ISSN: 2141 – 3290
www.wojast.com

DIAGNOSTIC EVALUATION OF HEPATITIS UTILIZING FUZZY CLUSTERING MEANS

IMIANVAN A.A[1]. AND OBI J.C.[2]
Department of Computer Science,
University of Benin, Benin City. Nigeria.
E-mail: tonyvanni@yahoo.com[1]
and tripplejo2k2@yahoo.com[2]
+234 (0)7069742552[1] and +234(0)8093088218[2]

ABSTRACT: Hepatitis is a general name for a group of serious diseases that affect the liver or inflammation of the liver. Hepatitis can be transmitted from person to person by eating and drinking contaminated food and water, sharing sharp needles, unprotected sex and blood transfusion. It is usually caused by Immune cells in the body attacking the liver, Infections from viruses (such as hepatitis A, B, or C), bacteria or parasites, liver damage from alcohol, and certain medication. Hepatitis is characterized by life threatening symptoms such as malaise, biliousness, lethargy, orange urine, pales feces and nausea. The application of Fuzzy Cluster Means (FCM or Fuzzy C-Mean) analysis to the diagnosis of different forms of hepatitis is the focal point of this paper. Application of cluster analysis involves a sequence of methodological and analytical decision steps that enhances the quality and meaning of the clusters produced. The uncertainties often associated with analysis of a hepatitis test data are eliminated by the proposed system.

INTRODUCTION

Hepatitis often refers to infectious hepatitis, includes hepatitis A, hepatitis B, hepatitis C, hepatitis D, hepatitis E. Each of these types of hepatitis is caused by the hepatitis virus of the same name, such as the hepatitis A virus. The most common forms of hepatitis are hepatitis A, hepatitis B and hepatitis C, Healthline, (2011) and Wrongdiagnosis, (2011). Hepatitis is a general name for a group of serious diseases that affect the liver or inflammation of the liver, Medicinenet, (2011) and Wrongdiagnosis, (2011). It is not a condition but is often used to refer to a viral infection of the liver, Righthealth, (2011).

Hepatitis is usually caused by certain medications (for example Tylenol (acetaminophen), mononucleosis and the cytomegalovirus), immune cells in the body attacking the liver, Infections from viruses (such as hepatitis A, B, or C), bacteria or parasites and liver damage from alcohol and poisonous mushrooms, Healthline, (2011), Medicinenet, (2011) and Wrongdiagnosis, (2011). All forms of hepatitis cause an inflammation of the liver. The liver is a vital organ, and normal functioning of the liver is crucial to health and life. Hepatitis can reduce the liver's ability to do its vital job in helping the body to fight infection, stop bleeding, clear the blood of toxins, store energy, produce healthy blood, digest food and remove waste.

There are differences in the way that the various forms of infectious hepatitis are passed from person to person and how they progress. Example of such are eating and drinking contaminated food and water, sharing sharp needles, unprotected sex and blood transfusion. Symptoms of hepatitis also differ between individuals depending on such variables as the specific type of hepatitis, age, medical history, the presence of complications and general health.

Symptoms common to hepatitis include flu-like symptoms, fever, headache, nausea, muscle aches and jaundice, a yellowing of the skin and whites of the eyes. Complications can be serious, even life-threatening, and include the development of chronic hepatitis, cirrhosis, and increased risk of liver cancer, and liver failure.

Fuzzy Logic presents an inference morphology that enables appropriate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic utilizes mathematical strength to capture the uncertainties associated with human cognitive processes. This paper presents a Fuzzy Cluster Mean (Fuzzy C-Mean or FCM) knowledge-based model for the diagnosis of Hepatitis. Fuzzy C-means clustering algorithm is being used as the problem solving and reasoning algorithm in the inference engine of the knowledge base system for the evaluating, classification and matching of patterns to more than one class of hepatitis. Statistics, Neural network and Physiology are also incorporated. The main objective is to classify and match any individual independently with more than one cluster depending on the degree of membership.

