ONLINE FUZZY-LOGIC KNOWLEDGE WAREHOUSING AND MINING MODEL FOR THE DIAGNOSIS AND THERAPY OF HIV/AIDS

Igodan C. E.[1] Akinyokun O.C.[2] and Olabode Olatubosun[3]

[1]Department of Computer Science, University of Benin, Benin City, Nigeria
[1] charlesigodan@yahoo.com
[2,3]Computer Science Department, Federal University of Technology, Akure, Nigeria
[2]akinwole2003@yahoo.co.uk, [3]olabode_olatubosun@yahoo.co.uk

ABSTRACT

The need for a reliable and efficient way of storing and mining data about people living with HIV/AIDs with the intent to monitor the health status for effective therapy is on the increase. This paper presents a model of a web-based system for knowledge warehousing and mining of diagnosis and therapy of HIV/AIDs using Fuzzy Logic and data mining approach. A model was developed, using the predictive modeling technique, for predicting HIV/AIDs and monitoring of patient health status. The fuzzy inference rule and a decision support system based on cognitive filtering was employed to determine the possible course of action to be taken. A case study of some data of PLWH was used and the result obtained shows that the developed system is efficient. The system uses XAMP on Windows OS platform. The system was tested and evaluated with satisfactory results.

KEYWORD

Knowledge warehousing, data mining, HIV/AID, Fuzzy logic

1. INTRODUCTION

Medical diagnosis is the ability to form valuable opinions and make good decisions about the condition of a patient, while therapy is the administration of drugs to patients. The major problem in the medical field is the diagnosis of disease which is due to human error [1]. In [2] a medical diagnosis involves the state space search of medical knowledge of a particular disease, patient history, patient examination and drug. Medical diagnosis and therapy involves the determination of the nature of a disease from observation of signs and symptoms and subsequent curative treatment of such a disease. These are the work of a medical expert (Physicians) who possesses a wealth of knowledge in the field. According to [3], the African AIDS epidemic continues to pose severe public health and developmental challenges for many African nations. A primary impediment in the fight against AIDS is lack of information and communication about the disease. Information and Communications Technology (ICT) hold vast potential to hinder the spread of the disease, as well as key elements of all aspects of HIV/AIDS strategies, including prevention, treatment, care and protection to human rights. In [3], the role of information and communications in the fight against AIDS in the African continent. It is intended to increase

understanding of how information and communication can offer key resources in relation to HIV/AIDS, and to suggest key actions information and communications can take to assist in combating HIV/AIDS.

In [4], it is observed one of the many challenges of access to treatment through anti-retroviral medication is the difficulty of receiving accurate and up-to-date information at the planning level. It was noted that Information and Communication Technology (ICT) has been used in a number of pilot studies to address this bottleneck, by providing an infrastructure for telemedicine and e-health applications. It was also suggested that Anti-Retroviral Treatment (ART) delivery to poor communities can be significantly improved by integrating local primary health-care information with centralized databases to allow national monitoring.

In [5], a framework for the interventions to close the digital divide in AIDS care was developed. It describes the development and pilot testing of a workshop-style intervention designed to close the digital divide in AIDS care. [6], evaluated the efficacy of LifeWindows, a theory-based, computer-administered antiretroviral (ARV) therapy adherence support intervention, delivered to HIV plus patients at routine clinical care visits. In that study, 594 HIV plus adults receiving HIV care at five clinics were randomized to intervention or control arms. The result indicated that intervention versus control impact in the intent-to-treat sample - including participant who's ARVs had been entirely discontinued, who infrequently attended care, or infrequently used Life Windows - did not reach significance. Intervention impact in the On Protocol sample (328 intervention and control arm participants whose ARVs were not discontinued, who attended care and were exposed to Life Windows regularly) was significant. On Protocol intervention various control participants achieved significantly higher levels of perfect 3-day ACTG-assessed adherence over time, with sensitivity analyses maintaining this effect down to 70% adherence. That study supports the utility of Life Windows and illustrates that patients on ARVs who persist in care at clinical care sites can benefit from adherence promotion software.

A software engine for the mining and processing, periodic update of the knowledge warehouse and statistical analysis of findings was developed in [7] for NITDA on its intervention in capacity building for Nigerian tertiary institutions and by extension the education sector of Nigeria economy. Infection with HIV is now endemic in several African countries; estimates of the number of infected people is in millions. Zero prevalence rates of HIV infection vary from less than 1% up to 20% of the general adult population, and from 27% to 88% of female prostitutes in the cities of Central and East Africa. It is evident that, in many parts of the continent, HIV infection has created a major public health problem, which may have reached the same scale as malaria, diarrhea, respiratory diseases, and malnutrition. In one major hospital in Central Africa, up to 35% of the children and adults admitted having were HIV infection, and this figure is likely to increase. Thus, many African clinicians are taking care of HIV-infected patients, and they will be confronted with an increasing number of AIDS cases in the near future.

