IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

DISEASE PREDICTION USING MACHINE LEARNING

Mithun.V.Mhatre [1], Ranjeet.R.Pawar[2]

HOD, Department of Computer Technology, Bharati Vidyapeeth Institute of Technology[1]

HOD, Department of Information Technology, Bharati Vidyapeeth Institute of Technology[2]

Abstract: We are designing a system called Disease Prediction using Machine Learning which isdedicated for medical sector. We see that there are many online diagnosis systems available but we can encounter one thing that this system has a search bar for searching the expertise on the basis of symptoms or diseases and guide to the user for nearby location of hospitals in theemergency situation. Many peoples are facing problems and challenges regarding disease and they are looking for online helps.

If this type of system is made for recommendation system used by doctors' medicines it will beneficial for real time use of system. Now coming to technical things for the prediction of disease we are used four different algorithms namely Decision Tree Classifier, Random Forest, K-Nearest Neighbor, Naive Bayes Classifier.

All these algorithms are directly connected to tanning and testing dataset and GUI. Users of the system are layman so they are unable to use and understand this type of system so in future we are going to use some Human Interaction concepts in same project or system like local language can be added which is understandable by users of different regions. So as a software engineers we are developed a system called Disease Prediction using Machine Learning which takes five symptoms as an input and gives prediction of disease from which user is suffering.

We are also tried to use feedback system in our application to improve system accuracy, currently all algorithms give up to 94% of accuracy.

Keyword: Decision Tree Classifier, Random Forest, K-Nearest Neighbor, Naive Bayes Classifier

I. INTRODUCTION

Number of patients are increasing day by day which result to rush to words OPD's, need of increasing hospital staff and many more. "Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap.Prediction of disease by looking at the symptoms is an integral part of treatment. In our project we have tried accurately predict a disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%.

Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interfaceto facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project." The proposed system solves all these real-life problems and gives provides solution on the same.

II. MOTIVATION

Health is very important for everyone. "Unfortunately, almost everyone now a days are suffering from at least one type of illness, like Blood Pressure, Diabetes, Asthma, Back pain to name a few of them. Since there are so many patients everywhere, more medical diagnosis experts are needed. Computer-based methods are increasingly used to improve the quality of medical services. Mostly the remote areas, the population are basic thing of the facilities of having experts to diagnosedisease. So, it is the need of problem statement that they can consult the specialist doctor if it is necessary or serious. For example, from the latest statistics, the ratio for a doctor and a patient is 1: 600.The government is finding it difficult to have enough medical experts to get a good ratio amongmedical experts and patients.

Furthermore, the doctors have to consider lots of issues before they could subscribe an accurate treatment for a patient. In addition, computer technology also has gone through very extreme changes and the changes have also influenced many fields and to change completely the way they function. One of the major fields which have been completely changed by computer technology is medicine and health care field. It has brought many changes to maintain thepatient, diseases, doctor, medicals."

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

III. PROBLEM STATEMENT

To develop Computer Based Disease Prediction System using Machine Learning to improve quality of medical services.

Objectives

- To design a System for the better care of patient.
- To provide a primary prediction of diseases.
- To classify major or minor disease.
- Reduce hospital operating costs.
- To provide a support the diagnosis process of

Scope

1. The scope for the system is typically for regular patient, Primary diagnosis of disease and for doctors to give immediate and free diagnosis of the diseases.
2. Introduction of database reusability in this environment with a purpose to increase the quality of an intelligent medical system.

IV. LITERATURE SURVEY

In the introduction part we seen the problem statement and objectives that we areachived in this project now we will see the research papers that we are studied for medical system that we are developed.The medical diagnosis system using machine learning published by Nirma University in2014 by Dhaval raval and in this paper they have proposed an algorithm to predicate the swine fludiesease based on several attribute.

The improving the accuracy of medical diagnosis with casual machine learning by nature communication in 2020 by Jonathan riches and Ciaran Lee in this reformulate diagnosis as a counterfactual inference task and derive counterfactual diagnostic algorithm.From Expert System to machine-Learned Diagnosis models by Xavier Amratian Cornell University in 2018 by Murali Ravari and Anitha Kannan in that they Present a method to merge bothapproaches by using expert systems as generative models that create simulated data on which modelscan be learned.

