

Automated Health Care System for Rural Areas

Devanand Galande¹, Richard Roy², Akshay Goyal³, Tejas Deshmukh⁴, Prof. A. J. Kadam⁵

UG Student, Department of Computer Engineering, AISSMS, Pune Maharashtra, India¹

UG Student, Department of Computer Engineering, AISSMS, Pune Maharashtra, India²

UG Student, Department of Computer Engineering, AISSMS, Pune Maharashtra, India³

UG Student, Department of Computer Engineering, AISSMS, Pune Maharashtra, India⁴

Asst. Professor, Department of Computer Engineering, AISSMS, Pune Maharashtra, India⁵

ABSTRACT: Earlier as well as nowadays also, the doctors are using trial and error approach for predicting the diseases based on clinical investigations available. To predict the diseases is one of the major challenges in past years and today also. There is great need of some system that predicts the diseases early on the basis of available symptoms and patients health. Because of this it will become possible to cure the people from diseases which may lead the humans to death for e.g. Cancer, AIDS etc. We are proposing system which is based on combination of different locations, doctor's details, disease symptoms and disease treatment that are useful to predict the patient's disease. The patient's disease states can be finding out by formalizing the disease based on symptoms of the patient before recommending treatments for the prevailing diseases by using Support Vector Machine (SVM) and Naïve Bays Classification techniques respectively. The basic aim of our system is to assist doctors in diagnosing the patient by analyzing disease symptoms and relevant information. As our project's main focus is providing medical assistance to people living in rural areas to provide them with free treatment for general diseases which they tend to neglect.

KEYWORDS: Diseases, Disease Prediction, Symptoms, Treatment, Support Vector Machine, Naïve Bays Classification.

I. INTRODUCTION

As we know there are numerous kinds of diseases are there in environment. Some diseases can be easily get cured but some are there which have bad impact on human body and may lead to death. Diseases like Bulimia Nervosa (BN), Diabetes, Chronic Kidney Disease (CKD), Skin Cancer (SC), Leukemia, Colds-Flu-Gripe (CFG), Dengue (De), Malaria (M), Cholera (Cl), Leptospirosis (L), Chikungunya (CG), Chicken pox (CP), Diarrhea (Di) have similar characteristics at early stage which makes clinician job difficult in diagnosing such life-threatening disease. Similarly, in 2009, H1N1 was spreading fast around the world. It is contagious type of disease and it spreads same as seasonal flu. Its symptoms are same as seasonal flu like cough, fever, sore throat, body ache, and headache. So at early stages it seems like a seasonal flu to Doctors but later on when these symptoms lead to more serious complications, including pneumonia and respiratory failure Doctors came to know this is not seasonal flu and something else. Similarly consider Dengue fever. It is transmitted by the bite of an Aedes mosquito infected with a dengue virus. The mosquito becomes infected when it bites a person with dengue virus in their blood. It can't be spread directly from one person to another person. Sometimes, symptoms are mild and can be mistaken for those of the flu or another viral infection by the victims or patients. However serious problems develop if this cannot be treated early like damage to lymph and blood vessels, bleeding from nose and gums. Many of these diseases are preventable if detected at early stages. It may be difficult for clinician as well as the person who is suffering to distinguish between these prevailing diseases due to medical symptom similarity at early stages. Today's such variety of diseases makes us to feel the need of implementing the system which can predict the diseases as early as possible based on patient health status. If it is possible to predict the

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

disease at early stage based on available symptoms we can provide respective medicinal treatment to patient who is suffering.

II. RELATED WORK

In this work, an effective heart disease prediction system (EHDPS) is developed using neural network for predicting the risk level of heart disease. The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction. The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, e.g, relationships between medical factors related to heart disease and patterns, to be established [1].

The medical environment is still information rich but knowledge weak. There is a wealth of data possible within the medical systems. However, there is a lack of powerful analysis tools to identify hidden relationships and trends in data. Disease is a term that assigns to a large number of health care conditions related to the body. These medical conditions describe the unexpected health conditions that directly control all the body parts. Medical data mining techniques like association rule mining, classification, clustering is implemented to analyze the different kinds of general body based problems [2].

This work supports an end user and online consultation. Here author propose a framework that enables clients to get moment direction on their medical problems through an astute social intelligent health care system online. The framework is bolstered with different symptoms and the disease or illness associated with those systems. Also the system allows user to share their symptoms and issues. Data Mining as a field of research has already well proven capabilities of identifying hidden patterns, analysis and knowledge applied on different research domains, now gaining popularity day by day among researchers and scientist towards generating novel and deep insights of these large biomedical datasets also. Uncovering new biomedical and healthcare related knowledge to support clinical decision making, is another dimension of data mining [3].

