Comprehensive Healthcare System Using ML

[1]Prof. Aparna Kalaskar,[2]Abhijit Shendage,[3]Saurabh Gore, [4]Yash Patil, [5]Samyak Vaidya

[1]Professor,[2]Student,[3]Student, [4]Student,[5]Student,
Computer Engineering Department,
Sinhgad College of Engineering, Pune, India.

Abstract : People often fall sick, though everyone may not be able seek immediate medical help. As seen in the Covid-19 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 because of long queues and lengthy processes. It has been observed that the people in rural and remote areas face the same difficulties. As well, due to the newly normalized helter-skelter lifestyle, we choose to ignore our medical problems, or sometimes we may overvalue them. In such situations, many people take advice from unprofessional. This all might lead to some serious concerns. Therefore, this paper proposes a web application that 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. Our proposed framework implements 3 different machine learning techniques Naïve Bayes, K-Nearest Neighbors, Decision Tree Classifier. This may create a path of enormous development in the health care sector.

***IndexTerms* – Disease Prediction, Django, Doctor, K-Nearest Neighbors, decision tree, healthcare, Machine Learning, Pharmacists.**

I. INTRODUCTION

The Covid-19 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 under-privileged may not always be able to seek the medical help and take necessary medication. So, it is the need of the century 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.

The proposed 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.

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

3.1 System Architecture

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 in data pe-processing phase. Hence, changing dataset leads to changing accuracy of algorithms. The symptoms similar to user input are also suggested to user using decision tree classifier. 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 differs from 70%-100%. The result of most accurate algorithm is shown to user in probabilistic percentile format in final stage.

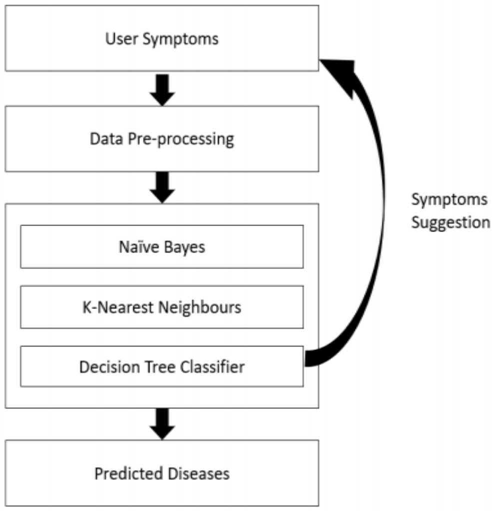

Fig.1: Diagnosis Module

As shown in the Fig 2, 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. And system predicts the output as discussed. Furthermore, if the user is not convinced, they can consult 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 pharmacist can produce a bill and schedule a delivery. The system also has a communication module and orthopaedic analysis module.

