

Disease Prediction System

Saksham Shukla

Submission: 11-05-2022

Revised: 22-05-2022

Acceptance: 25-05-2022

ABSTRACT:

This project aims to develop software for disease prediction using Random Forest Machine Algorithm. The Objective of making this project called "Disease Prediction System" is to predict the accurate disease of the patient using all their general information and also the symptoms. Using this information, there we will compare it with our previous datasets of the patients and predicts the disease of the patient he/she is been through. If this Prediction is done at the early stages of the disease with the help of this project and all other necessary measures are taken then the disease can be cured at the very early stage.

KEYWORDS: Random Forest, Machine Learning, Disease prediction, symptoms.

I. INTRODUCTION

A disease prediction system using Machine Learning is a system that predicts the disease based on the information provided by the user. It also predicts the disease of the patients or the user based on the information or the symptoms he/she enters into the system and provides accurate results based on that information. It is a system that provides accurate results about the disease by which the patient is suffering. Now a day's health industry plays a major role in curing the diseases of the patients so this is also some kind of help for the health industry. This Disease prediction using

Machine Learning is completely done with the help of Machine Learning using Python Programming language and also using the dataset that is available using that we will predict disease. The project disease prediction system using machine learning is developed to overcome diseases in earlier stages. There are 40% of people who ignores the general disease which leads to harmful disease later. The main reason for ignorance is laziness to consult a doctor and time concerns the people have involved themselves so much that they have no time to make an appointment and consult the doctor which later results in fatal disease. According to research there are 70% of peoples in India suffers from general disease and 25% of peoples face death due to early ignorance the main motive to developing this project is that a user can sit at their convenient place and have a check-up on their health the UI is designed in such a simple way that everyone can easily operate on it and can have a check-up.

Proposed Work

We are predicting whether a person is suffering from the diseases like diabetes, heart disease, Pneumonia, Malaria, chronic kidney, breast cancer, and liver disease depending upon the symptoms he or she is suffering. Here we take multiple symptoms from the patient and evaluate them by using an algorithm such as Random Forest.

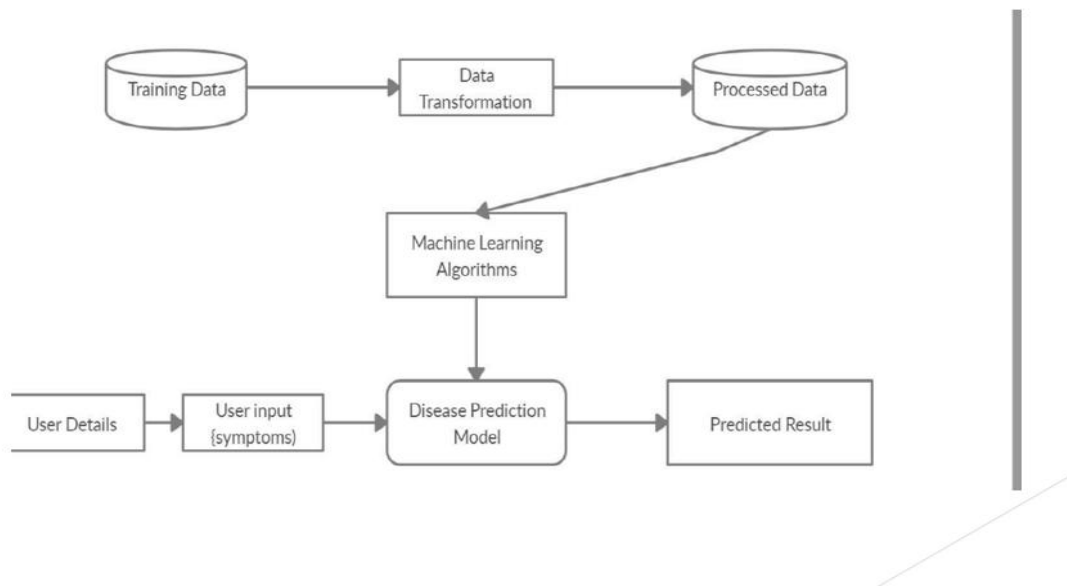


FIG 1 Block Diagram

II. METHODOLOGY

Random Forest

It is based on the concept of "Ensemble Learning". It can be used for both classifications as well as regression problems. Since Random forest is a supervised machine learning algorithm so it is a very good option for this work, this methodology is being opted to decide whether a patient has a particular disease based on the symptoms provided by him. The performance of Random Forest is comparatively better in classification in comparison to regression problems and the accuracy we achieved by this algorithm in our work is also very high.

