Multi Diseases Prognosis System using Machine Learning

4 authors, including:

Sahil Pandita
MIT Academy of Engineering
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Hitesh Bulani
Maharashtra Institute of Technology
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   MULTI DISEASES PROGNOSIS SYSTEM USING MACHINE LEARNING View project

Multi Diseases Prognosis System using Machine Learning

Hitesh Bulani[1], Ajay Bhilare[2], Sahil Pandita[3], Ms. Farhana Shaikh[4]

[1, 2, 3] Department of Computer Engineering, MIT Academy of Engineering, Pune, India
Email – [1]hgbulani@mitaoe.ac.in, [2]asbhilare@mitaoe.ac.in,
[3]spandita@mitaoe.ac.in, [4]fjshaikh@mitaoe.ac.in

***Abstract** – Smart healthcare is a growing need in India. The covid pandemic has shown us the worst of Indian Healthcare. Innovation in the healthcare industry would bring about an important change in managing a patient. There are virtual doctors who choose to consult patients online via video and phone appointments, rather than in-person appointments but this is not possible during the case of an emergency. Machines are always considered better than humans as, without any human error, they can perform tasks more efficiently and with a good accuracy. A Disease Prediction system can act as a saviour where the rush of patients is at its par. So, we decided to design a multi diseases prognosis system based on machine learning. We have used compared Navies Bayes Algorithm for prediction of diseases. Based on the symptoms, the prognosis system gives the output as the disease that the individual might be suffering from.*

Keywords – Multi Disease Prediction, Machine Learning, Navies Bayes, Symptoms

I. INTRODUCTION

Smart healthcare is a booming industry. Technological advancements in the healthcare industry are one of the most significant callings of the 21st century. Medicine and healthcare are the most important parts of human life.

Our main objective of the project was to develop a disease prognosis system using models of machine learning which will help a user to identify his diseases on the basis of whatever symptoms he has provided to the system. This system will help the user to prognose some high-risk diseases they have during emergency cases. So here we are going to develop a Multi-Disease Prognosis system in which Navies Bayes classification will be used to predict the various other diseases based on the symptoms the patient is dealing with. So, by using Machine Learning the proposed system will detect the diseases in a better way. Types of disease like Allergy, GERD, Fungal Infection, Ulcers, Migraine, Diabetes, Jaundice, Malaria, Common Cold will be covered under this model.

II.LITERATURE REVIEW

Dr. Balaji, Dr. Suresh, Dr. Veer Raju [1] worked on symptoms based disease prediction using big data analytics. They used data collected from various repositories and store the entire

data on Hadoop Distributed file HDFS. They used various classification algorithm such as Decision Tree, Random Forest and Navies Bayes. Among this model the Navies Bayes Algorithm provide the highest accuracy.

Ayan Mir, Sudhir Dhage [2] designed a model for Diabetes Diseases prediction Their aim was to build a classifier model using WEKA tool to predict diabetes disease by using Naive Bayes, Support Vector Machine, Random Forest and Simple CART algorithm. They wanted to find out the best algorithm based on performance result for the prediction of diabetes disease. Experimental results of each algorithm used on the dataset was evaluated. It is observed that Support Vector Machine performed best in prediction of the disease having maximum accuracy

P.Sreevalli, keerthana [3] used the random forest machine-learning algorithm to predict the disease based on the symptoms. The system resulted in low time consumption and minimal cost for the prediction of diseases. The algorithm resulted in an accuracy of 84.2 %.

Bhanuteja, Venkata, Poornachand [4] worked on symptoms based diseases prediction. The entire system was designed by them in such a way to predict the diseases. They used three Algorithms i.e., Decision Tree model, LightGBM model and Random Forest classifier model, so that the predictive analysis study is proposed at the end of the study by exploring its speed, efficiency and performance of the various algorithms for the input dataset.

Nishant Yede, Ritik Koul, Chetan Harde [5] used Machine learning Decision tree map, Navie Bayes, Random forest algorithm for diseases prediction. They used structured and unstructured data from hospital. The accuracy of their proposed algorithm was 94.5%.

S Radhika, S Ramya Shree, A Ranjitha [6] worked on symptom based diseases prediction where they used decision tree algorithm and Electronic Health Record Analysis. The Decision Tree Classifier algorithm helped them to detect the patient's health condition after receiving their symptoms by giving the predicted disease. Their dataset contained physiological measurements with 40 Diseases and 132 Symptoms. Additionally, the respective patient's EHR is also collected for summarizing the prescription/test report using NLTK.

III. PROPOSED METHODOLOGY

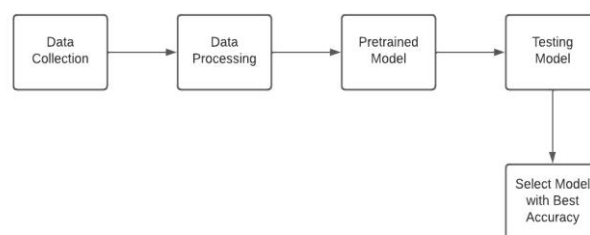The proposed methodology consists of 6 main parts as shown in Fig 1

Fig 1 Proposed Block Diagram

A. Data Collection
The dataset was downloaded from kaggle which consisted of 2 parts, train and test. Total unique diseases present in dataset are 41 and total unique symptoms present are 132.