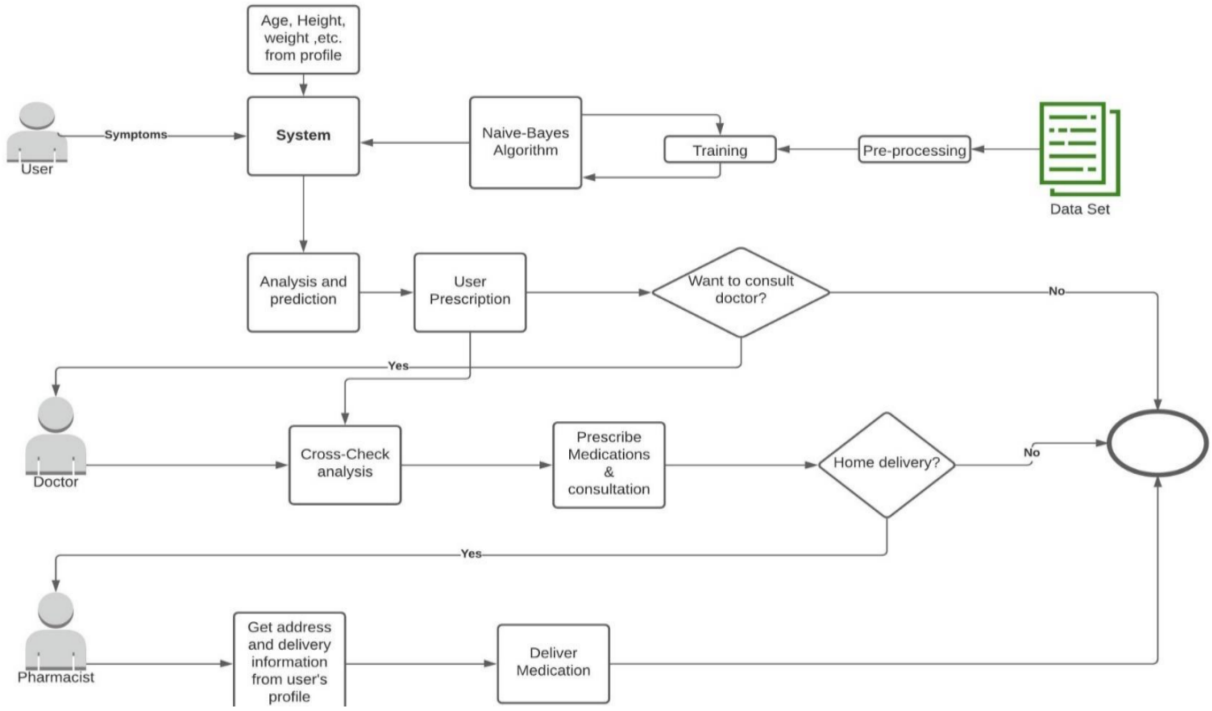

Fig.2: System Flow

3.2 Mathematical Model

3.2.1 Model for AI Diagnosis

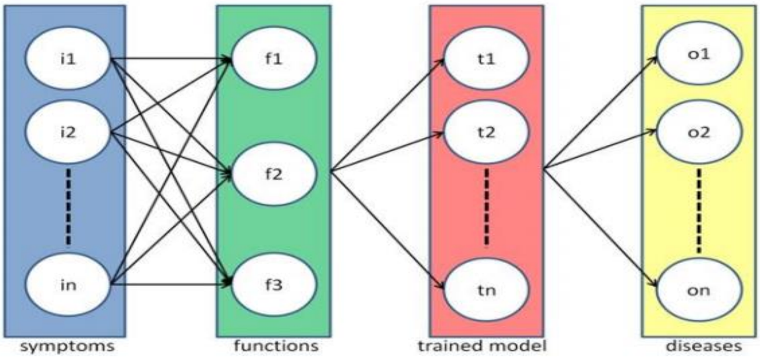

Fig.3: Mathematical Model for AI diagnosis

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.