In that the main Approach to Develop Expert Systems in Medical Diagnosis Using Machine Learning Algorithms (Asthma) and A Performance Study by International Journal on Soft Computing in 2011 by BDCN Prasad, P. E. S. N Krishna Prasad and Y Sagar in that they Design the expert system for diagnosis of asthma. Advancements in Computational Model-based Medical ImageAnalysis by Hindawi Journal of Healthcare Engineering in 2021 by Naresh Kumar, Nripendra Narayan Das & Deepali Gupta in that they Predict the risk of COVID-19, heart disease, and diabetes in an individual based on answering a few questions related to various factors like travel history, age,gender, and blood pressure.

“In Cancer Diagnosis by International Conference on Computer Communication and Informatics in 2016 by Chandra Arya in that the the goal of this survey paper is to determine the current state of research in breast cancer and to help extract the key features and problems with existing expert systems. Expert System for Diagnosing Shortness of Breath in Infants and Children by International Journal of Engineering and Information Systems in 2017 by J Y Abuel-Reesh and SS Abu Naser in that we learn the system will help to diagnosis the shortness of breath in children.”

V. SYSTEM DESIGN

In previous chapter discussed about Literature survey and that findings from them now in this chapter we will see how the system is actually work, how the users and different componentsare integrated to each other and there working with the help of some diagrams.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

1. Proposed system setup

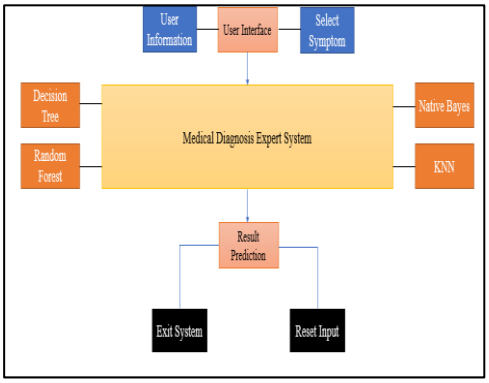

Fig1.1 Proposed system setup

Description

System will work mainly in three parts which includes requirement gathering. Disease prediction and finishing. In requirement gathering we will take name of the user as a personal data and remaining selection of symptoms are selected from dropdown list. Minimum two symptoms are mandatory in prediction of diseases. In the next phase of system, we are used four algorithms with their own buttons to predictions of disease from inputsymptoms. And in last step of system, we use exit system button for exiting the system and reset input button for new prediction of user.
Here we are used some python inbuild libraries for input and output handling so the flow of input is like from user interface to directly algorithms used in system one by one and then the output generated is again transferred to GUI.

2. Use Case Diagram

In the block diagram we seen systems architecture for handling data operations now inuse case diagram we will discuss about user interaction with the help of use case diagram. “Use case diagram describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.”

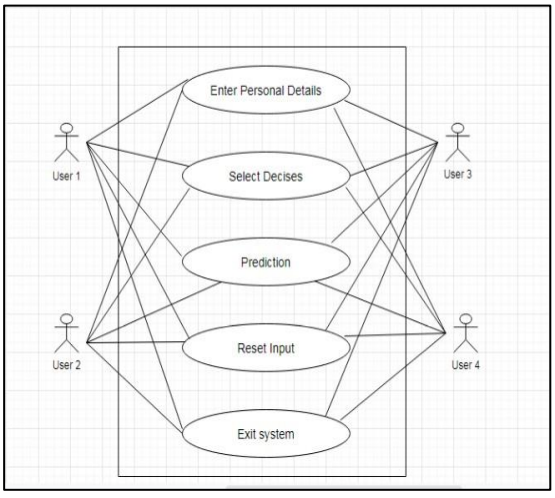

Fig2.1 Use Case Diagram

Here important thing is only one user can interact with system at a time.
In disease prediction system the main things which user having access to operate in Entering of personal details, Selection of Symptoms from the dropdown list, prediction of disease and last one is exit and reset of inputs.
Here we are given all access to the user but the condition is only one user can access the system at a time.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

3. Activity Diagram

In previous diagram we discussed about user of system with use case diagram new in next we see how the control of system is transferred from one to other activity. “Activity diagram is basically a flowchart to represent the flow from one activity to another activity. This activity is described as an operation of each system. The main point is control flow is drawn one operation to another operation.”

