Comprehensive Healthcare System Using ML

Prof.Aparna Kalaskar
Assistant Professor
Department Of Computer Science and Engineering
Savitribai Phule Pune University, Pune, India
asnakil.scoe@sinhgad.edu

Abhijit Shendage
Department Of Computer Science and Engineering
Savitribai Phule Pune University, Pune,
abhijit1191999@gmail.com

Saurabh Gore
Department Of Computer Science and Engineering
Savitribai Phule Pune University, Pune, India
saurabhgore38@gmail.com

Mr. Yash Patil
Department Of Computer Science and Engineering
Savitribai Phule Pune University, Pune, India
yashpatil17290@gmail.com

Mr. Samyak Vaidya
Department Of Computer Science and Engineering
Savitribai Phule Pune University, Pune, India
samyakvaidya2020@gmail.com

***Abstract*- People fall sick often, though everyone can't seek immediate medical help. So many people take advice from unprofessionals. As seen in the Covid situation, it may not always be possible to consult the doctor. Persons with disabilities or people having serious medical conditions face difficulties in visiting doctors and standing in queues. It has been observed that the people in rural and remote areas face the same difficulties. As well, due the newly normalized helter-skelter lifestyle we choose to ignore our medical problems or sometimes we may overvalue them.This all might lead to some serious concerns.Therefore, this paper proposes a web application which uses modern machine learning algorithms to diagnose user symptoms.The proposed system also includes all the generalized healthcare features which can be virtualized. The system connects all healthcare related stakeholders with one another, so as they can communicate with others to solve their respective problems. Patients can take medical opinions from medical professionals, order pharmaceuticals from nearby pharmacies.We investigate our proposed framework by using 3 different machine learning techniques. Performance of the framework for identifying appropriate medical departments under the machine learning techniques is thoroughly investigated and compared. This framework can be used for telemedicine platforms or in the automated hospital management sector. This may create a path of enormous development in the health care sector..**

I. Introduction

The Covid Pandemic has taught us nothing but how much we lag in providing good healthcare facilities to people. Due to the increasing population, it is not possible for everyone to be admitted in the hospitals and take treatment as we have limited medical resources. This tells us how important it is to remotely monitor the patient.Also, the underprivileged may not always be able to seek the medical help and take necessary medication.So, it is the need of the century to to digitalized healthcare infrastructure which can be accessed by everyone in need.

The ongoing advancements in computer fields such as machine learning and artificial intelligence can surely be used to improve the current situation.Nowadays ML is one of the most rapidly developing field of AI which is used in many areas of life, like healthcare, cyber security, commerce and many other fields. In the past few years, large amounts of data have been collected and stored because of monitoring and many other data collection devices are available in modern hospitals which enable researchers to use ML in the healthcare sector more efficiently.

As these devices are being used every day, and enormous amounts of data are being generated. It is quite hard or even impossible for human beings to derive useful information from these massive amounts of data sets. This is why to analyze these data and diagnose problems in the healthcare field ML is used widely. A simplified explanation ML is it will learn from previously diagnosed cases of patients and utilize it for future diagnosis. The resulting classifier can be helpful for the doctors to diagnose new patients faster and efficiently and train students and non-specialists to diagnose patients.Precisely for that reason ML can be exploited in disease symptom analysis based department selection.

The Comprehensive healthcare system includes all the generalized healthcare features which can be virtualized. The system connects all healthcare related stakeholders with one another, so as they can communicate with others to solve their respective problems. Patients can take medical opinion from medical professionals, order pharmaceuticals from nearby pharmacies.Patient can share their experience with doctor through feedback, so that other users can use it for better experience of the system.The system has introduced a chatting system for better communication. The Doctors and Pharmacists can be filtered based on location to access nearby facilities. Proper tabular format of pending, in-progress, completed requests are stored and presented. Doctors can write prescriptions with different fields such as type of medicine, quantity, time of intake etc. Patients can order these medicines from nearby pharmacies as soon as they receive the prescription. Pharmacists can verify the request for medicine, create a bill in advance, and schedule a day for delivery of medicines. Users can verify the bill, and confirm the request for delivery of medicine.

