

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

Machine learning based disease prediction website using symptoms of a patient

Varinder Garg¹, Harish Kumar², Surinder Rana³, Bikramjeet Singh Kalsi⁴, Siddhant Mukherjee⁵, Shashvat Bargale⁶, Shivani Suhal⁷

¹ MBBS, MD Principal Investigator, ICMR-Centre for Innovation and Bio-Design, Room No. 2022 and 2027, Advanced Cardiac Centre, PGIMER, Sector 12, Chandigarh, 160012. Email: me_dvg@yahoo.co.in

² B.E., M.E., PhD CSE Dept, UIET Panjab University, Chandigarh 160014.

³ M.D., D.M Gastroenterology, Professor, Department of Gastroenterology, PGIMER, Sector 12, Chandigarh, 160012. Email: rana.surindersingh@pgimer.edu.in

⁴ B. Tech., Computer Science, UIET Panjab University, Chandigarh 160014.

⁵ B.Tech., M.Tech., Biomedical Instrumentation, Department of Instrumentation and Control, College of Engineering Pune (COEP), Shivajinagar, Pune.

⁶ B.Tech in Biomedical Engineering, Shri GS Institute of Technology & Science, Indore, M.P.

⁷ B.Sc (Non-Medical), DAV College, Sector-10, Chandigarh

ABSTRACT

We have been relying on doctors and physicians for generations to detect the ailments within us and give proper treatment according to the ailment detected. But no offence, this technique needs presence of the doctor in a nearby location and also in many circumstances, patients need to visit doctor which is very difficult in many situations, especially during disasters. There are many alternatives to bring doctors at home via video-conferencing and other technologies available but requires better internet connectivity. This paper describes a website in which user will add symptoms and website will predict diseases based on dataset prepared from different sources and good study and machine learning techniques applied to the dataset. Accuracy of the diseases predicted is 70-75% and we are working on increasing this accuracy. This method, if applied, will help users to predict the disease at the early stage of any ailment and they can consult doctor as early as possible.

Key Words: Website, video-conferencing, machine learning, dataset.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

INTRODUCTION

Doctors have been treating patients and giving suggestion since ages. Hippocrates, world's first ever recorded doctor cum physician, classified disease in four categories: acute, chronic, endemic and epidemic and laid the foundation of his oath through which doctors around the world need to serve the person and protect him/her irrespective of age, gender or race. The main point of this is to explain that doctors have been better at treating people and people trust their advice. Doctors use a variety of equipment and medicines to diagnose and treat their patients. These techniques have been modified keeping the traditional knowledge of such practices in place and combining it with the modern machinery and appliances which help them in the decision-making processes to find the root cause of the ailment and give suggestions or prescriptions based on them.

But in recent times, this practice of people meeting the doctor personally to give them information about their health- related problem has been challenged and solutions are provided to solve such situations where doctors can't meet their patients personally due to differences in the location, condition or the quality of treatment they can get from. There are many solutions implied in real world diagnostic and therapeutic practices like virtual conferencing between two such parties, virtual surgery of such patient in critical care scenarios etc.

There are many issues with such changes like network connectivity, patient interaction issues and many more but people are working towards solving such challenges.

So, as a result of this, we are presenting this piece of work which describes the website which predicts the disease based on the symptoms added using some machine-learning techniques. Section II will present the literature review to support the work done, Section III will reveal the work behind the website and some results achieved, Section IV will give impressions about the final notes and some discussions.

LITERATURE REVIEW

Doctor-patient relationship is a historical work which was developed in very early times. Doctor and patient interaction cause a huge social impact which is directed to improving lives of human society. Szasz and Hollender developed three models of such relationship i.e. active-passivity, guidance-co-operation and mutual participation. It developed from Ancient Egypt civilization to Medieval Europe period and enhanced more in the French Revolution period. [1]

In this internet age, people are increasingly accessing the internet for filling their information quota. While recording the patient's health history and family history, doctors and physicians readily inquire about the patient's use of internet so that they can understand amount of information users are getting fed from the internet. [2] This can in turn help in clearing doubts the doctors possess while understanding the negative and positive impact on doctor-patient relationship. [3][4][5]

