Applied Computing and Informatics (2017) 13, 27–37

Saudi Computer Society, King Saud University

Applied Computing and Informatics

(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

A Genetic-Neuro-Fuzzy inferential model for diagnosis of tuberculosis

Mumini Olatunji Omisore [a,*], **Oluwarotimi Williams Samuel** [b,1], **Edafe John Atajeromavwo** [c,2]

[a] *Centre for Information Technology & Systems, University of Lagos, Yaba-Akoka, Lagos, Nigeria*
[b] *Biomedical Engineering & Health Techonolgy, Shenzhen Institute of Advanced Techonolgy, CAS, Shenzhen, Guangdong, China*
[c] *Department of Computer Science, Delta State Polytechnic, Ogwashi-uku, Nigeria*

Received 4 March 2015; revised 2 June 2015; accepted 9 June 2015
Available online 24 June 2015

KEYWORDS

Medical diagnosis;
Mycobacterium tuberculosis;
Artificial intelligence;
Inference system;
Decision support

Abstract Tuberculosis is a social, re-emerging infectious disease with medical implications throughout the globe. Despite efforts, the coverage of tuberculosis disease (with HIV prevalence) in Nigeria rose from 2.2% in 1991 to 22% in 2013 and the orthodox diagnosis methods available for Tuberculosis diagnosis were been faced with a number of challenges which can, if measure not taken, increase the spread rate; hence, there is a need for aid in diagnosis of the disease. This study proposes a technique for intelligent diagnosis of TB using Genetic-Neuro-Fuzzy Inferential method to provide a decision support platform that can assist medical practitioners in administering accurate, timely, and cost effective diagnosis of Tuberculosis. Performance evaluation observed, using a case study of 10 patients from St. Francis Catholic Hospital Okpara-In-Land (Delta State, Nigeria), shows sensitivity and accuracy results of 60% and 70% respectively which are within the acceptable range of predefined by domain experts.

1. Introduction

Tuberculosis (TB) is a social, re-emerging infectious disease that has medical implications throughout the globe [1]. The largest single cause of adult illness and death from the communicable disease is caused by Mycobacterium Tuberculosis [2]. Nigeria has made great strides in increasing access to Directly Observed Therapy Short-course (DOTS) for TB yet, coverage, which was 45% in 1999, had reached 75% by 2005 while treatment success for 2005 cohort was 75% [3]. Although TB incidence in Nigeria is below the normal level for Sub-Saharan Africa, but it remains high at a rate of 311 cases per 100 grand

* Corresponding author. Tel.: +234 7031967847.
E-mail addresses: ootsorewilly@gmail.com (M.O. Omisore), timitex92@gmail.com (O.W. Samuel), edafejohn2006@yahoo.com (E.J. Atajeromavwo).
[1] Tel.: +86 15814491870.
[2] Tel.: +234 8064784094.

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2015.06.001

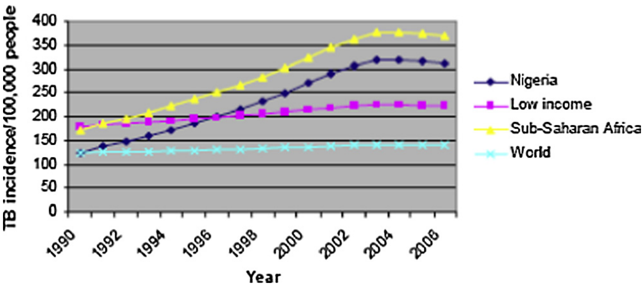

Figure 1 Incidence of TB in Nigeria and other Sub-Saharan African countries [39].

population members in 2006. The trends for both Nigeria and Sub-Saharan Africa, as depicted in Fig. 1, show a slight downward turn of TB incidence since 2003. Still with 250 grand new cases each year, a mortality rate of 81 deaths per 100,000 spells the disease as high burden on Nigeria [4].