The system has automated disease diagnosis application. Here the inbuilt module analyses previous records based on user symptoms to predict the disease they may be experiencing. While user provides inputs, the system provides suggestions to users comparable to given inputs using decision tree. If user is not sure about input, the user can select an option to answer the questionnaire asked by the system. Naïve Bayes, K-Nearest Neighbors, Decision Tree these algorithms are used for disease prediction. The accuracy of these algorithm is >90 most of the times. During prediction system varies the database based on user inputs, and includes only the relevant symptoms. This leads to varied database, which in turn leads to varied accuracy. Hence, we select the algorithm with highest accuracy to get the accurate result. The probabilistic prediction is used for better understanding. Also,extension of WebMD is added to the system to show the information of predicted disease.

Doctors can schedule appointments with patients and communicate with them to provide proper consultation to the user. Doctors can use an orthopedic classification module to analyze biomechanical values, useful to predict the disease the patient may be experiencing. This module uses Naïve Bayes and K-Nearest Neighbors to predict the result. The accuracy of the algorithms is tested with a confusion matrix. The accuracy of Naïve Bayes algorithm is 80%, of K-Nearest neighbors is 86%.

The system also has modules to display the current healthcare related affairs like ongoing Covid-19 pandemic. The graphical representation of these latest updates are displayed on the website. The graphical representation on the system dataset is also available to get an idea ofsystem's automated diagnosis module. The calendar is available to schedule, track daily events and receive proper notifications. The about page provides proper details of the system, contact information if the user is facing any problem while using the system.

The Interconnection between these modules is done smoothly considering user convenience and avoids any confusion that user may experience. The safety and integrity of data has been considered and has been handled well through Django Framework.The rest of the paper is organized as below,in section II of this paper Literature Review presented. Progressing to supervised machine learning and different algorithms that we have used. After the algorithms, the data is discussed. Section III presents our development of the ML model and Methodology. And section V describes the parameters for comparison of models. Section VI discusses the results and analyzes them properly. And at last section VII shows our conclusion and final remark about the algorithms we have applied and our future works.

II. Literature Survey

There are a lot of efforts already being taken to improvise the healthcare system.Though, most of them focus only on increasing the disease prediction accuracy. Some of them were able to reach the accuracy of 60% to 80%. The main drawback is that there isn't any single existing system which truly connects all the aspects of the healthcare system.

The Sneha Grampurohit and Chetan Sagarnal's work [1], presents comparative study of the results obtained by algorithms such as Decision tree classifier & Random Forest classifier.
They were able to achieve the accuracy of 93 and 95 % but that was on a limited dataset.Also the system had any changes in the database and was not tolerable to unknown symptoms.

The Md. Latifur Rahman, Rahad Arman Nabid ,Md. Farhad Hossain's work[2], presents disease symptom analysis based department selection using machine learning for medical treatment. The idea was to present the comparative study of nearly all classifying algorithms based on their accuracies. As the dataset considered was very small in size the possibility of accuracy of algorithms working on large dataset was out of the thinking box.

The Feng-Jen Yang's work[3], presents a twisted method to implement Naive Bayes Classifier. Although this paper is not directly related to healthcare, it discusses the mathematical model of Naive Bayes Classifier in detail.He took the advantage of the powerful built-in programming constructs in Python programming language, this implementation of Naive Bayes classifier is done without much intensive coding.

The Tejal P. Burange, Dr. P. N. Chatur's work[4], presents analysis of symptoms wise disease inference systems using data mining techniques. They implemented an automatic question-answer system in which the user will ask any question and the system will process that question using data mining algorithms to find out the proper answer. The paper was solely focusing on symptoms discovery instead of disease prediction and analysis.

