Abstract

Data mining is applied in medical field since long back to predict disease like diseases of the heart, lungs and various tumors based on the past data collected from the patient. In India, though the data collection of medical patient is not streamlined, we made an effort to predict the most widely spread disease in India named tuberculosis. Using data collected from various TB centers, we made an effort to fetch out hidden patterns and by learning this pattern through the collected data for tuberculosis we can diagnose and predict the disease. In the research work
we are comparing naïve bayes classifier and KNN, two the most effective techniques for data classification (especially for medical diagnoses), implemented using C language and using Weka tool respectively and classify the patient affected by tuberculosis into two categories (least probable and most probable). We have used 19 symptoms of tuberculosis and collect 154 cases. We have achieved nearly 78% accuracy with low false negative.

**Reference**

2. Intelligent Heart Disease Prediction System Using Data Mining Techniques by Sellappan Palaniappan and Rafiah Awang IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008
4. An Introduction to Data Mining by Prof. S. Sudarshan CSE Dept, IIT Bombay
9. Inductive and Bayesian Learning in Medical Diagnosis Igor Kononenko University of Ljubljana.

**Index Terms**

Computer Science  
Data Mining

**Key words**

Data mining  
naïve bayes  
KNN  
pattern recognition
Comparative study of Naïve Bayes Classifier and KNN for Tuberculosis

tuberculosis

Machine learning