigm by using a webapp to simulate human interaction in medical contexts. Based on Machine Learning and Artificial Intelligence techniques, the webapp is able to overcome the limitation of classical human machine interaction, thus removing bias and allowing the

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

patient to a freer and natural communication. A webapp can successfully be designed to work as a helping tool in doctor-patient communication, but it must be emphasized that it should work as a supplement and never replacement.

All healthcare providers are always willing to help their patients and they understand how it is vital to be available if there is urgent need of medical attention. Unfortunately, doctors have limited time and a lot of patients which doesn't allow them to be available anytime. In their turn, webapp are there for those who need medical assistance at all time. Furthermore, virtual assistants may be responsible for reminding users to take their medicine and monitoring a patient's health status.

Chatbot as Health Bot

As chatbots in healthcare are highly in demand, medical institutions can offer various services from symptom checking and appointment scheduling to dealing with additional questions. People are able to get answers to their additional questions with the help of chatbot. There is no need for them to call the clinic to clarify some misunderstanding.

Nationally, medical attention was missing in the case of nearly one-third of all deaths. It is likely that most of those deaths warranted some kind of medical attention, and that such attention would have been provided if the households in which these deaths took place were as rich as the top 15%.

Heart Disease

A disease is an unnatural medical condition that negatively affects the functional state of an organism and is generally associated with certain signs of illness. As reported by World Health Organization (WHO), Heart Disease and Stroke are the world's biggest killers and have remained the leading causes of death globally in the last 15 years. In the direction of predicting heart disease, Machine Learning can present remarkable features that simplify the identification of unseen patterns, eventually providing clinical insights that assist physicians in planning and providing care. The Heart Disease dataset has been taken from Kaggle. This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. It has a total number of 303 rows and 14 columns among which 165 have a heart disease.

Liver Disease

Liver disease is any disturbance of liver function that causes illness. The liver is responsible for many critical functions within the body and should it become diseased or injured, the loss of those functions can cause significant damage to the body. Patients with Liver disease have been continuously increasing because of excessive consumption of alcohol, inhale of harmful gases, intake of contaminated food, pickles and drugs. This data set contains 416 liver patient records and 167 non liver patient records collected from North East of Andhra Pradesh, India. The "Dataset" column is a class label used to divide groups into liver patient (liver disease)

Breast Cancer

Breast cancer can occur in women and rarely in men. Symptoms of breast cancer include a lump in the breast, bloody discharge from the nipple and changes in the shape or texture of the nipple or breast. Its treatment depends on the stage of cancer. It may consist of chemotherapy, radiation, hormone therapy and surgery. The dataset has been taken from UCI machine learning repository. The main objective of the analysis is to perform classification of tumors i.e., benign(B) or malignant(M).

II. METHODOLOGY

1. Data Collection

This database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patients at New York Presbyterian Hospital admitted during 2004. The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of the disease, and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength of association. The method used the Med LEE natural language processing system to obtain UMLS codes for diseases and symptoms from the notes; then statistical methods based on frequencies and co-occurrences were used to obtain the associations. Dataset is taken from website using a method called Data Scrapping.

Data scraping, also known as web scrapping, is the process of importing information from a website into a spreadsheet or local file saved on your computer. It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0392680_shortness of breath
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall
		UMLS:C0039070_syncope
		UMLS:C0042571_vertigo
		UMLS:C0038990_sweat|UMLS:C0700590_sweating increased
		UMLS:C0030252_palpitation
		UMLS:C0027497_nausea
		UMLS:C0002962_angina pectoris
		UMLS:C0438716_pressure chest

Figure 1: Example of Dataset(Disease – Symptoms)

Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio
60	Male	0.5	0.1	500	20	34	5.9	1.6	0.37
40	Male	0.6	0.1	98	35	31	6	3.2	1.1
52	Male	0.8	0.2	245	48	49	6.4	3.2	1
31	Male	1.3	0.5	184	29	32	6.8	3.4	1
38	Male	1	0.3	216	21	24	7.3	4.4	1.5