III. Proposed Methodology

A. System Architecture :

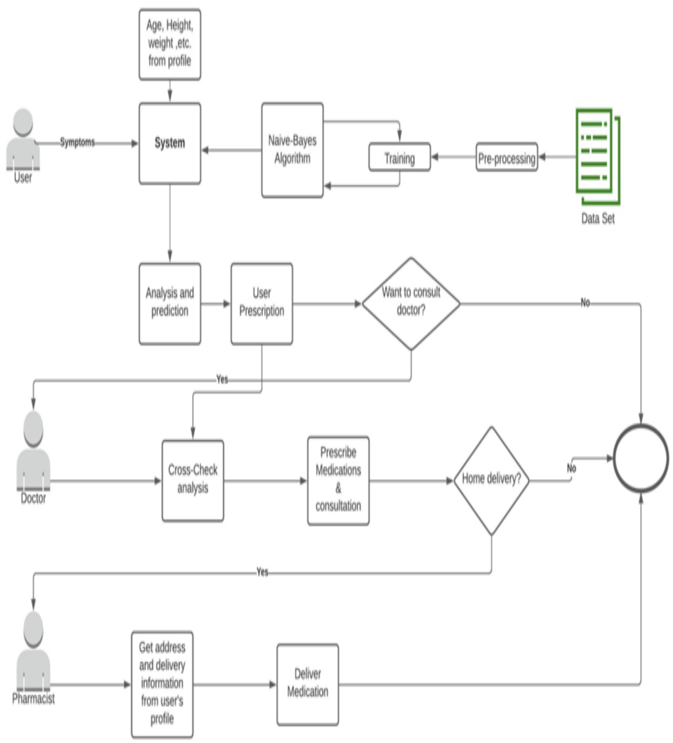

Fig. 1. General System Architecture

As shown in the figure, to access the system the user first has to authenticate himself or himself by providing required credentials. The user can then feed his or her symptoms to the system. The system, at the same time suggests the other symptoms using a decision tree classifier. The user has the option to identify it if he or she has it by answering the questions. The user also has the option to enter their symptoms manually. After that the system with the help of algorithms such as Naive Bayes Classifier, K-Nearest Neighbour Classifier, and the Decision Tree Classifier suggest to a user a probable disease that he or she may have. Furthermore if the user is not convinced, they can consult6 the medical professional by scheduling the appointment. The doctor then can see and analyze the diagnosis and can prescribe required medication to the user. Users can immediately access this prescription and get access to nearby pharmacists to order the prescribed medication. The system also has a communication module and orthopedic analysis module.

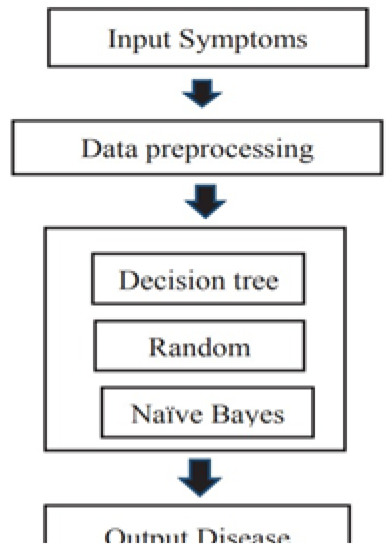

Fig. 2. Detailed Modules Description

The system predicts the disease based on user input i.e. symptoms. The system filters the previously available dataset based on given input and considers only the relevant data for training. Hence, changing dataset leads to changing accuracies of algorithms. The system uses 3 algorithms Naïve Bayes, K-Nearest Neighbors, and Decision tree classifier for training. And shows user data of the most accurate result. The accuracy of these algorithms differ from 70%-100%.

B. Mathematical Model :

a) Model for AI Diagnosis

AD = {I, O, F, T

Where,

$I = \{i_1, i_2, i_3, \ldots i_n\}$

I is the user symptoms

$O = \{o_1, o_2, o_3, \ldots o_n\}$

O is the set of diagnosed diseases

$F = \{f_1, f_2, f_3\}$

f_1= function for data cleaning

f_2= function to retrieve user details

f_3= function for symptoms recommendation

$T = \{t_1, t_2, t_3, \ldots t_n\}$

T is the collection of symptoms and disease data to train the model for Diagnosis

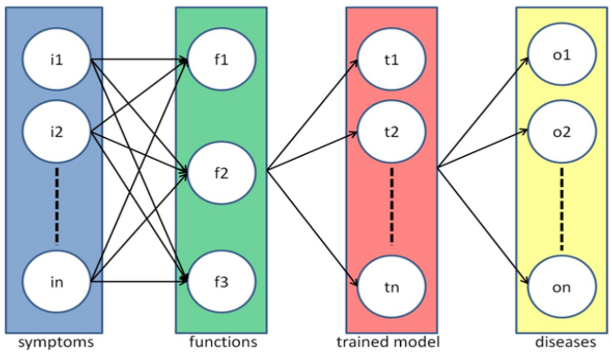

Fig.3. : Mathematical Model for AI Diagnosis

b) Model for Scheduling Appointment

$SA = \{I, O, F\}$

Where,

$I=\{i_{1,2}\}$

i_1= User selects a doctor for consultation.

i_2=Doctor selects feasible date and time for an appointment.

$O = \{o_1, o_2\}$

o_1 = Date and Time of scheduled appointment
o_2 = Information of stakeholders involved

$F = \{f_1, f_2\}$

f_1= Function to recommend doctor to user
f_2 = Function to fix appointment and convey its details to user

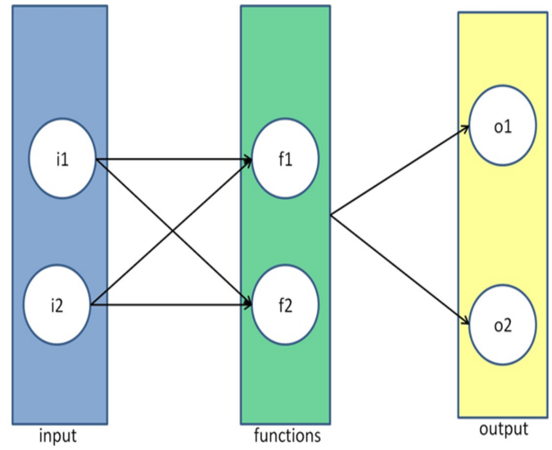

Fig.4. Model for Scheduling Appointment

c) Model for Stakeholder Communication

$SC = \{I, O, F\}$

Where,

$I = \{i_1\}$

i_1 = Information of stakeholders communicating

$O = \{o_1, o_2\}$

o_1 = Meeting conclusion

o_2 = User feedback

$F = \{f_1, f_2, f_3\}$

f_1 = function for remainder if appointment scheduled.

f_2 = function to provide a platform for communication.

f_3 = function to collect feedback after meeting

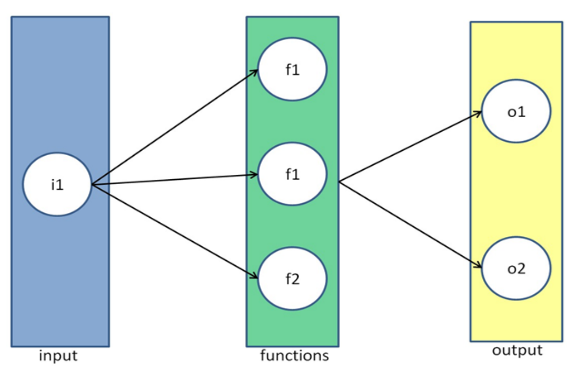

Fig 5. Model For Stakeholder Communication

C. *Algorithms:*

a)Decision Tree algorithm for symptoms suggestion:

The classification models built by decision tree resemble the structure of tree. By learning the series of explicit if-then rules on feature values (symptoms in our case), it breaks down the dataset into smaller and smaller subsets that results in predicting a target value(disease).

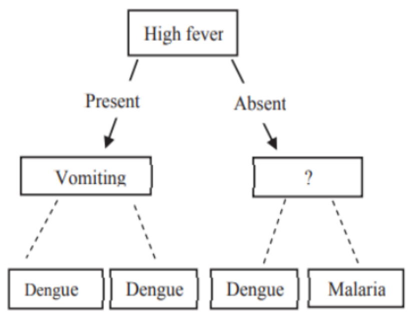

Fig 6.Decision Tree Example

A decision tree consists of the decision nodes and leaf nodes.

• Decision node: Has two or more branches. In our work presented, all the symptoms are considered as decision nodes.

• Leaf node: Represents the classification that is, the Decision of any branch. Here the Diseases correspond to the leaf nodes. After reducing to smallest dataset entropy is each symptom is calculated to get top 5 symptoms for user suggestion.

b) Naïve Bayes Algorithm: :

Naïve Byes algorithm is based on Bayes theorem given by:

$$P(s/h) = \frac{P(h/s)P(s)}{P(h)}$$

In the formula above 's' denotes class and 'h' denotes features. In P(h), the denominator consists the only term that is a function of data(features)- it is not a function of the class we are currently dealing with. Thus, it will be same for all the classes. Traditionally in naïve Bayes Classification, we ignore this denominator as it does not affect the result of the classifier in order to make the prediction:

$$P(s/h) \propto P(h/s)P(s)$$

c) K-Nearest Neighbors:

The KNN algorithm finds the K nearest records from available data. As dataset is variable the value of is determined dynamically. And the diseases with highest number records in these k records has high probability and with lowest records has low probability.

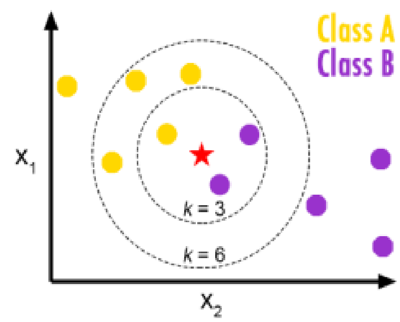

Fig.7 K-Nearest Neighbours Example

IV. Conclusion

From the historical development of machine learning and its applications in the medical sector, it can be shown that systems and methodologies have emerged that have enabled sophisticated data analysis by simple and straightforward use of machine learning algorithms. This paper presents a comprehensive comparative study of three algorithms performance on a medical record each yielding an accuracy >90 percent. The performance is analyzed through confusion matrix, accuracy score and visualization aids such as graphs. Artificial Intelligence will play an even more important role in data analysis in the future due to the availability of huge data produced and stored by modern technology.The proposed system is also able to connect all the stakeholders and provide crucial medical assistance sitting at any place.

References

[1] 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science "Disease Symptom Analysis Based Department Selection Using Machine Learning for Medical" by *Md. Latifur Rahman, Rahad Arman Nabid*, and *Md. Farhad Hossain*.

[2] 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020 "Disease Prediction using Machine Learning Algorithms" by *Sneha Grampurohit and Chetan Sagarnal*.

[3] 2018 International Conference on Computational Science and Computational Intelligence (CSCI) "An Implementation of Naive Bayes Classifier " by *Feng-Jen Yang*.

[4] Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) "Analysis of Symptoms Wise Disease Inference System Using

Data Mining Technique" by *Tejal P. Burange* and *Dr. P. N. Chatur*

[5] Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) "An Intelligent Framework for Health Estimation with Naïve Bayes Approach" by *Rahul Katarya*

[6] International Journal of Intelligent Systems and Applications in Engineering ISSN:2147-67992 "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification" by *Mücahid Mustafa Saritas and Ali Yasar*