International Journal of ChemTech Research
CODEN (USA): IJCRGG ISSN: 0974-4290
Vol.8, No.10 pp 341-345, 2015

Identifying new regions infected by mycobacterium tuberculosis(TB) :Case Study in Kumbakonam Region,Thanjavur District

*Durga Karthik[1], K. Vijayarekha[2] and V. Dhivya[3]

[1]CSE, SRC, SASTRA University, Kumbakonam. India.
[2]School of EEE, SASTRA University, Thanjavur, India
[3]CS, SRC,SASTRA University,Kumbakonam,India

Abstract: Tuberculosis a communicable disease that spreads from an infected person to a vulnerable person. The spread of disease is usually from an infected person hence require early identification of TB infected person to protect the nearby location people. Nearly 2 million people are newly infected by TB in India and National TB control programme has enforced various measures to control the spread of the disease. Our aim is to identify new regions that are spread by the disease using data mining techniques and can be used to create awareness among the people of that region.
Keywords: Tuberculosis, regions, Data mining, clustering, awareness.

Introduction

Tuberculosis is a communicable disease [1] caused by infection of mycobacterium tuberculosis can be classified as pulmonary , extra pulmonary and recently Multi drug resistant TB or also known as Extreme Drug Resistant TB is prevailing across the world. The disease spreads when an infected person coughs in open air. The initial symptoms is the presence of persistent cough at night for a longer period, night sweat, fatigue, loss of weight and appetite. Government imposes measures to control the spread of disease using various programs.

Data mining methodologies [1] such as SVM,C4.5, PLS-DA,KNN can be employed for prediction of TB and can be used for diagnosis of survivability of the disease.Bayesian,neural meta, lazy modeling [2] and various other models and works are used for diagnosing TB. Cloud models [3] are popularizing for chronic diseases so as to create a self care for patients and cloud software [4] is available for disease diagnosis. Association rule generation[5] an important data mining technique is generated for predicting the disease to a greater accuracy.

Patient data contains various parameters hence parameter reduction [6,7] can employed for identifying principal components[8]. In this work, data mining [9,10] clustering algorithm is applied for identifying the new TB infected clusters in Kumbakonam region, Thanjavur district. The work can be used to create awareness among people of the region.

Materials and Methods

Region selected for the study is Kumbakonam 10.9700° N, 79.4200° E in Tamilnadu, India. Data from regional TB centre was collected to identify the new cases reported for infected TB. Two months data was

collected (Aug-Sep) from nearly 68 patients were diagnosed with TB. The various attributes such as patient details, gender, age, type (Pulmonary/Extra pulmonary), Category(I/II Curable/Persistent), Type(N/D/R/O ,New/Relapse/Default/Others) ,New patient is in initial stage, Relapse is the second stage with 40 Years above category and active. Default is next stage where TB is positive even after treatment, Others TB in regions other than lungs like cervical node etc., Given below in Table.1 that shows location wise statistics of patients affected by TB and Figure 1 shows the chart describing the statistics.

Table 1: Statistics of Patients Affected By TB in Kumbakonam Region age and Gender Wise

	Age					Gender	
Location	<3	03 - 20	20-40	40-60	>60	MALE	FEMALE
Thiruvidai Maruthur	1		1	2		4	
Srirengarajapuram			1				1
Maruthanalloor				2	1	3	
Kathiramangalam		1		1		1	1
Nagakudi				1		1	
Swamimalai					2	2	
Thirupuram			1				1
Senganoor				1		1	
Dharasuram				2	1	2	1
Nachiyarkovil			2		1	2	1
Puliyam Pettai					1	1	
Sakkotai			1	1	1	3	
Patteswaram			1				1
Pozhakudi			1			1	
Kumbakonam		1	9	7	3	14	6
Thirunageswaram			1			1	
Pandhanallor				1		1	
Ayyanallur				1		1	
Thirumangaicheri			1			1	
Narasingampettai				1			1
Veppathur			1	1		2	
Ammachathiram				1			1
Karupoor					2	2	
Valaiyapettai			1			1	
Kothankudi			1			1	
Neelathanalloor			1			1	
Murukangudi				1		1	
Sundharaperumal kovil			1				1
Sholapuram	1					1	
Keelakottaiyur			1			1	
Thirupanandhal				1	1		2
Asoor			1				1
Thenampadugai					1	1	

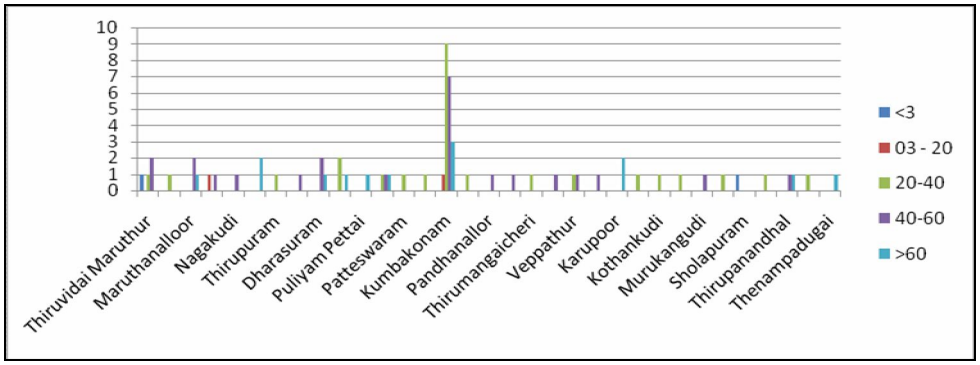

Fig 1: Patients affected by TB near Kumbakonam Region age wise

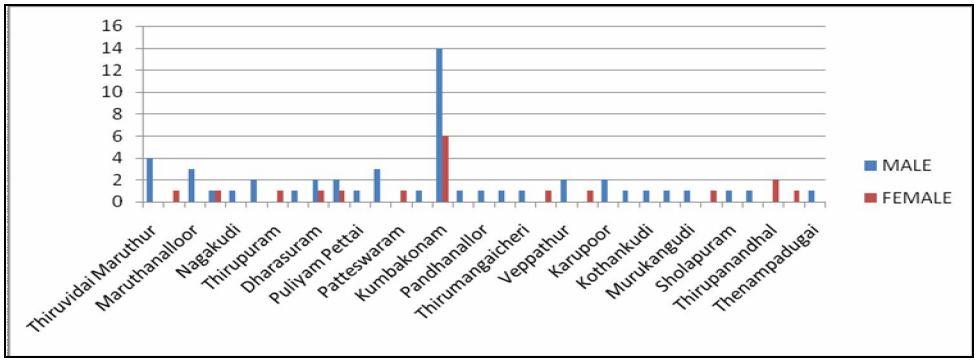

Fig 2: Patients affected by TB near Kumbakonam Region Gender Wise

Results and Discussion.

Data Mining Clusters:

Weka Data mining tool was employed to classify the patient data. ZeroR was used to classify the 66 patient details and the results of classification shows that 4 types were identified as N,D,O,R meaning New,Default,Others, Relapse.

```
Correctly Classified Instances          55               83.3333 %
Incorrectly Classified Instances        11               16.6667 %
=== Confusion Matrix ===
 a  b  c  d   <-- classified as
55  0  0  0 |  a = N
 2  0  0  0 |  b = D
 3  0  0  0 |  c = O
 6  0  0  0 |  d = R
```

Fig 2: Classification of TB patient data using ZeroR classifier.

The output shows that New patient statistics is 55 for the recent two months and for the other types the statistics is much lesser. The above classifier data is employed to identify the clusters using EM method. The output of EM is given in Fig.3.

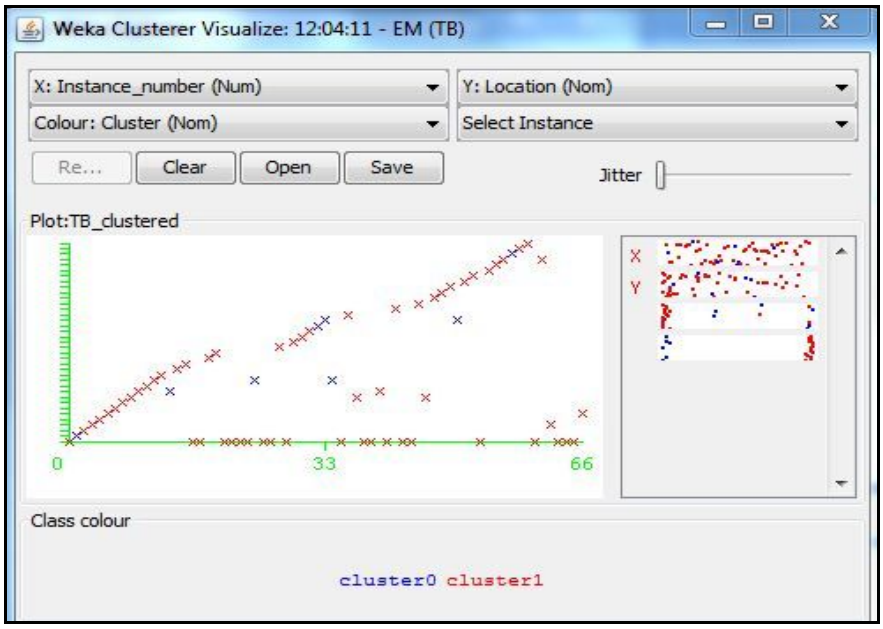

Fig 3: EM clusters for TB patient data of Kumbakonam Region

Cluster 0 is the formed from type O,R,D and Cluster 1 shows the new location affected by Mico Bacterium Tuberclosis. The new regions of spread are near each other when compared to the cluster 0.

Conclusion

Tuberculosis is contagious diseases that spreads through air and our work is to identify new regions that are affected by it. Data mining is found to be a great tool for faster and accurate method for identifying such regions that are reported to be affected by TB. The above method can be used for identifying clusters of region so as to create awareness to the people of that region and can be extended to identify the socio economic background of the people getting newly infected by the disease.

References:

1. K.R.Lakshmi,M.Veera Krishna,S.Prem Kumar,"Utilization of Data Mining Techniques for Predicition and Diagnosis of Tuberculosis Disease Survivability",I.J.Modern Education and Computer Science, 2013,8,8-17.
2. Utah Rusdah, Edi Winarko "Review on Data Mining Methods for Tuberculosis Diagnosis", Information Systems International Conference ;2013,563-568.
3. John D. Piette,Ph.D,Milton O,Mendoza-Avelaress,M.D@ etc.,"A premilinary study of a cloud computing model for chronic illness self-care support I an underdeveloped country ", 40(6),2011,1-14.
4. Ria lestari, Munawar Ahmad,Batchti Alisjahbana,Tjahjono Djatmiko,"The Lung disease Diagnosis software:Influenza and Tuberculosis case studies in the Cloud Computing environment".
5. T.Asha,S.Natarajan, K.N.B.Murthy, "Data mining Techniques in the Diagnosis of Tuberculosis", www.intechopen.com,333-352.
6. DurgaKarthik and K. VijayaRekha, 2012. SATTVA-Statistical Affirmative Testing Tool for Various Adulterants. Research Journal of Applied Sciences, Engineering and Technology, 4(24): 5357-5360.
7. Karthik, D.; Vijayarekha, K.; Manjula, K.,2012. "Multivariate analysis for detecting adulteration in edible oil: A review," Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on , vol., no., pp.272,277, 30-31 March 2012.
8. Karthik D, K. Vijayarekha, S.Sekar, 2014, "Profiling Water Quality Using Multivariate

a. Chemometric Method",Polish Journal of Environmental Studies,Vol 23,No:2.573-576.

9. Karthik, D., Vijayarekha, K., Raga Subha, S.,2014. "Classifying the suitability of river cauvery water for drinking using data mining classifiers and authenticating using fuzzy logic ",International Journal of ChemTech Research, 7 (5), pp. 2203-2207.

10. Karthik, D., Vijayarekha, K. 2015, "Simple and Quick Classification of Soil for Sunflower Cultivation using Data Mining Algorithm" International Journal of ChemTech Research, 7 (6), pp. 2601-2605.

*****