In this research work, authors have comprehensively compared different data classification techniques and their prediction accuracy for chronic kidney disease. Authors have compared J48, Naïve Bayes, and Random Forest, SVM and k-NN classifiers using performance measures like ROC, kappa statistics, RMSE and MAE using WEKA tool. Authors have also compared these classifiers on various accuracy measures like TP rate, FP rate, precision, recall and f-measure by implementing on WEKA. Experimental result shows that random forest classifier has better classification accuracy over others for chronic kidney disease dataset [4].

Author proposing a system which is based on combination of different data mining techniques such as clustering, classification etc. that are useful to predict the patient's disease state. The patient's disease states can be finding out by formalizing the hypothesis based on test results and symptoms of the patient before recommending treatments for the prevailing diseases. The basic aim of our system is to assist doctors in diagnosing the patient by analyzing his available data and relevant information [5].

Data mining techniques are used for a variety of applications. In healthcare industry, data mining plays an important role in predicting diseases. For detecting a disease number of tests should be required from the patient. But using data mining technique the number of tests can be reduced. This reduced test plays an important role in time and performance. This work analyzes data mining techniques which can be used for predicting different types of diseases. This work reviewed the research papers which mainly concentrate on predicting heart disease, Diabetes and Breast cancer etc [6].

Medical data mining techniques like association rule mining, classification, clustering is implemented to analyze the different kinds of heart based problems. Classification is an important problem in data mining. This work given a database contain collection of records, each with a single class label, a classifier performs a brief and clear definition for each class that can be used to classify successive records. A number of popular classifiers construct decision trees to

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

generate class models. The data classification is based on MAFIA algorithms which result in accuracy, the data is estimated using entropy based cross validations and partition techniques and the results are compared [7].

Chronic Kidney Disease (CKD) is a gradual decrease in renal function over a period of several months or years. Diabetes and high blood pressure are the most common causes of chronic kidney disease. The main objective of this work is to determine the kidney function failure by applying the classification algorithm on the test result obtained from the patient medical report. The aim of this work is to reduce the diagnosis time and to improve the diagnosis accuracy using classification algorithms. The proposed work deals with classification of different stages in chronic kidney disease according to its severity. The experiment is performed on different algorithms like Back-Propagation Neural Network, Radial Basis Function and Random Forest [8].

III. PROPOSED SYSTEM

In this proposed system which is based on combination of different locations, doctor's details, disease symptoms and disease treatment that are useful to predict the patient's disease. The patient's disease states can be finding out by formalizing the disease based on symptoms of the patient before recommending treatments for the prevailing diseases. We have collected the different disease dataset. In this dataset include multiple disease lists with multiple symptoms of each disease. Also we have collected the doctor's location dataset which include multiple locations and doctors names and contact no. We used association rule with Support Vector Machine (SVM) and Naïve Bays classification techniques to predict the exact disease name. As our project's main focus is providing medical assistance to people living in rural areas to provide them with free treatment for general diseases which they tend to neglect. This proposed system also recommends the treatment according to particular disease name.