Figure 2: Example Dataset of Liver Disease

age	sex	cp	trestbps	restecg	thalach	exang	oldpeak	slope	ca	thal
41	1	1	120	1	182	0	0	2	0	2
38	1	2	138	1	173	0	0	2	4	2
67	1	0	160	0	108	1	1.5	1	3	2
67	1	0	120	0	129	1	2.6	1	2	3
62	0	0	140	0	160	0	3.6	0	2	2

Figure 3: Example Dataset of Heart Disease

texture_mean	perimeter_mean	smoothness_mean	compactness_mean	concavity_mean	concave_points_mean	symmetry_mean	radius_se	compactness_se	concavity_se	concave_points_se
20.68	108.1	0.117	0.2022	0.1722	0.1028	0.2164	0.5692	0.02501	0.03188	0.01297
22.15	130	0.09831	0.1027	0.1479	0.09498	0.1582	0.7582	0.01893	0.03391	0.01521
14.36	87.46	0.09779	0.08129	0.06664	0.04781	0.1885	0.2699	0.0146	0.02387	0.01315
15.71	85.63	0.1075	0.127	0.04568	0.0311	0.1967	0.1852	0.01898	0.01698	0.00649
12.44	60.34	0.1024	0.06492	0.02956	0.02076	0.1815	0.2773	0.01432	0.01985	0.01421
14.26	102.5	0.1073	0.2135	0.2077	0.09756	0.2521	0.4388	0.05328	0.06446	0.02252

texture_worst	smoothness_worst	compactness_worst	concavity_worst	concave_points_worst	symmetry_worst	fractal_dimension_worst
31.48	0.1789	0.4233	0.4784	0.2073	0.3706	0.1142
30.88	0.1512	0.315	0.5372	0.2388	0.2768	0.07615
19.26	0.144	0.1773	0.239	0.1288	0.2977	0.07259
20.49	0.1312	0.2776	0.189	0.07283	0.3184	0.08183
15.66	0.1324	0.1148	0.08867	0.06227	0.245	0.07773
19.08	0.139	0.5954	0.6305	0.2393	0.4667	0.09946

Figure 4:Example of Breast Cancer

2. Naïve Bayes Classifier Algorithm

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems

Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, which can be described as

Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of colour, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.

Bayes: It is called Bayes because it depends on the principle of Bayes Theorem

)

Bagging Classifier

A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction. Such a meta-estimator can typically be used as a way to reduce the variance of a black-box estimator (e.g., a decision tree), by introducing randomization into its construction procedure and then making an ensemble out of it.

Random Forest

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. (Liver Disease

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

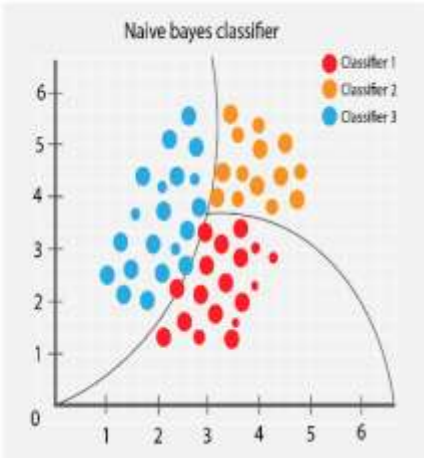


Figure 5: Naïve Bayes Classifier Formula

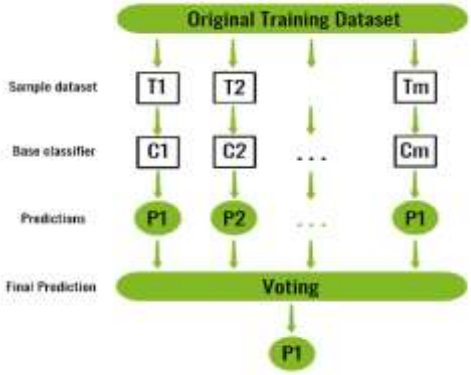


Figure 6: BaggingClassifier

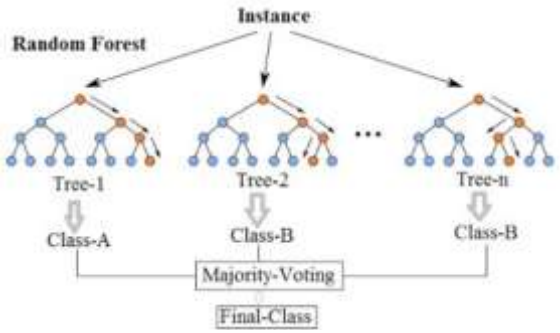


Figure 7: Random Forest

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

3.Web-application Deployment

Flask micro-framework would be used to make to web application, for easy integration with Naïve Bayes Model in the backend. Flask Rest API would act as a bridge between frontend and the backend. User will be given some dropdown options to select the symptoms they are facing. Their response will be sent to the backend, analyzed by the Machine Learning model and the result will be sent back.

III. DESIGN AND IMPLEMENTATION

The proposed system will interact with user via web-based platform. After that the application asks for symptoms that the patient is facing. The user inputs the symptoms to which the application responds and provides the disease it predicts the patient can have.

Software will take symptoms into account and then try to predict the disease. System may ask for more symptoms from the user or will suggest some symptoms just to make sure the disease the patient is having is the same disease the system is predicting.

After conformation of symptoms system will classify the disease as major or non-major, if the disease is major then system will try to recommend doctors in that pin area

The webapp can answer to the users based on their disease related queries based on symptoms, causes,and prevention or medicine suggestion.

1. Data Scrapping:

Data scraping, also known as web scrapping, is the process of importing information from a website into a spreadsheet or local file saved on your computer. It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website. This method has been used to get data from the Website and save and process it in local computer after pre-processing

2.Preprocessing of Data:

Data pre-processing in Machine Learning is a crucial step that helps enhance the quality of data to promote the extraction of meaningful insights from the data. Data pre-processing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models.

In this the data pre-processing steps that are done are:

- Data Cleaning
- Feature Scaling
- Data Transformation
- Data Reduction
- Feature Selection Based on p-value

3. Using Sklearn:

After checking various algorithms, the Naive Bayes algorithm gives the best results in this main Dataset. In machine learning, Naïve Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong(naive) independence assumptions between the features.

Training this Naïve Bayes model on dataset took about 5-10 minutes when trained on CPU and accuracy above 92% was achieved.

In Liver Disease Dataset,RandomForestClassifiergives the best results with an accuracy of 90% using sklearn In Heart Disease Dataset Bagging Classifier gives the best results with an accuracy of 98% using sklearn

In Breast Cancer Dataset Bagging Classifiergives the best results with an accuracy of 95% using sklearn

4.Saving and Deploying the model

This model is deployed as a webapp using Flask, HTML and CSS on Heroku platform.

Heroku is a platform as a service based on a managed container system, with integrated data services and a powerful ecosystem, for deploying and running modern apps

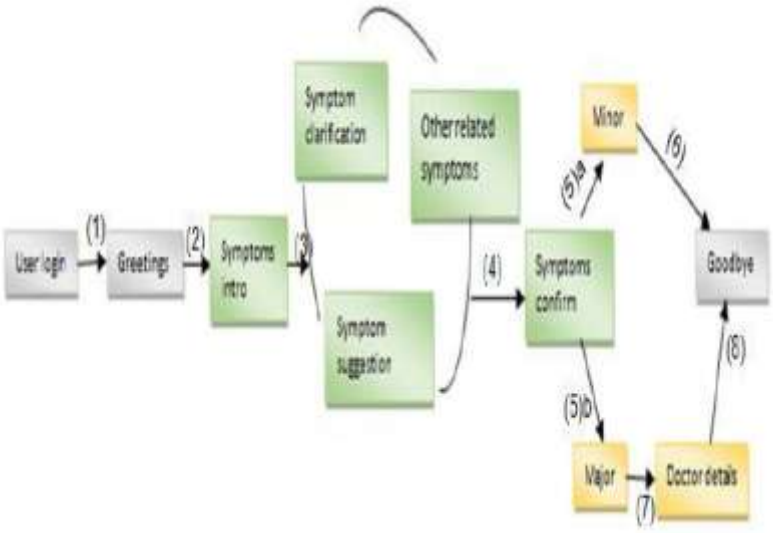
5.Proposed Workflow

Idea is to provide user with all the set of symptoms and predict disease from it, Symptoms will be like: Fever, chills, Pimples, Cough, Sneezing etc. There will be more than 500 symptoms to choose from.