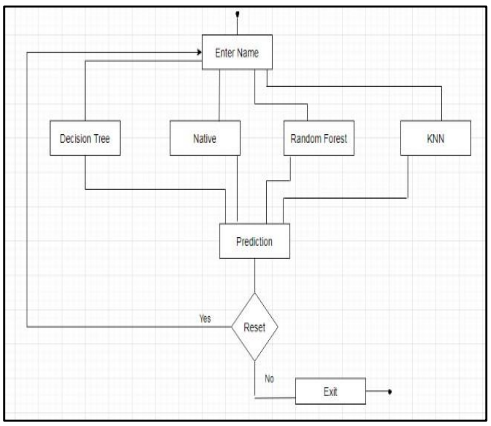

Fig3.1 Activity Diagram

Here the all four algorithms are working at same position and level, there is no dependency of one algorithm to another one but for speedup of system we have to run one algorithm at one time so for that the sequence of all algorithms is like Decision Tree-> NativeBayes->Random Forest-> K-Nearest Neighbour. All the data including symptoms and name are then given to those algorithms from that the predict the disease for each algorithm separately and again given to GUI.

4. Mathematical Modelling

Now we will see how we can represent whole system in mathematical model. “In mathematical modelling, Mathematical modelling is the process of describing a real-world problem in mathematical terms, usually in the form of equations, and then using these equationsboth to help understand the original problem, and also to discover new features about the problem. We are all very familiar with one application of mathematical modelling.”

Sets= {K Nearest Algorithm, Random Forest Algorithm, Gaussian Naïve Bayes Algorithm,Decision tree Algorithm}

Prediction1= {Prediction1,Prediction2, Prediction3, Prediction4}

Q ={S0,S1,S2,S3,S4,S5,S6,S7,S8}

∑= {Start/Open System, *Enter Name, Select *Symptoms1, Select *Symptoms2, SelectSymptoms3, Select Symptoms4, Select Symptoms5, Prediction, Exit / Reset}

$^{\delta}$= S0 X Access system by user ->S1.

S1 X Enter user name ->S2.
S2 X Selection of first Symptoms ->S3.
S3 X Selection of second Symptoms ->S4.
S4 X Selection of third Symptoms ->S5.
S5 X Selection of forth Symptoms ->S6.
S6 Xs Selection of fifth symptom ->S7.
S7 X Selection of sixth symptom ->S8.
S8 X Exit system or reset inputs ->S0.

q0= {Access to system, Enter Name, Select *Symptoms1, Select *Symptoms2}

F= S7->Exit system or Reset Inputs.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

VI. IMPLEMENTATION DETAILS

We researched total 20 research papers in that we find different algorithms and in those 8 algorithms used for medical diagnosis system for prediction of diseases so from that these 4 algorithms with the higher accuracy and in that for random forest Algorithm is used for both classification and regression and the accuracy of the algorithm is 95%. The decision tree is used for classification because it is highly adaptable and gives more accuracy, it gives up to is 95%. In KNN algorithm it finds extensive use in pattern findings anddata mining with 92% accuracy. Another one algorithm we use the naive bias. These algorithm passes common principle that is every pare of prediction is independent from each other and theygive up to 95% of accuracy for prediction.

VII. ALGORITHM AND FLOWCHARTS

Basically, we used four different kinds of algorithm in our project to predict the disease based on symptoms and all gives up to 94% of accuracy. Now we will discuss it one by one.

1. Decision tree Algorithm

"It is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for image. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes and leaf."

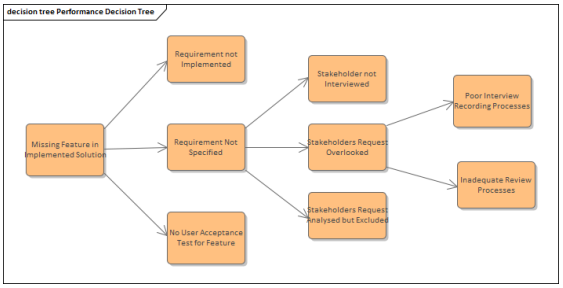

fig1.1 Decision Tree Algorithm

By using decision tree algorithm in disease prediction using machine learning in which prediction of disease based on symptoms we achieved up to ~95% of accuracy. Roots consistsof attributes in decision tree which has most effect on the on the final outputs.