The World Health Organization (WHO) estimated in 2006, that each year, more than 8 million new cases of TB occur and approximately 3 million persons die from the disease [4,5] and estimated that between 19% and 43% of the world's population will be infected with Mycobacterium Tuberculosis. Within the past decade it has become clear that the spread of HIV infection and the immigration of persons from areas of high incidence have resulted in increased numbers of TB cases. It has always occurred disproportionately among disadvantaged populations such as the homeless, malnourished, and overcrowded [6]. Today, several methods for the diagnosis of TB have been proposed. Tuberculin Test, Radiological Examination, and Sputum Smear Microscopy are common conventional approaches however in the last 10 years, several molecular methods have been developed for direct detection, identification and susceptibility testing of mycobacteria [7].

Orthodox methods of diagnosing TB are primarily through physical examination and laboratory tests. The former involves asking patients certain questions for prognosis purposes while tests are carried out to affirm physical examination. Diagnosis can be stopped if medical practitioner is totally convinced after physical examination however, this is not advised. This orthodox method is currently faced with a number of challenges such as lack of medical facilities in most medical centers and as a result, inhibiting the management of TB in developing countries.

The strength of IT in providing an effective and efficient solution to real life problems has been explored to aid scientific discoveries and advancement of different fields of medicine [8]. Hence, to reduce the morbidity and mortality rates in human as a result of TB, there is need to incorporate IT into its diagnostic approach. This study, therefore, proposes a decision support model for intelligent diagnosis of TB using Genetic-Neuro-Fuzzy Inferential technique. The model is aimed at providing a decision support platform that can aid medical practitioners in administering accurate, timely, and cost effective diagnosis of TB in developing countries.

2. Literature review

This section presents a review of literature on the concept of Expert System (ES). Description of major tools for building such adaptive systems including is briefed while review of hybrid and decision support systems is also presented.

2.1. Artificial intelligence

Research on Artificial Intelligence (AI) in the last two decades has greatly improved performance of both manufacturing and service systems [9]. AI, first coined by John McCarthy in the fifties is concerned with the 'hows' and the 'whys' of human intelligence; however, it has become an important area of research in virtually all fields including engineering, science and education, and as well as its applications in accounting, marketing, stock market and law, among others [10,11].

AI is the intelligence deployed by machines to handle complex imprecise tasks that require intelligence if done by humans. The central problems of AI include reasoning, programming, artificial life, belief revision, knowledge representation, machine learning, natural language understanding, and theory of computation [12,13,38]. It achieved greater feats in practical application despite, some setbacks, its success is being revived with the commercial success of ES [39]. Fuzzy Logic, Neural Networks, and Genetic Algorithms, are such techniques used in modeling intelligence.

2.2. Expert systems

Expert Systems (ESs) is a branch of AI that employs the use of human knowledge to solve problems that require human's expertise and it helps to solve complex problems by reasoning about knowledge rather than following developers' procedures as in the case of conventional programming [14]. ES continues to evolve for specific applications in medical diagnosis due to influx of new and massive information that requires experts to be specialized.

The basic steps in ES development have been reported in [15]. Many AI systems have been developed for the purpose of enhancing healthcare delivery, providing better healthcare facilities, and reducing the cost associated with quality healthcare services. Early studies in intelligent systems have been shown to outperform manual practices of medical diagnosis. Examples of such systems are as follows: INTERNIST, a rule-based expert system for the diagnosis of complex problems in general internal medicine; MYCIN, a rule-based expert system to diagnose and recommend treatment for certain blood infections; CASNET, an expert system for the diagnosis and treatment of glaucoma, EXPERT, an extension generalization of the CASNET formalism which was used in creating consultation systems in rheumatology and endocrinology [16].

2.3. Soft computing tools

The intervention of soft computing tools (techniques) in medical analysis has greatly reduced the cost of human support and medical diagnosis, with increase in accuracy of diagnosis results. Fuzzy Logic, Neural Networks, and Genetic Algorithm are common tools adopted in developing ESs [17].

2.3.1. Fuzzy logic

Fuzzy Logic (FL) is one of AI techniques that deals with uncertainty in knowledge and simulates human reasoning in an incomplete or fuzzy data. FL is defined as a nonlinear mapping of an input data set to a scalar output data set, basically it is aimed at providing approximate reasoning. This technique has attracted growing attention and interest in modern IT, pattern recognition, and decision making among others [18]. FL theory provides a mathematical strength to capture uncertainties associated with human cognitive processes, such as thinking and reasoning. It is a suitable and applicable basis for developing knowledge-based systems in varying sectors of life such as health. It has been applied to interpret sets of medical findings; syndrome differentiation in eastern medicine and diagnosis of diseases in western medicine; and for real-time monitoring of patient data [19].