A fuzzy rule-based framework for the management of tropical diseases with main focus on malaria is presented in [8]. The system involves fuzzzification, inference and defuzzification. The system is able to diagnose malaria case as either mild, moderate, several or very several. The symptoms and signs of the diseases were fuzzified. With the help of fuzzy rules developed and the assistance of some medical personnel, the fuzzified variables were composed, inferred and later defuzzified to give final diagnosis. [9], implemented an evidence based medicine principles via a computerized clinical decision support system (CDSSs). In [10] a model of asynchronous collaboration between people with very different knowledge of medicine in which a computer framework attempts to mediate between patients and physicians to help reduce some of the differences in communication was designed.

[11], developed a medical consultation system, CADIAG-4, for general internal medicine, which is intended to assist the differential diagnostic process by indicating all possible diseases which might be the cause of patient's pathological findings, with special emphasis on rare diseases. It uses the formal framework of fuzzy set theory and fuzzy logic to represent the uncertainty in medicine. In [12], a fuzzy expert system for heart disease diagnosis was designed. The system has 13 input fields, one output field and rule base. The mamdani inference method – fuzzy expert system designing, membership function, fuzzy rule base, fuzzification and defuzzification methods were used. [13], focused on the use of ICT to design a web-based fuzzy expert system for the management of hypertension using the fuzzy logic method. In the study, systolic blood pressure, diastolic blood pressure, age and Body Mass Index (BMI) were taken as input parameters to the fuzzy expert system and hypertension risk as output parameter. The resultant hypertension risk was based on fuzzy rules that were developed for the expert system. The defuzzification method was Root Sum Square.

This paper attempts to develop a fuzzy-logic, knowledge warehouse and mining model for HIV/AIDs diagnosis, therapy and preventive measure. Data were sourced from the University of Benin Teaching Hospital.

2.   SIGNS, SYMPTOMS AND CLASSIFICATION OF HIV/AIDS

According to [14], AIDS in an adult is defined by the existence of at least two major signs associated with at least one minor sign, in the absence of known causes on immunosuppression such as cancer, severe malnutrition or other recognized etiologies. However, the presence of generalized Kaposi sarcoma or cryptococcal meningitis is sufficient by itself for the diagnosis of AIDS. The major signs are:

a.   Weight loss   10% of body weight
b.   Chronic diarrhea (for more than 1 month)
c.   Prolonged fever (for more than 1 month) (intermittent of constant)

The minor signs are:

a.   Persistent cough (for more than 1 month)
b.   Generalized pruritic dermatitis
c.   Recurrent zoster
d.   Oropharyngeal Candidiasis
e.   Chronic progressive and disseminated herpes virus infection
f.   Generalized lymphadenopathy

[14] and [15], it is opined that epidemiological studies have demonstrated that HIV has three major modes of transmission: sexual, parenteral, and perinatal. No evidence exists to support any other routes of transmission, including through food, water, vectors, or casual contact. HIV has been identified in the whole blood, serum, tears, saliva, cerebrospinal fluid, brain tissue, urine, semen, breast milk, amniotic fluid, and vaginal/cervical secretions. However, the low titre of virus and infrequency of virus isolation from most of these fluids suggest that with the exception of blood, serum, semen, cervical/vaginal secretions, and breast milk, they play an insignificant role in HIV transmission.

In the National Guideline for HIV and AIDS Treatment and Care in Adolescent as reported in [16], diagnosis of HIV infection is based on the demonstration of antibody in plasma or serum, and of virus in the blood. The virus can be demonstrated in the blood with nucleic acid-based tests

(PCR for proviral DNA and RT-PCR for plasma viral RNA), culture and p24 antigen assay. With the technology that is available at present, HIV antibodies are detectable within four to six weeks of infection, and within 24 weeks in virtually all affected individuals. However, virus can be detected in plasma at least one week earlier. This period of absent antibody in the presence of virus in plasma is called the "window period".

Antiretroviral Therapy has the following components:

a.   Pre-treatment Evaluation includes:

   i.   Complete history and physical examination (always think of Pregnancy, Anaemia, Hepatitis and TB)
   ii.   Clinically and Immunological classification of the patient
   iii.   Check Laboratory results (FBC with differential, ALT, Creatinine, CD4+ cell count, pregnancy test)
   iv.   Evaluation of nutritional and psychosocial status
   v.   Assessment of readiness for therapy
   vi.   Development of patient-specific adherence strategy

b.   Criteria for ART
   Initiation of therapy depends on CD4+ cell count and WHO clinical staging:

   i.   WHO Stage IV disease irrespective of CD4+ cell count
   ii.   WHO Stage III disease with CD4+ cell counts $< 350/mm^3$
   iii.   WHO Stage I or II disease with CD4+ cell counts $< 200/mm^3$
   iv.   Any WHO Stage with CD4+ cell count 200-350 mm^3 – consider for therapy

c.   Counseling for ART

   There are many patients who know their HIV sero-status through VCT and other testing services but are yet to consider starting ART. However, when a decision is reached to commence ART, additional counseling is required to address the following issues:

   i.   ARVs do not offer a cure. HIV may be suppressed but is not eradicated from the body
   ii.   Use of ARVs is associated with impaired quality of life and long term survival
   iii.   ARVs need to be taken daily for life to prevent development of resistance and treatment failure
   v.   ARVs, like other medications, are associated with side effects
   vi.   Better results are obtained with good adherence to the treatment regime
   vii.   Some patients may fail to respond to treatment and may required changes in their treatment regime [16]