2. Random Forest Algorithm

The beauty of random forest algorithm is it works classification and regression that's why we are used it to predict purpose in system at second number. Random Forest Algorithmworks on 4 following basic steps–

1. At the very firstly it chooses random samples of data from the given dataset.
2. Then secondly it constructs the decision tree for every data sample chosen from dataset.
3. In the third step the predicted result from step two is compiled and voted on.
4. Last step involves or check for the voted samples and then it presents mostly votedprediction as a result.

By using Random Forest Algorithm in our project, we got up to ~95% of accuracy with providing 100 random samples.

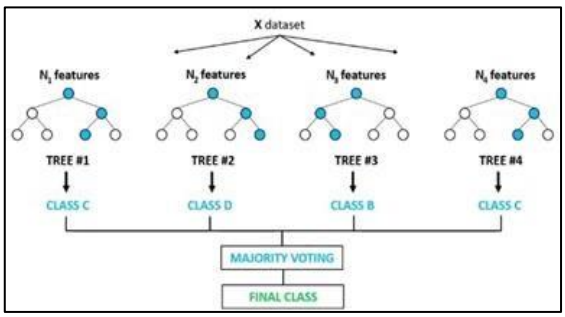

fig2.1 Random Forest Algorithm

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

"Random forest algorithm can be used for both Classification and Regression problems in Machine Learning. The reasons for using Random Forest algorithm in medicaldiagnosis system is It takes less training time as compared to other algorithms Also It predicts output with high accuracy, even for the large dataset it runs efficiently and last oneis it can also maintain accuracy when a large proportion of data is missing.

3. K nearest Neighbor Algorithm

One of the important models in prediction of disease because it having higher accuracyas well as it is one of the best supervised learning algorithms used in machine learning. It is one of the basic yet essential algorithm for prediction. Also, it gives extensive use in pattern recognition in case of pattern findings and data mining.The working of K Nearest Neighbor Algorithm is like "it works by finding a pattern indata provided by system which links data to results and it improves upon the pattern recognition with every iteration of prediction. We have used K Nearest Neighbor to class if your data set and achieved ~92% accuracy which is quite good**".**

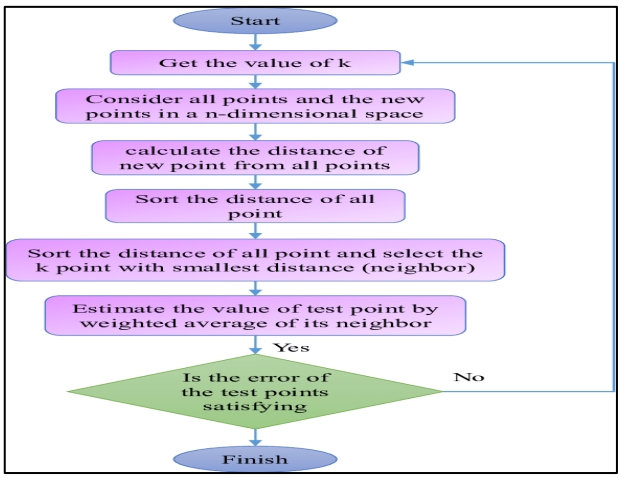

Fig3.1 K Nearest Algorithm

4. Naive Bias Algorithm

Native Bias algorithm is nothing but the family of all algorithms. Native bias algorithmshares the common principals that are every prediction is independent from each other based on every pair. By using this algorithm for prediction of disease in out system we got ~95% ofaccuracy. Native bias makes an assumption that each and every feature gives there equal and individual contribution for the prediction of disease.