Making a diagnosis of hepatitis includes performing a complete medical evaluation, history and physical examination. This includes questioning about risk factors for contracting hepatitis, such as being sexually active outside of a monogamous relationship, living in or travelling to places with poor sanitation.

Diagnosis of hepatitis may include test. Blood tests include tests that can check for the viruses that cause hepatitis. Liver function tests are blood tests that can help to determine the level of severity of hepatitis by checking level of functioning of the liver and if there is any damage to the liver, Healthline, (2011), Medicinenet, (2011) and Wrongdiagnosis, (2011). Imaging tests that create a picture of the liver and may be performed in some cases include an ultrasound, CT, and/or a nuclear liver scan. It is possible that a diagnosis of hepatitis can be missed or delayed because symptoms can be vague or there may be no symptoms in some people in some forms of the disease.

Treatment of hepatitis includes rest and ensuring good nutrition. For serious cases, especially if there is liver damage, hospitalization may be necessary. Treatment in the hospital may include medications, a liver biopsy, and other diagnostic testing and treatment. There are different forms of hepatitis (A, B, C, D and E). This paper focuses on identifying Hepatitis A and B.

Hepatitis A is one of the most common forms of hepatitis. Hepatitis A is an infectious form of hepatitis that is caused by the hepatitis A virus. The liver is a vital organ, and normal functioning of the liver is crucial to health and life. Hepatitis A can reduce the liver's ability to do its vital job in helping the body to fight infection, stop bleeding, clear the blood of toxins, store energy, produce healthy blood, digest food, and remove waste. In rare cases, hepatitis A can be fatal. The hepatitis A virus is found in the stool of people who are infected with the virus. Hepatitis is most common in developing countries and in places where sanitation is poor, such as in refugee camps and is closely contacted by children, Wrongdiagnosis, (2011).

Hepatitis B is an infectious form of hepatitis that is caused by the hepatitis B virus. The hepatitis B virus is spread by having contact with the blood, semen, and vaginal secretions of a person infected with the hepatitis B virus. High risk activities include having unprotected sexual activity, having multiple sexual partners, sharing contaminated needles, or masking a tattoo or body piercing using unsterilized needles. A baby born to an infected woman can also contact an infection of hepatitis B. Any person who comes into frequent contact with blood, such as healthcare workers, is also at risk for hepatitis B, HealthLine, (2011).

Fuzzy Logic (FL) helps computers paint pictures of uncertain world. Fuzzy sets were introduced by Zadeh (1965) as a means of representing and manipulating data that are not precise, but rather fuzzy. Fuzzy logic provides an inference morphology that enables appropriate human reasoning capabilities to be applied to knowledge-based systems. The theory of fuzzy logic provides a mathematical strength to capture the uncertainties associated with human cognitive processes, such as thinking and reasoning, Nascimento, (1997). A fuzzy

set A is called trapezoidal fuzzy number (Figure 1) with tolerance interval [a, b], left width α and right width β if its membership function has the following form

$$A(t) = \begin{cases} 1-(a-t)/\alpha & if\ a-\alpha \le t \le a \\ 1 & if\ a \le t \le b \\ 1-(t-b)/\beta & if\ a \le t \le b+\beta \\ 0 & otherwise \end{cases}$$

and we use the notation A = (a, b,α,β). It can easily be shown that

$[A]^{\gamma}$ = [a − (1 − γ) α, b + (1 − γ) β], ∀ γ ε [0, 1].
The support of A is (a − α, b + β (Nascimento, 1997).

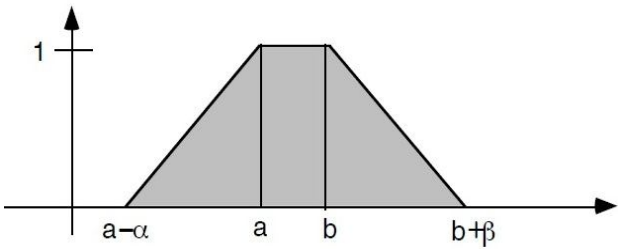

Figure 1: Trapezoidal fuzzy number.

Fuzzy Clustering

Clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible, Yang and Wang (2001). Clustering can also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative prototypes or clusters, Giles and Draeseke (2001). Depending on the data and the application, different types of similarity measures may be used to identify classes, where the similarity measure controls how the clusters are formed, Inyang (2005). Some examples of values that can be used as similarity measures include distance, connectivity, and intensity, Berks et al (2000).

In non-fuzzy or hard clustering, data is divided into crisp clusters, where each data point belongs to exactly one cluster, Albayrak and Amasyali (2003). In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades which indicate the degree to which the data points belong to the different clusters, Nascimento, (1997).

OVERVIEW OF FUZZY CLUSTERING MEANS (FCM) ALGORITHM

The FCM algorithm is one of the most widely used fuzzy clustering algorithms. The FCM algorithm attempts to partition a finite collection of elements $X=\{X_1, X_2,...,X_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centers V, such that $V = V_i$, i=1, 2,..., c
and a partition matrix U such that $U = U_{ij}$, i =1,..., c , j =1,..., n
where U_{ij} is a numerical value in [0, 1] that tells the degree to which the element X_j belongs to the i-th cluster.

The following is a linguistic description of the FCM algorithm, which is implemented by fuzzy Logic. The algorithm is as follows;

Step 1: Select the number of clusters c ($2 \le c \le n$), exponential weight μ ($1 < \mu < \infty$), initial partition matrix U^0, and the termination criterion ε. Also, set the iteration index 1 to 0.

Step 2: Calculate the fuzzy cluster centers $\{V_i^1 | I = 1, 2,..., c\}$ by using U^1.

Step 3: Calculate the new partition matrix U^{1+1} by using $\{V_i^1 | i=1, 2,..., c\}$.

Step 4: Calculate the new partition matrix = $|| U^{1+1} - U^1|| = | U_{ij}^{1+1} - U_{ij}^1 |$. If $>\varepsilon$, then set $l = l + 1$ and go to **step 2**. If $\leq\varepsilon$, then stop.

The initial cluster centers are computed in two ways; Arithmetic means of all the data points or running FCM several times each starting with different initial cluster centers. In this work the first method is adopted.

METHODOLOGY

This expert system which employs fuzzy C-Means for the diagnosis of hepatitis is developed in an environment characterized by Microsoft Window, XP professional Operating System, Microsoft Access Database Management system, Visual Basic Application Language and Microsoft Excel. Neuro Solutions and Crystal reports were used for neural network analysis and graphical representation respectively. An approach for analyzing clusters to identify meaningful pattern for determining whether a patient suffers from hepatitis or not is presented. The system provides a guide for diagnosis of hepatitis within the decision-making framework.

The process for the medical diagnosis of hepatitis starts when an individual consults a physician and presents a set of complaints. The physician then requests further information from the patient or from others close to him who knows about the patient's symptoms in severe cases. Data collected include patient's previous state of health, living condition and other medical conditions. A physical examination of the patient condition is conducted and in most cases, a medical observation along with medical test(s) is carried out on the patient prior to treatment.

From the symptoms presented by the patient, the physician narrows down the possibilities of the illness that corresponds to the apparent symptoms and make a list of the conditions that could account for illness. These are usually ranked in possibility order (low, moderate and high). The physician then conducts a physical examination of the patient, studies his medical records and ask further questions, as he goes in an effort to rule out as many of the potential conditions as possible. When the list has been narrowed down to a single condition, it is called differential diagnosis and provides the basis for a hypothesis of what is ailing the patient. Until the physician is certain of the condition present; further medical test are performed or schedule such as medical imaging, scan, X-rays in part to conform or disprove the diagnosis or to update the patient medical history. Other physicians, specialist and expert in the field may be consulted for further advices.

Despite all these complexities, most patient consultations are relatively brief because many diseases are obvious or the physician's experience may enable him to recognize the condition quickly. Upon the completion of the diagnosis by the physician, a treatment plan is proposed, which includes therapy and follow-up (further meeting and test to monitor the ailment and progress of the treatment if needed). Review of diagnosis may be conducted again if there is failure of the patient to respond to treatment that would normally work. The procedure of diagnosing a patient suffering from hepatitis is synonymous to the general approach to medical diagnosis. The physician may carry out a precise diagnosis, which requires a complete physical evaluation to determine whether the patient have hepatitis. The examining physician accounts for possibilities of having hepatitis through an interview, physical examination and laboratory test. Many primary health care physicians use screening tools for hepatitis evaluation.

A thorough diagnostic evaluation may include a complete history of the following:

a. When did the symptoms start?
b. How long have the symptoms lasted?
c. How severe are the symptoms?
d. Have the symptoms occurred before, and if so, were they treated and what treatment was received?

RESULTS AND DISCUSSION

To design the FCM knowledge base system for diagnosis of hepatitis, we designed a system which consists of a set of parameters needed for diagnosis (using 10 basic and major parameters Table 1).

Table 1: Symptoms of Hepatitis

s/n	Parameter
1.	**Malaise:** a feeling of general discomfort or uneasiness.
2.	**Biliousness:** Characterized by an excess secretion of bile.
3.	**Lethargy:** the quality or state of being lazy and sluggish
4.	**Orange urine:** Dark orange urine in nature
5.	**Pale feces:** pale or grey feces may be caused by insufficient bile
6.	**Liver failure:** poor functioning of the liver
7.	**Kidney failure:** Bad functioning of the kidney
8.	**Yellowish eye :** Caused by the hepatitis infection
9.	**Diarrhea:** Greater looseness of stool.
10.	**Nausea:** Sensation of unease and discomfort in the upper stomach

Figure 2 presents the model of the FCM system for the diagnosis of Hepatitis. It comprises knowledge base system, fuzzy c-means inference engine and decision support system. The knowledge base consists of the database engine. The knowledge base system holds the symptoms for hepatitis.

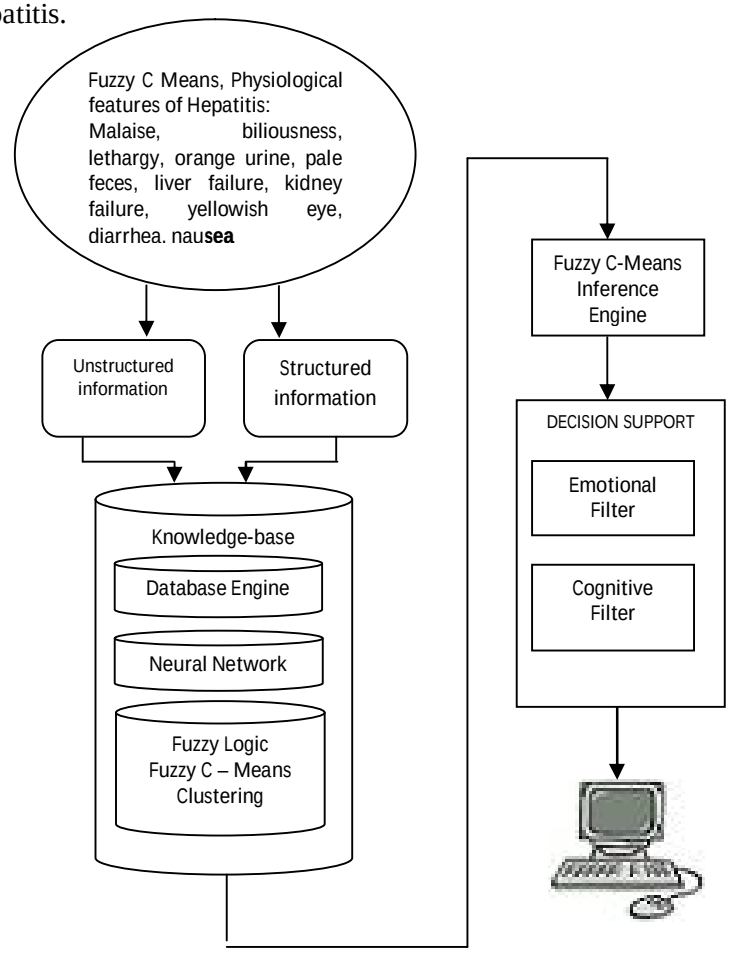

Figure 2: Architecture of the Fuzzy Cluster Means (FCM) knowledge base system for the diagnosis of Hepatitis

The values for the diagnosis of hepatitis symptoms are not precise hence the adoption of fuzzy logic as a means of analyzing these information. The different forms of hepatitis symptoms, therefore, constitute the fuzzy parameters of the knowledge base system. The fuzzy set of parameters is represented by 'P', which is defined as $P = \{P_1, P_2 \ldots, P_n\}$

where P_i represents the j^{th} parameter and n is the number of parameters (in this case n =10). Neural network provides the structure intelligent learning for all forms of hepatitis symptoms, which serves as a platform for the inference engine. The inference engine consists of reasoning algorithms, driven by production rules. These production rules are evaluated by using the forward chaining approach of reasoning. The fuzzy logic and fuzzy C-means algorithm provides the rules for the partitioning of patients into a number of homogenous clusters with respect to a suitable similarity measure.

In this paper, the patients were classified into two form of hepatitis. Hepatitis A refers to patient experiencing five or more of the listed symptoms after dwelling in poor sanitary area. Hepatitis B refers to patients experiencing at least four of the listed symptoms after frequent contact with blood.

Table 2: FCM membership grade of all patients in all clusters of Hepatitis

CODES	**DEGREE OF MEMBERSHIP**	
	CLUSTER 1 (A)	**CLUSTER 2 (B)**
P01	0.60	0.40
P02	0.65	0.35
P03	0.25	0.75
P04	0.20	0.80
P05	0.62	0.38
P06	0.90	0.10
P07	0.17	0.83
P08	0.46	0.54
P09	0.77	0.23
P10	0.11	0.89

Each of the symptoms highlighted in Table 1 is represented with P (starting from 01 – 10 i.e. $P_{01} - P_{10}$). In addition, we form two clusters namely, Hepatitis A and Hepatitis B.

Table 2 represents a typical scenario of the degree of membership, for instance, P_{10} in cluster 1, we notice it has 0.11. In term of percentage it can be represented as 11% and in cluster 2, 89%. This means that the degree of symptoms of P_{10} matches 11% of Hepatitis A and 89% Hepatitis B.

The FCM clustering distribution (Figure 3) a graphical representation of Table 2 depicts a total of five symptoms with high degree of membership of Hepatitis A and five symptoms of Hepatitis B.

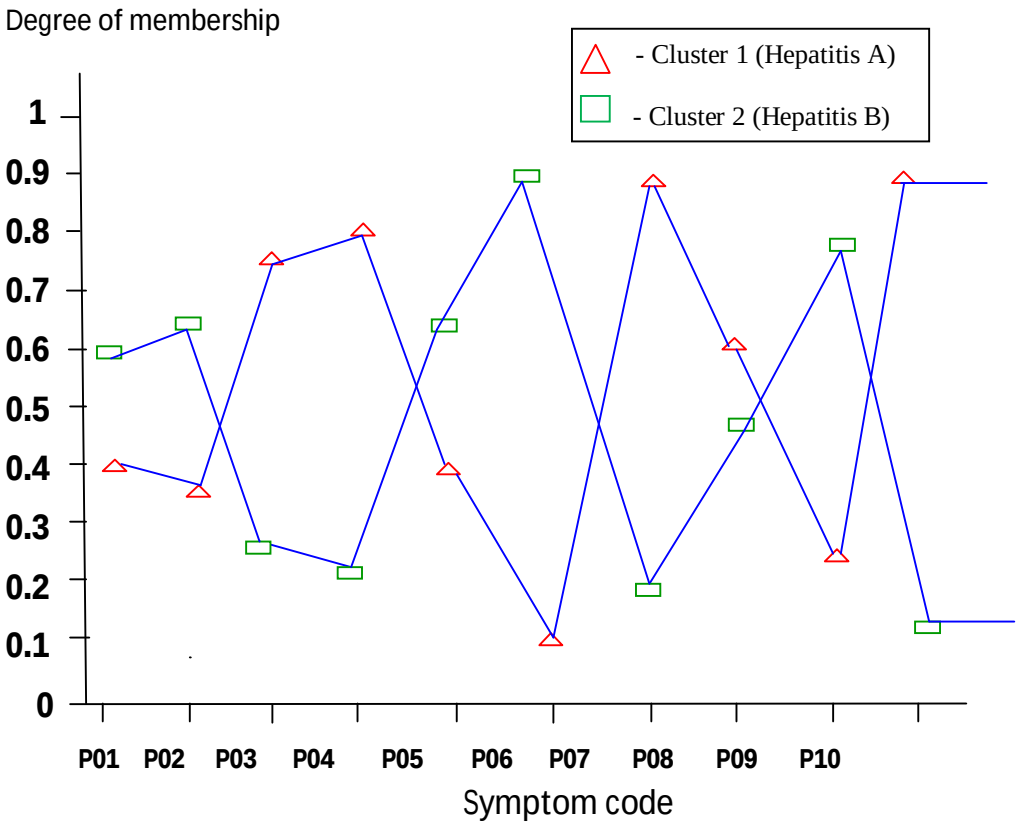

Figure 3: Graphical representation of Membership Grades of Hepatitis symptoms.

CONCLUSION

The need to design a system that would assist doctors in medical diagnosis has become imperative and hence cannot be over emphasized. This paper presents a diagnostic fuzzy cluster means system to help in diagnosis of Hepatitis using a set of symptoms. This advanced system which uses clustered data set is more precise than the traditional system. The classification, verification and matching of symptoms to the two groups of clusters (Hepatitis A and Hepatitis B) was necessary especially in some complex scenarios. This paper demonstrates the practical application of Information Technology in the domain of diagnostic pattern appraisal in medical by determining the extent of membership of individual symptoms. The model proposed allows for the classification of and matching of cluster groups to hepatitis symptoms. The fuzzy-cluster means model proposed in this paper appears to be a more natural and intelligent way of classification and matching of symptoms to hepatitis groups.

REFERENCES

Albayark S. and Amasyali F. (2003) “Fuzzy C-Means Clustering on Medical Diagnostic Systems”, International XII Turkish Symposium on Artificial Intelligence and Neural Networks.

Berks G., Keyserlingk D.G.V., Jantzeen J., Dotoli M. and Axer H. (2000): Fuzzy Clustering – A Versatile Mean to Explore Medical Database”, ESIT 2000, Aachen, Germany.

Giles D. E. A. and Draeseke R. (2001), "Economic Modelling Based on Pattern recognition via fuzzy C-Means clustering algorithm", Department of Economics Working Paper EWP0101, University of Victoria, Canada.

Healthline, (2011), "Hepatitis" retrieved from http:// healthline.com

Inyang U. G. (2005), "Neuro Fuzzy Expert System for Recognition of the Behavioural Patterns of some Selected Ethnic Groups in Nigeria", M.Tech Thesis, Department of Computer Science, Federal University of Technology Akure, Nigeria.

Medicinenet, (2011), "Hepatitis", retrieved from http:// hepatitis.com

Nascimento L. (1997), "A Genetic Approach to Fuzzy Clustering with a Validity measure Fitness function", retrieved from http://www.springerlink. com/content/ uw53x5444n320j29/.

Righthealth (2011), "Righthealth result for causes of hepatitis" retrieved from Read more: http://www.righthealth.com/topic/Causes_Of_Hepatitis#ixzz1KWw3hiau

Wrongdiagnosis (2011), "Hepatitis: Introduction, Causes and Symptoms", retrieved from wrongdiagnosis, com/l/ hepatitis/Introduction/causes.htm#symptoms_list.

Yang X. and Wang W. (2001), "GIS Based Fuzzy C-Means Clustering Analysis of Urban Transit Network Service", the Nanjing City Case Study, Road and Transport Research China.

Zadeh L.A. (1965): "Fuzzy sets and systems". In: Fox J, editor, System Theory. Brooklyn, NY: Polytechnic Press, 1965: 29–39.