

Prediction on Lung Disease Using K means Algorithm

Priyanka D.¹, Ms. S. Shahar Banu²

¹Student, Department of Computer Applications,

²Assistant Professor, Selection Grade, Department of Computer Applications,
B S Abdur Rahman University.

Abstract- Lung cancer is one of the major causes of death in both genders when compared to all other cancers. Lung cancer has become the most hazardous types of cancer in the world. Early detection of lung cancer is essential in reducing life losses. This paper presents prediction on lung disease using K means algorithm. This project comprises of three modules. First, admin module which is administrator's login there the details of the patient will be generated. Now the user will authenticate based on their credentials. The Second module is User module there the patient enters his username and password to predict cancer. Third module is Cancer prediction module in which the result will be predicted at the last stage with the help of K means algorithm. The K means will classify the input features into two classes of cancer type (benign and malignant). This project is implemented in java as the front end and mysql as the back end. This project aims to implement an effective prediction on lung cancer with the help of K means algorithm user can know the cancer status. From this project we infer that the K means is suitable for lung cancer prediction

Index Terms— lung cancer; K means Clustering, Wisconsin lung Cancer Database (WBCD).)

I. INTRODUCTION

Lung cancer accounts for more deaths than any other cancer in both men and women. An estimated 158,040 deaths are expected to occur in 2015, accounting for about 27% of all cancer deaths. Death rates began declining in 1991 in men and in 2003 in women. From 2007 to 2011, rates decreased by 2.9% per year in men and by 1.9% per year in women. Gender differences in lung cancer mortality reflect historical differences in patterns of smoking uptake and cessation over the past several decades.

Early detection of lung cancer is essential in reducing life losses. However earlier treatment requires the ability to detect lung cancer in early

stages. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign lung tumors from malignant ones. The International Journal of Innovations in Engineering and Technology (IJET) automatic diagnosis of lung cancer is an important, real-world medical problem. Thus, finding an accurate and effective diagnosis method is very important.

II. RELATED WORKS

Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi, Alok Choudhary [1] Different methods for lung cancer detection are explored and their accuracies are compared various data mining techniques are implemented to compare the accuracies

T.Karthikeyan et al [2] This paper presents an approach for using PCA-NB Algorithm to enhance the predictive accuracy for lung cancer disease, The drawback is that work some measure of evaluating performance has to be introduced.

V. Krishnaiah et al [3] Diagnosis of lung cancer prediction system using data mining classification techniques, Non-clinical symptoms and risk factors are some of the generic indicators of the cancer diseases.

III. PROPOSED SYSTEM

K means clustering technique, is proposed to help in the diagnosis of lung cancer. The database used in our study is the Wisconsin lung cancer database. It has been done in the University of Wisconsin by Dr. William H. Walberg. A lung cancer neural network is optimized to improve the success of diagnosis using Fine Needle Aspirates (FNA). The calculator can be very useful to not only accurately estimate survivability of a lung disease patient, but also aid doctors in decision making and improve informed patient consent by providing a better

understanding of the risks involved in a particular treatment procedure, based on patient-specific attributes. Accurate risk prediction can potentially also save valuable resources by avoiding high risk procedures that may not be necessary for a particular patient.

3.1 Clustering Techniques

Clustering is the process of making a group of abstract objects into classes of similar objects. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. In the field of biology, it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations.

3.2 K means Clustering

This is a simple k-means procedure. It can be viewed as a greedy algorithm for partitioning the n samples into k clusters so as to minimize the sum of the squared distances to the cluster centers. It does have some weaknesses:

- The way to initialize the means was not specified. One popular way to start is to randomly choose k of the samples.
- The results produced depend on the initial values for the means, and it frequently happens that suboptimal partitions are found. The standard solution is to try a number of different starting points.
- It can happen that the set of samples closest to m_i is empty, so that m_i cannot be updated. This is an annoyance that must be handled in an implementation, but that we shall ignore.
- The results depend on the metric used to measure $\|x - m_i\|$. A popular solution is to normalize each variable by its standard deviation, though this is not always desirable.
- The results depend on the value of k.

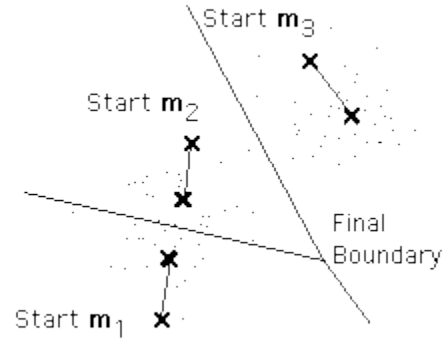


Figure 1 K mean clustering

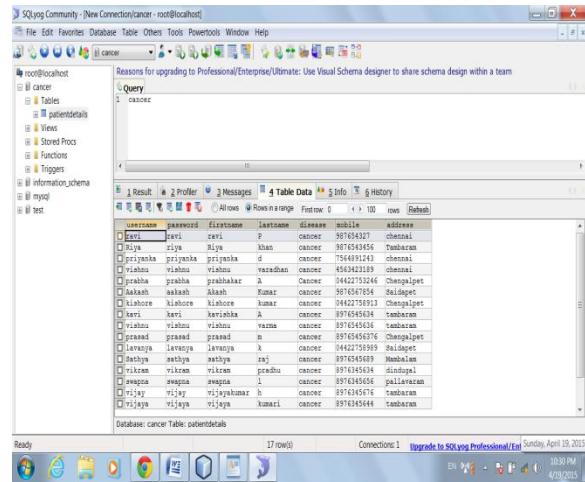


Figure 2 Dataset Table

IV. RESULT AND DISCUSSION

The user will authenticate based on their credentials. The patient enters his username and password in patient login prompting them to view the results after the process complete. Once admin enter the details, all details are entered into the local database. If the patient gives his username and password and presses submit button he will be directed to the patient details from there the next module cancer prediction has to be selected.