B. Data Pre-processing

First of all, we created two lists ; one of symptoms and other one of disease. Replaced all the values of diseases with numbers starting from 0 to 40 as there are 40 diseases. Then we made the dataframe containing all the symptoms as colums and one last column of prognosis in which the name of the disease corresponding that symptoms will be present.

C. Selecting Pre-trained ML Models

There are four different kind of models present in our project to predict the disease these are
• Decision tree
• Random forest tree
• KNN
• Gaussian Navies Bayes

Decision tree is an effective classification algorithm. In decision tree classification technique, first of all a root attribute is selected on the basis of various measures such as information gain and C4.5. Decision tree is made from to down in divide and conquer recursive manner. Partition of attributes is repeated until there are no samples left or no attributes are remaining for further partition.

Random Forest Algorithm is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps –
1. It chooses random data samples from dataset.

2. It constructs decision trees for every sample dataset chosen.
3. At this step every predicted result will be compiled and voted on.
4. At last most voted prediction will be selected and be presented as result of classification.

In this project we have used random forest classifier with 100 random samples

K Nearest Neighbour is a supervised learning algorithm. It converts the datasets into certain datapoints and then estimates the chance of the particular datapoint to be part of the certain group having nearest distance to it.

Naïve Bayes algorithm is a probabilistic model based on naïve bayes theorem.
It works on conditional probability. The attributes used are likelihood probability, prior class probability, prior object probability and posterior probability.

D. Testing the Model

After running the model on training data, we ran all four model on testing data and obtained accuracy of each and every model. With Decision tree we obtained an accuracy of ~92%. Random Forest Algorithm had an accuracy of ~89%. KNN had an accuracy of ~93% and Navies Bayes had the highest accuracy ~95%

E. Select the model with the Best Accuracy

After, comparing all the four machine learning algorithm we found that naive bayes had the highest accuracy of ~95% in the prognosis of the diseases on the basis of the symptoms. So, we used naive bayes algorithm in the proposed system for prognosis of the diseases.

IV. RESULT AND DISCUSSION

Different machine learning models were used to examine the prediction of disease for available input dataset. We used 4 different ML models for the prediction. Out of the 4 models we used Desicion Tree had an accuracy of ~92%. Random Forest Algorithm had an accuracy of ~89%. KNN had an accuracy of ~93% and Navies Bayes had the highest accuracy ~95%. Navies Bayes had the Highest Accuracy among all the four algorithms. So, our proposed system consist of navies bayes algorithm for the prediction of the diseases.

V.CONCLUSION AND FUTURE WORK

We set out to create a system which can predict disease on the basis of symptoms given to it. Such a system can decrease the rush situations at hospitals and reduce the workload on medical staff. We were successful in creating such a system. We compared the result accuracy of 4 ML models. We used the ML Algorithm which has the highest accuracy in our proposed system. We achieved accuracy of ~95%. Creating this system we also added a way to store the data entered by the user in the database which can be used in future to help in creating better version of such system. Our system also has an easy to use interface. Currently, our system can prognose 42 common diseases but in future we can add more diseases in system for prognosis.

VI. REFERENCES

1. S. S. Prakash and K. Visakha, "Breast Cancer Malignancy Prediction Using Deep Learning Neural Networks," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 88-92, doi: 10.1109/ICIRCA48905.2020.9183378.
2. S. A. El Rahman, A. Al-montasheri, B. Al-hazmi, H. Al-dkaan and M. Al-shehri, "Machine Learning Model for Diseases Prediction," 2019 International Conference on Fourth Industrial Revolution (ICFIR), 2019, pp. 1-8, doi: 10.1109/ICFIR.2019.8894777.
3. Maram, Balajee & Dr, Suresh Kumar K & Gampala, Veerraju. (2021). SYMPTOMS BASED DISEASE PREDICTION USING BIG DATA ANALYTICS. 32. 3228-3234.
4. Keniya, Rinkal & Khakharia, Aman & Shah, Vruddhi & Gada, Vrushabh & Manjalkar, Ruchi & Thaker, Tirth & Warang, Mahesh & Mehendale, Ninad. (2020). Disease Prediction From Various Symptoms Using Machine Learning. SSRN ElectronicJournal.10.2139/ssrn.3661426.
5. Ayman Mir, Sudhir Dhage "Diabetes Disease Prediction using Machine Learning on Big Data of Healthcare" 2018 Fourth International Conference on

Computing Communication Control and Automation (ICCUBEA) 978-1-5386-5257-2

6. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm

7. Talasila Bhanuteja, Kilaru Venkata Narendra Kumar, Kolli Sai Poornachand "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-10 Issue-9, July 2021

8. Nishant Yede, Ritik Koul, Chetan Harde ,Kumar Gaurav "General disease prediction based on symptoms provided by patient" Open Access International Journal of Science & Engineering. Volume 6 June 2021 DOI 10.51397/OAIJSE06.2021.0032

9. S Radhika, S Ramiya Shree and A Ranjitha "Symptoms Based Disease Prediction Using Decision Tree and Electronic Health Record Analysis" European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 4, 2020