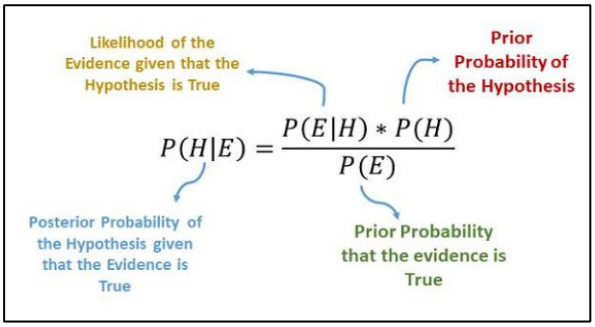

fig4.1 native bias algorithm

Basically, Naive Bayes classifier assumes the "features we use to predict the target areindependent and do not affect each other. While in real-life data, features depend on each otherin determining the target, but this is ignored by the Naive Bayes classifier. Though the independence assumption is never correct in real-world data, but often works well in practice.so that it is called Naive. Some of the other applications of Native Bias Algorithm are Real- time Prediction, Multi-class Prediction, Text classification/ Spam Filtering/ Sentiment Analysis and Recommendation Systems."

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

VIII. RESULTS

1. Initial State

fig1.1 Initial Stage of System

The file which is used to create the interface of our system. GUI stands for Graphical User Interface and to create it we have used Tkinter which gives a software kind of view to our project where user can directly interact with the system by entering the symptoms of diseases and he/she willget the disease through various algorithms.

2. Taking Personal Data

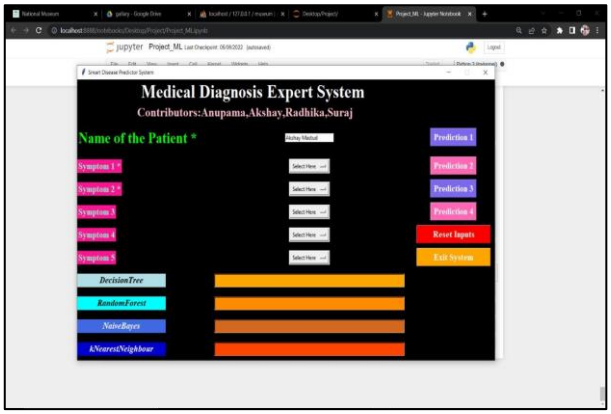

Fig2.1 Taking Personal Data

Enter the name in the provided space in front of the label as "Name of the Patient". It is the mandatory field which user have to enter in order to get result. Select 5 Symptoms from the dropdown menu which are labelled as Symptom 1, Symptom 2, Symptom 3, Symptom 4, and Symptom 5 respectively.

If user is not aware of 5 symptoms, then it is mandatory for him to enterat least 2 starting systems, otherwise the result will not come and a message box will pop up for the same as per user interest, he/she can predict the disease using different algorithms such as Decision tree algorithm, Random Forest algorithm, Naive Bayes algorithm and K..

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

3. Selecting Symptoms

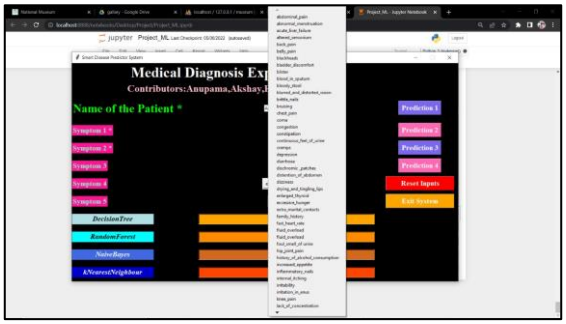

fig3.1 dropdown list of symptoms

Fig3.2Selection of first symptom

fig3.3 Selection of second symptom

4. Prediction for Decision Tree

Fig4.1Prediction by decision tree

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

"It is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for image. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes and leaf."

By using decision tree algorithm in prediction of disease based on symptoms we achieved up to ~95% of accuracy. In decision tree roots consists of attributes which has most effect on the on thefinal outputs.

5. Prediction for Random Forest

Fig5.1Prediction by Native Bias

Native Bias algorithm is nothing but the family of all algorithms. Native bias algorithm sharesthe common principals that are every prediction is independent from each other based on every pair.By using this algorithm for prediction of disease in our system we got ~95% of accuracy. Native biasmakes an assumption that each and every feature gives there equal and individual contribution for the prediction of disease.