The result will be tested the performance of the neural networks based on the Wisconsin lung Cancer Database (WBCD). The problem of lung cancer detection led researchers and experts in this field to focus on other trends, such that new technologies other human to address this social problem. The objective of our work is to create a new approach that allows whether a patient has a benign cancer or malignant following several descriptors. The neural

network is trained with lung cancer data by using feed forward neural network model and back propagation learning algorithm with momentum and variable learning rate. The performance of the network is evaluated.

The work of this paper will be presented in different sections. In the second section, we introduce a general overview about lung cancer anatomy and different types of cancers to be classified. Adaptation of parameters to the neural network and the results obtained will be presented in the third module. And finally, we present a conclusion that talk about the performance of the method used in the purpose of lung cancer classification.

Algorithm

Step 1: Begin with a decision on the value of k = number of clusters.

Step 2: determine the centroid coordinates.

Step 3: Take each sample in sequence and compute its distance from the centroid of each of the clusters. If a sample is not currently in the cluster with the closest centroid, switch this sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample.

Step 4 . Repeat step 3 until convergence is achieved, that is until a pass through the training sample causes no new assignments.

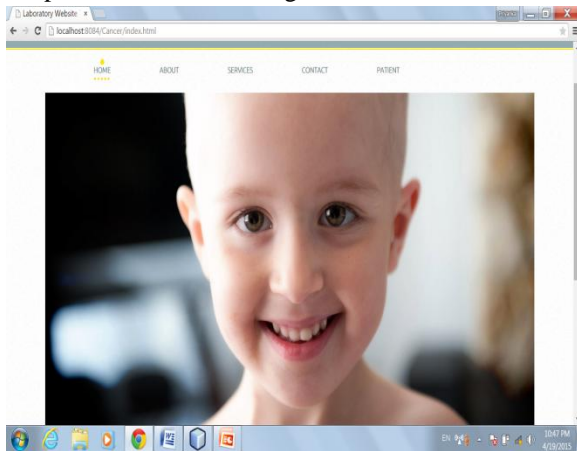


Figure 3 Homepage

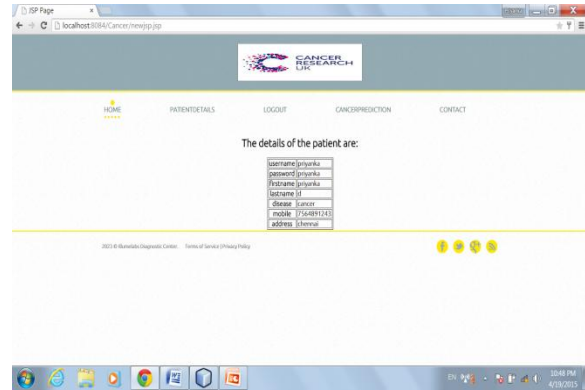


Figure 4 Information page

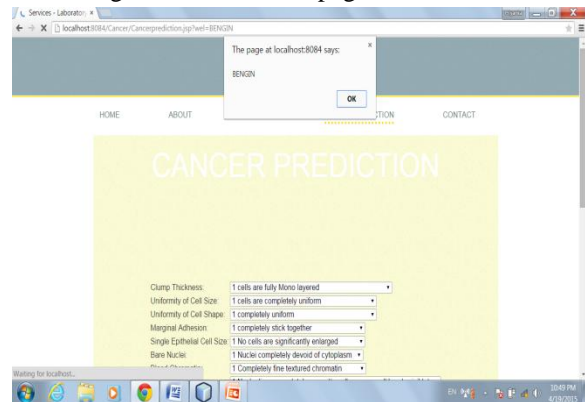


Figure 5 result page

V. CONCLUSION

In this research implemented the K-Means Clustering algorithms has been implemented. This performs certain number of iterations randomly which access the nearest observations into k , so as to attain the high speed time consumption and offers stability of the accurate result. Here, this research approaches the Compactness and Connectedness for accuracy result.

The compactness and connectedness for complementary measures are used and it is found that the efficiency and effectiveness of the method for predicting Heart Disease is better than the other three techniques through software prototype.

REFERENCES

[1]Ankit Agrawal, Sanchit Misra, Ramanathan Narayanan, Lalith Polepeddi, Alok Choudhary, “A Lung Cancer Outcome Calculator Using Ensemble Data Mining on SEER Data”(2011).
 [2] Rajneet Kaur, “A Study of Detection of Lung Cancer Using Data Mining Classification

Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering (2013).

[3] S.Sudha, “Disease Prediction in Data Mining Technique – A Survey”, International Journal of Computer Applications and Information Technology (2013).

[4] T.Karthikeyan,“PCA-NB Algorithm to Enhance the Predictive Accuracy”, International Journal of Engineering and Technology (IJET) (2014).

[5] V. Krishnaiah et al, “Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques”/ (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 - 45