# “Detection of Diseases using Machine Learning”

Raj Ghadi
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information Technology
Mumbai, India
r.ghadi@somaiya.edu

Shrejay Patil
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information Technology
Mumbai, India
shrejay.p@somaiya.edu

Rohit Desai
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information Technology
Mumbai, India
rohit.desai@somaiya.edu

Prof. Pankaj Deshmukh
Department of Electronics Engineering
K.J.Somaiya Institute of Engineering and Information Technology
Mumbai, India
pankaj@somaiya.edu

***Abstract*—The world's growing population has put enormous pressure on the healthcare sector to offer high- quality treatment and accommodations. Artificial intelligence and Machine Learning are no exceptions in the healthcare industry, which has long been a vigorous adherent of cutting-edge technology. We have developed a web application using flask framework. It consists of web pages designed for different functionality. It is a disease prediction system which can be deployed on any network for communication among ecumenical users. Report is generated that can be subsidiary in-order to keep records. The report generated can be downloaded locally on user’s device as well as provided on user’s personal emails.***

***Keywords—prediction, MySQL, python, session, admin, machine learning,***

## I. Introduction

World is suffering because of Many Diseases and to surmount it we can utilize artificial intelligence and machine learning. The purport of artificial intelligence and machine learning is to make the machine more prosperous, efficient, precise and reliable than afore. However, in a healthcare system, it will definitely avail medicos a lot in critical situations and decisions. To minimize the pressure of the healthcare system and to avail the medicos and society we have engendered the project which will predict the particular disease, detect the disease in earlier stage and predict the diseases very accurately and efficiently. This will avail medicos to attest or cross- check their postulation and analysis. It will avail them in critical situations and decisions. The interface has a navigation-bar-driven programme that enables facile utilizer interaction with some GUI applications. Login and Signup forms are a component of user authentication. All details acquired during signup process are stored in the database which can only be accessed by Admin. Admin here is the one who manages the website and works in the backend maintaining all the data extracted during the user’s session. Sessions are engendered that avails maintain users state and data all over the application. Different sections such as contact, FAQ, feedback and analysis are present on the webpage. Especially the analysis section provides transparency to the users about how our model works in the backend as healthcare is a consequential issue and just can’t be ignored for.

## II. Literature survey

### A. *UCI researchers create model to calculate COVID-19 health outcomes*

University of California, Irvine health sciences researchers have engendered a machine-learning model to predict the probability that a COVID-19 patient will require a ventilator or ICU care. The implement is free and available online for any healthcare organization to utilize.

"The goal is to give an earlier alert to clinicians to identify patients who may be vulnerably susceptible at the onset," verbally expressed Daniel S. Chow, an assistant pedagogia in residence in radiological sciences and first author of the study, published in PLOS ONE. The implement predictions whether a patient's condition will worsen within 72 hours.

Coupled with decision-making concrete to the healthcare setting in which the implement is utilized, the model utilizes a patient's medical history to determine who can be sent home and who will require critical care. The study found that at UCI Health, the implement's predictions were precise about 95 percent of the time.

### B. *Disease Prediction Using Machine Learning*

Computerized systems are currently considered to be much more efficient than the traditional ones, similarly adapting these systems in the healthcare sector would yield better results comparatively. The concept of supervised machine learning algorithms holds enormous potential for disease diagnosis. Huge amount of data is required in such systems in order to gain high precision output. There are many types of algorithms available, selection of these algorithms is very crucial at the time of designing the machine learning model. In this literature, the aim is to apperceive trends across various types of supervised ML models in disease detection through the examination of performance metrics.

There are some algorithms such as Naves Bayes (NB), Decision Trees (DT), And K-Nearest Neighbor (KNN) etc. is considered to be most prominent among others. According to the research Support Vector Machine (SVM) was found to be most eligible at detecting Kidney and Parkinson’s

diseases. Similarly Logistic Regression (LR) for heart disease, Random Forest Classifier (RFC) and Convolutional Neural Networks (CNN) for breast and common diseases were selected respectively.

## III. PROPOSED SYSTEM

Proposed Diseases prediction System to Improve quality and efficiency of current healthcare system. "It is introduced with the aid of web creation. In order to prevent pitfalls in the current system, the proposed system is built as a web-based system where it can be accessed at anytime and anywhere on their mobile/PC. The user can use their credential to access the server system. The website faculties will now keep track of user's health by providing insights of health issues if any. Additional features include generation of report.

Figure 1 shows the flowchart of proposed web-based system. A login page is a web page that needs user identification and authentication by entering a username and password combination on a regular basis. Logins can provide access to the entire website. Logging in not only gives the user access to the site, but also enables the website to monitor user activities and behaviour. Logging off a website or site may be a manual for the user or may occur automatically when such circumstances (such as page closure, device shutdown, long delay, etc.) occur.

### *A. Diseases Parameters*

Every disease has different parameters on the basis of which it is detected. Below are few most significant parameters tabulated in table 1 after data pre-processing.

Table I. Most prominent parameters for disease detection

| Disease | Parameters | Algorithm Used | Accuracy in (%) |
|---|---|---|---|
| Diabetes | BP (mm Hg), Glucose, Insulin (mu U/ml), BMI (kg/m2), Diabetes Pedigree Function, Age | Random Forest Classifier | 82 |
| Heart | Cholesterol, Fasting Blood Sugar, Chest Pain type | Logistic Regression | 80 |
| Liver | Proteins, Albumin, Bilirubin, Albumin and Globulin Ratio | Random Forest Classifier | 79 |
| Kidney | Sugar, Red Blood Cells, Blood Urea, Hyper Tension | Random Forest Classifier | 99 |
| Parkinson's | range of biomedical voice measurements (Hz) | Random Forest Classifier | 94 |
| Breast Cancer | Radius, Perimeter, Area, Concavity, Concave Points | Logistic Regression | 97 |

### *B. Units*

- The number of cycles per second is called frequency. The SI unit for frequency is the hertz (Hz).
- A millimetre of mercury is a manometric unit of pressure, described as the extra pressure generated by a column of mercury one millimetres high, defined as exactly 133.322387415 pascals. It is denoted mmHg.
- Bbody mass index defined by the square of the body height, and is expressed in units of kg/m2, formulated from mass in kilograms and height in meters.
- A litre is a measure of volume that is a bit greater than a quart.
- A millimeter abbreviated as mm is a small unit of displacement in the metric system.
- All the parameters utilized in Parkinson's diseases are the measure of voice in Hertz and db.(decibel) of healthy person and some of unhealthy person.

### *C. Developing Tools*

The developing tools are utilized for web page plan and database building. To begin with, the internet pages of Diseases Prediction Framework were outlined by the HTML5, CSS3, J Query and JavaScript, because the software is simple to induce and simple to use. We have utilized PYTHON for the programming, since the internet page designed by it is more proficient in preparing the complex working environment providing vast range of libraries. Flask is a web application framework written in Python.

The application can be deployed on desktops with any operating systems. About the database apparatuses, the MySQL server is used.

Several algorithms were implemented, Figure 10 shows comparison between them on basis of accuracy. Hence considering accuracy two most prominent algorithms were shortlisted and deployed mentioned below:

Random Forest Classifier:

It's a type of supervised learning algorithm that focuses on ensemble technique. Multiple decision trees make a forest and are trained with the "bagging" method. Combining various learning models increases the accuracy relatively. Decisions are provided individually by the trees and the most voted one becomes the prediction of our models.

Logistic Regression:

Another most popular supervised learning algorithm is the Logistic Regression. It utilizes the famous sigmoid function providing a S-shaped graph. It basically works similar to that of mathematical probability resulting in classification of categorical data.

## IV. IMPLEMENTATION OF PROPOSED PROJECT

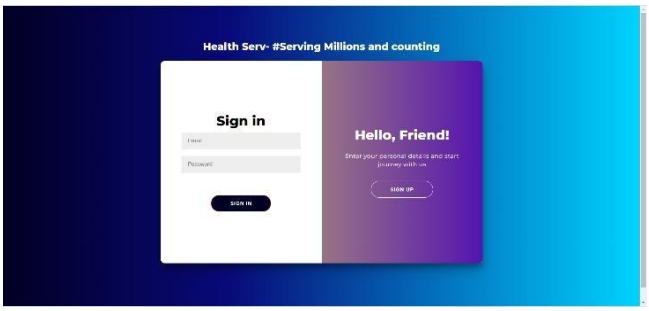


Fig. 2. Authentication Login form

Figure 2 shows the diseases prediction system login page and it is password-protected. A unique email and password have to be entered. If the correct email and password are entered, it will lead to a Home Page of the website.

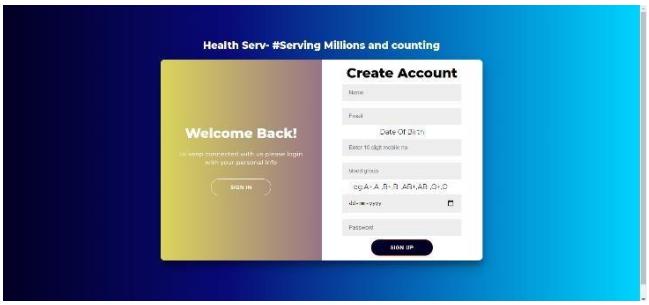


Fig. 3. Sign Up Form

Figure 3 shows the diseases prediction system Sign Up page in which the user needs to fill its credentials which is stored in a database for further processes. It contains fields such as name, email, contact, blood group, date of birth and password.

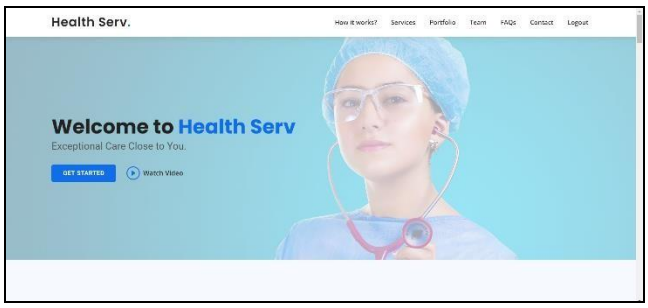


Fig. 4. Home page

Top navigation bar consists of links to each section present on the webpage providing better user experience as shown in figure 4. It also consists of a professional welcome to the user logged in.

Sessions are created for every user in order to maintain data of a specific user throughout the web application.

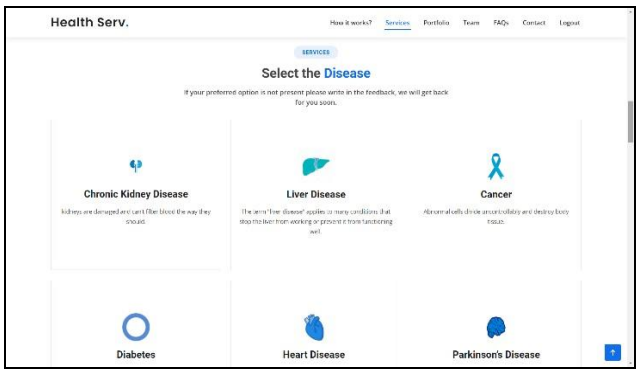


Fig. 6. Diseases section

In Figure 6 these are the various diseases that are available for the users for check-up using their medical data. It consists of kidney disease, liver disease, breast cancer, diabetes, heart disease and Parkinson's disease.


Fig. 9. Disease prediction page

Figure 9 shows a disease prediction page. Each disease has their separate prediction page. As soon as the user clicks on the disease that he/she prefer to check will land on this page where the user need to enter the medical data and click on predict button. The output generated will be displayed on the same page itself.

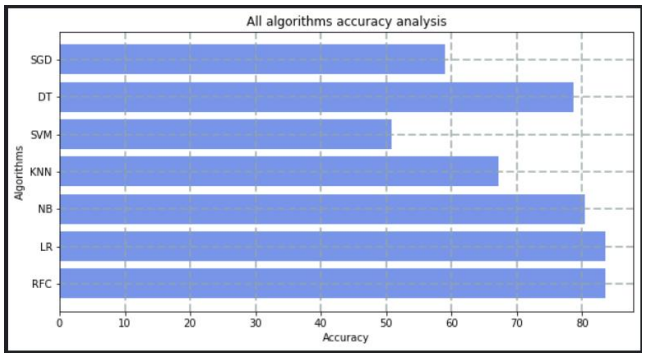


Fig. 10. Algorithm comparison

Algorithms and their analysis are provided to the user as shown in figure 10.

## V. CONCLUSION

The overall aim is to define various data mining techniques utilizable in efficacious disease prediction. Efficient and precise prediction with a lesser number of attributes and tests is our primary goal. In this study. The data were pre-processed and then utilized in the model. We found the precision after implementing algorithms to be above 70 percent.

Another crucial goal we are looking forward to is to soothsay the disease in its early stage which affects the patients salubrity a lot. As we all might be aware of the fact that early stages of any diseases are too tough to detect and if detected will open gates for lots of treatment processes.

We all might have wondered utilizing online algorithm for these sensitive issues is not reliable, we can verbalize that it's partly veritable as it is a machine and we cannot plenarily rely on its prediction, but still if it provides you with even a slightest insight of your health that might get worse in future if ignored than its likely propitious for us.

## REFERENCE


[1] Daniel S. Chow, Justin Glavis-Bloom, Jennifer E. Soun, Brent Weinberg, Alpesh N. Amin, Peter D. Chang. Development and external validation of a prognostic implement for COVID-19 critical disease. PLOS ONE, 2020 ; 15 (12): e0242953 DOI: 10.1371/journal.pone.0242953.

[2] Ferjani, Marouane. (2020). Disease Prediction Using Machine Learning. 10.13140/RG.2.2.18279.47521.

[3] M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.

[4] A. C. Lyngdoh, N. A. Choudhury and S. Moulik, "Diabetes Disease Prediction Using Machine Learning Algorithms," 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 517-521, doi: 10.1109/IECBES48179.2021.9398759.

[5] A. Grover, A. Kalani and S. K. Dubey, "Analytical Approach towards Prediction of Diseases Using Machine Learning Algorithms," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 793-797, doi: 10.1109/Confluence47617.2020.9058120.

[6] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic Diseases Diagnosis using Machine Learning," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1-6, doi: 10.1109/ICCSDET.2018.8821235.

[7] Goel, Rati, Heart Disease Prediction Using Various Algorithms of Machine Learning (July 12, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021, Available at SSRN: https://ssrn.com/abstract=3884968 or http://dx.doi.org/10.2139/ssrn.3884968.