Proposed System Architecture

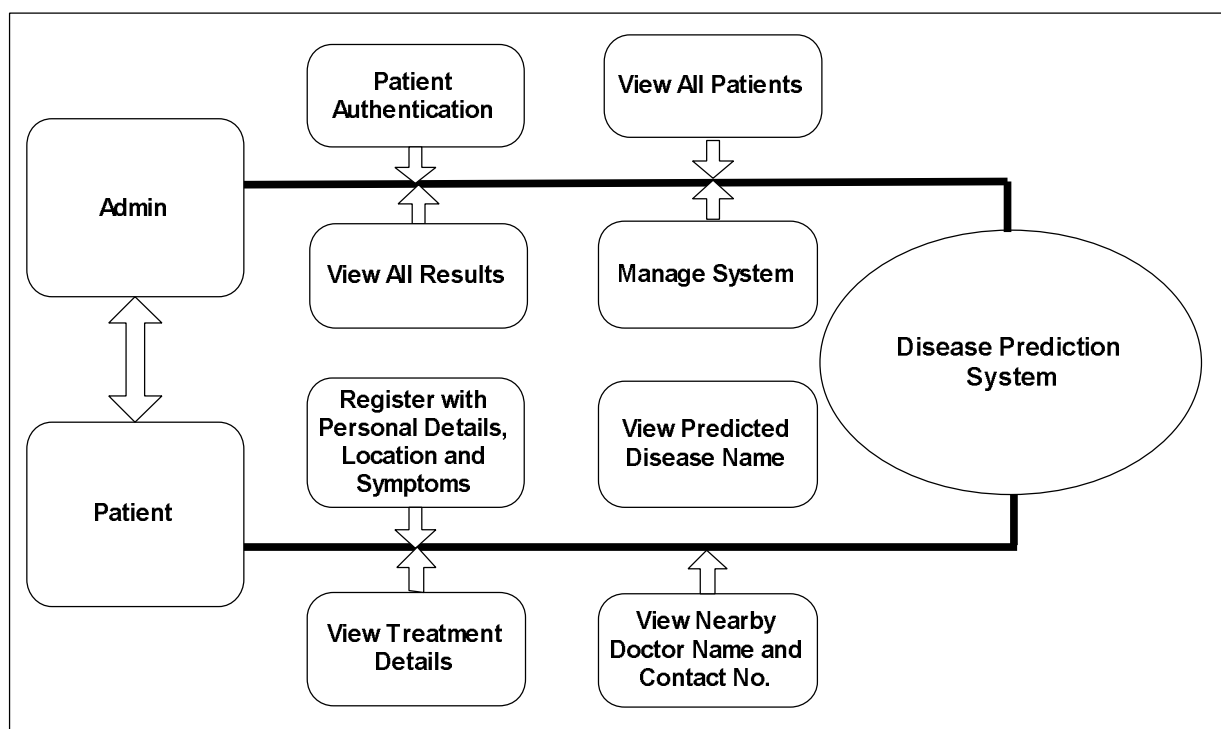


Fig: Proposed System Architecture

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

IV. ALGORITHMS

1. Support Vector Machine

- Support Vector Machine (SVM) is used for both classification and regression problems.
- The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.
- The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class

2. Naïve Bays Classification

- Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.
- The Naive Bayes algorithm is called “naive” because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features.

V. MATHEMATICAL MODEL

1. The Mathematics of the Support Vector Machine

We have k sub-spaces so that there are k classification results of sub-space, called CL_SS1, CL_SS2, ..., CL_SS_k. Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$CL = \frac{1}{k} \sum_{i=1}^k CL_{SS_i}$$

Or weighted mean value:

$$CL = \frac{1}{k} \sum_{i=1}^k w_i CL_{SS_i}$$

Where w_i is the weight of classification result of subspace SS_i, and satisfies:

$$\sum_{i=1}^R w_i = 1$$

The centroid of a data is calculated as follows:

$$\bar{X} = \frac{\sum_{i=0}^k x_i}{k}, \bar{Y} = \frac{\sum_{i=0}^k y_i}{k}$$

Where (\bar{X}, \bar{Y}) represents the centroid of the hand, x_i and y_i are x and y coordinates of the i^{th} pixel in the hand region and k denotes the number of pixels that represent only the input data.

In the next step, the distance between the data values was calculated. For distance, the following Euclidean distance was used:

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Where (x_1, x_2) and (y_1, y_2) represent the two co-ordinate values.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

2. The Mathematics of the Naive Bayes Algorithm

The basis of Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. More formally, Bayes' Theorem is stated as the following equation:

$$P(A/B) = \frac{P(B/A) P(A)}{P(B)}$$

Let us understand the statement first and then we will look at the proof of the statement. The components of the above statement are:

P(A/B): Probability (conditional probability) of occurrence of event A given the event B is true

P(A) and P(B): Probabilities of the occurrence of event A and B respectively

P(B/A): Probability of the occurrence of event B given the event A is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

A is called the proposition and B is called the evidence.

P(A) is called the prior probability of proposition and P(B) is called the prior probability of evidence.

P(A/B) is called the posterior.

P(B/A) is the likelihood.

This sums the Bayes' theorem as

$$\text{Posterior} = \frac{(\text{Likelihood}) \cdot (\text{Proposition prior probability})}{\text{Evidence prior probability}}$$

VI. RESULT AND DISCUSSION

The mean and standard deviation values of the input data are computed in each spectral channel as the feature. We let n be the number of data, and v_{ij} denotes the j th band value of the i th data. The mean (mean_j) and standard deviation (std_j) of the patch are calculated according to

$$\text{Mean}_j = \frac{\sum_{i=1}^n v_{ij}}{n}$$
$$\text{Std}_j = \sqrt{\frac{\sum_{i=1}^n (v_{ij} - \text{mean}_j)^2}{n}}$$

Table I is a summary of classification accuracies among different classifiers based on the feature for classifiers. Note that the Support Vector Machine and Naive Bayes-based classifier outperform other classifiers. The classification accuracy for SVM and NB is 77.5% and 77.2% on average, respectively.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Table I: Table of Classification Accuracy

Classifier	Mean (Average)	Standard Deviation
Support Vector Machine	77.5	7.4
Naive Bayes Classification	77.2	7.1

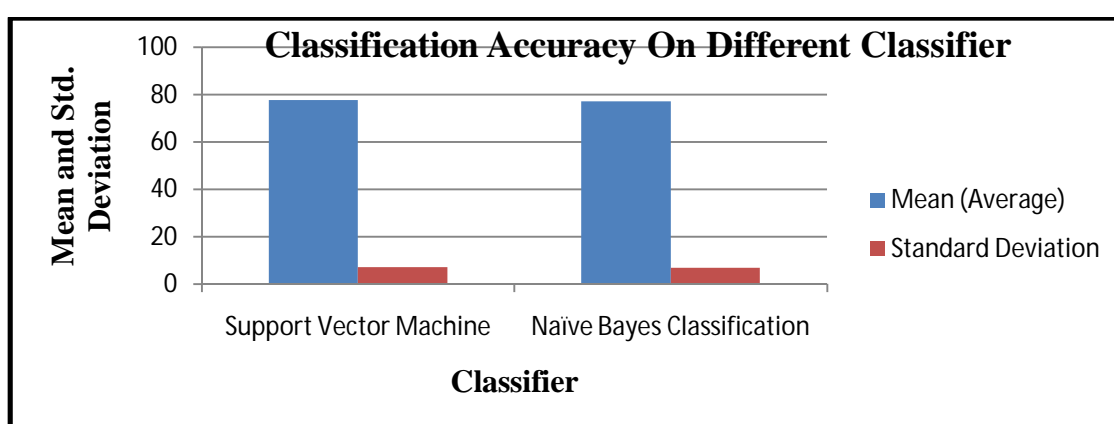


Fig2: Classification Accuracy Graph

VII. CONCLUSION

Prediction of accurate human disease based on patient's basic symptoms is one of the essential topics in medical diagnosis. The proposed work is to classify the different disease according on symptoms of the patient before recommending treatments for the prevailing general diseases. This will also provide remedy for calculated particular diseases. People can self-analyze their health state and can take precautions as per the results. It would help the practitioners / Doctors to analyze the health state of the patient and based on that the manual diagnosis of the disease can also be possible by using the disease prediction system. We used Support Vector Machine (SVM) and Naive Bays classification techniques to predict the exact disease name. As our project's main focus is providing medical assistance to people living in rural areas to provide them with free treatment for general diseases which they tend to neglect. This will provide early diagnosis of the disease which is not possible with the help of manual diagnosis.

REFERENCES

- [1] Poornima Singh, Sanjay Singh, Gayatri S Pandi-Jain, "Effective heart disease prediction system using data mining techniques". International Journal of Nanomedicine March 2019
- [2] Shratik J. Mishra, Albar M. Vasi, Vinay S. Menon, Prof. K. Jayamalini, "GDPS - General Disease Prediction System". International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 03 | Mar-2018
- [3] Puja Sarage, Trupti Sudrik, Kalyani Zodage, "Health Prediction System by using Data Mining", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887 Volume 5 Issue XII December 2017
- [4] Narander Kumar, Sabita Khatri, "Implementing WEKA for medical data classification and early disease prediction", 3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017)
- [5] Rahul Patil, Pavan Chopade, Abhishek Mishra, Bhushan Sane, Yuvraj Sargar, " Disease Prediction System using Data Mining Hybrid Approach". Communications on Applied Electronics (CAE) – ISSN: 2394-4714 Foundation of Computer Science FCS, New York, USA Volume 4 – No.9, April 2016 – www.caeaccess.org
- [6] K.Gomathi, Dr. D. ShanmugaPriyaa, "Multi Disease Prediction using Data Mining Techniques". 17 September 2017.
- [7] M.A.Nishara Banu, B Gomathy, "DISEASE PREDICTING SYSTEM USING DATA MINING TECHNIQUES". International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 1, Issue 5 (Nov-Dec 2013), PP. 41-45
- [8] Shivane Pandey et al., "Diagnosis and Classification of HypothyroidDisease using Data Mining Techniques", Vol. 2 Issue 6, pp. 3188-3192,ISSN: 2278-0181 ISSN: 2278-0181, 2013