3.2.2 Model for Scheduling Appointment

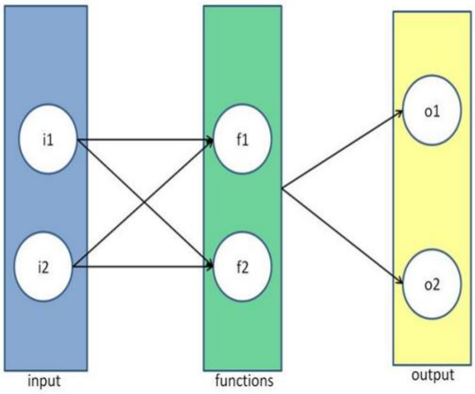

Fig.4: Mathematical Model for Scheduling Appointment

SA = {I, O, F}
Where,

I= $\{i_1, i_2\}$
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

3.2.3 Model for Stakeholder Communication

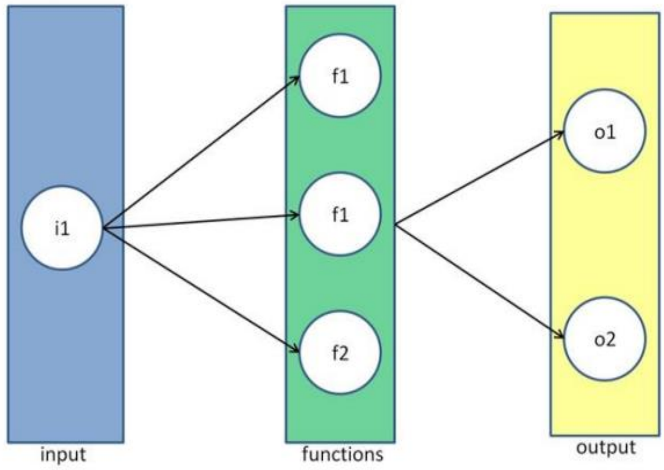

Fig. 5: Mathematical Model for Stakeholder Communication

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

3.3 Algorithms

3.2.1 Decision Tree algorithm for symptoms suggestion

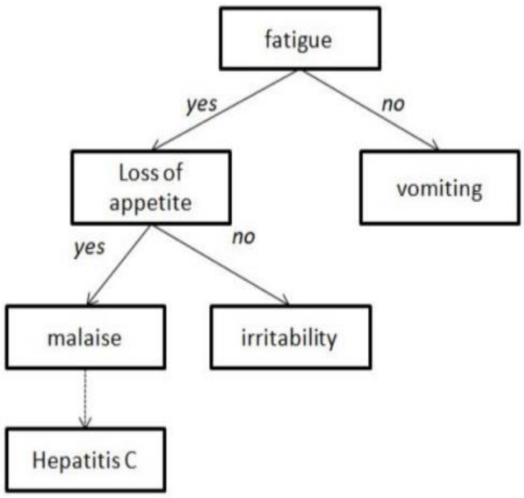

Fig. 6: Implementation of decision tree

The classifier works as its name suggest, it constructs tree based on the user's input. The classifier distributes data in two parts i.e. present or absent and creates smaller and smaller dataset, and calculates information gain of remaining symptoms to get top 5 symptoms for user suggestion. The internal nodes of tree are user symptoms (input), the leaf nodes are Disease (output). In this way, the process of suggestion goes on till the tree reaches its limit i.e. leaf node.

3.2.2 Naïve Bayes Algorithm

Naïve Bayes algorithm uses Bayes's theorem to calculate probability of certain result.

$$P(Dis|Sym_1,Sym_2.....,Sym_n) = \frac{P(Sym_1|Dis) \text{ X } P(Sym_2|Dis) \text{ X } .... P(Sym_n|Dis)}{P(Sym_1) \text{ X } P(Sym_2) \text{ X } .... P(Sym_n)}$$

Where,

Sym = Symptoms that user is experiencing
Dis = Disease predicted by system
P (|) = Conditional Probability

This interpretation of Bayesian theorem is used for probability calculation of each disease. The denominator of equation remains constant for each disease hence it is ignored and result is proportional to numerator.

3.2.3 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 have high probability and with lowest records has low probability.

IV. IMPLEMENTATION AND RESULTS

4.1 System Modules

4.1.1 Disease Prediction

The system has automated disease diagnosis module. Here system 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, they can select an option to answer the questionnaire asked by the system.