5.1 Collecting Data

This is the important part of work. Because if we have a good amount of data then we can train the model more accurately and we can make it more accurate in terms of predicting disease. Data has been collected in the form of excel sheets.

5.2 Data Preprocessing

Preprocessing of data is done to transform the information into a format that will be more easily and effectively processed for the aim of the user.

5.3 Model Training

This step involves choosing the appropriate algorithm for the model. We have chosen Random forest algorithm for our project. Then the preprocessed data is split into two parts i.e training data and testing data. Here we have splitted our dataset into 80:20 ratio which means 80 percent data for training purpose and 20 percent data for testing purpose. The first part (training data) will be used for developing the model and the second part (testing data) will be used as a reference.

Random Forest

It is based on the concept of "Ensemble Learning". It can be used for both classifications as well as regression problems. Since Random forest is a supervised machine learning algorithm so it is a very good option for this work, this methodology is being opted to decide whether a patient has a particular disease based on the symptoms provided by him. The performance of Random Forest is comparatively better in classification in comparison to regression problems and the accuracy we achieved by this algorithm in our work is also very high.

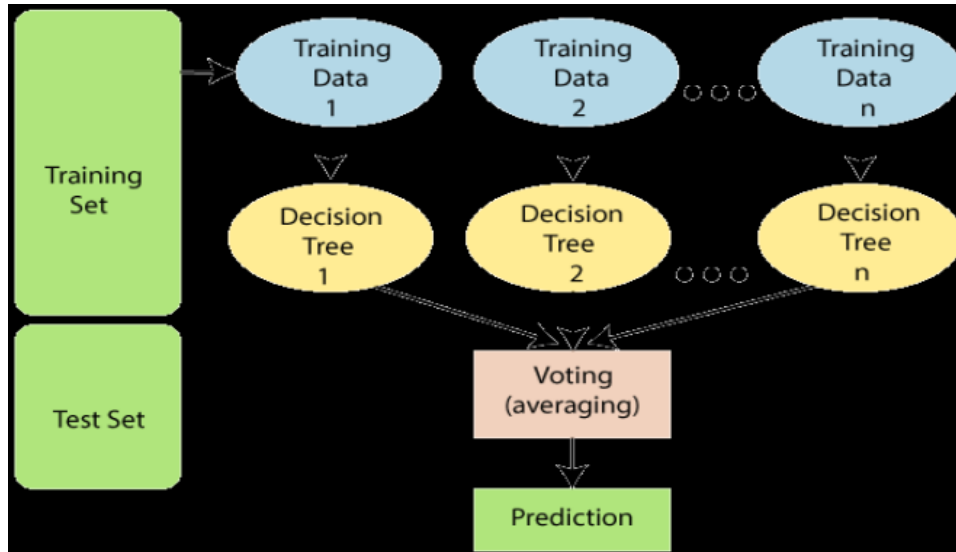


FIG 2 RANDOM FOREST ALGORITHM

III. RESULTS

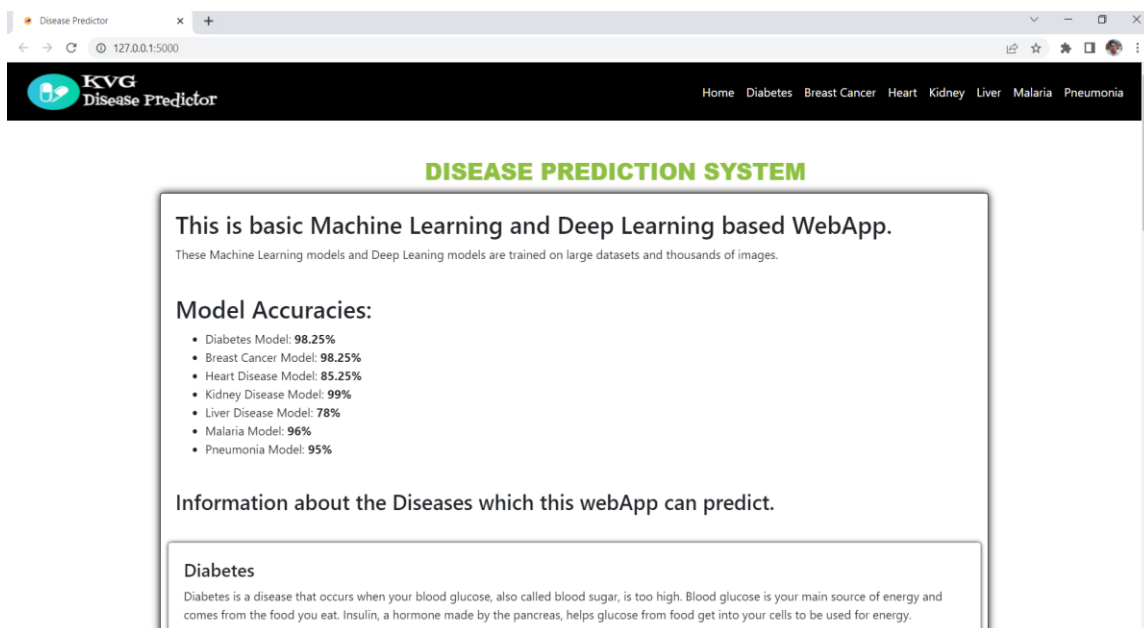


FIG 3 HOME PAGE

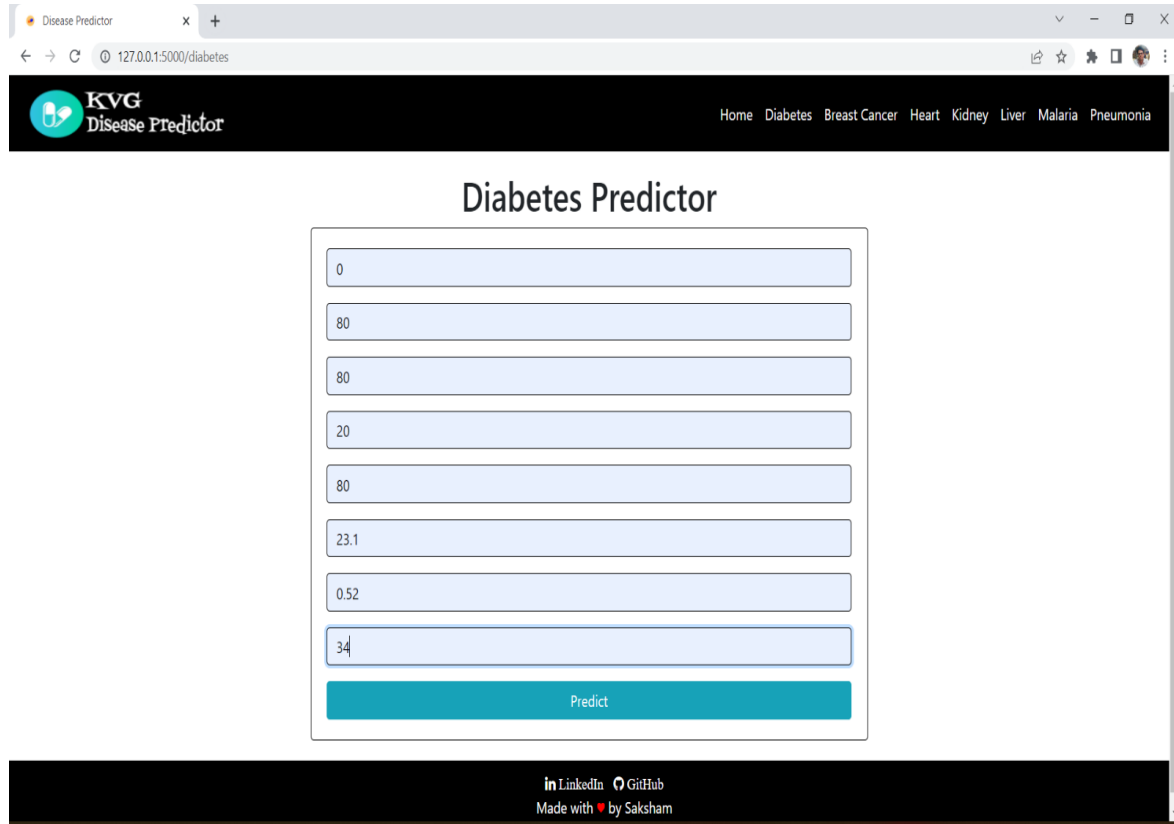


FIG 4 SYMPTOMS PAGE

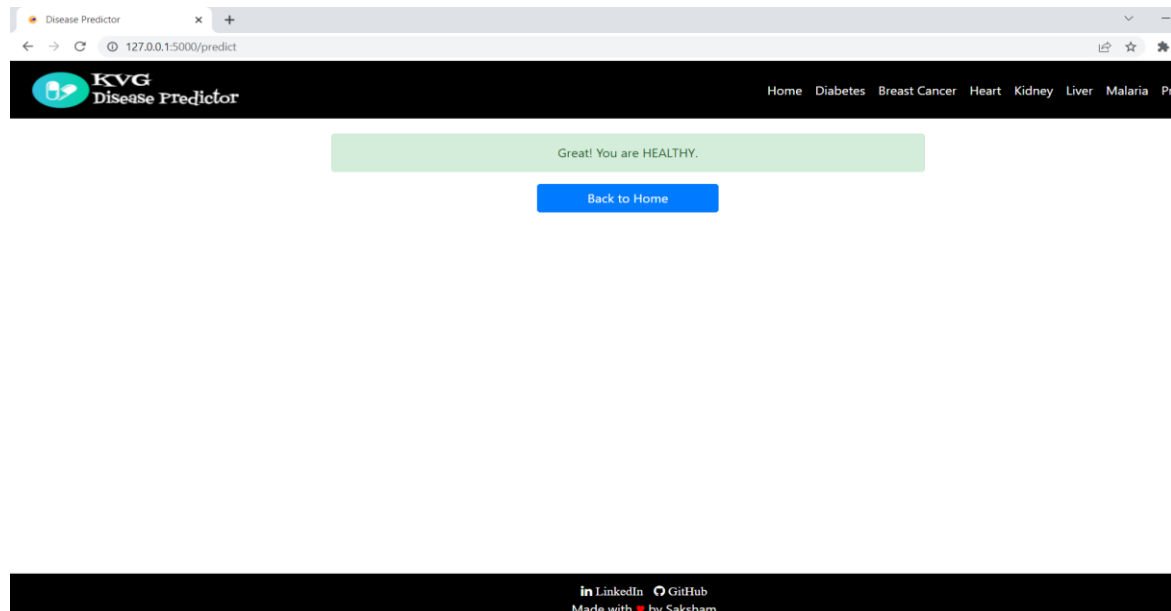


FIG 5 RESULT PAGE

IV. CONCLUSION

Disease prediction system using machine learning is very much useful in everyone's day to day life and it is mainly more important for the healthcare sector because they are the one that daily

uses these systems to predict the diseases of the patients based on their general information and their symptoms that they are been through. The Prediction Engine provides optimal performance with the right dataset and efficient training of the

classifier models considering all aspects and a lot of learning from the previous experiences. The implemented Prediction Engine is capable of predicting the presence of Diabetes with an accuracy of 98.25%, Breast Cancer with an accuracy of 98.25%, Heart Disease with an accuracy of 85.25%, Kidney Disease with an accuracy of 99%, Liver Disease with an accuracy of 78%, Malaria Disease with an accuracy of 96% and Pneumonia with an accuracy of 95%.

Future work

- To enhance the functionality of the prediction engine providing the details of the 5 nearest hospitals or medical facilities to the user input location.
- Provide a user account that allows the user to keep track of their medical test data and get suggestions or support to meet the right specialists or the tests to be taken

REFERENCES

- [1]. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", , IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.
- [2]. Disease Prediction using Machine Learning Kedar Pingale, Sushant Surwase, Vaibhav Kulkarni, Saurabh Sarage, Prof. Abhijeet Karve .
- [3]. T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent parkinson disease prediction using machine learning algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568 (2013)
- [4]. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm