

# Health Care & Disease Prediction Webapp Using AI/ML

Akshay Neurgaonkar, Tejas Inamdar, Ravikant Rathod

Department of Computer Department, Zeal College of Engineering & Research, Pune, Maharashtra, India

## ABSTRACT

### Article Info

Volume 8, Issue 2

Page Number : 449-453

### Publication Issue :

March-April-2022

### Article History

Accepted: 10 April 2022

Published: 22 April 2022

Health Care and Disease Prediction Using AI/Machine Learning is an online application that predicts the condition based on information or numeric symptoms given by the patient and gives reliable findings based on the input. If the patients aren't serious and only want to know what kind of ailment they've had. It is a system that provides the user with advice and tactics for maintaining their health as well as disease information utilizing this prediction. In the current era, the health industry plays an important role in curing the disease of the patients, so that is additionally a type of assistance for the health industry to tell the patient, and it is also useful for the patients in case they do not want to go to the hospital or any other clinics, so simply by entering the numerical data of symptoms and all other useful information, the user can get to know the disease they are suffering from, and the health industry can also get the information. This Health Care & Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with a Website Interface for it and also using the dataset that is available previously by the hospitals using that we will predict the disease.

**Keywords:** Random Forest, Machine Learning, Deep Learning, Disease Prediction.

## I. INTRODUCTION

Health Care and Disease Prediction using Machine Learning is an online application that predicts disease depending on the patient's input. It also forecasts the patient's or user's disorder based on the information or numerical data of symptoms they submit into the system, and offers reliable results based on that information.

If the sufferers aren't too concerned and just want to know what kind of disease they've had. It is a system that provides the user with techniques to maintain the

user's health as well as disease information using this prediction.

In current era health industry plays an important role in curing the illness of the sufferers so that is additionally a some type of assist for the health industry to tell the patients and also it is useful for the patients in case they don't want to visit the hospital or any other clinics, so just by entering the symptoms and all other information the users can get to know the disease they are suffering from and the health industry can also get benefit from this system by asking the numeric symptoms data from the patient and entering in the system and in the few seconds they

can tell the exact and up to that extent the accurate diseases.

This DPUML is previously done by some organizations but our intention is to make it different and beneficial for the users who are using this system. This Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with website Interface for it and also using the dataset that is available previously by the hospitals using that we will predict the disease.

## II. OVERVIEW OF PROJECT MODULES

This web application basically predicts the disease of person by taking the input from user in the form of medical parameters or X-rays/images and further provides information about the disease which is very useful in healthcare industry.

- **Cancer model:** In this model we have to provide all the parameters asked and after submitting these details, the webpage will show that the person is affected with cancer or not with the accuracy percentage & can also give the detailed information about the Cancer & its symptoms below it.
- **Diabetes model:** In this model we have to provide all the parameters asked and after submitting these details, the webpage will show that the person is affected with Diabetes or not with the accuracy percentage & can also give the detailed information about the Diabetes & its symptoms below it.
- **Heart model:** In this model we have to provide all the parameters asked and after submitting these details, the webpage will show that the person has heart disease or not with the accuracy percentage & can also give the detailed information about the heart disease & its symptoms below it.
- **Liver model:** In this model we have to provide all the parameters asked and after submitting these details, the webpage will show that the person has

liver disease or not with the accuracy percentage & can also give the detailed information about the liver disease & its symptoms below it.

- **Kidney model:** In this model we have to provide all the parameters asked and after submitting these details, the webpage will show that the person has kidney disease or not with the accuracy percentage & can also give the detailed information about the kidney disease & its symptoms below it.
- **Malaria model:** In this model we have to provide or upload the maximized microscopic image of affected cell and after submitting it, the web page will process that image & will show that the person is affected with Malaria or not with the accuracy percentage & can also give the detailed information about malaria & its symptoms below it.
- **Pneumonia model:** In this model we have to provide or upload the X-ray of lungs of patient and after submitting it, the webpage will process that image & will show that the person is affected with Pneumonia or not with the accuracy percentage & can also give the detailed information about Pneumonia & its symptoms below it.
- **Home page:** In this module page we will display the introduction, details about the app.

## III. ALGORITHM USED

### ➤ What is random forest?

Random forest is a machine learning technique that focuses on supervised learning. It creates a "forest" out of an ensemble of decision trees, which are normally trained using the bagging approach. The bagging method's basic premise is that combining learning models improves overall result accuracy. Random forest combines the results of numerous decision trees to get a more accurate and consistent prediction. The more trees in the forest, the more accurate it is and the problem of over-fitting is avoided.

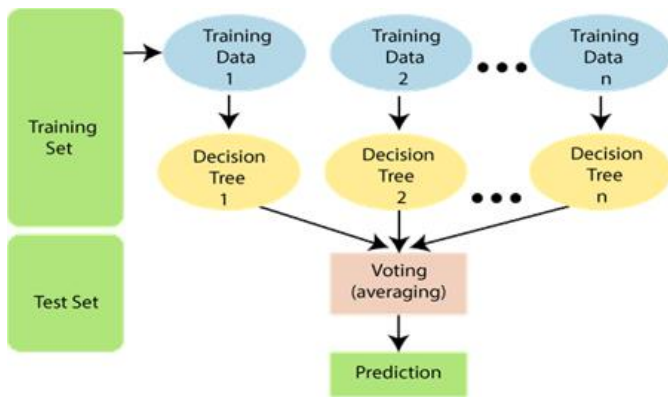


Figure 1: Random Forest

➤ Why use Random Forest?

Below are few points that explain why we should use the Random Forest algorithm:

- It requires shorter training time than other algorithms.
- It accurately predicts output, and it runs quickly even with large datasets.
- It can also maintain accuracy when a considerable amount of the data in the dataset is missing.

IV. SYSTEM ARCHITECTURE

Machine learning-based disease prediction forecasts the presence of a disease for a user based on numerous symptoms and information provided by the user, such as sugar level, hemoglobin level, and other general information via symptoms. The architecture of the system for disease prediction using machine learning consists of several datasets that will be used to compare and predict the user's symptoms, after which the datasets will be transformed into smaller sets and categorized using classification algorithms. After that, the classified data is fed into machine learning algorithms, which process the data and feed it into a disease prediction model that incorporates all of the above-mentioned user inputs. The system then combines and compares the aforesaid information and total processed data in the prediction model, and eventually forecasts the disease. An architecture diagram is a visual representation of a set of architectural concepts, including their principles,

elements, and components. The diagram depicts the system software in the context of a system overview.

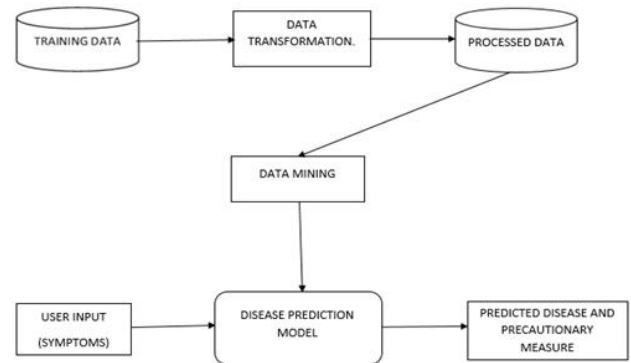


Figure 2: System Architecture

- **Training Data:** Data is used to train machine learning algorithms. They use the training data to form relationships, gain understanding, make judgments, and assess their confidence. The model works better when the training data is good.
- **Data Transformation:** The process of changing data from one format to another, often from the format of a source system to the required format of a destination system, is known as data transformation. Most data integration (essential for reliable output) and data management operations, such as data wrangling and data warehousing, need data transformation.
- **Processed Data:** Information refers to the data that is processed. Data processing is the transformation of raw data into useful data through a series of steps. Data is analyzed to produce results that aid in the resolution of a problem or the improvement of a present situation. Raw data is information that has not yet been processed for use.
- **Data Mining:** Data mining is the technique of predicting outcomes by identifying anomalies, trends, and correlations in huge data sets. You may use this information to enhance sales, lower costs, strengthen customer connections, reduce risks, and more using a variety of approaches.

- **Disease Prediction Model:** We employ the (Random Forest) Random Forest machine learning technique for disease prediction. The model is trained with the required disease numeric symptoms dataset for disease prediction.

## V. RESULT & DISCUSSIONS

### Experimental Setup

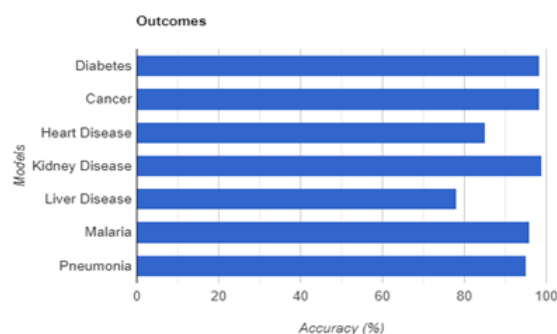
All the experimental cases are implemented in Python in conjunction with Machine learning & Deep learning tools and flask as backend, algorithms and strategies, and the Random Forest approach along with different feature extraction techniques, and run in the environment with System having configuration of Intel Core i5/i7, 2.30 GHz Windows 10 (64 bit) machine with 8GB of RAM

### Dataset Description

Patient disease dataset downloaded from Kaggle machine learning website.

### Result

This section presents the performance of the Random Forest and deep learning algorithms in terms of performance measure such as accuracy.



**Figure 3: Model Accuracies**

So, we are getting the accuracy above 95% for 5 models out of 7 models.

## VI. CONCLUSION

Finally, we can state that this project Health care & Disease prediction using machine learning is highly beneficial in everyone's day-to-day life, but it is especially significant for the healthcare sector, because they are the ones who use these systems on a regular basis to predict the diseases of patients based on their information and the symptoms that they have experienced. In today's world, the health industry plays a critical role in curing patients' diseases, so this is a useful tool for the health industry to inform users, and it's also useful for users if they don't want to visit a hospital or other clinic, because by simply entering numeric symptom data and other relevant information, the user can learn about the disease they're suffering from, and the health industry can benefit from this system by asking. If the health business embraces this initiative, doctors' workload will be decreased, and they will be able to forecast the disorder of patients more simply.

## VII. REFERENCES

- [1]. Akkem Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning and Flask API", IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1, 2020.
- [2]. Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities", IEEE transaction, 2017, pp 8869-8879.
- [3]. W. Yin and H. Schutze, "Convolutional neural network for paraphrase identification", in HLTNAACL, 2015, pp. 901-911.
- [4]. L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for tele-health in cloud computing", in Smart Cloud (Smart Cloud), IEEE International Conference on. IEEE, 2016, pp. 184-189.

- [5]. Siwei Lai, Xu Kang Liu, Jun Zhao,” Recurrent Convolutional Neural Networks for Text Classification”, in proceeding of the twentyninth AAAI Conference on Artificial Intelligence 2015.
- [6]. Xingyou Wang, Weijie Jiang, Zhiyong Luo, “Combination of Convolutional and Recurrent Neural Network for Sentimental Analysis of Short Texts”, International Conference on Computational Linguistics: technical papers, 2016, pg 2428-2437
- [7]. Ying Wen, Weinan Zhang, Rui Luo, Jun Wang, “Learning text representation using recurrent convolutional neural network with highway letters”, Neu-IR 16 SIGIR Workshop on Neural Information Retrieval, July 21,2016, Pisa, Italy.
- [8]. N. Nori, H. Kashima, K. Yamashita, H. Ikai, and Y. Imanaka, “Simultaneous modelling of multiple diseases for mortality prediction in acute hospital care”, in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.
- [9]. Liqiang Nie, Xiaochi Wei, Dongxiang Zhang, Xiang Wang, Zhipeng Gao, and Yi Yang,” Data-driven Answer Selection in Community QA Systems”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING JUNE 2016.
- [10]. Kanchan M. Tarwani, Swathi Edem, “Survey on Recurrent Neural Network in Natural Language Processing”, International Journal of Engineering Trends and Technology (IJETT), Volume 48 Number 6, June 2017.
- [11]. Jan Koutnik, Klaus Greff, Faustino Gomez, Jurgen Schmidhuber, “A Clockwork RNN”, IDSIA, USI SUPSI, Manno Lugano, CH-6928, Switzerland.