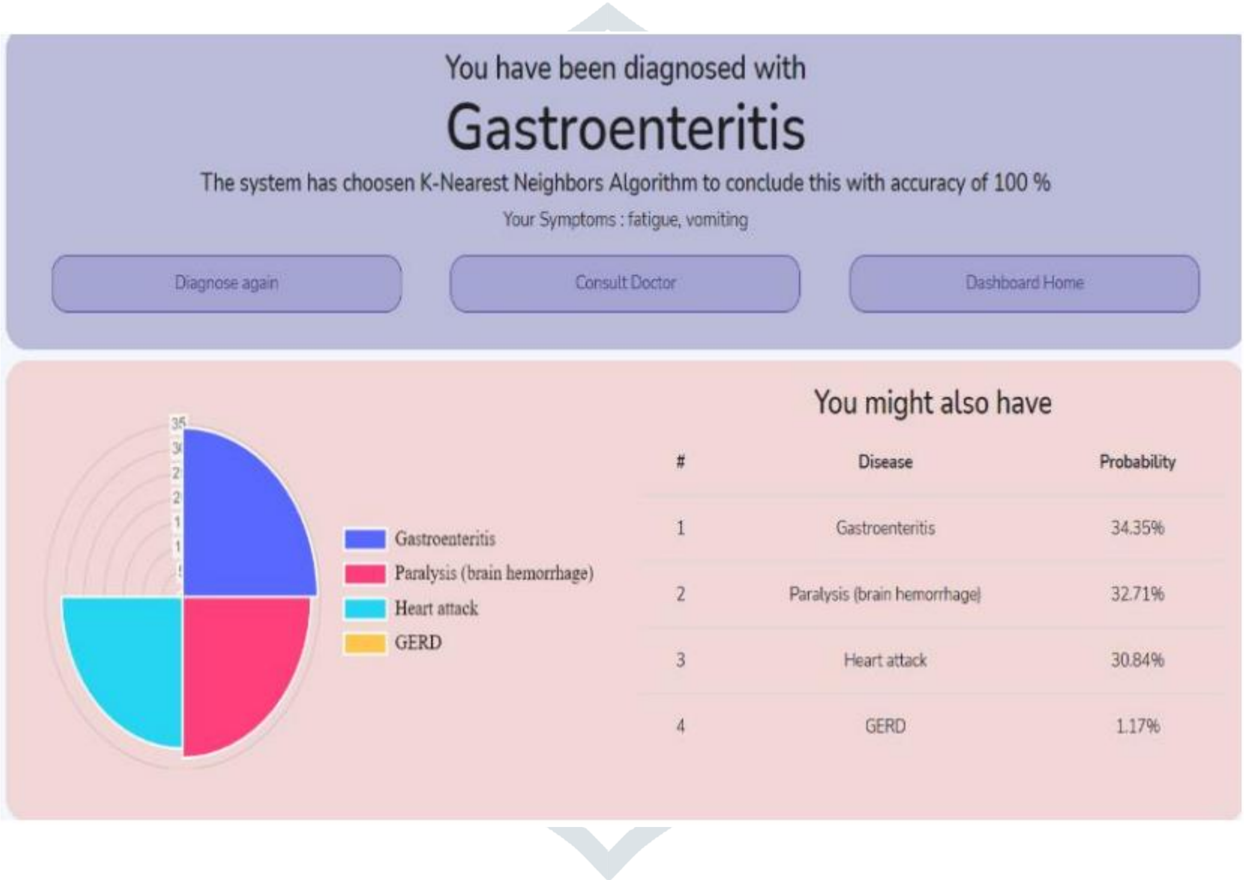

Fig.7: Disease Prediction

Naïve Bayes, K-Nearest Neighbors, Decision Tree these algorithms are used for disease prediction. The accuracy of these algorithm is 60%-100% most of the times. During prediction system considers the data relevant to user given input, and includes only the related symptoms. This leads to varied database, which in turn leads to varied accuracy. Hence, we select the algorithm with highest accuracy to get the final result shown to the user. The probabilistic prediction is used for user's better understanding. Also, extension of WebMD is added to the system to show the information of predicted disease.

4.1.2 Stakeholder Communication model

The various stakeholders of the system can communicate with each other using this model. The patient, after scheduling an appointment with the doctor, can communicate by sending text messages to them and then doctors can reply or ignore the same. The same is applicable with the pharmacist. They can discuss the rates of medicines, and date and time of delivery of medicines.