A Data Warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of relational/ multidimensional database optimized for decision support which is separated from the database [17], [18], [19]. It is an information repository in which components (knowledge components) are cataloged and stored for reuse [20], and has logical structures to store knowledge that are analogous to the system tables that implement data storage in the data warehouse [21]. The architecture of the knowledge warehouse and mining of diagnosis and therapy of HIV/AIDS consists of the Knowledge base, Database model, Weighted Averaging Method, Fuzzy logic, Inference engine, Decision Support Engine, Cognitive filter and Emotional filter as presented in [7].

A three tier web based system which has the following elements is proposed:

a.   **Client-Tier:** Provides a Graphical User Interface (GUI) to the end-users, which is achieved using a web browser, such as Internet Explorer, built using HTML with PHP.

b.   **Web-Tier:** Provides the dynamic content to users; presents the data model to clients; accepts and analyze user inputs; presents users input to ES for processing and forwards responses back to user.

c.   **Database:** Stores the object data and handles system logic. It deals with systems rules and data.

The database model considered for the proposed system employs an object-oriented approach and is viewed as a network of semantically related structured and unstructured knowledge of the HIV/AIDS and its diagnosis. Relational database modeling technology is adopted [22]. The general form of a relation is given by:

$$R[a_i, a_{i+1}, a_{i+2}, a_{i+3}, a_{i+4} , \dots, a_{n-1}, a_n,]$$

Where R represents the name of the relation, the set $\{a_i\}$, i = 1, 2… n, represents the attributes, properties and decision variables of R. The database is composed of nine relations and some presented below

Patient [patientNo, patientName, PatientAddr, patientAge, patientSex, Nationality, MaritalStatus, BloodGroup, NameAddrofNextofKin

Signs [PatientNo, diseaseID, description-of-sign]

Symptoms [PatientNo, symptomID, description-of disease]

PatientHist [PatientNo, datehist, timeHist, HistLog, patientAddr, occupation, sex**]**

LabTest [date, patientNo, description-of-lab-test]

HIVDIAGNOSIS [patientNo, Diag-No, Diag-Log**]**

PatientConsult [patientNo, medical-Pract, dataofConsult**]**

The activities of the user on the database are described as transaction. A transaction may be classified as an update or query type:

a.   **Update Transactions:** The update transactions are the activities that are responsible for the insertion of new records, modification of existing records and deletion of unwanted records in the database. Update transactions on any database may be carried out by the Database Administrator (DBA) or some other privileged users. The DBA will be responsible for the initial creation of the database and modification of the database from time to time to reflect the current state exhibited in real life. The database is the visual image of real life of HIV/AIDS and diagnosis of the disease. The database should, therefore, constantly exhibit the true position of current trends and real life possibilities in HIV/AIDS and the diagnosis of the disease, to maintain integrity and credibility. The security and confidentiality of the database can be guaranteed if appropriate right of human expert access and update database is adequately provided, controlled and monitored. In this study, the right of privileged users to access and update database is provided, monitored at the following three levels: Insertion of new records, Modification of old records, Deletion of unwanted (see Figure 2.1)

b.   **Query Transaction:** query transactions are activities that are responsible for questioning the database and generate useful reports. In practice, query transactions are carried out on a database by any authorized user of the database in the category. Database expert users, Database end user, database casual users.

In order to create the security of the database and the confidentiality of the transactions carried out on it, each category of users is provided with the appropriate right into the

database. In the web based system proposed in this study, only the first two access levels are defined because all end users can query the database and also perform update transaction if they are interested.
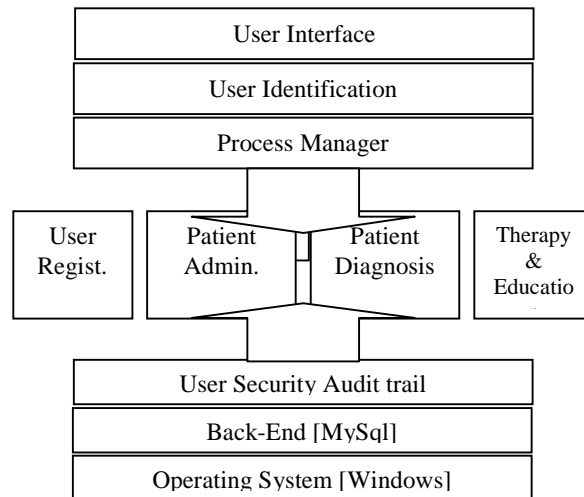

Figure 2.1. Data Processing Module

Algorithm for processing some transactions

* Display general information about HIV/AIDS
Open the File
SELECT About-HIV/AIDS File
    FROM File
        Filter out all information about HIVAIDS
        If there is any   Display
        Else Return current page
        Endif
    Close File

* Display a page for prototype diagnosis for HIV/AIDS and proffers possible advice based on results
Open the File
    SELECT Diagnostic File
        FROM File
        Enter Patient Variables
Perform Fuzzfication to generate Fuzzy value
Click Defuzzify to generate crisp values
if crisp value is POSITIVE
    Display information about the level of infection and possible Advice
Elseif Value is Negative
    Display information about how to prevent HIV/AIDS and stay safe
Else "Indeterminate"
    Display information on Advice to retake the test in few months time
End if
Close File

2.1   Data Warehouse and Mining

This stage involves the conception, acquisition, cleaning and loading of data into the data warehouse. Specifically, patient's record was collected from the Introviral Human Virology, Nigeria (IHVN), U.B.T.H. The data was carefully analyzed. The analysis involves demographics, study results, problems, therapies, allergies, subjective and physical findings, and encounter summaries. The data warehouse was created on a centralized server dedicated to fielding data mining queries. The clinical data was mapped from data structure into relational tables in the personal computer environment. MySQL was chosen as the Database Engine installed into the Computer System. For the purpose of this study, a few sample dataset from the data warehouse was created that was mined for knowledge discovery. Multiple SQL queries were run on the data warehouse to create the dataset. As each variable is added to the dataset, it is cleansed of erroneous values, data inconsistencies and formatting discrepancies.

In this study, the predictive model was used to mine the data. The predictive models use data with known results to develop a model used to explicitly predict values for different data [23]. For example, a model can be built using the payment history of people to whom you have given loans to help identify people who are likely to default on loans. The model in this study was built from the patients' history which helped to identify the various classes and ranges of infections about patients. Using the Classification models requires you to identify those characteristics of cases that indicate to which group each case belongs. This pattern can be used both to understand the existing data and to predict how new instances will behave. The Data mining creates classification models by examining already classified data (cases) and inductively finding the predictive pattern. In some data mining problems, an expert classifies a sample of the database and this classification is used to create the model which will be applied to the entire database.

2.2   Fuzzy Logic for Diagnosis of HIV/AIDS

The logic of the diagnosis and therapy of HIV/AIDS involves the following phases: Fuzzification, Fuzzy Inference (Fuzzy Rules) and Defuzzification. Fuzzy logic requires knowledge in order to reason. The knowledge is provided by the expert who knows the process and stores it in the fuzzy system. The knowledge repository for HIV/AIDS contains both the qualitative and quantitative decision variables. These variables are used in the fuzzification, inference, and defuzzification as discussed in the following sections.

Fuzzy Rule Base

The rule base is defined according to expert's experience. As used in this research work, Figure 2.2 describes the action to be taken given the states of the inputs variables. Fuzzy logic offers a possibility to update the knowledge base continuously [24] and by this, previous diagnosis is improved. After proposition (rules), the next phase is to perform defuzzification in order to evaluate the crisp output.

The fuzzy logic analysis and control method involves:

a.   Receiving of one, a large number, of measurement or assessment of conditions existing in some system we wish to analyze.
b.   Processing all these inputs according to human based, fuzzy "*If-Then" **rules***, which can be expressed in plain language words, in combination with traditional non-fuzzy processing.

c.   Averaging and weighting the resulting outputs from all the individual rules into one single output decision or signal which decides what to do or tells a controlled system what to do.

The output signal eventually arrived at is a precise appearing, defuzzified, "crisp" value. Fuzzy Logic Control/Analysis Method diagram is shown in Figure 2.2.
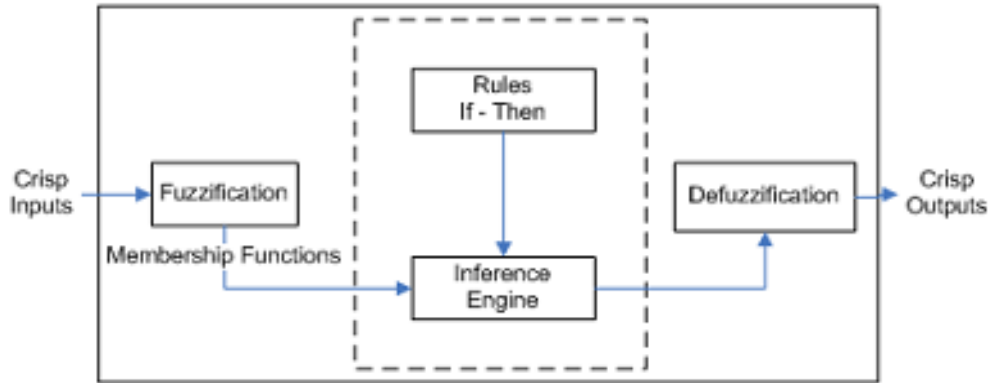

Figure 2.2: Fuzzy Logic Control System

Defuzzification is the process of producing a quantifiable result in fuzzy logic, given fuzzy sets and corresponding membership degrees. Defuzzification is interpreting the membership degrees of the fuzzy sets into a specific decision or real value. In this study, the center of gravity (centroid method) is used for defuzzification given in equation (1):

$$y = \frac{\sum_i^k \mu(x_i)x}{\sum_i^k \mu(x_i)}$$ …………………………………………………………….. (1)

where: x_i represent the ith. symptoms, k is the number of items in the fuzzy set and y is the crisp value.

The average is probably the easiest and the most widespread solution to aggregate information. The average is defined as the single value which all the individuals of a population should have so that their total is unchanged. When the sample contains several times the same individuals, it is possible to reformulate the average as a weighted average by introducing coefficients related to each individual. So a weighted average is an average in which a weight, w_i, is assigned to each quantity x_i to be averaged. The weights determine the relative importance of each quantity on the average [25].

A physical interview session with some medical doctors working closely with patients living with the disease was carried out to gather information about the various symptoms of the disease so as to enable the construction of the analytical membership function shown below. Some of the diseases are measured by days, weeks or months to ascertain the level of severity, while some are measured in degrees centigrade. Also some of these symptoms are quantitative while others are qualitative in nature. Each symptom has its membership function constructed and represents its degree of severity denoted with letter "X". Using set theory, the symptoms are represented as: {X | X is a set of symptoms}

Where x_1=chronic-cough, x_2=diarrhoea, x_3=swollen-gland, x_4=weight-loss, x_5=frequent-fever, x_6= frequent-yeast-infection, x_7=skin-rashes, x_8=memory-loss, x_9=body-sore, x_{10}=severe-headache, x_{11}=energy-loss, x_{12}=appetite-loss, x_{13}=pelvic-abdominal-cramp and $x_1, x_2, x_3,$ …, x_{13} X

The analytical representation of membership functions for the various input parameters to the fuzzy expert system are listed

$$low \quad x = \begin{cases} 0, & x<7 \\ x/14, & 7 \le x < 14 \\ 1 & 14 \le x < 20 \end{cases} \quad moderate \quad x = \begin{cases} 0, & x=3 \\ x/5, & 3 \le x < 5 \\ 1, & x=5 \end{cases} \quad high \quad x = \begin{cases} 0, & x=6 \\ x/6, & 6 \le x < 11 \\ 1 & x \ge 11 \end{cases}$$

The input variables *for swollen gland* is measured in centimeters (cm), *chronic cough* and *diarrhea* are measured by the number of days/weeks/ or months of infections.

The analytical representation of membership functions for the ***weight loss*** input parameters are represented:

$$low \quad x = \begin{cases} 0, & x<1 \\ x/2, & 1 \le x < 2 \\ 1, & x=2 \end{cases} \quad moderate \quad x = \begin{cases} 0, & x=3 \\ x/5, & 3 \le x < 5 \\ 1, & x=5 \end{cases} \quad high \quad x = \begin{cases} 0, & x=6 \\ x/6, & 6 \le x < 11 \\ 1, & x \ge 11 \end{cases}$$

The analytical representation of membership functions for the ***frequent fever*** (measured in ^{o}C degree) input parameters to the fuzzy expert system is as presented:

$$x = \begin{cases} 0, & x<6 \qquad low \\ x/37, & 11 \le x < 37 \quad moderate \\ 1, & x \ge 37 \qquad severe \end{cases}$$

The analytical representation of membership functions for the ***frequent yeast infection, Skin rashes, Short-term memory and Body sores*** (measured in days/weeks or months of infections) input parameters to the fuzzy expert system is presented:

$$low \quad x = \begin{cases} 0, & x<7 \\ x/12, & 7 \le x < 12 \\ 1, & 12 \le x < 20 \end{cases} \quad moderate \quad x = \begin{cases} 0, & 20 \le x < 25 \\ x/30, & 25 \le x < 30 \\ 1, & 30 \le x < 35 \end{cases} \quad high \quad x = \begin{cases} 0, & 35 \le x < 40 \\ x/50, & 40 \le x < 50 \\ 1 & x=50 \end{cases}$$

The analytical representation of membership functions for the ***Headache, Lack of Energy, Loss of appetite and Pelvic/Abdominal Cramps*** (measured by doctor's physical diagnosis of infections) input parameters to the fuzzy expert system is as presented:

$$low \quad x = \begin{cases} 0, & mild \\ 1, & severe \end{cases}$$

The input fields of the following membership functions above are just two values (0, 1) sets. Value '0' means that the illness is absent while '1' meaning the existence of the illness in the patients. The membership function shapes for the input/output variables were elicited from the knowledge acquisition sessions with a medical expert in UBTH. The range of fuzzy values for

HIV/AIDS risk is shown in Table 1. The membership function of negative and positive are trapezoidal and membership function of Indeterminate is triangular. In this study, the indeterminate fuzzy set explains the inability to detect the AIDS virus because the patient is in his 'window period', so a retest is advised after 3 months. The window period is the period of absent antibody in the presence of virus in plasma [16].

Table 1. Range of fuzzy values for HIV/AIDS risk

Output	Ranges	Fuzzy Sets
Results	0.0-0.50	Negative
	0.51-0.52	Indeterminate
	0.53-1.00	Positive

Fuzzy logic resembles human decision making process which allows the use of approximate data to find precise solution. It allows one to express knowledge with subjective concepts as "very severe", "severe", and "mild" which are mapped into exact numeric ranges which can be described as the membership function of a fuzzy set for the classification of pattern. For each quantitative and qualitative variable which do not allow for a precise outcome, a psychometric scale is drawn to measure its total solution space which is bounded by the smallest value to the largest allowable value. Given a solution space for the attribute of "HIV/AIDS risk" using a psychometric scale defined within the range of 1 to 10. This attribute could be broken down to three fuzzy sets "Negative", "Indeterminate", "Positive". The fuzzy sets are not necessarily symmetric because the ranges of values they cover are not necessarily equal but always overlapping to some degree. This is because a value may be a member of more than one set. The inference procedure employs the averaging method of drawing inference. In this procedure, each function is clipped at the average and the "fuzzy" centroid of the composite area is computed.

Most of the therapies are qualitative in form, thus they form fuzzy and linguistic variables. The quantitative variables of the drug therapy can be transformed to fuzzy and linguistic variables taking the minimum and maximum quantity of drugs prescribed by the system. The CD4+ T-Cell count helps to monitor disease progression and ART therapy. In this study, the patient's histories are mined to determine the stages of the HIV/AIDS treatment that should be and therefore the drugs regimen are prescribed according the Doctors prescription.

The Antiretroviral Therapies include the following:

a. Pre-treatment Evaluation
b. Initiation of ART: Initiation of therapy depends on CD4+ T-Cell count and WHO clinical staging.
c. Counseling for ART

These issues should be thoroughly discussed by the Counselor or any health worker who is directly involved with the patient. They should also do follow-up visits whenever the opportunity arises [16]. Thus, there is a very dynamic situation in HIV-infected people involving continuous viral replication and destruction and replacement of CD4+ T cells. While the CD4+ T-cells count is a less expansion and less technical measure of HIV disease progression, quantifying the viral load is currently the most direct measurement of the HIV disease process. It has also been used to assess the risk of disease progression and the response to antiretroviral therapy (ART). As the disease progresses, CD4+ T-cell count declines but may rebound if therapy is efficacious; however, this parameter alone is an incomplete marker for clinical assessment of a patient. Nevertheless, in resource poor settings (which include a large proportion of the most affected

countries), the CD4+ T- cell count is a more affordable and hence more practical yardstick for monitoring disease progression and ART efficacy [16]. The CD4+ T-cell count in table 1 shows the various stages of HIV disease progression:

Table 2.2: CD4 T-Cell Count

Stages of Disease	CD4+ T-cell Count	Interpretation
Acute Sero-conversion Syndrome	>1000/mm^3	Low
Early Disease	>500/mm^3	Mild
Middle-Stage Disease	200-500/mm^3	Moderate
Late Disease	50-200/mm^3	Severe
Advanced Disease	<50/mm^3	Very Severe

The therapy datasets are fuzzified using the following functions with the linguistic variables: Low, Mild, Moderate, Severe and Very Severe. The analytical representation of membership functions of the therapy for HIV/AIDS is as presented below:

$$x = \begin{cases} \geq 1000, & \text{low} \\ 500 < x \leq 1000, & \text{mild} \\ 200 < x \leq 500 & \text{moderate} \\ 50 < x \leq 200 & \text{severe} \\ \leq 50 & \text{very severe} \end{cases}$$

There are four classes currently available for treatment based on the site and mechanism of action. Other classes are at various stages of development but are not yet widely available for clinical use. Different classes and their mechanisms of actions:

a.   Non-nucleoside Reverse Transcriptase Inhibitor (NNRTIs)
b.   Nucleoside Reverse Transcriptase Inhibitor (NRTIs)
c.   Nucleotide Reverse Transcriptase Inhibitor (NtRTIs)
d.   Protease Inhibitors (PIs)

Some of the new classes include the Entry Inhibitors which prevent the virus from gaining access into the cytoplasm of the T-cell. There are categories of these inhibitors:

a.   The Chemokine receptors antagonists
b.   Attachment inhibitors
c.   Fusion inhibitors

According to the [16], it is recommended that combinations of drugs from at least two different classes be used so that drugs can act on at least two different points in the HIV life cycle. Monotherapy or dual therapy is not recommended for treatment as the risks of resistance development are greater. The more the number of sites on the life cycle of the virus the drugs act, the less the likelihood of development of drug resistance. A minimum of drug including an NNRTI or a PI is typically used. Based on availability, efficacy and ease of administration, the following are recommended:

Preferred first line Regiment
ZDV or TDF + 3TC or FTC + EFV or NVP
Other first line Option
ABC or D4T + 3TC or FTC + EFV or NVP
where ZDV = Zidovudine, TDF = Tenofovir, 3TC = Lamivudine, FTC = Emtricitabine,
EFV = Efavirenz, NVP = Nevirapine, ABC = Abacavir and D4T = Stavudine

The Decision Support System module is used to fine-tune and filter results from the data mining engine. This consists of two sub-components namely:

a.   **The Cognitive Filter**: this sub-module helps to analyze the alternative output reports of the Data Mining engine using the objective analyses presented to it by the Data Warehouse, and
b.   **The Emotional Filter:** helps to analyze the output reports of the Cognitive filter on the basis of the subjective feelings of the Data Warehouse.

The Decision Support System also helps to construct the unstructured information which have been passed to the data mining engine and have been processed via a "Keyword Analysis" to determine the degree of relevance to the search query. Also, text analytics technique which is a process of text mining is used to construct relevant data from unstructured information. When the results have been adequately filtered by the Decision Support System, they are passed to the user through an Interface. The end of the DSS is the User Interface which allows the user to interact with the system by specifying a data mining query or task, providing information about processes and data warehouse schemas or data structures, evaluate alternative patterns of the data mining, and visualize the patterns in different forms like tables, Cross Tabs, graphs etc.

3.   CONCLUSION

HIV/AIDS is one of the greatest health pandemic in the world over. The social, moral, economic, psychological and mental state of a person who has the disease can be completely destroyed due to stigmatization from friends and families except well managed through social therapy. The continuing spread of the HIV/AIDS pandemic is mostly as a result of ignorance and lack of acceptance of the fact of the existence of the disease amongst people. The disease is mostly transmitted through sex, use of infected needles, blood transfusion and contact with infected blood.

Today's Internet technology is helping in the dissemination of information about the existence and realities of the disease; and is also a tool for creating awareness, educating, and helping to reach out to large number of people on how to control and stop the spread of the disease. Apart from the use of the Internet, other media like television, radio, tabloid and seminars are also being used.

The study has been able to make use of open source tools to develop a system which provides an easy access and storage of large database of people living with HIV/AIDs which are mined so as to enable prediction of the disease. The system also helps educate about the diagnosis, symptoms, curative measures and prevention of the disease. Also, data gathered from people living had been analyzed for effective decision support for the control and management of the disease. The essence of the analysis is to further create awareness and foster major stakeholders in curbing the spread of the world's most dreaded disease.

REFERENCES

[1]   Norsarini, S (2004), "Medical Diagnosis Using Neural Networks", Faculty of Information Technology Universiti Utara Malaysia, Sintok, Kedah. Retrieved 28th April, 2011 from http://www.generation5.org/content/2004/MedicalDiagnosis.asp

[2]   Akinyokun, O.C, and Adeniji, O.A (1991), "Experimental study of Intelligent Computer Aided Medical Diagnosis and Therapy", Modeling, simulation & Control, C, AMSE Press Vol. 27, No 3, 1991, pp.1-20

[3]   Johnstone, J. (2005), "Knowledge, Development and Technology: Internet Use Among Voluntary-sector AIDS Organizations in KwaZulu-Natal", Submitted for the degree of Doctor of Philosophy. Development of Information Systems, London School of Economics and Political Science.

[4]   Serensen, T., Rivett, U. and Fortuin, J. (2010), "A Review of ICT Systems for HIV/AIDS and Anti-Retroviral Treatment Management in South Africa", Online Journal of Space Communication. Retrieved 24th May, 2011 from www.spacejournal.ohio.edu/issue14/health_hivAids.html.

[5]   Kalichman, S.C., Weinhardt, L., Benotsch, E and Cherry, C. (2002), "Closing the Digital Divide in HIV/AIDS Care: Development of a theory-based intervention to increase Internet Access".

[6]   Fisher, J.D., Amico, K.R., Fisher, W.A., Cornman, D.H., Shuper, P.A., Trayling, C., Redding, C., Barta, W., Lemieux, A.F., Altice, F.L., Dieckhaus, K., Friedland, G. (2011), "Computer-Based Intervention in HIV Clinical Care Setting Improves Antiretroviral Adherence", The LifeWindows Project. Retrieved 25th May, 2011 from http://www.springerlink.com/content/388615r0331217k8/

[7]   Akinyokun, O.C. & Angaye, C.O. (2011), "Development of Web-Based System for Knowledge Warehousing and Mining of I.T Human Capital and Infrastructure of Nigerian Tertiary Institution", International Journal of Computer Science and Applications. A Publication of Nigerian Computer Science (NCS), Vol. 17, No 2. pp 10-25.

[8]   Obot, O.U. and Uzoka, F.M.E (2008), "Fuzzy-rule based Framework for the Management of Tropical Disease", Int J Medical Engineering and Informatics. Vol 1. http://inderscience.metapress.com/ media/acecbdayxq7uyvbwd4u/contributions15/6/4/0/564023q746568m75_html/fulltext. Retrieved 14th March, 2011.

[9]   Anitha, B & Rajagopalan, S.P (2011), "Computer Decision Support System for Evidence-Based Meicine", – An Overview. European Journal of Scientific Research ISSN 1450-216X Vol. 50 (3), pp 444-452. Retrieved 18th March, 2011 from www.eurojournals.com/ejsr.htm.

