iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42985

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

A Machine Learning based Healthcare Diagnostic Model

Prakhar Mittal[1], Udit Parasher[2], Ruhi Khanna[3], Parv Yadav[4], Sunil Kumar[5], Mariya Khurshid[6]
[1, 2, 3, 4, 5, 6]*Department of Computer Science & Engineering, Meerut Institute of Engineering and Technology, Meerut*

Abstract: ***This To live a healthy life, healthcare is considered very important but it is terribly difficult to get consultation from a doctor nowadays. The basic idea and motive of the project is to build a chatbot system by using AI (Artificial Intelligence) that can predict the health problems and give details of the disease before consulting the Doctor. The system provides text assistance to coordinate with the chatbot.***
Based on user's symptoms chatbot will provide what kind of disease user have and provide doctor's details according to the diagnose. Based on the symptom that user have chatbot will tell user the disease with all the other possible symptoms. The chat will inform whether the disease is major or minor. Hence, user can gain the maximum benefit of this AI based chatbot only after it diagnoses maximum kinds of illness and provide with all the required information(Abstract)
Keywords: Online Diagnosis, Chatbot, Healthcare, Diseases (key words)

I. INTRODUCTION

Nowadays, health care is very important in our life. Today's people are busy with their work at home, once work and are more addicted to the Internet. They are not concerned about their health. So they avoid hospitals for small problems It may become a major problem. So we can provide an idea, which is to develop a health care chatbot system using AI that can analyze the disease and give idea of basic information about the disease before consulting a doctor. Which helps the patients know more about their illness and improves their health. Users can achieve all kinds of disease information. The system uses question and answer procedure to diagnose, chatbot communicate with the user. Bot will diagnose which type of disease you have, based on user symptoms that user will provide in the form of Yes or No and chatbot also gives the details of the doctor for particular diseases. It may reduce their hustle to find a doctor and disease by using this application system. This system is developed to control and reduce the healthcare cost and time of the users as it is hard for the patients to visit the doctors or experts immediately whenever needed and it may save their time if they are in panic for no reason as they don't have any idea about the disease.

II. RELATED WORK

In [1] the author suggested this method, the methods used to support vector machine algorithms, NLP, Word similarity between Word orders and sentences. A heart disease database is used. Its advantage is that SVM can be used to solve complex problem classification & additional training. Its disadvantage is the poor interpretation of NLP. The author of [2] proposed a model that uses Ensemble reading in which classifier is automatically trained and the final decision is made by a majority vote, a group-based collection that includes multiple votes as a whole. The general health database and the Pima Indian diabetes database are used. The advantage of this model is that there are no dominant dividers present and your disadvantage is that the calculation and design time are high. The author [3] used the Information Graph & the categorized bi-directional attention approach in which the architecture of the cross QA model Framework, integrates the information graph to control the medical database and HBAM to understand the text. A data set of 3,500 organizations (including 674 diseases and 2824 symptoms) and 4,500 contacts were used. Its advantage is that it uses a systematic retention to help ease the retention and acquisition of certain domain information and its disadvantages that complex function exists. In [5] NLU, NLTP, Multinomial Naive have used methodologies where Emotional Analysis, Tokenization, Named Entity Recognition, Normalization, Dependency Parsing are used. The Corpus_words dataset and class_words data set are in use. Its advantage is that it is easy to build and your downside is that no data is provided about this disease. In [6] RNN, NLP, verbal and textual methods are used in this model. Sequence-to-Sequence model, the Apriori algorithm is used. This model is based on a database found at New York Presbyterian Hospital. This first principle can minimize the number of items needed to be checked. This model requires a lot of training time even though the hardware is capable of handling it. This model [7] used the method of forcing Teachers into their proposed system. Medical advice maker and a common sympathetic general generator with four identical LSTM layers tracked by Concatenation, a Facebook AI Empathetic Dialogue data set and a medical question response set.

The accuracy of the target separator was 98.5% and the emotion detector was 92.4%. Paper [8] proposes a model using a text mine via Wit.ai and uses API Medic methods. Glove vectors and API Medic algorithms are used for this. Demographic information survey, description of the natural language of the symbols, additional description of symbols and hypothetical diagnoses and the API Medic website. It removes the site from the medic API and it is easy to check for symptoms. It does not give accurate results. [11] Incorporates graph-based discussion information based on factoid medical questions, using algorithms such as native language translator, chat manager, and natural language generator. An RDF data set is applied to it. It manages the conversation, marks the missing information, and produces accurate and contextual answers.

This [13] prototype has a Facebook chatbot for sexual health statistics on HIV / AIDS. It uses NPC Editor to drive answers to chatbot, chat manager, and plug-in to Face-book. The layout and NLP algorithms are helpful for this purpose. It uses an online survey, QA in SHIH bot Domain as a database. Live chats will showcase the SHIH boat's ability to understand new questions, the chatbot's ability to deal with pointed questions without background information, and the full flow of the conversation. A study [14] on the use of chatbot in health care using NLTK and an algorithm such as Natural language analysis is performed. Dataset is a QA record for it. It is easy to use and can be used by anyone who knows how to write in their own language in a mobile app or desktop version, and provides a personalized diagnostic based.

III. MODEL ASSUMPTION

The next section discusses the major libraries and databases used in modeling. Chatbots are the helpers of the new age of the people so that the demands can be met in a timely manner, and the waiting time for the real helpers can be reduced. A health-based model that integrates such an approach can help frustrated health care systems to function better.

A. Platform Structure

Python language is being used in the paper. It is a standard programming language for a general purpose program. The syntax of python is very simple compared to the English language. Use of python differs from the software field to create workflow, it can be used to connect to a database, manage big data and perform complex calculations. To create a GUI again, using the python language.

B. Model Libraries

1) Numpy – A basic computer science package. It is ued as an efficient multi-dimensional data container for collected or comprehensive data.
2) Pandas – It is an open library that provides efficient data structures and data analysis tools.
3) Matplotlib-It is a Python 2D editing library that produces publishing quality statistics in a variety of hardcopy formats and interactive scenarios across all platforms.
4) Text Blob- It is a Python Library for processing text data. It provides a simple API for logging into NLP activities.
5) Tkinter- Tkinter is a standard Python GUI library. Python when combined with Tkinter provides a quick and easy way to build GUI applications. Tkinter provides a strong visual-based interface to the Tk GUI tool kit.

C. Dataset Generation

The sample data we used to predict disease according to user-selected indicators, from now on the data set includes 4921 rows and 133 columns. This database is collected from Kaggle [15]. This includes 41 different types of diseases and their several symptoms. Also, we have a Doctor database [16] that will help recommend the right doctor for the user according to his or her condition. She can visit a Doctor's office and book an appointment whenever she wishes to see a doctor. Table 1 contains a description of the data in the Doctor Dataset. However, the symptom set data contains 133 columns in the Boolean form for all symptoms and disease.

Table 1. Data Description of Doctor's Dataset

S.No.	Data Column	Data Type
1.	Doctor"s Name	String
2.	Website Name	String

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

IV. IMPLEMENTATION AND INTERPRETATION

This segment focuses on the basic construction and operation of the planned model for the purpose diagnosing health care.

A. *Basic Framework*

1) As mentioned, this paper discusses data based on health care. The first step here will be data cleaning and then the machine learning process data training using the Decision Tree Classifier.
2) Then, a chatbot is built whose response will be based on a pre-trained class divider.
3) After a successful diagnosis the chatbot also gives the patient the recommendation of a specialist doctor.
4) To visualize the whole model, a GUI is created that will help the patient to use the model easil

B. *Procedure*

1) *Data Training:* The project involves two different kinds of data sets that is: training data set and a doctor data set, where the first is a training data set consisting of 133 columns or we can say 133 different types of features based on what we are. I will get the resulting prediction. Predictions are class labels and for all 41 diseases from now on. Therefore, based on the given indications we will conclude with some prognosis. This training data contains data for 4921 different types of people suffering from different types of symptoms. These symbols are labeled with Boolean values which means that the person has or does not have that mark. So here 1 means that the person has that mark and 0 means that the person does not have that mark. All of this information is related to the training database. It comes in another database which is a doctor's database. The Doctor Database list consists of 2 separate columns where the first column contains the Doctor's name and the second column contains a website link from where patients can book their appointment. This database contains 41 participating doctors from now on. All of these links have their origins in Practo.com which is a health information technology provider based in Bengaluru, India. It is a website / doctor consulting app that offers complete telemedicine solutions. We are talking about the training of our decision tree division model dividing our core data into 75% training and 25% test database using the train_test_split () method available in scikit-learn. This module is a free Python editing software for reading software. Includes algorithms for segmentation, deceleration and random integration**.** Rest, support vector equipment, k-means, gradient boosting and DBSCAN, designed to work with Python number libraries and science known as NumPy and SciPy.
2) *Dimensionality Reduction:* Variable inputs or datasets are referred to as their size. Size reduction refers to a technology in which high-density data space can be converted to low-density data or may mean reducing the number of dynamic input objects in the database. A set of data with multiple input features often makes modeling tasks more challenging in modeling, often referred to as a curse of size. The performance of machine learning algorithms can reduce the level with a wide range of input variables. If your data is represented by rows and columns, such as in a spreadsheet that means the variable input columns are fed as inputs to the model to predict the target variable. Therefore, we can consider a column representing the size or space of the n-dimensional element. It is therefore often helpful to reduce the size by exposing the data to a smaller size that captures the sum of the data. There are many strategies that can be used to reduce the following size: Feature Selection Methods, Matrix Factorization, Learn More, Automatic Copy Methods etc.
3) *Classification:* It is a two-step process namely, learning and predicting. In the first step, the model is created on the basis of training data provided, and in the second step, the model predicts the response from the set of provided data. The Decision Tree is one of the most popular and simple mathematical equations for understanding and interpreting. This algorithm is from a family of supervised learning algorithms that are being used to solve setup and re-editing problems. The target of this breakthrough is to construct a training model that can predict a class or amount of target flexibility by learning the simple decision rules found in training data. There are also some basic assumptions that need to be considered when designing a decision tree as initially the entire training set is considered root and then the feature values are selected to be categorized. If prices continue then they are divided before modeling. Records are repeatedly distributed on the basis of qualifications. Even placing attributes like the root or inner nodes of a tree is done in a certain mathematical way. Decision trees are easy to visualize and to easily capture Indirect patterns. It needs a few pre-processing data from users so there is no requirement of creating standard columns. They are also used to feature engineering such as pre-naming shortages are appropriate for variable options. The main disadvantage of the decision tree is that it is sensitive to noisy data as it can be overloaded with noisy data so it is suggested that you should balance the data before constructing a resolution tree.

4) *Training Bot with Classification Effects:* The Training Bot is a python-based menu application that communicates with the user and asks for individual clues. On the basis of a 'yes' or 'no' answer the training board will break down the tree structure. If the user answers "no" with a particular symbol the bot will come back again and again with other symbols and so on. But if the user answers “yes” to a specific disease the bot will diagnose all related diseases to better understand prognosis and also analyze the doctor's database of the name of a specialist in the field and provide. user and all details including a website link where the user can book an appointment with a doctor. GUI Creation- Python has many GUI frameworks, but Tkinter is the only framework built into the standard Python library. Cross-platform therefore the same code applies to Windows, macOS and Linux. Tkinter is lightweight and painless compared to other components. There are various widgets like button, check button, screen, input, etc. used to build python GUI applications. The GUI starts with a window frame that requests login or registration. Based on the selected selection, the appropriate window opens. The login window contains two text fields, one for the username and one for the password. If the user information is correct the system will allow the user to log in to the system otherwise the user will need to first register or create an account and after that he will be able to log into the system . In the same way the registration window also contains two text fields that allow the user to create his own username and password. Once the user has logged into the system after proper confirmation, a symbols window appears for the user to enter symbols in the fields. After that, when the model it is fed with the right amount of ingredients, and produces a response in the form of a predicted disease, given symptoms, confidence intervals and a recommendation that the doctor visit next. The signage window also provides a booking link for an appointment with the relevant physician which can be downloaded and pasted into the user's browser for further information. Fig1 and Fig 2 describe the processes diagrammatically