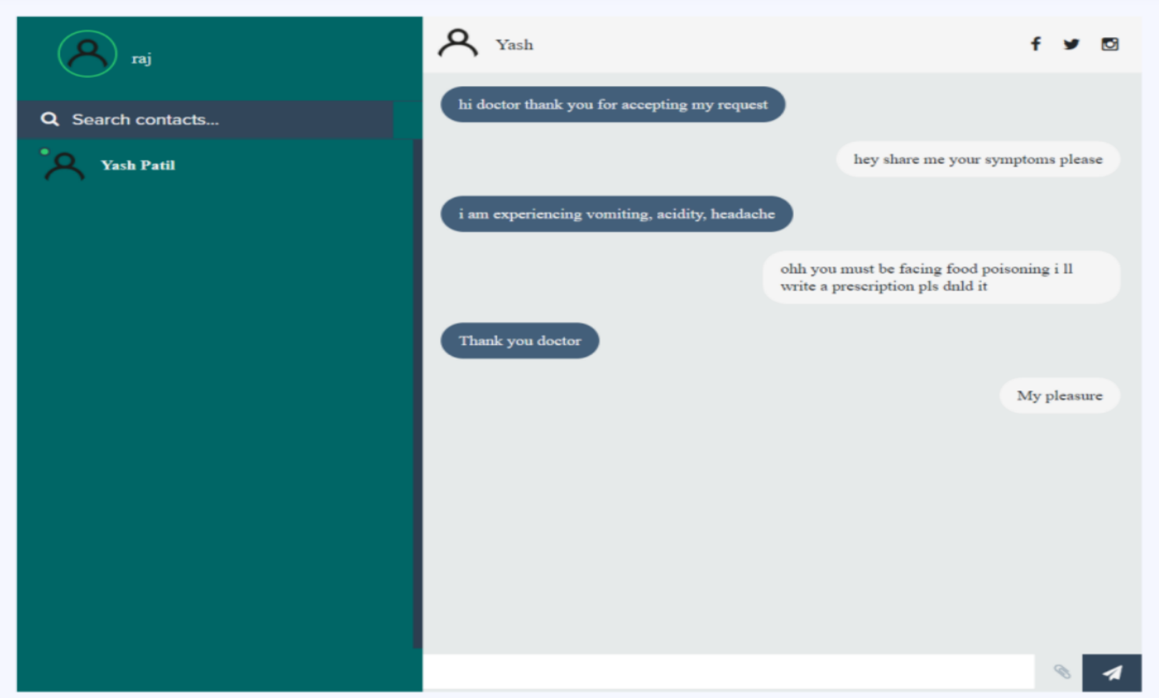

Fig.8 Stakeholder Communication Interface

4.1.3 Prescription and Bill Generation Model

After diagnosis and communicating with the patient, the doctors then can prescribe the required medication to the user using the system itself. The medications are classified into 3 categories viz. Tablet, syrup and ointment. The doctor can also specify the quantity, the time of having the medicine and when they should be taken i.e. before or after meal. This prescription is then mailed to the user and sent to the pharmacist who is nearby to the patient and can deliver the medication to the patient.

Bill

For,
Name : Yash Patil
Address : C-703, Akashdeep Apt., Ganeshnagar, Dhayri, Sinhgad Road
Pune, 411041

From,
Name : Dr. om Patil
Address : C-703, Akashdeep Apt., Ganeshnagar, Dhayri, Sinhgad Road
Pune, 411041

#	Meditype	Name	Quantity	Price
1	Tablet	crosin	10	130
2	Tablet	aspirin	2	40
3	Ointment	moov	1	56
			Sub total :	**226**
			Other :	**0**
			Grand Total :	**226**

Thank you & Visit Again!!

Fig. 9: Bill from Pharmacist

The pharmacist is able add the bill and additional costs that they may charge to provide this service. The user can see this prescription on the dashboard or in his mailbox.

V. Conclusion

With time, the innovation in digitalization of healthcare facilities, many applications are being developed. The paper proposes a way of developing the centralized system connecting all the different aspects for digitalization of healthcare infrastructure. Based on the approach of using machine learning algorithms, the paper presents a comprehensive comparative study of three algorithms analyzing the user symptoms and providing a proper diagnosis to the user. Each yielding an accuracy >70 percent. The performance is analyzed through confusion matrix, accuracy score and visualization aids such as pie charts. The proposed system is able to connect all the stakeholders related to healthcare through a proper communication platform. So that crucial medical assistance can be provided at any time, from anywhere.

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