Journal of Indian Research (ISSN: 2320-7000)
Vol.2, No.1, January-March, 2014, 115-117

Medical Science

UTI DIAGNOSIS: A COMPUTATIONAL APPROACH

Tariq Rasool*
S.M.K Quadri**
Farooq Iqbal****

ABSTRACT

Urinary tract infections are the most common type of infections affecting the urino-genital tract of human beings. Although most of the times urinary tract infections are not life threatening but if not diagnosed and treated well in time, can lead to severe complications. UTIs should not be difficult to diagnose but at times symptoms of some other diseases of Urino-Genital tract overlap with the symptoms of UTIs to a greater extent. This overlapping makes the diagnosis challenging and chances of misdiagnosis increase considerably particularly in case of junior physicians. In this paper, we explore the applicability of a modified version of K-Nearest Neighbor technique from the field of data mining to diagnose UTIs on the basis of clinical symptoms presented in a health care facility.

Keywords: UTI, Urino-genital, Data Mining, K-NN.

INTRODUCTION

According to an estimate UTI burden accounts for as many as seven million out-patient department visits and 100,000 hospitalizations. Of all nosocomial infections UTI alone accounts for 35% and is the second largest cause of bacterimia in hospitalized patients [1]. A major chunk of the work load in clinical microbiological laboratories is contributed by UTI [2].Clinical diagnosis of UTI is challenging as most of the times the clinical symptoms on presentation overlap with other diseases to a considerable degree. Clinical diagnosis of diseases on the basis of symptoms and clinical examination requires extraction of relevant clinical leads and discovering hidden connections in the physiological change patterns on the part of a Clinician. This task of clinical diagnosis becomes challenging as the symptoms of various diseases most of the times overlap to a larger extent. With the advent and exponential growth in the knowledge base of mathematical and computational sciences, researchers have applied tools and techniques from these fields to healthcare and medical sciences in order to have a deeper understanding of medical phenomena and attempt to solve real world health problems.

Data mining is concerned with digging deep into large data sets in order to find hidden and previously unknown connections, relationships in data patterns and hence data mining results in knowledge discovery. Researchers have applied the tools and techniques from the field of data mining to a wide range of disciplines such as financial forecasting, analysis of organic compounds, weather forecasting and healthcare [3]. Data mining in medical sciences is a relatively new and upcoming field. Medical data mining aims at solving clinical problems of diagnosis and treatment [4]. Previously tools and techniques from data mining have been used in diagnosis of heart diseases [5], diabetes [6], stroke [7] and cancer [8].One of the most widely used techniques from data mining for classification problems is the K-Nearest Neighbor approach[9].

***Tariq Rasool** is a Research Scholar, Mewar University, Gangrar, Chittorgarh,Rajasthan.

****M.K Quadri** is Faculty at the Department of Computer Sciences, University of Kashmir.

*****Farooq Iqbal**, MD is a Physician Specialist based in Bangalore, Karnataka.

MATERIALS AND METHODS

Study setting

This retrospective study was conducted using the medical records from the W.C medical center, a super-specialty nephrology hospital in Kashmir.

Study design

Medical records of randomly selected 130 patients who were treated at WC Medical center, a super specialty nephrology health care facility from Jan 2010 to June 2012 were perused. After going through their medical records in consultation with a resident doctor, it was found that after clinical examination 80 patients had been diagnosed and treated for UTI, whereas the rest of the study population had overlapping symptoms with UTI, however their laboratory investigations had excluded UTI. In consultation with the resident doctor, a set of classical cardinal symptoms of UTI was prepared and appropriate weight was given to every individual clinical symptom. This weighted set of clinical symptoms constituted our data set and it was used to design the UTI partition of our knowledge-base. The set of clinical symptoms of the rest 50 patients constituted the non-UTI partition of our knowledge-base.

Modified K-Nearest Neighbor Algorithm

The following algorithm was used to design our knowledge-base, logically separated into several separate classes to predict the classification of a new sample vector into one of the two partitions (UTI, Non-UTI). It works as under:

Step1. Design a knowledge base 'Z' partitioned into well defined 'N' classes,

Where,

$$Z=\bigcup_{i=1}^{n} X_i \quad [n >=2]$$

And

$$\bigcap_{i=1}^{n} X_i = \Phi \quad \text{Where, } Xi = x[j];\ j= \{1\ldots m\}$$

Step2. Input=Y[q],

Where, $q = \{1\ldots r\}$

Step3. Calculate K–nearest neighbors of Y using Euclidean distance between vectors

$$d\ (x_{ij,}\ y_q) = \sqrt{\sum^{n}}_{i=1}(x_{ij\ ,}\ y_q)^2$$

Step4. Determine $C(y) = f\ (d\ (x_{ij,}\ y_q))$

Step5. Return $C(y) = X_i\ \ 1<= i<= n$

The algorithm works by mapping the input symptom vector to one of the two possible partitions (either the UTI partition or the non-UTI partition) of our knowledge base.

RESULTS

In addition to the derivation vector set (consisting of 130 patients), an additional set of 50 vectors containing clinical symptoms from the medical record department was prepared. Out of these fifty, 35 had been confirmed as having UTI and treated for the same and 15 had overlapping symptoms with UTI but investigations had revealed some other diseases. This constituted our validation vector set.

Out of this combined test vector set (115 UTI vectors +65 Non-UTI vectors) our technique correctly identified 102 patients as having UTI (out of the UTI vector set) and out of the 65 who did not have UTI, correctly categorized 53 in the Non-UTI class. These results indicated a sensitivity of 88.7%, specificity of 81.54% and an overall accuracy of 86.11% (**Table** 2).

CONCLUSION

This paper stresses on the need to utilize tools and techniques from the field of computational sciences in order to understand and better diagnose various disease conditions. While it is imperative to mention that the technological advances are no replacement for trained clinicians, however this study demonstrated that computational techniques can serve as a valuable tool to clinicians in general, and junior physicians in particular.

Table 1: Clinical Symptoms

Lower Abdominal pain
Hematuria
Dysuria
Frequency
Fever
Flank Pain
Vaginal irritation
Vaginal discharge
Back pain
Costovertebral Angle Tenderness
Dipstick Urinalysis

Table 2: Performance evaluation

Sensitivity	Specificity	Accuracy
88.7%	81.54%	86.11%

REFERENCES

1. Need Akhtar(2000). Urinary tract bacterial pathogens, their antimicrobial susceptibility patterns at Bahawalpur. *The Professional*, vol. **7(**2):131-137.
2. Sklar, AH, Caruana, RJ, Lammers, JE. And Strauser, G O.(1987). Renal infections in Autosomal Polycystiuc Kidney diseases. *Am J. Kidney Dis*., Vol.**10**, pp. 81-88.
3. D. Ashby and A. Smith(2005). The Best Medicine? , *Plus Magazine Living Mathematics.*
4. Liao, S. C. and I. N. Lee(2002). "Appropriate medical data categorization for data mining classification techniques", *MED. INFORM*, vol. **27(**1): 59-67.
5. Das, R., I. Turkoglu, A. Sengur(2009). Effective diagnosis of heart disease through neural networks ensembles, Expert systems with Applications, *Elsevier,* pp.7675-7680.
6. T. Porter and B. Green(2009). "Identifying Diabetic Patients: A Data mining Approach." *Americas Conference on Information Systems*, 2009.
7. S. Panzarasa et al.(2010). "Data mining techniques for analyzing strike care processes," in Proc. Of Thirteenth World Congress on Medical Informatics, 2010.
8. Li, L, T. H., Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark(2004). "Data mining techniques for cancer detection using serum proteomic profiling", *Artificial Intelligence in Medicine*, Elsevier.
9. Moreno-Seco, F., L Mico, and J. A. Oncina(2003). "Modification of the LAESA Algorithm for Approximated K-NN Classification," *Pattern Recognition Letters,* pp. 47-53.