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

In such times, many sources of free videoconferencing apps such as Google Meet, Skype etc. and many other doctor consultation apps like Intelehealth as done by Goel et. al [6] and Practo [7] have emerged to establish doctor-patient relationship via medically legit processes and also government-initiated apps like Arogya-Setu which proved effective to record patient's daily update of his/her situation during COVID-19 times and catered to around 90 million people to effectively curb COVID-19 patients on rise [8] and is continuing to inspire other states in India to publish their apps for doctor and patient interaction. Due to Google and Apple's policy for restricting the apps not approved by the governing institutions worldwide, some small developers have switched to developing websites that provide information about epidemics and tele-consultation. [9]

These systems have disadvantages related to requiring better internet connection with higher ping rate (at least 30 MS), especially for video-conferencing apps for better audio and video quality and hence, it will be extremely difficult to operate these apps where internet connection is low and ping rate is unstable and low. There are security issues with such apps like Zoom who have been questioned constantly for not providing end-to-end encryption for their video calls which doctors have used for contacting patients. [10] HIPAA i.e. Health Insurance Portability and Accounting Act needs that healthcare information needs to be encrypted and covers only registered entities like healthcare workers and not patients and also FDA checks medical devices and not such apps or any technology and so patient privacy is always in question. [11][12][13]

Medical encyclopedias also are being used by patients and doctors to gain valuable medical information. Now in this age of internet and modern multimedia, there have been many such encyclopedias available online like Medline Plus, WebMD, and Merck Manual. Also, some use medical Wikipedia built on Wikipedia software and WebMD also offers symptom checker. But there are issues with such encyclopedias also like how symptom checker from WebMD is a hypochondriac's worst nightmare as searching for bloating in the lower abdomen suggest many diseases and disorders and confuses users with medical information and also shows contents for medical information that are restricted and penalized by governing organizations. [14] Despite all of this, such sites display at the end of the webpage as the information from these sites are not medical advice and concerned medical personnel should be concerned which raises questions on the reliability of such information.

It is the need of the hour to develop new and effective strategies to at least increase accuracy of the prediction of the disease based on user's basic information excluding location and unnecessary information using well-defined and updated websites with correct medical information and machine learning techniques. [15] The above written article explains the literature review to support work done which will be elaborated in the next sections.

IMPLEMENTATION OF WEBSITE AND RESULTS

This paper proposes a website that needs to be designed in a way that users can add symptoms and get results within seconds and the prediction of the disease should be specific and shouldn't confuse the user.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

For such a website to be implemented, we followed certain steps which were as follows as shown in fig. 1:

1. *Collection of medical information from reliable sources*
2. *Segregation and classification of data acquired*
3. *Conversion of such data into one hot key encoding format or format that the machine learning technique can detect*
4. *Decision Tree Classifier and training the model*
5. *Website developed and results achieved*

1. *Collection of medical information from reliable sources:*

Our aim was to make sure that the information which we gather are from sources that are correct and from books and are specific. So, we relied on team of medical researchers to search information from the internet which was verified and the diseases were collected and classified in different categories. We also referred the journals like A-Z of lung diseases, Oxford book of General Medicine etc.



Fig. 1 Steps taken for implementation of website

2. *Segregation and classification of data acquired*

All the medical information collected was segregated into different sheets of the comma separated value file. The diseases were categorized based on the site or organ which was affected through that disease. After this segregation and classification, symptoms of such diseases along with actions used to confirm a particular disease were added in the sheet. The doctor recommendations and suggested medications were incorporated in the file. A dataset of such files with diseases covered under different organ and region of infection was developed.

3. *Conversion of such data into one hot key encoding format or a format that the machine learning technique can detect*

Complete dataset was converted into one-hot encoding format in which the symptoms and diseases were categorized into row-column format and symptoms that corresponds to that disease in the file were marked as '1' and others that don't correspond as '0'. This was done so that model can be prepared in a way that it would be easier to train the model.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

4. *Decision Tree Classifier and training the model*

The model is trained using the coordinates obtained from the categorization of diseases and symptoms and this happened using python's dedicated libraries to test the data by splitting arrays into random train and test subsets given that number of test samples are added in the code. Also decision tree classifier (as a python library) was used in the code which uses a divide-conquer algorithm to split the data into subsets and then further classified into smaller subsets until the process stops when the algorithm will determine the accurate data within the subset which are homogenous or another criteria has been satisfied. After labelling the predicted data, the user will get an output (in the form of a disorder or disease) which can be stated as to something that can touch the highest level of accuracy as it came after constant training of the model and classification technique used.