The user will choose the symptoms he/she is facing from the dropdown. The symptoms are passed on to a Naive Bayes trained machine learning model. This model will then predict the disease on the provided symptoms and all the necessary details about the disease.

The webapp asks the user to choose the symptoms from the dropdown provided to which the user responds and the webapp provides all the information regarding the disease it predicted based on the symptoms provided by the user. The system will identify the symptoms by providing the symptoms to the trained machine learning model which is trained on Naive Bayes algorithm which in turn will predict the disease according to symptoms.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252



IV. RESULT

According to the model given above, the final accuracy is stable at more than 92%. The following figure shows the accuracy of my model

```
In [35]: mnb.score(X,y)*100
Out[35]: 92.10526315789474
```

Accuracy can be further increased if we increase the number of symptoms to be taken by user, the more symptoms we take the better the model will get. If Symptoms taken are less then

Prediction accuracy is also less are there will be many diseases with a single symptom

Hence if we increase the symptoms the better result, we will get. Here accuracy of 92% is achieved when the symptoms taken by user are only three every time

In Liver Disease Dataset, RandomForestClassifiergives the best results with an accuracy of 90% using sklearn

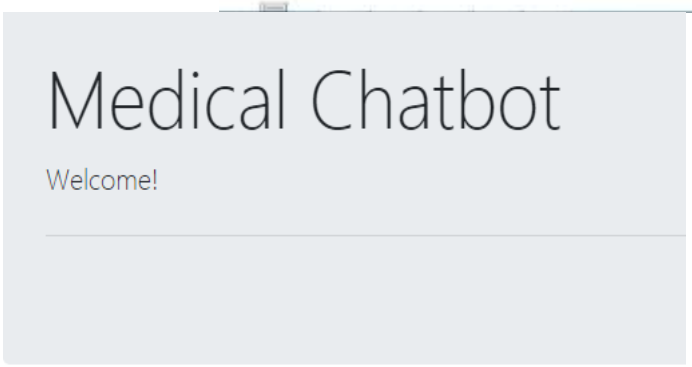
In Heart Disease Dataset Bagging Classifier gives the best results with an accuracy of 98% using sklearn

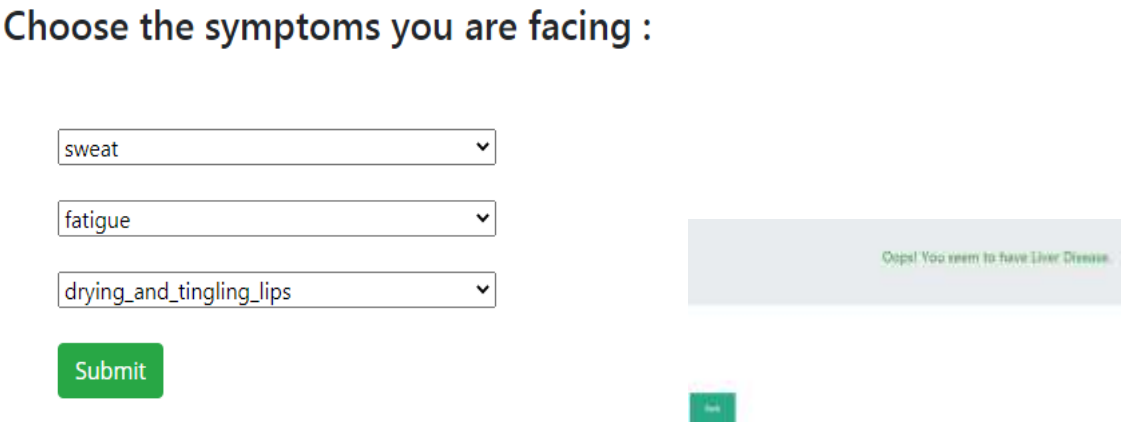
In Breast Cancer Dataset Bagging Classifiergives the best results with an accuracy of 95% using sklearn

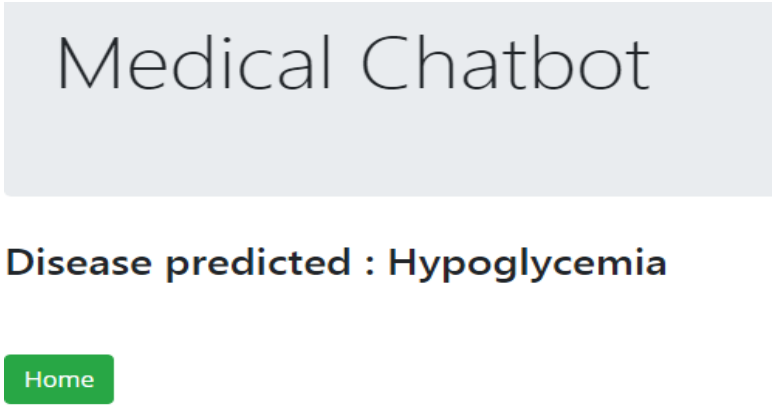
-Web Application Result

The user will choose the symptoms he/she is facing from the dropdown. The symptoms are passed on to a Naive Bayes trained machine learning model. This model will then predict the disease on the provided symptoms and show it on another page.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252







International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1283-1289 www.ijaem.net ISSN: 2395-5252

V. CONCLUSION

A medical bot provides personalized diagnoses based on symptoms. In the future, the bot's symptom recognition and diagnosis performance could be greatly improved by adding support for more medical features, such as location, duration, and intensity of symptoms, and more detailed symptom description. I am working on adding a new chatting system in the webapp which will make webapp more interactive to use. At last, the implementation of personalized medicine would successfully save many lives and create a medical awareness among the people. As said before, the future era is the era of messaging app because people going to spend more time in messaging app than any other apps. Thus, medical bot has wide and vast future scope.

VI. FUTURE SCOPE AND IMPROVEMENTS

This diseaseprediction method is fast enough so that it can be used by all the websites for Healthcare process. Also, it provides overall better user experience.

It can be further improved by implementing following:

• Saving user past history of disease and Medication
• Increasing the dataset and refining it more
• Including more symptoms.
• Including the duration of major symptoms user have

VII. ACKNOWLEDGMENT

This project is carried out by Sumit, studying Engineering in Information Technology from Maharaja Agrasen Institute of Technology.

REFERENCES

[1]. SimonHoermann, Kathryn L McCabe, David N Milne, Rafael A Calvo1,"**Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review**", Journal of Medical Internet Research ,volume: 19 , issue 8 , August 2017.

[2]. DivyaMadhu,Neeraj Jain C. J, ElmySebastain, ShinoyShaji, AnandhuAjayakumar**," A Novel Approach for Medical Assistance Using Trained Chatbot**",International Conference on Inventive Communication and Computational Technologies(ICICCT 2017).

[3]. Abbas SaliimiLokman, JasniMohamadZain,FakultiSistemKomputer, KejuruteraanPerisian," **Designing a Chatbot for Diabetic Patients**",ACM Transactions on Management Information Systems (TMIS), Volume 4, Issue 2, August 2015 .

[4]. Marr, B. (2016). **How Machine Learning, Big Data And AI Are Changing Healthcare** Forever.

[5]. Retrieved from https://www.forbes.com/sites/bernard marr/2016/09/23/how-machine learning big-data-and-ai-are changing-healthcare-forever/ - 70f924e11a1c.

[6]. Gillian Cameron, David Cameron, Gavin Megaw ,RaymondBond,MauriceMulvenna ,Siobhan O'Neill, Cherie Armour, Michael McTear,"**Towards a chatbot for digital counselling**",Journal of Medical Internet Research, 4(1), pp. e3.