Disease Prediction System Using Fuzzy C-Means Algorithm

T. Bala Ramya
Student
Dept. Of Computer Application
Kongu Engineering College
Perundurai

Abstract:- In today's era, each and every human-being on earth depends on medical treatment and medicines. Every day we can hear some new diseases or new symptoms of the existing disease being discovered. But with the growing number of diseases and their symptoms, everyone cannot manage to be updated with it. To predict the diseases is one of the major challenges in past years and today also. . People tend to get suffered to or sometimes even die from certain diseases which could easily be cured, if those were known beforehand. This lack of knowledge sabotages the health of a person and can create deeper repercussions. This shows the importance of predicting the diseases early on the basis of available symptoms. Because of this it will become possible to cure the people from hazardous diseases which may lead the humans to death.
The main objective of this paper is to predicting the disease of a patient based on the symptoms they enter using FCM or Fuzzy C Means algorithm. FCM is an unsupervised clustering algorithm, which allows one piece of data to belong to two or more clusters.

Keywords: Clustering, FCM, Symptoms

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. 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 pneumonia and respiratory failure. Doctors came to know this is not seasonal flu and something else. 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 disease at early stage based on available symptoms we can provide respective medicinal treatment to patient who is suffering. .
The main objective of this paper is to develop a prototype intelligent disease prediction system with Fuzzy C Means

Clustering algorithm. Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is most widely used.

This system used to predict the diseases based on the input symptoms provided by user and help mankind or society people to analyze and understand their health status. 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.

The proposed system "Disease prediction system using fuzzy c-means" which has a list of large number of diseases, their symptoms, their treatment and medicines required to cure it.

II LITRATURE SURVEY

In this section the previous related studies are reviewed. Here the lists of symptoms are required medication for every possible disease consists of both accurate and inaccurate. **Wikipedia,** List of medical symptoms [1] describes the features of such diagnosis and related symptoms for such diseases

A Novel Method for Disease Recognition and Cure Time Prediction Based on Symptoms [2] in this paper they have assign severity rating to each symptom of the disease and based on that they have calculated cut off factor for each disease They also predicted the cure time for the disease by using pre-defined medical dataset they have created.

Analysis of Data mining prediction technique in healthcare management system [3] is the paper in which they have discussed the importance of data mining techniques in medical field as well as the need of automated medical diagnosis system. They have also done comparison of data mining techniques like decision tree, Bayesian classifier, and neural network, support vector machine on

Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCICCT - 2018 Conference Proceedings

specific disease like heart disease, diabetics, eyes disease and cancer.

Dengue disease prediction using weka data mining tool [4] is the paper used to predict dengue disease for this they have use different data mining algorithm like Naïve Bayes, J48, SMO, REP Tree and Random tree in weka and compared it to find out which algorithm gives better result. According to their survey they have concluded that Naïve Bayes and J48 are the best performance algorithms with accuracy= 100% takes less time taken to build and shows maximum ROC area = 1, and had least absolute error.

Symptom's based diseases prediction in medical system by using k-means algorithm [5] in which they have used k-means algorithm, Large Memory Storage and Retrieval (LAMSTAR) and medical diagnosis methodology. Their system used System Oriented Architecture (SOA) whereby the system elements of diagnosis, data portal and alternative miscellaneous services are provided. In their architectural model they have two databases one is Patient Record Database and another Disease/Symptoms Database. For each symptom they have assign weights to the symptom.

A Proposed Fuzzy Framework for Cholera Diagnosis and Monitoring [6] developed a proposed fuzzy framework for cholera diagnosis and monitoring. The designed system can be increased to any number of inputs. The following membership functions like No cholera, mild cholera, moderate cholera and severe cholera all depend on the inputs diarrhea, vomiting and dehydration. We can define this system for any number of inputs. To achieve more efficient human diagnose and monitoring result, the system can be defined with more than three inputs

III PROPOSED SYSTEM

There are numerous kinds of disease in the environment. Some diseases can be easily cured. Some have bad impact on human body. Many of the diseases are preventable if detected at early stage. It may be difficult for clinician as well as the person who is suffering to distinguish between these diseases at early stage. 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 the symptoms.