5. *Website developed and results achieved*

A dedicated website was developed using this model and FLASK was used as a platform to develop one. In the website, different edit boxes were provided with drop- down option to provide user flexibility of choosing their symptoms. After clicking on submit button, user will get an accurate prediction of the disease in a different dialogue box. Also, a separate section containing Google Maps was developed which will provide the location of the doctors present in the nearby location which will provide an extended help to the user and will incorporate a feature for enlisting of the drugs

The whole website was deployed using either Heroku or other reliable web-deployment services. Results of the website developed are shown in Fig.2, Fig.3, Fig.4, Fig.5.

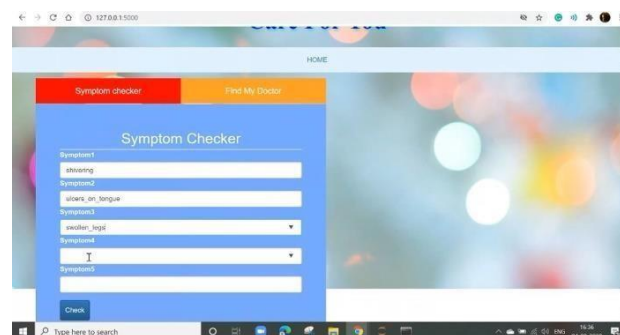


Fig. 2 Website in which symptoms can be added

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

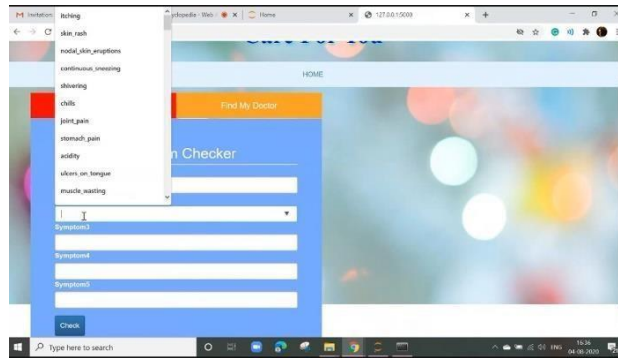


Fig. 3 Dropdown box showing different option of symptoms to the user

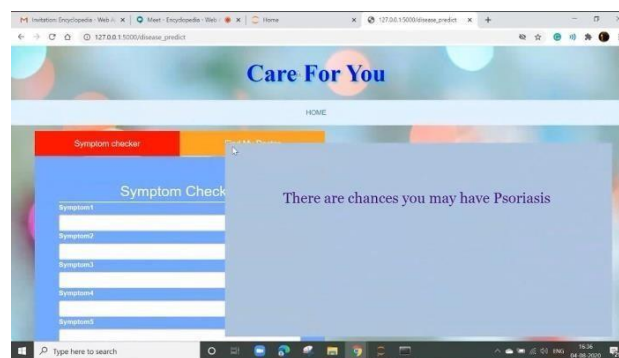


Fig. 4 After adding symptoms and clicking submit, you get a single output

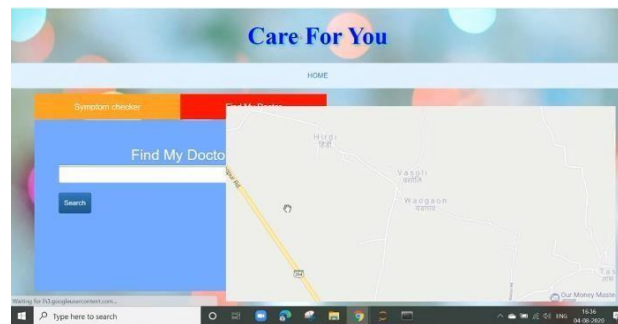


Fig. 5 Google maps section which helps user to find doctors nearby

The website was developed and results above show the rate of prediction that the website shows.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

FINAL NOTES AND DISCUSSIONS

The website was developed and tested with different symptoms for different diseases. However, more features need to be added and website's user interface needs to be modified so that user gets a better experience of the website. This website was made for the purpose of serving the user right medical information they need before they can consult the doctor. Google Maps feature will add more taste to the website. Upon analysis, it was seen that this website was satisfyingly giving better predictions than WebMD which used to give more predictions and suggestions such that the user gets confused.