In fuzzy set theory, linguistic terms are used to illustrate the correlation of Membership Function (MF) which describes the membership of an element within the base of a fuzzy set. Each element has a unit value that characterizes the grade of membership of a set and such element can simultaneously belong to another set, possibly, at varying degrees. Ref. [20] emphasize that a number of different types of MFs have been proposed for fuzzy control systems though [21] concluded triangular and trapezoidal MFs as the mostly used. Triangular MF is a particular case of MF that is specified by three parameters (a, b, c) and shows the degree of membership of each class of a linguistic term as possibility distribution [22]. Fig. 2a represents a typical Triangular MF of input and output variables while Fig. 2b uses four parameters to describe the membership of an element in a fuzzy set using Trapezoidal MF.

The trapezoidal MF of a fuzzy set F with each element having tolerance interval $[a,b]$, left width α and right width β is determined using Eq. (1). If notation $F = (a, b, \alpha, \beta)$ is used, then $[F]\gamma = [a-(1-\gamma)\alpha, b+(1-\gamma)\beta]$, $V\gamma\varepsilon[0,1]$ hence the support of A is $(a-\alpha, b+\beta)$.

$$F(x) = \begin{cases} 1-(a-x)/\alpha & \text{if } a-\alpha \leqslant x \leqslant a \\ 1 & \text{if } a \leqslant x \leqslant b \\ 1-(x-b)/\beta & \text{if } a \leqslant x \leqslant b+\beta \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

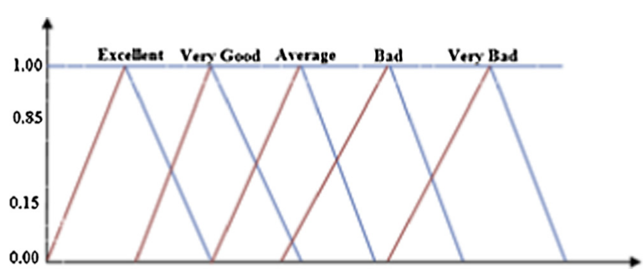

Figure 2a Triangular MF of input and output variables.

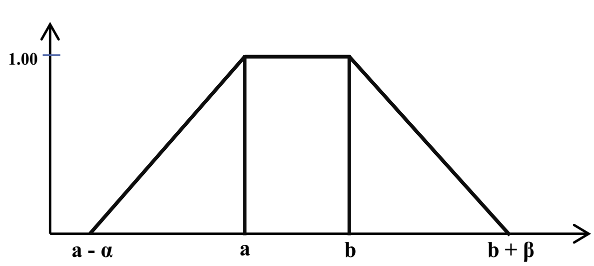

Figure 2b Trapezoidal MF of input and output variables.

In fuzzy set, an element can belong to both its set and its compliment set or to neither of them. This principle preserves the structure of the logic and avoids the contradiction of element. However, fuzzy logic is highly abstract and employs heuristic requiring human experts to discover rules about data relationship [19]. FL has been widely adopted in developing ESs for health management. For instance, [18] proposed model ES for typhoid fever, while [23] for malaria diagnosis, and lastly, [24] developed a diagnostic ES for cardiovascular diseases. In [19], the use of Fuzzy Cluster Means was applied to diagnose HIV/AIDS shortly after [25] proposed the use of fuzzy sets for diagnosing low back pain in computer users.

2.3.2. Neural network

Neural Network (NN) is a group of interconnected artificial neurons that mimic the properties of biological neurons. It follows analog and parallel computing system made up of simple processing elements that communicate through a rich set of interconnections with varying contributory weights. Artificial Neural Network (ANN), is synthetic nervous systems loosely inspired to simulate functions of human brain [26]. ANN attempts to abstract the complexity of biological nervous system so as to focus on what may hypothetically matter most from an information processing point of view.

Medicine has always benefited from forefront of technology as it has boosted medicine to extraordinary levels of achievement. ANN has been successfully used in various areas of medicine such as biomedical analysis, imaging systems, and drug development but extensively used in diagnosis to detect ailments such as cancer and heart problems in human [27]. The term **network** in ANN arises because of the function $f(x)$ defined as a composition of other function $g_i(x)$ which are further used as composition of more functions.