Earlier as well as nowadays, 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 something that predicts the diseases early on the basis of available symptoms.

We are proposing a system **"DISEASE PREDICTION SYSTEM"** which can predict the disease of patients or victims based on his/her symptoms. Data mining technique clustering is helpful for prediction purposes. Main objective is medical care to any place that health care is needed. Using fuzzy c-means clustering algorithm we are predicting the

possibilities of having disease. This algorithm is better when the datasets are noise.

It is an expert system which is use for simplifying the task of doctors. It is a system that checks a patient at initial level and suggests the possible diseases. It starts with asking about symptoms to the patient, if the system is able to name the disease then it provide the name and the corresponding medicines, If the system is not sure enough, It ask some queries to the patients, still of the system is not sure then it will display some test to the patient. On the basis of available cumulative information, the system will display the name and the prescribed medicines of the disease. This system not only simplifies task of the doctors but also helps the patients by providing initial medicines for small diseases in emergency.

This paper is for such system that can predict the disease state i.e. the person is suffering with disease or not. The working flow for the prediction system is as below.
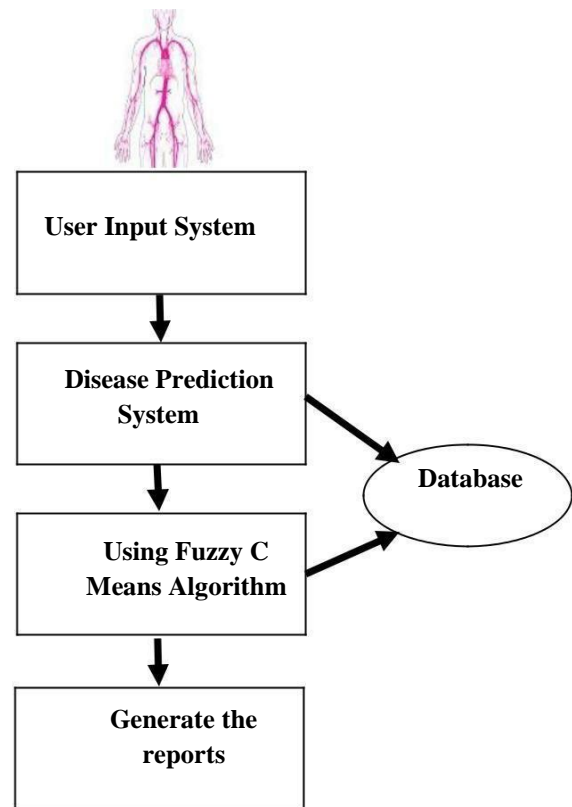

Fig 1: Workflow for disease prediction system

The User will enter their symptoms according to the disease states he / she is suffering from to the disease prediction system. After entering the symptoms by using the fuzzy c means algorithm the disease will be predicted and displayed to the user.

IV METHODOLOGY

The main objective of this research is to develop a prototype Intelligent Disease Prediction System Using Fuzzy C Means

Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCICCT - 2018 Conference Proceedings

Clustering algorithm. In these fuzzy c means clustering algorithms, the membership degree is associated to the values of the features in the clusters for the cluster centers instead of being associated to the patterns in each cluster. Fuzzy c means clustering is useful when the dataset are noise.

Fuzzy C Means (FCM) classifier

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be *in the cluster* to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available. Any point x has a set of coefficients giving the degree of being in the kth cluster $w_k(x)$. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

The degree of belonging, $w_k(x)$, is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. Assign randomly to each point coefficients for being in the clusters. Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than , the given sensitivity threshold) :

• Compute the centroid for each cluster, using the formula above.

• For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance.