The future development of this website will include addition of video calling application interfaces or voice recording based application interfaces where user can send their recording about the problem they are facing once they located their nearest doctor on maps. Also, work might happen regarding the addition of messaging services on the website where user can engage with the doctor via messages and images can be transferred with keeping in mind end-to-end encryption is ensured.

By using machine learning based techniques to develop a website that predicts any disease based on symptoms increases the doors of possibility of using machine learning algorithms to solve more problems related to prediction of certain diseases. Many companies have used machine learning techniques to quantify the amount of infection the certain organ has gone through and we should also proceed in the same direction to protect human lives.

ACKNOWLEDGMENT

This project was developed under the guidance and help from Indian Council of Medical Research (ICMR) – Center for Innovation and Bio-Design (CIBioD) and Post Graduate Institute for Medical Education and Research (PGIMER), Chandigarh.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

REFERENCES

PRIMARY SOURCES

- [1] SZASZ TS, HOLLENDER MH. A Contribution to the Philosophy of medicine: The Basic Models of the Doctor-Patient Relationship. *AMA Arch Intern Med*. 1956;97(5):585–592. doi:10.1001/archinte.1956.00250230079008
- [2] Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. *J Gen Intern Med* 2002; 17:180-5.
- [3] Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: Patient perceptions. *Arch Intern Med* 2003; 163:1727-34.
- [4] Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: Qualitative study. *J Med Internet Res* 2010;12: e70.
- [5] Chung JE. Patient-provider discussion of online health information: Results from the 2007 health information national trends survey (HINTS). *J Health Commun* 2013; 18:627-48.
- [6] Goel, Neha & Alam, Amal & Eggert, Emily & Acharya, Soumyadipta. (2017). Design and development of a customizable telemedicine platform for improving access to healthcare for underserved populations. Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2017. 2658-2661. 10.1109/EMBC.2017.8037404.
- [7] Kumar, Amit & Kumar, Amit. (2014). Practo Technologies: the online way of life! *Emerald Emerging Markets Case Studies*. 3. 1-19. 10.1108/EEMCS-06-2013-0127.
- [8] "9 crore people have downloaded Aarogya Setu app: Government", *The New India Express*, May 6, 2020. [Online]. Available: <https://www.newindianexpress.com/nation/2020/may/06/9-crore-people-have-downloaded-aarogya-setu-app-government-2139926.html>. [Accessed August 28, 2020]
- [9] N. Statt, "Apple and Google are cracking down on coronavirus apps to combat misinformation", *The Verge* at Mar 5, 2020. [Online]. Available: <https://www.theverge.com/2020/3/5/21167102/apple-google-coronavirus-iphone-apps-android-misinformation-reject-ban>. [Accessed August 28, 2020]
- [10] K. O'Flaherty, "Zoom Security: 100 New Features Later, Can You Trust Zoom?", *Forbes* at Jul 2, 2020. [Online]. Available: <https://www.forbes.com/sites/kateoflahertyuk/2020/07/02/zoom-security-100-new-features-later-can-you-trust-zoom-yet/#2dcc714336e4>. [Accessed August 28, 2020]
- [11] Dockser Markus A, Weaver C. "Heart gadgets test privacy-law limits." *Wall Street Journal*. 2012 Nov.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

- [12] Yang YT, Silverman RD. "Mobile health applications: the patchwork of legal and liability issues suggests strategies to improve oversight". *Health Aff (Millwood)*. 2014; 33 (2): 222 – 27.
- [13] Dwyer III, Samuel J., Alfred C. Weaver, and Kristen Knight Hughes. "Health insurance portability and accountability act." *Security Issues in the Digital Medical Enterprise* 72.2 (2004): 9-18.
- [14] J. Belluz, "The truth about WebMD, a hypochondriac's nightmare and Big Pharma's dream", *Vox* at Apr 5, 2016. [Online]. Available: <https://www.vox.com/2016/4/5/11358268/webmd-accuracy-trustworthy>. [Accessed August 28,2020]
- [15] D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT

MACHINE LEARNING BASED DISEASE PREDICTION WEBSITE USING SYMPTOMS OF A PATIENT