Fig. 3 shows a simple NN which comprises of three layers. The figure comprises of input units connected to hidden units which in turn is connected to a layer of "output" units. The activity of the input unit represented the raw information that is fed into the network; the activity of the hidden units is determined by the activity of the input units and the weights between the hidden and output units. The hidden units are free to construct their own representation of the input; the weights between the input and hidden units determine when each hidden unit is active and so by modifying the weights, a hidden unit can choose what it represents.

ANN employs learning paradigm that includes supervised, unsupervised and reinforced learning. One good thing is it does not require details on how to recognize disease but it has a self-learning and self-tuning feature which helps it to attain that [28]. Finally, it cannot handle linguistic information and vague information.

2.3.3. Genetic algorithm

Genetic Algorithm (GA) is simply a search algorithm based on the observation of sexual reproduction and principle of *survival of the fittest*, which enables biological species to adapt to their environment and compete effectively for resources. GAs are search algorithms which use principles inspired by natural genetics to evolve solutions to problems. The basic idea is to maintain population of chromosomes that represents candidate solutions to a problem, and the candidate will evolve

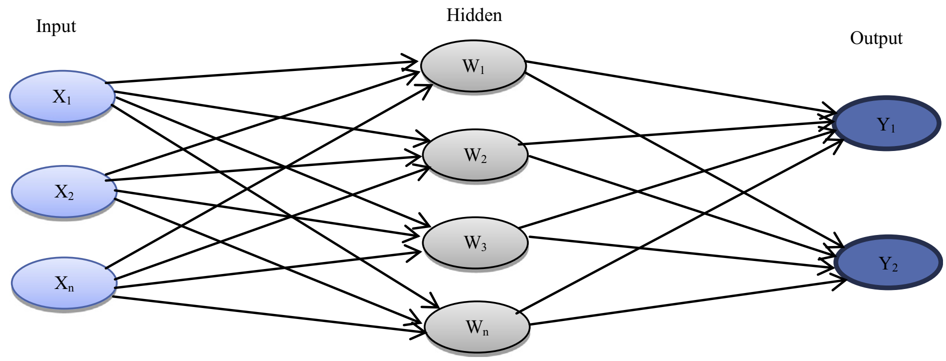

Figure 3 A simple neural network (Adapted from: Imianvan and Obi, 2012).

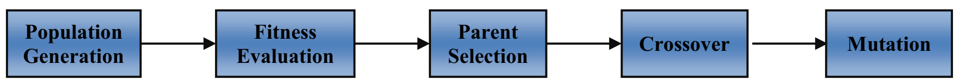

Figure 4 Basic operations of GA.

over a period of time through competition and controlled variation.

GAs have got a great measure of success in search and optimization problems. While the algorithm is relatively straight forward, it is an effective stochastic search method, proven as a robust problem solving technique that produces better than random results [29]. GAs are robust and powerful in difficult situations where the space is usually large, discontinuous, complex and poorly understood [30], it has been applied in a wide range of problem areas, though it just guarantee finding an acceptable solution in a quick and not a global optimum solution to a problem.

2.3.4. Review of hybrid and decision support systems

A Decision Support System is an interactive computer-based information system that utilizes database models to solve ill-structured problems and come up with a valuable decision. In the late 1980s, DSS was confirmed to have assisted in managerial positions using suitable available technologies to improve effectiveness of academic and professional activities [38] but over the years, it has migrated to an interactive system that assists users in taking quick appropriate decisions in any given context [31]. In a related study, Ref. [32] developed a fuzzy expert system as a platform for diagnostic support for hypertension management. The system is composed of four major components for fuzzy processes while Root Sum Square and Center of Gravity were employed as fuzzy inference and defuzzification methods respectively. A case study of 30 patients with tuberculosis was used to validate the ES. [33] combined NN, FL and Case Based Reasoning to model DSS for diagnosis of depression disorders. The NNs were constructed to imitate intelligent human biological processes of learning while FL provides a means for dealing with imprecision, vagueness and uncertainties in the medical data and CBR entails the use of past situations to solve new occurrences. Finally, this study proposes a Genetic-Neuro-Fuzzy inferential technique for diagnosis of tuberculosis.

3. Proposed decision support system

This section presents design of the system's architecture and procedures performed by each component of the architecture during diagnosis. Components of the architecture, as presented in Fig. 5, are Knowledge Base, Genetic-Neuro-Fuzzy Inference Engine, and Decision Support Engine.

3.1. Knowledge base

Knowledge base stores both static and dynamic interpreted information about the decision variables involved in the diagnosis of TB. The component, comprising of the Database, Fuzzy Logic, Neural Network, and Genetic Algorithm, serves as a repository for operational data that are to be processed.

3.1.1. Database

Structured database presents quantitative data about facts and the established rules in the field of medicine focusing on diagnosis of TB. The facts comprise of signs and symptoms of TB, while rules are patterns to draw deductions based on available information [18]. Unstructured database is heuristic in nature and hence gathered by experience, good practices, guesses, and judgments [34]. The database comprises of Patient-Bio-Data, Disease-Physical-Signs, Disease-Symptoms, Medical-History, Physical Examination, results of diagnostic tests and Patient Diagnosis.

3.1.2. Fuzzy logic

The diagnosis process harnesses the strength of fuzzy logic component in the following operational sequence:

3.1.2.1. Fuzzification of input variables. Given a fuzzy set A, defined as Eq. (2), represents TB diagnosis variables with element denoted by x_i, the fuzzification process involves transforming raw input value of each variable to a fuzzy term

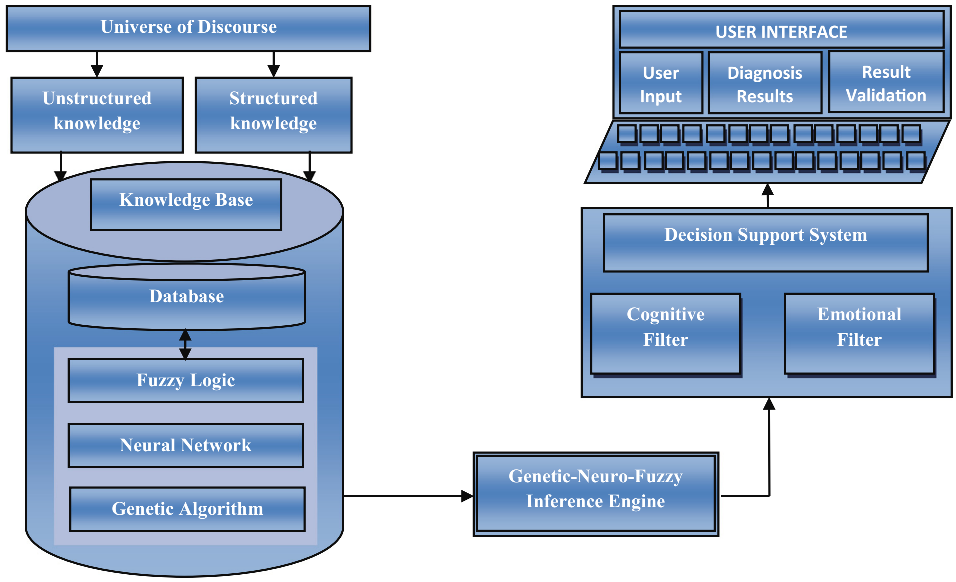

Figure 5 Architecture of the proposed Genetic-Neuro-Fuzzy expert system.

obtained from set [*very mild*, *mild*, *moderate*, *severe*, *very severe*] defined over the variables. That is, such values are derived from functions defined to determine the degree of membership of each variable in the fuzzy set.

$$A = \{(x_i, \mu_A(x_i))|x_i \in V, \mu_A(x_i) \in [0,1]\} \tag{2}$$

Fuzzification is done using function defined in Eq. (3)

$$\mu_A(x_i) = \begin{cases} 1 & \text{if } x_i < a \\ \frac{x_i-a}{b-a} & \text{if } a \leqslant x_i < b \\ \frac{c-x_i}{c-b} & \text{if } b \leqslant x_i < c \\ 0 & \text{if } c < x_i \end{cases} \tag{3}$$

where $\mu_A(x_i)$ is the MF of x_i in A using triangular MF while μ_A is the degree of membership of x_i in A. a, b and c are the parameters of the MF governing its triangular shape and each attribute is described with linguistic terms.

3.1.2.2. Establishment of fuzzy rule base. The rule base for TB diagnosis is characterized by a set of IF–THEN rules in which the antecedents (IF parts) and consequents (THEN parts) involve linguistic variables. The rules can be formulated with assistance of experts in the management of TB, or on consultation to existing standard literature. A rule can only fire if any of its precedence parameters such as *very mild*, *mild*, *moderate*, *severe*, and *very severe* evaluates to *TRUE*, otherwise it does not fire.

3.1.2.3. Fuzzy inference engine. This component controls the decision making logic by applying suitable composition procedure from rule base to values of variable inputs received. The inference engine applies composition procedure on the inputs to produce desired output, and Root Sum Square (**RSS**), is applied to scale the functions at their respective magnitude and computes a composite area. RSS is a method used to combine the effects of fired rules in order to draw relevant inference. It is computed with Eq. (4).

$$\mathbf{RSS} = \sum_{k=1}^{n}(R_k^2) \tag{4}$$

R_k is a fired rule where $k\forall 1,\ldots,n$ is the Id of fired rule

3.1.2.4. Defuzzification of output values. Defuzzification of output values involves translating result from the inference engine into crisp values which are, mostly, required by medical experts for proper analysis and interpretation, this aids efficient diagnosis. This research employs Centroid of Area (**CoA**) technique for its defuzzification. This interface receives the output of inference engine as its input and finalizes computation by applying Eq. (5).

$$\mathbf{CoA} = \frac{\sum_{i=1}^{n}\mu Y(x_i)x_i}{\sum_{i=1}^{n}\mu Y(x_i)} \tag{5}$$

where $\mu Y(x_i)$ is degree of $\boldsymbol{i}$ in a membership function and x_i is the center value in function.

The computational simplicity and intuitive plausibility of this approach gives rise to its adoption. For a complete medical evaluation of TB disease, the variables considered after consultations with medical experts and other standard literal sources are categorized as presented in Table 1.

3.1.3. Neural network

Neural Network has the capability of capturing domain knowledge from available indicators and can readily handle both continuous and discrete data. NN is used to train and test the designed fuzzy system to optimize the performance of the overall system. The NN component of Fig. 6 is made up of variables from *Physical Examination* (*PE*), *Medical History* (*MH*), *Laboratory Investigation* (*LI*), *and Chest Radiology* (*CR*) of patients. Each diagnosis variable has a weight W_i which shows its contribution in the diagnosis process.

The raw information obtained from patients is fed into NN via input layer and participation of each category of variables is determined at a hidden layer of the network using:

$$CAT_i = \sum_{i}^{n} A_i * W_{A_i} \tag{6}$$

CAT_i is ith category of variable, n is count of variables in CAT_i, and A_i is the ith diagnosis variable with weight W_{A_i}.

Table 1 Categorization of diagnosis variables for tuberculosis.

Category	Diagnosis variables	Code	Category	Diagnosis variables	Code
Physical Examination (PE)	Swollen lymph nodes	A_1	Medical History (MH)	Meningitis	B_9
	Blood pressure	A_2		Hoarseness	B_{10}
	Rale breathe	A_3	Laboratory Investigation (LI)	Sputum test	C_1
	Abnormal breast sounds	A_4		Cerebrospinal fluid test	C_2
Medical History (MH)	Loss of appetite	B_1		Pus test	C_3
	Confusion	B_2		Tuberculin skin test	C_4
	Cough	B_3		Blood test	C_5
	Fever	B_4		Biopsy test	C_6
	Chest pain	B_5	Chest Radiography (CR)	Plefus	D_1
	Weight loss	B_6		Pulede	D_2
	Night sweat	B_7		Cardrat	D_3
	Fatigue	B_8		Zanflo	D_4

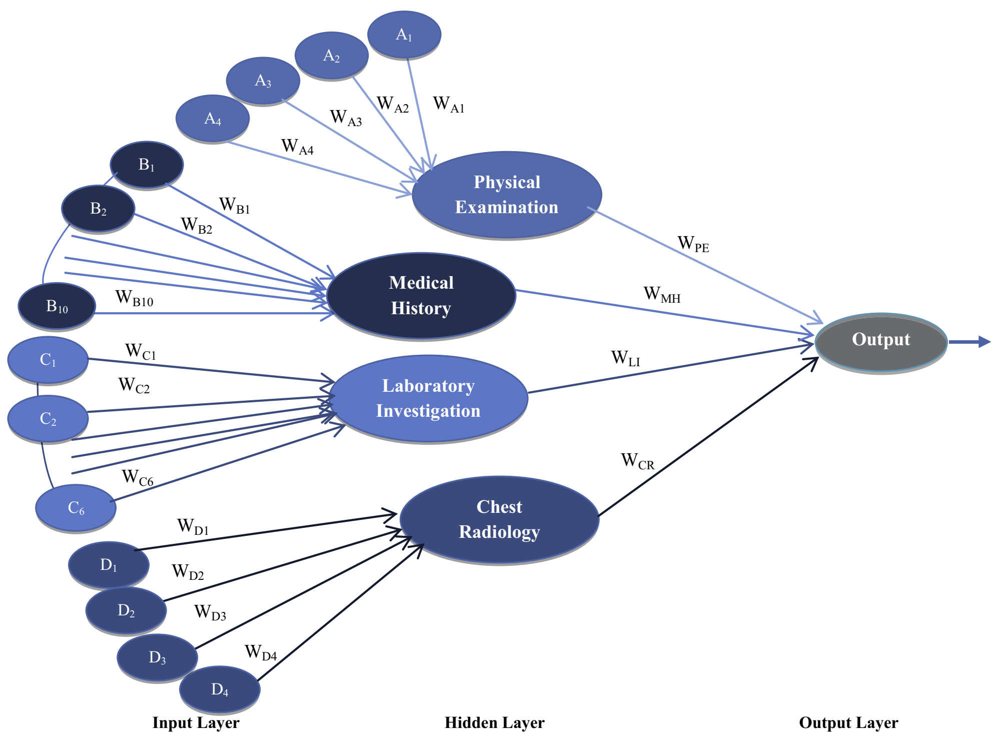

Figure 6 Block diagram of NN for the diagnosis of TB diseases.

Result of the output layer represents an overall output of diagnosis by the NN component of the architecture shown in Fig. 5. The output result is given by:

$$Output_{NN} = \sum_{i}^{n} CAT_i * W_{CAT_i} \tag{7}$$

where W_{CAT_i} is the connection weight of CAT_i

3.1.4. Genetic algorithm

Actually, NN provides a structure for combining the diagnostic parameters which could serve as a platform for the inference engine, but a specific issue with NN is lack of definite way of determining the connection weights for hidden layers when dealing with a particular problem. A number of medical diagnosis had been assisted by neuro-fuzzy systems though such systems had been built based on trial and errors, this increases computation cost. In this study, genetic optimization is performed to choose optimal values from a group of diagnostic parameters which serve as input. Fig. 6 shows there are 24 diagnostic parameters in the NN but the task is to decide which parameters are taken as input in order to minimize complexity.

An individual chromosome consists of 24 genes and each gene represents the connection weight of a diagnosis variable in a length of 1 bit. One feasible solution is to generate an initial population holding a set of possible solutions from random chromosomes. A chromosome is represented as a vector $C = (C_A, \ldots, C_X)$ of binary decision variables $C_{i=0,2,3}$; encoded in binary representation as string consisting $\{0, 1\}$ genes. A gene $C_i = 1$ if the ith variable is included in a solution set of a diagnostic process otherwise 0. Fitness function is used to optimize each chromosome by evaluating the genes that constitute the chromosome using their fitness value.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	**Fitness**
C 1	1	0	0	0	1	0	1	0	0	0	1	0	1	1	1	0	0	0	1	0	1	0	1	0	F1
C 2	1	1	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	1	0	1	1	0	0	F2
C 3	0	0	0	1	1	0	1	0	1	0	1	1	1	0	0	0	1	1	0	1	1				F3
C K	1	0	1	0	0	0	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	Fk

Figure 7 Chromosomes of the selected individual candidates and corresponding fitness value.