V   SYSTEM DESIGN

User Module

In this module, we can collect the details of the user

The data like the name of the user, address, phone number, sex, email id collect from the user to help with the creation of an individual user account Already registered user can directly start accessing the system with the help of the email id and password
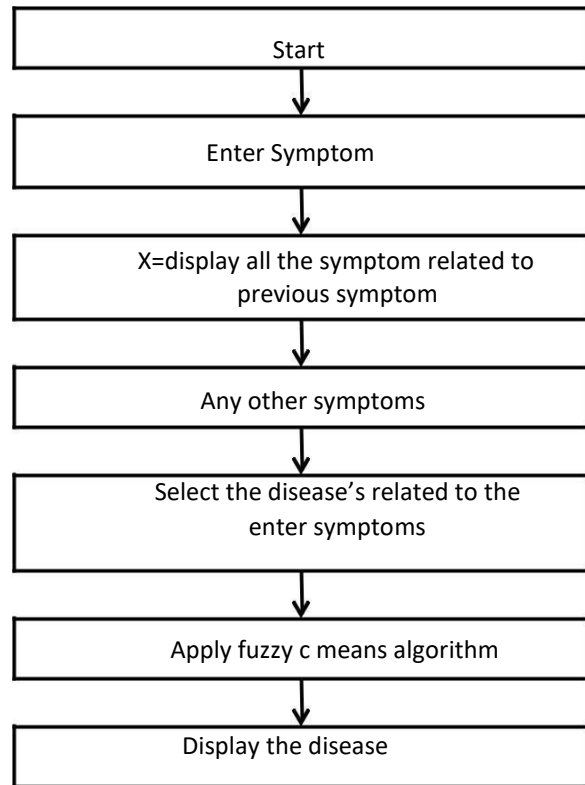

Fig 2: Workflow for user module

Admin Module

In this module, the admin can enter the new symptoms
For each symptoms, calculate the cut-off factor
Cut-off factor is calculated for each disease with that actual symptoms and it is stored in the database.
When user enters the symptoms it calculates the cut-off factor for each disease for those symptoms.
The nearest cut-off factor is taken for further process
Then the fuzzy c-means algorithm used to find the actual disease based on the symptoms

Dataset is obtaining from the Wikipedia List of medical symptoms, WebMD. Disease symptoms, Healthline.com websites.

VI CONCLUSION

As we have earlier mentioned, we solve the problem by predicting the disease by considering the symptoms entered by the user. The main focus of our paper is to identify the disease by calculating the numerical value based on the severity rating of the symptoms and also make our prediction more accurate by considering the previous history of the user. Although we have tested our system for viral and regular based diseases, it can be extended to larger settings. The future work can be focus on by considering various types of diseases with large number of symptoms in predicting more number of diseases. To improve the reliability of the system the test results for various medical conditions will be helpful. Since the results are dependent on the experience of previous users, it is important to isolate

Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCICCT - 2018 Conference Proceedings

genuine experiences from fake ones. This app has large scope as it has following features:

Automation of Disease Prediction

To save the environment by using paper free work
To increase the accuracy and efficiency so that patients can get direct help.
Management of disease related data It will useful in urgent cases where patient is unable to reach doctor, during late night emergencies

It is easy to handle

REFERENCES

[1] **Wikipedia**. List of medical symptoms, 2015. [Online; accessed 22-January-2015].

[2] Shankar M., Pahadia M., **"A Novel Method for Disease Recognition and Cure Time Prediction Based on Symptoms."** Advances in Computing and Communication Engineering (ICACCE), 2015 Second International Conference on. IEEE 2015.

[3] K. Rajalakshmi, Dr. S.S. Dhenakaram. **Analysis of data mining prediction technique in healthcare management system** Volume 5, Issue 4, April 2015.

[4] Kashish ara shakil, Shadma anis and mansaf alam. **Dengu disease prediction using weka data mining tool**.

[5] Sathyabama Balasubramanian, Balaji Subramani. **Symptom's based diseases prediction in medical system by using k-means algorithm** Volume 3, No. 2, February 201.

[6] Uduak A., and Mfon M., "**A Proposed Fuzzy Framework for Cholera Diagnosis and Monitoring.**" International Journal of Computer Applications 82.17 (2013).

[7] **WebMD**. Disease symptoms, 2015. [Online; accessed 22-January-2015].

[8] S Sudha. **Disease prediction in data mining technique–a survey**. IJCAIT, 2(1):17–21, 2013.