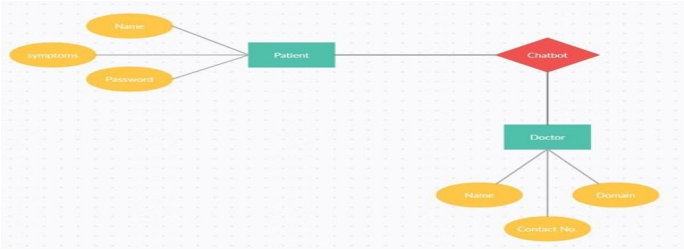
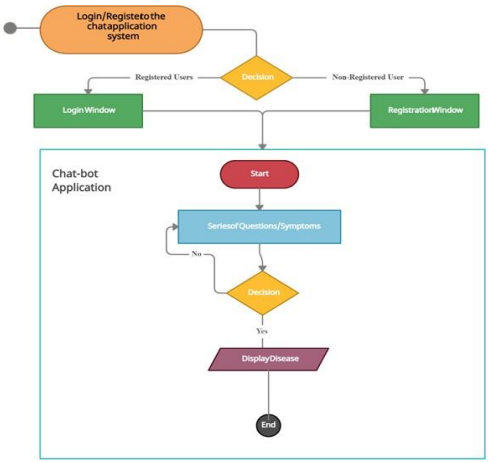


Fig. 1. Data associated with the chatbot



V. CONCLUSION

The proposed approach will use the diagnostic process and will also assist with medical infrastructure that is frustrated in times of crisis such as epidemics. The Decision-algorithm tree algorithm plays a role in predicting the disease and later the doctor's recommendation to patients makes it easier for the patient to find a complete cure for his or her disease. Automation in the medical field is an hour-long requirement. Chatbots and other AI tools will help create a way for mild and moderate disease symptoms to be treated at home and to leave hospitals for critically ill patients. This paper will be useful for that in automation so further modification of the method may be more useful and effective. This paper focuses on both the operating system and the GUI, so that it is easier for the user to diagnose his condition faster and, contacting the most qualified doctor also becomes easier.

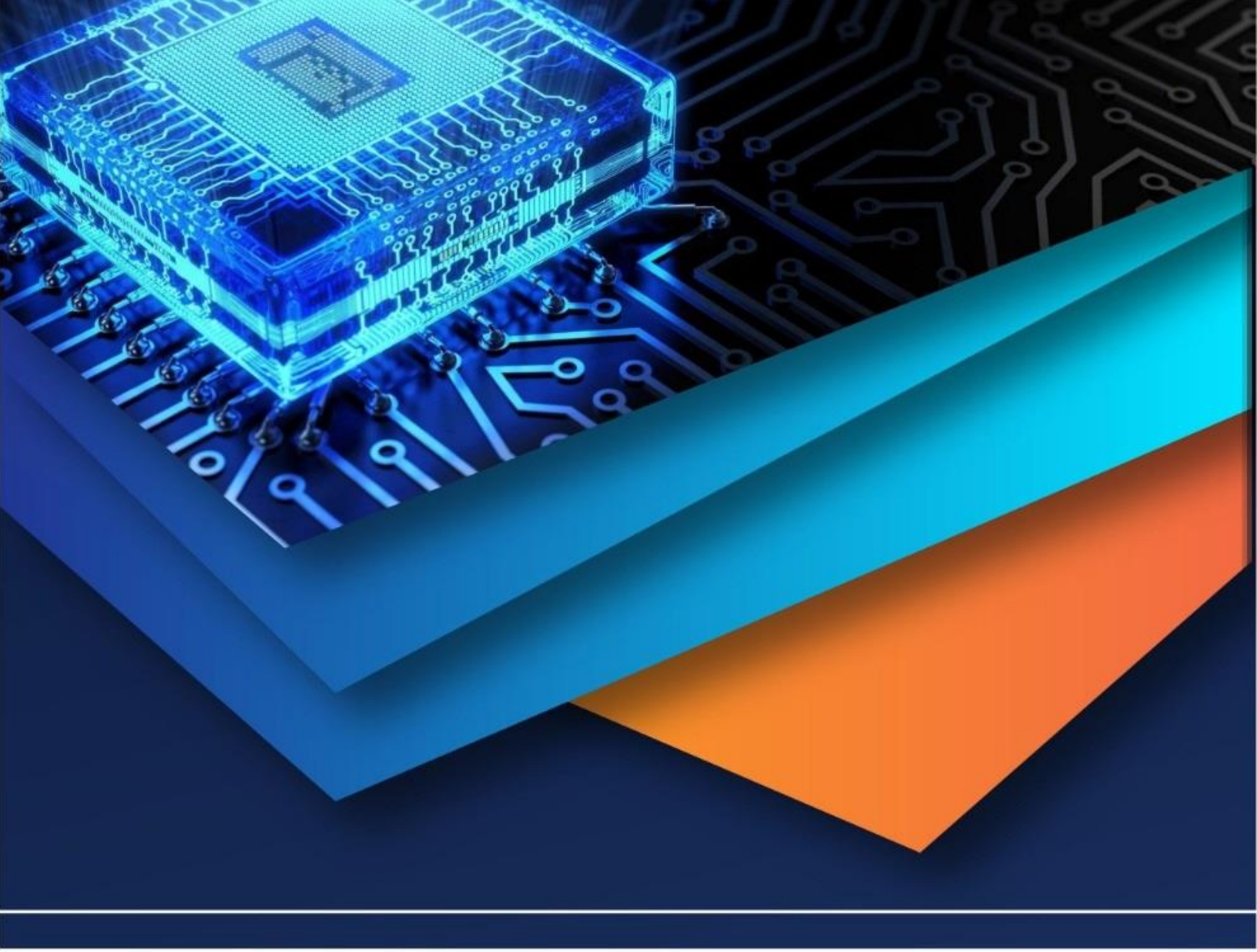
REFERENCES

[1] Dharwadkar, Rashmi, and Neeta A. Deshpande. "Medical chatbot." Int J Comp Trends Technol 60.1 (2018).
[2] Bali, Manish, et al. "Diabot: a predictive medical chatbot that uses integrated learning." Int. J. of Recent Technol. and Eng. (2019): 2277-3878.
[3] Bao, Q., Ni, L. and Liu, J., 2020, February. HHH: an online medical discussion system based on a graph of information and consecutive attention to two categories. At the Australian Computer Science Conference Multiconference (pp. 1-10).
[4] Kalla, D. and Samiuddin, V., 2020. Chatbot treatment using NLTK Lib. IOSR- J. Computer. Engi, 22.
[5] Chung, K. and Park, R.C., 2019. Chatbot-based healthcare service with basic cloud computing knowledge. Cluster Computing, 22 (1), pp.1925-1937.Harilal, N., Shah, R., Sharma, S. and Bhutan, V., 2020. CARO: sympathetic health dialogue for people with major depression. In the 7th ACM
[6] Aswini, D., 2019. Disclosure of medical information using crowdsourcing tech- nique. Int. Res. J. Eng. Technology, 6.
[7] Rarhi, K., Bhattacharya, A., Mishra, A. and Mandal, K., 2017. Automatic medical discussion. Available on SSRN 3090881.
[8] Yoo, S. and Jeong, O., 2020. Intelligent Chatbot Using BERT Model and Information Graph. Journal of the e-Business Studies Organization, 24 (3).
[9] All, S., 2018. "Plutchik": an artificial intelligence chatbot to search for NCBI databases. Medical Library Association Journal: JMLA, 106 (4), p.501.
[10] Brixey, J., Hoegen, R., Lan, W., Rusow, J., Singla, K., Yin, X., Artstein, R. and Leuski, A., 2017, August. Shihbot: Facebook chatbot for sexual health information on hiv / aids. In discussions of the 18th annual SIGdial meeting on speeches and interviews (pp. 370- 373).
[11] Sophia, J.J., Kumar, D.A., Arutselvan, M. and Ram, S.B., 2020. A survey of the use of chatbot in health care using NLTK. Int. J. Computer. Science. The crowd. Computer, 9.
[12] Zarouali, B., Van den Broeck, E., Walrave, M. and Poels, K., 2018. Predicting customer responses to chatbot on Facebook. Cyberpsychology, Behavior, and Social Networking-, 21 (8), pp.491-497.
[13] Dahiya, M., 2017. Chat tool: Chatbot. International Journal of Computer Science and Engineering, 5 (5), pp.158-161.
[14] https://www.kaggle.com/data/86712
[15] https://www.kaggle.com/pawanyalla/medical

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY