Global Journal of Computer Science and Technology
Volume 11 Issue 12 Version 1.0 July 2011
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Expert System for the Intelligent Diagnosis of Hiv Using Fuzzy Cluster Means Algorithm

By Imianvan A. A.,Anosike U.F , Obi J. C.
University of Benin, Benin

Abstracts - Human Immunodeficiency Virus (HIV) is a retrovirus that causes Acquired Immune Deficiency syndrome (AIDS) by infecting helper T cells or Lymphocyte of the immune system. HIV is transmitted primarily by exposure to contaminated body fluids, especially blood and semen. Other means of transmission of HIV include sharing contaminated sharp objects and blood transfusion. HIV symptoms can include: headache, chronic cough, diarrhea, swollen glands, lack of energy, loss of appetite, weight loss, frequent fevers, frequent yeast infections, skin rashes, pelvic/abdominal cramps, sores on certain parts of your body and short-term memory loss. The focal point of this paper is to describe and illustrate the application of fuzzy cluster means system to the diagnosis of HIV. It involves a sequence of methodological and analytical decision steps that enhances the quality and meaning of the clusters produced. The proposed system eliminates theuncertainties often associated with analysis of HIV test data.

Keywords : *Fuzzylogic, Clustering, FuzzyC-Means, HIV.*

GJCST Classification : *I.2.1*

Strictly as per the compliance and regulations of:

An Expert System for the Intelligent Diagnosis of Hiv Using Fuzzy Cluster Means Algorithm

Imianvan A. A.[α], Anosike U.F[Ω], Obi J. C.[β]

Abstract - Human Immunodeficiency Virus (HIV) is a retrovirus that causes Acquired Immune Deficiency syndrome (AIDS) by infecting helper T cells or Lymphocyte of the immune system. HIV is transmitted primarily by exposure to contaminated body fluids, especially blood and semen. Other means of transmission of HIV include sharing contaminated sharp objects and blood transfusion. HIV symptoms can include: headache, chronic cough, diarrhea, swollen glands, lack of energy, loss of appetite, weight loss, frequent fevers, frequent yeast infections, skin rashes, pelvic/abdominal cramps, sores on certain parts of your body and short-term memory loss. The focal point of this paper is to describe and illustrate the application of fuzzy cluster means system to the diagnosis of HIV. It involves a sequence of methodological and analytical decision steps that enhances the quality and meaning of the clusters produced. The proposed system eliminates the uncertainties often associated with analysis of HIV test data.

Keywords : *Fuzzy logic, Clustering, Fuzzy C-Means, HIV*

I. Introduction

Human immunodeficiency virus (HIV) is a retrovirus (a virus whose genetic information is contained in Ribonucleic acid instead Deoxyribonucleic acid) that causes Acquired Immune Deficiency Syndrome (AIDS) by infecting helper T cells or Lymphocyte (cells that defense the body at foreign bodies) of the immune system.

Antigen are substances that stimuli the production of antibody. A serotype or serovar is a group of bacteria that share a characteristic set of antigen. The most common serotype or serovar, HIV-1, is distributed worldwide, while HIV-2 is primarily confined to West Africa (Healthline, 2011). AIDS is a severe immunological disorder caused by the retrovirus HIV, resulting in a defect in cell-mediated immune response that is manifested by increased susceptibility to opportunistic infections and to certain rare cancers, especially Kaposi's sarcoma (Healthline and MedicineNet, 2011). It is transmitted primarily by exposure to contaminated body fluids, especially blood and semen. Other means include sharing contaminated sharp objects and blood transfusion. Everybody who has AIDS also has HIV, but not everybody with HIV is classified by the United States (U.S.) government as having AIDS. The U.S. government uses CD4 cell counts (part of the immune system) to make this distinction (Healthline, 2011).

The earliest known case of HIV-1 came from a human blood sample collected in 1959 from a man in Kinshasa, Democratic Republic of Congo (healthline, 2011). The method by which he became infected is not known; however, genetic analysis of his blood sample suggested that HIV-1 might have stemmed from a single virus in the late 1940s or early 1950s. HIV has existed in the United States since the mid to late 1970s. During 1979 to 1981, rare types of pneumonia, cancer, and other illnesses were reported by physicians in Los Angeles and New York among a number of male patients who had sex with other men. Los Angeles and New York among a number of male patients who had sex with other men. Since it is rare to find these diseases in people with a healthy immune system, public health representatives became concerned that a new virus was emerging (Healthline, 2011).

In 1982, the term AIDS was introduced to describe the occurrences of opportunistic infections, Kaposi sarcoma, and pneumonia (*Pneumocystis carinii*) in previously healthy persons and formal tracking of these cases in the United States began that year. The virus that causes AIDS was discovered in 1983 and named human or helper T-cell (lymphotropic) virus-type III/ lymphadenopathy associated virus (HTLV-III/LAV) by an international scientific committee who later changed it to HIV (Healthline, 2011 and MedicineNet, 2011). Many theories as to the origins of HIV and how it appeared in the human population have been suggested. The majority of scientists believed that HIV originated in other primates and was somehow transmitted to man. In 1999, an international group reported the discovery of the origins of HIV-1, the predominant strain of HIV in the developed world (Healthline, 2011). A subspecies of chimpanzees native to west equatorial Africa were identified as the original source of the virus. The researchers believe that HIV-1 was introduced into the human population when hunters became exposed to infected blood (Healthline, 2011; MedicineNet, 2011 and WrongDiagnosis).

Most scientists believe that HIV causes AIDS by directly inducing the death of CD4+ T cells (helper T cells in the immune system) or interfering with their

Author : *Department of Computer Science, University of Benin, Benin City. Nigeria.*
E-mail[α]*: tonyvanni@yahoo.com, +234(0)07069742552;*
E-mail[Ω] *: uchenna.anosike@aiesec.net, +234(0)8038031536*
E-mail[β] *: tripplejo2k2@yahoo.com, +234(0)8093088218*

normal function and by triggering other events that weaken a person's immune function. For example, the network of signaling molecules that normally regulates a person's immune response is disrupted during HIV infection, impairing a person's ability to fight other infections. The HIV-mediated destruction of the lymph nodes and related immunologic organs also plays a major role in causing the immunosuppressant seen in persons with AIDS (Healthline, 2011).

In the absence of antiretroviral therapy, the median time from HIV infection to the development of AIDS-related symptoms has been approximately 10 to 12 years (Healthline, 2011 and WrongDiagnosis, 2011). A wide variation in disease progression, however, has been noted. Approximately 10 percent of HIV-infected persons have progressed to AIDS within the first two to three years after infection, whereas up to 5 percent of persons have stable CD4+ Tcell counts and no symptoms even after 12 or more years (Healthline, 2011). Factors such as age or genetic differences among persons with HIV, the level of virulence of an individual strain of virus, and co-infection with other microbes may influence the rate and severity of disease progression. Drugs that fight the infections associated with AIDS have improved and prolonged the lives of HIV-infected persons by preventing or treating conditions such as *Journal of Infectious Diseases*. This approach is known formally as short-cycle structured intermittent antiretroviral therapy (SIT) or colloquially as the "7-7" approach (Healthline, 2011).

HIV symptoms can include: headache, chronic cough, diarrhea, swollen glands, lack of energy, loss of appetite, weight loss, frequent fevers, frequent yeast infections, skin rashes, pelvic/abdominal cramps, sores on certain parts of your body and short-term memory loss (MedicineNet, 2011).

Existing methods of medical diagnosis employed by physicians for the analysis of HIV infection uses manual methods characterized by the inability to handle uncertain or vague data existing between intervals. More so, those systems are not self-learning or adaptive in nature. This paper has chosen to solve these problems by employing the rich facilities of fuzzy cluster means. The proposed system which is self-learning and adaptive, is a time-capsule (a cache of information) to be preserved for ages to medical engineers for the diagnosis and analysis of HIV infection.

II. Literature Review

Cluster analysis is a statistical techniques used to classify objects into coherent categories based on a set of measurement, indicators or variables. A common use of cluster analysis in medicine is to categorize patients into subgroups or diagnostic categories based upon patterns of clinical signs and symptoms, in this case HIV infection (Brian et al., 2001). Two-way clustering techniques are frequently used to organize genes into groups or clusters with similar levels of expression across relevant subgroup of patient's tissues, sample or cell lines (Eisen et al., 1998)

In practice, a cluster analysis is the product of a series of analytical decisions. The analytical decisions made at each point in the series can significantly affect subsequent decisions, as well as the overall result of a cluster analysis (Everitt et al., 2003). This series of analytic decisions typically involves choices about what objects to cluster, unit of measurement to use for the variables, proximity measure and criteria for determining the number and quality of clusters within the data.

Likert scale is the most popular psychological measurement schemes that depend on human judgment. This scaling scheme assures that the human observer is good at quantitative observation and assignment of number or objects to reflect degrees of traits or statement being measured (Cartwrigh, 2003 and Chow, 2002). In this scoring scheme, subjects are asked to choose exactly one alternative that describe their substance (Yuan, 2008). However, this scheme disregard with human thinking as multi-valued, transitional and analogue, but rather clear-cut (precise) and digital.

The invention and application of Fuzzy Cluster Means (FCM) algorithm in pattern recognition allows entities (objects) to belong to many clusters or categories with different degrees of membership (Yi-ouyang et al.; 2007). In this paper a framework for partitioning, which proposes a model of how data are generated from a cluster structure is presented. The Fuzzy Logic and Neural networks of personnel performance within organizations has been studied with a view of evaluating them for productivity and promotion (Akinyokun and Uzoka, 2004). The application of Fuzzy C-means (FCM) algorithm to medical diagnostic expert systems is presented in (Albayrak and Amasyali, 2003 and Berk et al., 2000). This algorithm is used in assigning patients to different cluster of disease. The application of fuzzy C-means in clustering has been demonstrated in (Yang and Wang, 2001; De Fazio and Galeazzi, 2004 and Jantzen , 1998). In this paper, fuzzy C-means algorithm is used to assign patients with HIV conditions to clusters of HIV.

Overview of Fuzzy C-Means Clustering (FCM)

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 fuzzy logic linguistic description of the typical FCM algorithm is presented in Figure 1

Start

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

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

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

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.

Stop

Figure 1 : Typical FCM Clustering Algorithm

III. Methodology

The process for the medical diagnosis of HIV starts when an individual consults a physician (doctor) and presents a set of complaints (symptoms). 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 medical 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 what is wrong with the patient. The physician then conducts a physical examination of the patient, studies his or her 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 (sought) 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 HIV is synonymous to the general approach to medical diagnosis. The physician may carry out a precise diagnosis, which requires a complete evaluation to determine whether the patient is having HIV. The examining physician accounts for possibilities of having HIV through interview, physical examination and laboratory test. Many primary health care physicians use screening tools for HIV evaluation.

A diagnostic evaluation of HIV 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?

IV. Results and Discussion

To design the FCM Knowledge Base System for diagnosis of HIV, we design a system which consists of a set of parameters needed for diagnosis (here, we are using 13 basic and major parameters) presented in Table 1.

Table 1 : Symptoms of HIV

Symptom. Codes	Symptoms of HIV (Human Immunodeficiency Virus)
P01	Headache
P02	Chronic cough
P03	Diarrhea
P04	Swollen glands
P05	Lack of energy
P06	Loss of appetite
P07	Weight loss
P08	Frequent fever
P09	Frequent yeast infection
P10	Skin rashes
P11	Pelvic/ abdominal cramps,
P12	Sores on certain parts of your body
P13	Short-term memory loss

Figure 2 presents the model of the FCM system for the diagnosis of HIV. It comprises of knowledge base system, fuzzy c-means inference engine and decision support system. The knowledge base system holds the symptoms for HIV. The values of the parameters are vague and imprecise hence the adoption of fuzzy logic as a means of analyzing these information. Those parameters therefore constitute the fuzzy parameter of the knowledge base. 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 jth parameter and n is the total number of parameter (in this case n=13). The set of linguistic values which is modelled as a linker scale denoted by 'L' is given as L = (Low, Average and High).

Clustering of the data is achieved using the typical FCM algorithm presented in Figure 2. Neural networks provide the structure for the parameters 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 inference mechanism is fuzzy logic driven. The cognitive filter of the decision support engine takes as input the output report of the inference engine and applies the objective rules to rank the individual on the presence or absence of HIV infection. The emotional filter takes as input the output report of the cognitive filter and applies the subjective rules in the domain of HIV studies in order to rank individuals on the extent of the HIV infection.

The expert system is developed in an environment characterized by Microsoft XP Professional operating system, Microsoft Access Database Management System, Visual BASIC Application Language and Microsoft Excel.

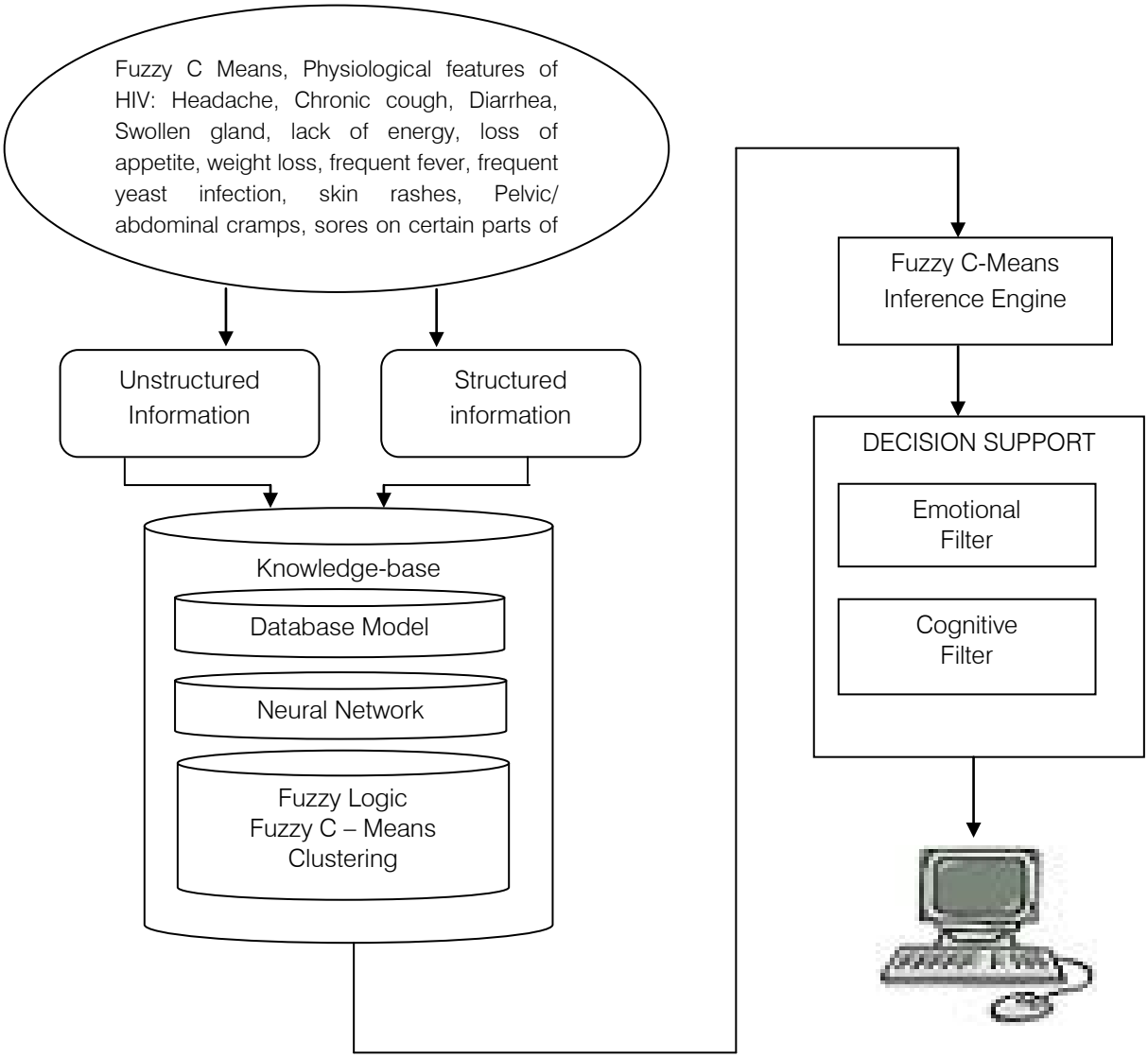

Figure 2 : Architecture of FCM Knowledge Base System for diagnosis/analysis of HIV

Table 2 : FCM membership grade table for Symptoms in all clusters Scale (0.00-1.00)

CODE	SYMPTOMS	DEGREE OF MEMBERSHIP		
		Cluster 1 (C_1)	Cluster 2 (C_2)	Cluster 3 (C_3)
P01	Headache	0.20	0.30	0.50
P02	Chronic cough	0.10	0.30	0.60
P03	Diarrhea	0.25	0.60	0.15
P04	Swollen gland	0.25	0.50	0.25
P05	Lack of energy	0.22	0.70	0.08
P06	Loss of appetite	0.20	0.80	0.00
P07	Weight loss	0.27	0.53	0.20
P08	Frequent fever	0.30	0.65	0.05
P09	Frequent yeast infection	0.10	0.80	0.10
P10	Skin rashes	0.56	0.44	0.00
P11	Pelvic/ abdominal cramps,	0.80	0.15	0.05
P12	Sores on certain parts of your body	0.50	0.37	0.13
P13	Short -term memory loss	0.70	0.15	0.15
RESULTS		Might be HIV Infected	HIV Infected	Not HIV Infected

A typical FCM membership grade table (Table 2) with 13 parameters and 3 clusters which shows the degree of membership of each parameter of HIV is represented using the graph in Figure 3.