6. Prediction for K nearest Neighbour

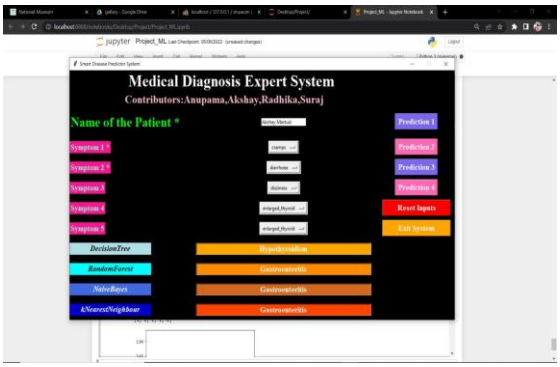

Fig6.1Prediction by K nearest Neighbour

One of the important models in prediction of disease because it having higher accuracy as wellas it is one of the best supervised learning algorithms used in machine learning. It is one of the basicyet essential algorithm for prediction. Also, it gives extensive use in pattern recognition in case of pattern findings and data mining.

The working of K Nearest Neighbor Algorithm is like it works by finding a pattern in data provided by system which links data to results and it improves upon the pattern recognition with every iteration of prediction. We have used K Nearest Neighbor to class if your data set and achieved good accuracy which is quite good.

7.System Reseting

Disease Recommendation will be available in front of the labels of algorithm of user's choice.Click on "Reset" button to predict the disease for any other patient or Press Shatter System" button to come out of the GUI.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

fig7.1 System Reset

8. Confirmation Message for Exiting System

fig8.1 Confirmation message showing for exitin system

At the end we are given Exit system button to move out from system. By using Human Computer Interaction concepts in mind button is colored with red color so it also used as a emergency exit also. To prevent error for the conformation we given dialog box with yes and no options. Because sometimes user clicks this button by mistake also.

IX. CONCLUSION

So here we are developed the system which predict the disease based on the symptoms from which user or patient suffering from. “These types of system are useful for reduce the rush at OPD’s of hospital and to reduce the workload of particular hospital staff. We were successful in creating such a system and use 4 different algorithms to do so. On an average we achieved accuracy of ~94%. Such a system can be largely reliable to do the job. 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 Hasan easy to use interface.”
It also has various visual representation of data collected and results achieved.

X. FUTURE WORK

1. The scope of our project is defined on number of prediction algorithms used and accuracy ofthat algorithms.
2. Currently we predict disease based on symptoms only but in future we are able to predict itbased on previous result also.
3. The GUI which we developed in future we are able to make it more attractive by HTML and CSS and host it on proper domain or we can also create mobile application for that also.

REFERENCES

[1].M. K. Chytil and Rolf Engel Brecht for key point which effect on medical expert system from book ‘Medical Expert Systems’ which originally published in 1987.
[2].Edward H. Short life for attributes of different object and its backend operation from its book ‘Computer-based Medical Consultations, MYCIN’ which originally publishedin 1976.
[3].Smitha V., 2010, An Expert System for Diabe-tes Diagnosis, an MPhil Thesis, Department of Computer Science,

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111113

Christ University, Bangalore.

[4]. An expert diagnostic tool for engineering systems by A K Verma & K Seetharam (journal of scientific & Industrial research, vol 53, pp 601- 603).

[5].Patra P.S.K., Sahu D.P. and Mandal I., 2010, An Expert System for Diagnosis of Human Dis-eases, International Journal of Computer Appli-cations Vol. 1 – No. 13.

[6].Brian R, Gaines 2010 Designing Expert System for Usability Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.414.583&rep=rep1&type= pdf.

[7].Jakob Nielsen. Usability 101: Introduction to Usability. 2014 Jan 2012. Available from http://www.useit.com/alertbox/20030825.html.

[8].Wingnjosoebroto S, Sudiarno A and Harenda D 2009 Designing Web Interface Prototype Based on Usability Aspects (Case Studi: Ergonomic and Work System Design Laboratory) p 7.

[9]. Karagiannis S., Dounis A. I., Chalastras, T., Tiro-panis P., and Papachristos D., 2007,Design of Ex-pert System for Search Allergy and Selection of the Skin Tests using CLIPS, Journal of Informa-tion and Communication Engineering 3:6.

[10].Giarranto J. C., 2007. CLIPS User Guide, Quick-silver Beta.

