ANFIS BASED FUZZY CLUSTERING SYSTEM FOR DIFFERENTIAL DIAGNOSIS OF CONFUSABLE DISEASES

ISSN: 2141 – 3290
www.wojast.com

OBOT, O. U. AND INYANG, U. G.,
Department of Computer Science,
University of Uyo, Nigeria
abatakure@yahoo.ocm, udoiinyang@yahoo.com

ABSTRACT

Some tropical diseases are no longer localized in tropical regions of the world and are increasingly seen in the emergency units of major medical centers in the developed world. A number of these tropical diseases present with overlapping symptoms that can be confusing to medical practitioners during the process of medical diagnosis. The combination of non-specific clinical manifestations that characterize these conditions and the probable lack of expertise and experience among primary care and emergency physicians exponentially increases the potential for mis-diagnosis and subsequent increased morbidity and mortality rates resulting from these diseases. This paper proposes an intelligent system based on fuzzy Logic and fuzzy clustering to assist physicians in investigating, diagnosing, and managing confusing symptoms for accurate and timely diagnosis of confusable diseases. Data on patients diagnosed and confirmed by laboratory tests of viral hepatitis, malaria, typhoid fever and urinary tract infection were collected from general hospitals in Nigeria. The data were used in training, testing and validation of the system. The validity of the clusters were also carried out. The results show that the system compares favorably with diagnosis arrived at by experienced physicians and in addition provides patient's level of severity in each confusable disease.

INTRODUCTION

Diseases that were considered endemic only in tropical regions of the world are increasingly being encountered at hospital emergency rooms in the developed world (Birnbanmer & Ratkowski, 2003; Keller *et al.*, 2008). This, in part, can be attributed to the globalization of travel, and the ease and efficiency of intercontinental flights. It is not uncommon in the developed world to observe that specialty medical practitioners in tropical conditions are in short supply, and most major health centers do not have tropical disease management units attached to them (Suh, *et al.,* 2004). As such medical doctors encounter difficulties in analyzing and diagnosing tropical diseases, especially when confronted with conditions that present non-specific manifestations or those illnesses that share similar clinical symptoms. In order to ascertain the etiology of the symptoms presented by patients, medical practitioners often depend on guess work, buy time for the symptoms to resolve on their own, or subject clients to unnecessary laboratory investigations with attendant delays. Such delays could result in unnecessary suffering by patients, predispose to disease complications and even death.

Konar (2000) suggests that there is a wide scope to many diseases with the same or similar symptoms; therefore doctors generally diagnose a disease from the relative strength and ranking of these symptoms. The main difficulty in doing the above is the inability of the physician to perform a relative ranking or quantification of the relative levels of these symptoms in order to arrive at the appropriate diagnosis in a timely manner. Artificial intelligence systems based on fuzzy logic enables a computer to make decisions that are more in line with human thinking. Computer based decision support systems, based on fuzzy logic have proven to be of great value in medical practice in the diagnosis of malaria (Uzoka, *et al.,* 2011) and in clinical documentation and record management, clinical diagnosis and treatment plan process, monitoring medication and laboratory test orders, and managing clinical complexity and details (Perreault and Metzger 1999). Fuzzy clustering can be used to resolve the ambiguity brought about by similar disease symptoms by assigning some degrees of influence on the manifestations of these symptoms, thereby associating any patient with varying degree of severity in all the Confusable Tropical Diseases (CTD). Having succeeded in doing so, medical practitioners are enabled to rank the confusable diseases based on the severity levels and can then call for specific laboratory investigations of the conditions with the highest severity level or attend to conditions that are most life threatening. Viral hepatitis, malaria, typhoid fever and urinary tract infections are common tropical conditions with similar symptoms, with the potential of delaying accurate diagnosis, or in worst scenarios, a wrong diagnosis that can cause death.

The objective of this research is to apply fuzzy logic technology comprising fuzzy clustering and fuzzy inference systems and neuro-fuzzy approaches to resolve the conflict or confusion arising from the common symptoms. It will also assist in streamlining the diagnostic process.

METHOD

Dataset Collection and Description

The datasets for the study were collected from general hospitals in Akwa Ibom State, Nigeria. Specialist doctors provided information on patients already diagnosed and confirmed as having any of these tropical diseases; Malaria, Urinary Tract Infection and Viral Hepatitis and Typhoid. From the physicians' physical examination of the patient and the patients self reported account of his symptoms, the medical doctor assigns linguistic values which represent the stage of the symptom. {Very Severe, Severe, Moderate, Mild} which were transformed into crisp values {4,3,2,1}. Data on a total of 67 patients were collected by the four physicians. The set of symptoms captured are presented in Table 1. From Table 1, a total of twenty two symptoms account for Confusable Tropical Diseases (CTD), "1" denotes the presence of a symptom in determining the disease while '0" implies absence of the symptom in that disease.

Table 1: Symptoms of confusable tropical diseases

S/N	Symptom Description	Symptom Code	Tropical Diseases Malaria	Viral Hepatitis	Typhoid	Urinary Tract Infection
1	Abnormal Breathing	a1	1	1	1	1
2	Appetite	a2	1	1	1	1
3	Backache	b1	1	0	1	0
4	Chill	c1	1	1	1	1
5	Cough	c2	1	1	1	1
6	Dehydration	d1	1	1	1	1
7	Delirium	d2	1	1	1	1
8	Diarrhea	d3	1	0	1	0
9	Dizziness	d4	1	0	0	0
10	Excessive Sleeping	e1	1	0	1	0
11	Fever	f1	1	1	0	0
12	Headache	h1	1	1	1	1
13	Joint Pain	j1	1	1	1	1
14	Muscle Ache	m1	1	1	1	0
15	Nausea	n1	0	0	1	0
16	Paleness	p1	0	1	1	0
17	Shivering	s1	0	1	1	1
18	Stomach Discomfort	s2	0	0	1	0
19	Sweating	s3	1	0	1	0
20	Tiredness	t1	1	1	1	0
21	Vomiting	v1	0	1	1	0
22	Yellowish Eye	y1	0	1	1	0

Fuzzy Clustering Means (FCM) and Adaptive Neuro-fuzzy Inference System (ANFIS)
The classical set representation of the four Confusable Tropical Diseases (CTD), Malaria (M), Typhoid (T), Urinary Tract Infection (U) and Viral Hepatitis (H) (Figure 1), there are fifteen labeled sections each representing a subset of any one or more CTDs. The shaded portions represent the overlapping symptoms while the non-shaded portions A,C,D and F are specific symptoms of respective CTD. The intersecting subsets give the degree or measure of contribution of the overlapping symptoms to each of the CTD.

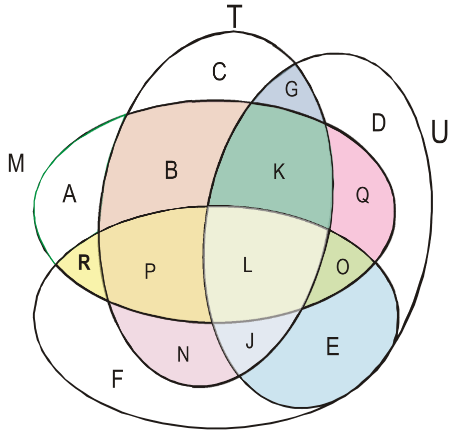

Figure 1: Classical set diagram of CTD

In this paper, the problem of Confusable Tropical Disease Diagnosis (CTDD) can be formalized as a classification problem, where a set of overlapping and non-overlapping symptoms are measured and a set of partial diagnostic values (μ_{d_i}) are defined for the tropical disease set X= {$d_1, d_2, ..., d_c$}, where μ_{d_i} represents the degree (proportion) to which any patient suffers from the *ith* disease or the chances (probability) that a patient is suffering from the *ith* disease, i=1,2,3…, n. Such that $\mu_{d_i}(x): X \rightarrow [0,1]$ and $\sum_{i=1}^{n} \mu_{d_i} = 1$. Let $\mu_M(x)$ represents the degree to which any patient suffers from malaria, $\mu_U(x)$ represents the degree to which a patient suffers from Urinary Tract Infection, $\mu_H(x)$ and $\mu_T(x)$ represent the degree to which a patient suffers from Viral Hepatitis and Typhoid Fever respectively.

This paper considers a method by which fuzzy membership functions are created for four clusters of CTD. The major components of the system are Knowledge Base (KB) and Inference Engine. The KB contain symptoms, fuzzy linguistics values, fuzzy rules and fuzzy membership values. The Inference engine is driven by two clustering algorithms (FCM and k-means algorithms) and ANFIS. Fuzzy Inference Systems structures were generated for ANFIS. Takagi-Sugeno-type was generated using subtractive clustering in determining the number of rules and antecedent membership functions; and linear least squares estimation method for determining each rule's consequent. The FCM parameters were set as; m = 2, ε = 0.01, and k = 200. The training dataset consists of 49 records while 18 records were used as the test dataset. The mean of all the data points was the initial cluster center.

RESULTS

The performances of the FCM objective function while partitioning the training data set into four clusters are shown in Figure 2. The initial value of the objective function is 418.55, and decreases to 323.89 at the 4th iteration without further changes. The properties of the ANFIS model are presented in Table 2 while the graph of the testing error is presented in Figure 3. The

Root Mean Square Error (RMSE) for training and testing are 0.2020 and 0.2000 respectively. The patient's severity level in each confusable disease is presented in Table 3 and Figure 4 respectively.

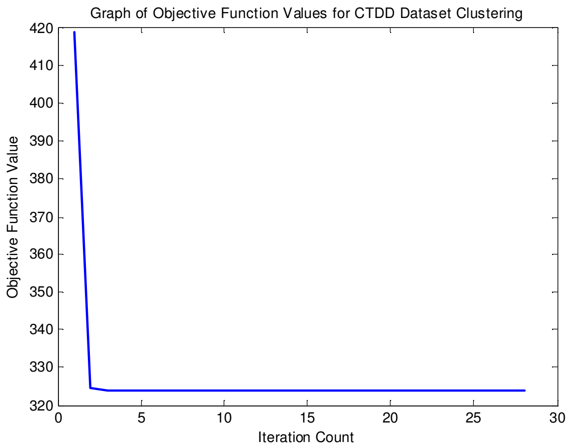

Figure 2: Objective Function of CTDD Dataset Clustering

Table 2: Description of Training Parameters for ANFIS

ANFIS Parameter	Values
Number of nodes	1681
Number of linear parameters	828
Number of nonlinear parameters	1584
Total number of parameters	2412
Number of training data pairs	49
Number of checking data pairs	18
Number of fuzzy rules	36

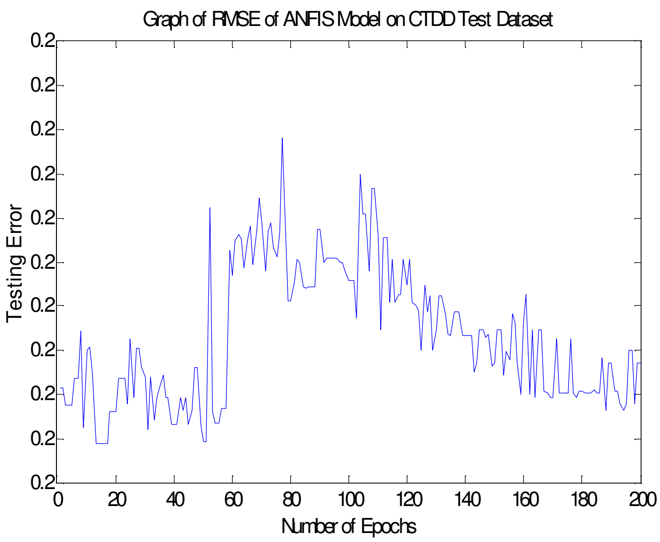

Figure 3: Graph of Testing error

Table 3: FCM degree of membership of CTD for each patient

Patient ID	Degree of Membership				Patient ID	Degree of Membership			
	M	H	T	U		M	H	T	U
01	0.18	0.2	0.53	0.09	26	0.25	0.31	0.38	0.06
02	0.68	0.12	0.19	0.01	27	0.16	0.24	0.27	0.33
03	0.28	0.21	0.45	0.06	28	0.64	0.19	0.15	0.02
04	0.03	0.05	0.05	0.87	29	0.25	0.46	0.26	0.03
05	0.46	0.18	0.33	0.03	30	0.43	0.24	0.26	0.07
06	0.19	0.21	0.4	0.2	31	0.21	0.51	0.23	0.05
07	0.42	0.18	0.38	0.02	32	0.72	0.12	0.15	0.01
08	0.56	0.14	0.27	0.03	33	0.28	0.22	0.43	0.07
09	0.83	0.06	0.1	0.01	34	0.14	0.42	0.32	0.12
10	0.17	0.54	0.19	0.1	35	0.15	0.27	0.38	0.2
11	0.03	0.04	0.05	0.88	36	0.68	0.12	0.19	0.01
13	0.16	0.25	0.28	0.31	38	0.87	0.05	0.05	0.03
14	0.6	0.19	0.18	0.03	39	0.39	0.19	0.34	0.08
15	0.25	0.46	0.26	0.03	40	0.15	0.21	0.3	0.34
16	0.48	0.22	0.26	0.04	41	0.34	0.22	0.38	0.06
17	0.26	0.47	0.24	0.03	42	0.18	0.2	0.53	0.09
18	0.55	0.21	0.21	0.03	43	0.06	0.83	0.1	0.01
19	0.27	0.24	0.35	0.14	44	0.17	0.54	0.19	0.1
20	0.14	0.33	0.36	0.17	45	0.03	0.04	0.05	0.88
21	0.43	0.16	0.38	0.03	46	0.18	0.33	0.35	0.14
22	0.57	0.13	0.27	0.03	47	0.17	0.22	0.29	0.32
23	0.06	0.83	0.1	0.01	48	0.64	0.19	0.15	0.02
24	0.17	0.54	0.19	0.1	49	0.25	0.46	0.26	0.03
25	0.07	0.09	0.11	0.73					

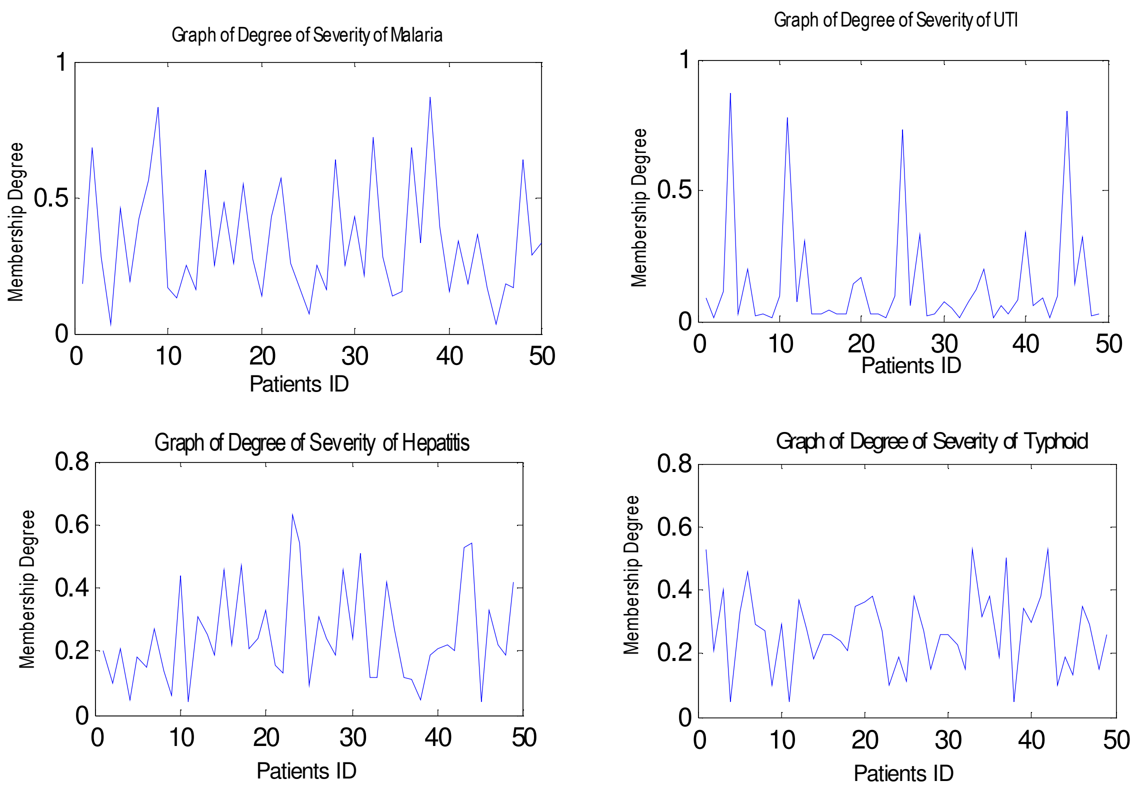

Figure 4: Patients severity levels in each of the Confusable Tropical Diseases

DISCUSSION AND CONCLUSION

The FCM Clustering results (Table 3) show that patients suffer from the four confusable diseases with varying degree of severity. Seventeen patients had highest degree of severity in malaria while eleven patients had in Hepatitis. The degree of severity of thirteen patients is highest in Typhoid while eight patients showed highest degree of severity in Urinary Tract Infection. This information provide guides to medical experts on which CTD a patient may be treated at a particular time. For patients with competing degree of severity in more than one CTD, (For example, Patient P12, P23 and P46 possess competing degree of severity level in Typhoid and Hepatitis), the physician may treat the patient in such diseases based on his experience or may consider the most life threatening one.

Partition Coefficient (PC) given in Trauwaert, (1988) and Partition Entropy Coefficient (PE) presented in Hung *et al,* (2008), were indices used in validating the clusters. The cluster validation analysis yields 0.396 and 0.795 for PC and PE respectively. The PC value is in the range [¼,1] and closer to the lower bound (0.25) showing the existence of fuzzy partitions. Similarly, the PE value of 0.795 lies within the required range of [0, $\log_4 4$] (0,1). The PE value is closer to the upper bound which explains the existence of fuzzy clusters. The results from both validity indices (PC, PE) validate the clustering results.

The results are satisfactory and the system will assist physicians in arriving at accurate diagnosis in no time. This confirms the suitability of the hybridization of fuzzy clustering and ANFIS in the analysis and diagnosis of Confusable Tropical Diseases.

REFERENCES

Birnbaumer, D. M. and Rutkowski, A. (2003) Malaria: A comprehensive review for the emergency physician. *Topics in Emergency Medicine*, 25(1), 2-12.

Hung , Wen-Liang, 1, Yang, Miin-Shen (2008):On Similarity Measures between intuitionist Fuzzy Sets. *International Journal of Intelligent Systems*, (23) pp 364–383

Keller, A., Frey, M., Schmid, H., Steffen, R., Walker, T. and Schlagenhaut, P. (2008). Imported tropical fever in switzerland 1993-2004. *Journal of Travel medicine,* 15(4), 248-251.

Konar A (2000) Artificial Intelligence and Soft Computing; Behavioural and Cognitive Modeling of the Human Brain, *CRC press*, New York.

Perreault L and Metzger J. (1999); A pragmatic framework for understanding clinical decision support. *Journal of Healthcare Information Management.* 13(2):5-21

Ryan, E. T., Wilson, M. E., & Kain, K. C. (2002). Illness after international travel. New England Journal of Medicine, 347(7), 505-506.

Suh, K. N., Kain, K. C. and Keystone, J. S. (2004). Malaria. *Canadian Medical Association Journal*, 170(1), 1693-1702.

Trauwaert, E., 1988. On the meaning of Dunns partition coefficient for fuzzy clusters. *Fuzzy Sets Systems* 25, 217–242.

Uzoka, F. M. E., Osuji, J. C. and Obot, O. (2011). Clinical decision support system (DSS) in the diagnosis of malaria: A case comparison of two soft computing methodologies. *Expert Systems with applications,* 38(3), 1269-2922