From Figure 3, it is shown that there are no unitary (crisp) coefficients, indicating that each data point belongs to more than one cluster. For example the parameter " Frequent yeast infection" has its fuzzy set membership function as $\{0.1/C_1 + 0.8/C_2 + 0.1/C_3\}$ where C_1, C_2 and C_3 are clusters and in this study represents "Might be HIV infected", "HIV infected" and "Not HIV infected". This represents the degree of membership in terms of percentage as 10%, 80% and 10% match with "Might be HIV infected", "HIV infected" and "Not HIV infected" respectively.

Finally, Table 2 presents membership grades of parameters in all clusters whereas the degree of membership of the clusters is presented in Figure 3.

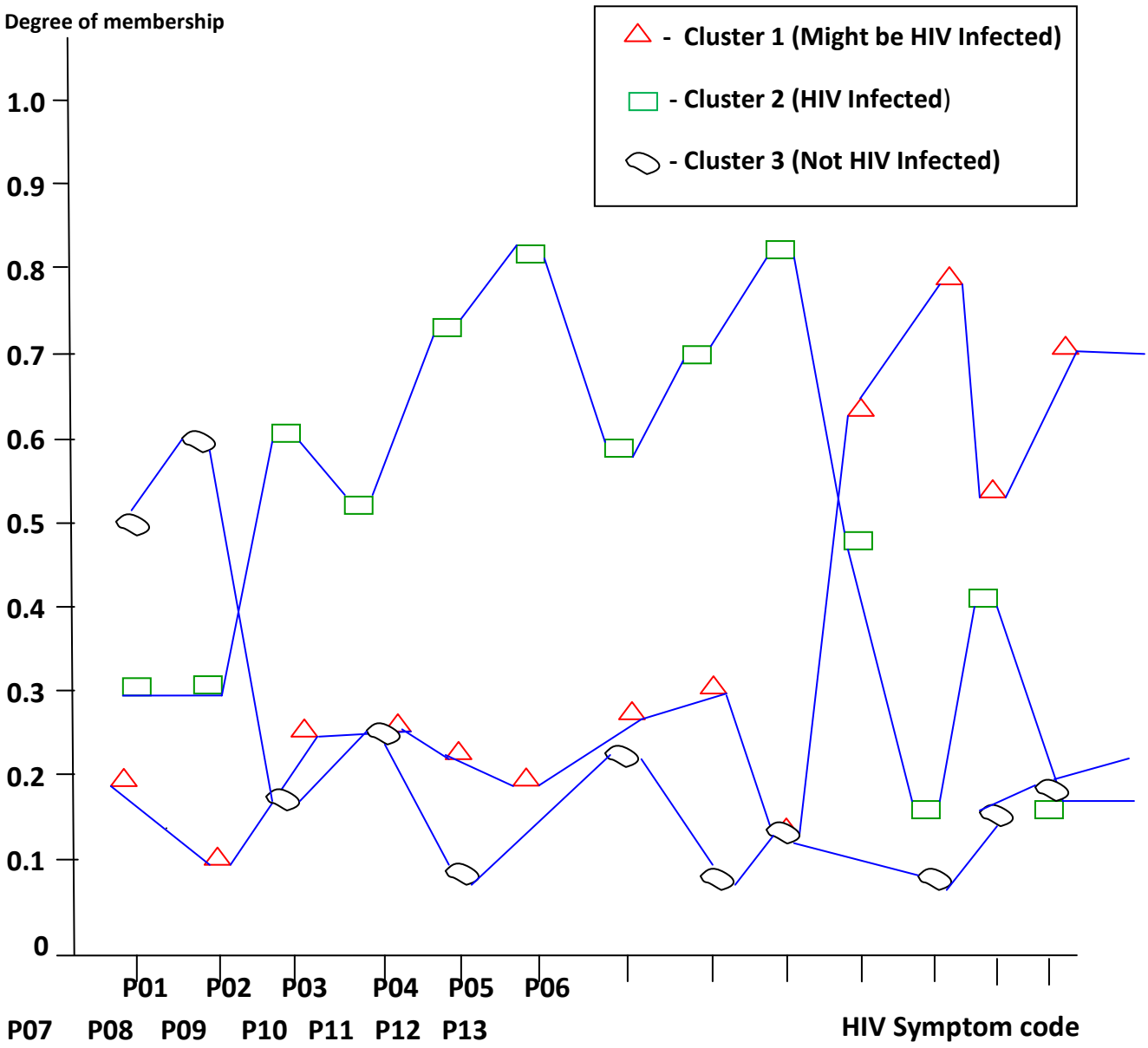

Figure 3 : Graphical representation of Membership Grades of HIV symptoms.

Figure 3 above shows the degree of membership of the various clusters. The parameters are grouped into various clusters using the degree of membership of each parameter to any particular cluster.

V. 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 present a diagnostic fuzzy cluster means system to help in diagnosis of HIV using a set of symptoms and demonstrates the practical application of ICT (Information and Communication Technology) in the domain of diagnostic pattern appraisal by determining the extent of membership of individual symptoms. This advanced system which uses a set of clustered data set is more precise than the traditional system. The classification, verification and matching of symptoms to the three groups of clusters was necessary especially in some complex scenarios. The fuzzy- cluster means system proposed and tested in this paper appears to be a more natural and intelligent way of classification and matching of symptoms to HIV.

References References Referencias

1. Akinyokun O.C. and Uzoka F. M.(2004), "Analysis of the Effect of HR profile on the Investment Portfolio of an Organization" retrieved from http//:research. mtropal.ca /research.
2. Albayrak S. and Amasyali M.F., (2003) "Fuzzy C-Means Clustering on medical Diagnostic system" retrieved from hhtp//: Ce.yildiz.edu ./r/en /

myindex.php
3. Berk et al. (2000), "Moving folded proteins across the bacterial cell membrane" Retrieved from mic.sgmjournals.org/./547
4. Brian E., Landau S. and Leese M. (2001), "Cluster Analysis" retrieved from http//: Amazon. co. uk/.../0340761199.
5. Cartwrigh N.L. (2003), "A Theory of Measurement", retrieved from www2.lse.ac.uk/philosophy /Newsand Events /.../theoryOfMeasurement.doc.
6. Chow S.L. (2002), "Methods in Psychological Research", retrieved from http:// *Cogprints.org / 2643/1/EOLSSrm.pdf*
7. De Fazio L. and Galeazzi G.M. (2004), "Women victim of stalking and helping Profession: Recognition and intervention in the modern group in stalking Retrieved from http//: stalking. medlegmo. unimo.it/gruppo.
8. Eisen M.B., Spellman P.T., Brown P.O., and Botstein D. (1998), "Cluster Analysis and Display of genome-wide expression pattern", Depart. Of Genetics, Stanford University School of Medicine, United Kingdom.
9. Everitt et al. (2003), "Data from the Nurses' Health Study on Women and the health Professional" retrieved from http//:biomedsearch.com/ attachment/00.
10. HealthLine (2011),"HIV diagnosis and symptoms", retrieved from http:// healthline.com.
11. Jantzen G. M. (1998), "Becoming divine: towards a feminist philosophy of Religion. Bloomington: Indiana University retrieved from http/: en.wikipedia. Org/wiki/grace-jantzen.
12. MedicineNet (2011), HIV: introduction/ diagnosis and symptoms", retrieved from http:// medicine Net.com
13. WrongDiagnosis(2011), "HIV epidemic: Introduction /symptoms and causes retrieved from http:// wrongdiagnosis.com.
14. Yang L. and Wang X.Z. (2001), "Lee circle threom for an ideal pseudopin-1/2 Bose gas In an external magnetic field" retrieved from http//:link.aps.org/./ Phys RevE.63.04
15. Yi-ouyang Y.; Yun-Lung Y. and AnDing Zu A (2007), "EHM-Based Web Pages Fuzzy Clustering Algorithm retrieved from http://www.computer.org /portal/web /csdl/doi/10.1109/MUE.2007.123
16. Yuan C.D. (2008), "A novel mutation (INS TCCG) in the TSC2 gene in a Chinese Patient" retrieved from http//:ncbi.nlm.nih.gov/pub-med/17888633.

Global Journals Inc. (US) Guidelines Handbook 2011

www.GlobalJournals.org