[10]   Buchanan, B.G., Carenini, G., Mittal, V.O., and Moore, J.D. (1998)," Designing Computer-based framework that facilitate doctor-patient collaboration", Artificial Intelligence in Medicine 12 (1998) 169-191. Retrieved 28th April, 2011 from www.cs.ubc.ca.

[11]   Kolousek, G., Adlassnig, K., Bogl, K & Leitich, H. (1995), "An Overview of CADIAG-4: A Medical Diagnostic and Therapeutic consultation System".

[12]   Adele, A. & Neshat, M. (2010), "A Fuzzy Expert System for Heart Disease Diagnosis", Proceedings of the International Multi Conference of Engineers and Computer Scientists. Vol. 1, pp. 1-11.

[13]   Djam, X.Y & Kimbi, Y.H. (2011), "Fuzzy Expert System for the Management of Hypertension", The Pacific Journal of Science and Technology. Vol. 12, No. 1, pp 390-402.

[14]   Piot, P., Kapita, B.M., Ngugi, E.N., Mann, J.M., Colebunders, R., and Wabitsch, R. (1992), "AIDS in Africa: A Manual for Physicians", World Health Organization, Geneva.

[15]   Adler, M.W. (2000), "ABC of AIDS, Department of Sexually Transmitted Diseases", University of College London, Medical School, London.

[16]   Lambo, E. (2007), "National Guidelines for HIV and AIDS Treatment and Care in Adolescents and Adults", Federal Ministry of Health, Federal Secretariat, Abuja, Nigeria.

[17]   Eom, B.S. (2001), "Decision Support Systems," International Encyclopedia of Business and Management, 2nd Edition, Edited by Malcolm Warner, Internation Thomson Business Publishing Co., London, London, England, 2001, forthcoming.

[18]   Stolba, N., Tjoa, A. (2005), "The Relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making".

[19]   Sen, A & Sinha, A.P. (2005), "A Comparison of Data Warehousing Methodologies: Using a common set of attributes to determine which methodology touse in a particular data warehousing project." COMMUNICATIONS OF THE ACM. Vol. 48, No 3. pp79-84.

[20]   Yacci, M. (1999), "The Knowledge Warehouse: Reusing Knowledge Components". Vol. 12, No. 3.

[21]   Dymond, A. (2005), "The Knowledge Warehouse: The Next Step Beyond the Data Warehouse."

[22]   Codd, E.C. (1970), "Relational Model of Data for Large Shared Data Banks", Communication of ACM; Vol. 13, No 6, pp. 377-387.

[23]   Two Crows Corporation (1998), "Introduction to Data Mining and Knowledge" Retrieved 15th August, 2011 from www.twocrows.com.

[24]   Mihaele, U. (2003), "Internet-Enabled soft Computing Holarchies for e-Health Applications". In: New Directions in Enhancing the power of the Internet. L.A. Sadeh and M. Nikravesh (editors). Springer Verlag: Berlin, Germany.

[25]   Galichet, S. & Boukezzoula, R. (2009), "Optimistic Fuzzy Weighted Average", IFSA-EUSFLAT 2009.

Authors:

Igodan Charles Efosa, obtained his B.Sc. degree in Computer Science in 2004 from University of Benin (UNIBEN) where he works, and M.Tech degree in Computer Science in 2012 from Federal University of Technology, Akure, Nigeria. He is currently working on his Ph.D. His interests are in Data mining, Soft Computing, and Machine Learning.

Professor Oluwole Charles Akinyokun had BSc. First Class (Combined Honours) in Computer Science and Mathematics from Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria in 1979 and PhD in Computing Studies with specialization in Distributed Database System from the University of East Anglia, Norwich, United Kingdom in 1984. He was a Lecturer I in 1985, Senior Lecturer in 1987, Reader in 1992 and Professor of Software Engineering in October 1995. He is a Commonwealth Scholar and Fellow, Fellow of Nigerian Computer Society (NCS), Member of the Computer Professional Registration Council of Nigeria (CPN), British Computer Society (BCS) and American Computing Machinery (ACM). He has been actively involved in Research, Innovation and Development (RID) in Knowledge Engineering in computer network environment since 1985. He has featured prominently as a Reviewer to many reputable Journals and Conference Proceedings, External Examiner and Assessor to many Nigerian and overseas universities. His publications have featured prominently in many reputable national and international conference proceedings and journals.

Dr. Olabode Olatubosun holds a B.Tech. degree in Ind. Maths, M.Tech. in Computer science and a PhD in Computer science from the Federal University of Technology, Akure Nigeria (1991, 1999 and 2005). He is a researcher of both local and international repute. He has many articles published in recognized journals and conference proceedings. Presently, he is a senior lecturer at the Federal University of Technology, Akure. His research interest includes, Machine learning, AI, Database and ecommerce. Email:olabode_olatubosun@yahoo.co.uk. Tel:234-8033511257