As evolutionary algorithm continues through its cycle, fitness value of each chromosome keeps improving till it reaches an optimum value when it can no longer improve. Fig. 7 shows chromosomes of some candidates and their fitness values. A number of constraints have been considered in carrying out appropriate management of disease in medical diagnosis, therefore fitness evaluation of chromosome must be done with proper constraint validation. Constraints can be termed as objectives that must be achieved in which some render most of the solutions from the search space hence, its application in GA is problem specific. Ordering constraint proposed in [35] is adopted in this study. The fitness evaluation of an individual $F(i)$ is done as:

$$F_i = \left(1 + \sum_{i=0}^{n} W_i * C_i(p)\right)^{-1} \tag{8}$$

where n is the number of diagnosis variables, W_i is the weight associated with ith variable and $C_i(p)$ is the number of violations for ith constraint at solution p.

This fitness function has a range of $[0, 1]$ and an optimal solution occurs when we have 0 violations thus $\sum_{i=0}^{n} W_i * C_i(p)$ which results in $F_i = 1$. Chromosomes with higher fitness value are selected as parents for mating in order to produce outstanding candidates and maximize the fitness function. The probability of choosing an individual for genetic operation is proportional to its fitness, that is, if the fitness value of an individual is F_i, then the probability, P_i, of choosing the individual is:

$$P_i = \frac{F_i}{\sum_{i=0}^{n-1} F_i} \tag{9}$$

This process is repeated until an optimal connection weight is achieved.

3.2. Genetic-Neuro-Fuzzy Inference System (GENFIS)

Genetic-Neuro-Fuzzy Inference System (GENFIS) is an inferential technique proposed to integrate GA, NN and FL components of Fig. 4 to provide a self-learning and adaptive system for handling uncertain and imprecise data for diagnosis of tuberculosis. The inference system employs feed forward propagation learning technique made up of seven layers of neurons as shown in Fig. 8. Both hidden and output layers consist of active nodes where computations take place, while the nodes at input layer are passive.

The inference engine consists of reasoning algorithm driven by the production rules based on Mamdani's Inference Mechanism. Of the seven layers, the first one consists of active nodes which denote inputs to the system. The inputs are numeric values representing how severe a patient feels the diagnosis variables. The output of this layer is the linguistic labels corresponding to each input value. The second layer is made up of adaptive nodes that receive the output of preceding layer as input, and produce their corresponding membership grade determined as:

$$L_2(x_i) = \mu_{Ai}(X_i) \tag{10}$$

The fuzzy value of each variable is computed using triangular MF, given as:

$$\mu_{Ai}(x_i) = \frac{x_i - b}{a - b} \tag{11}$$

where a and b are the variables of the triangular MF that bounds its shape such that $b \leqslant \mathbf{x_i} \leqslant a$.

Third layer act as multipliers and their operations are fixed and labeled as M. These nodes compute the firing strengths of associated rules as:

$$L_3(X_i) = \mu_{Ai}(X_i) * \mu_{Bi}(X_i) * \mu_{Ci}(X_i) \tag{12}$$

In the fourth layer, nodes fixed but they do normalize the firing strength of each rule. The normalized strength of a kth rule is determined as:

$$L_4(X_i) = \frac{W_k}{\sum_{j=1}^{3} W_j} \tag{13}$$

The product of normalized firing strength of a rule and its corresponding output value is observed in the fifth layer to determine the variable's contribution to the diagnosis processes. This is done with Eq. (14).

$$L_5(X_i) = L_4(X_i) * L_3(X_i) \tag{14}$$

The sixth layer consists of a single fixed node labeled Y which represents the GENFIS's final output. It obtains the cumulative sum of all incoming signals as shown in Eq. (15).

$$Y = \sum_{i=1}^{n} L_5(x_i) \tag{15}$$

Finally, we employed Eq. (16) to classify the crispy numeric value in Eq. (15) as the system's output, which represents the patient's diagnosis result.

$$\text{Output} = \begin{cases} \text{Very Mild} & Y \leqslant 0.2 \\ \text{Mild} & 0.2 \leqslant Y < 0.4 \\ \text{Moderate} & 0.4 \leqslant Y < 0.6 \\ \text{Severe} & 0.6 \leqslant Y < 0.8 \\ \text{Very Severe} & 0.8 \leqslant Y \leqslant 1.0 \end{cases} \tag{16}$$

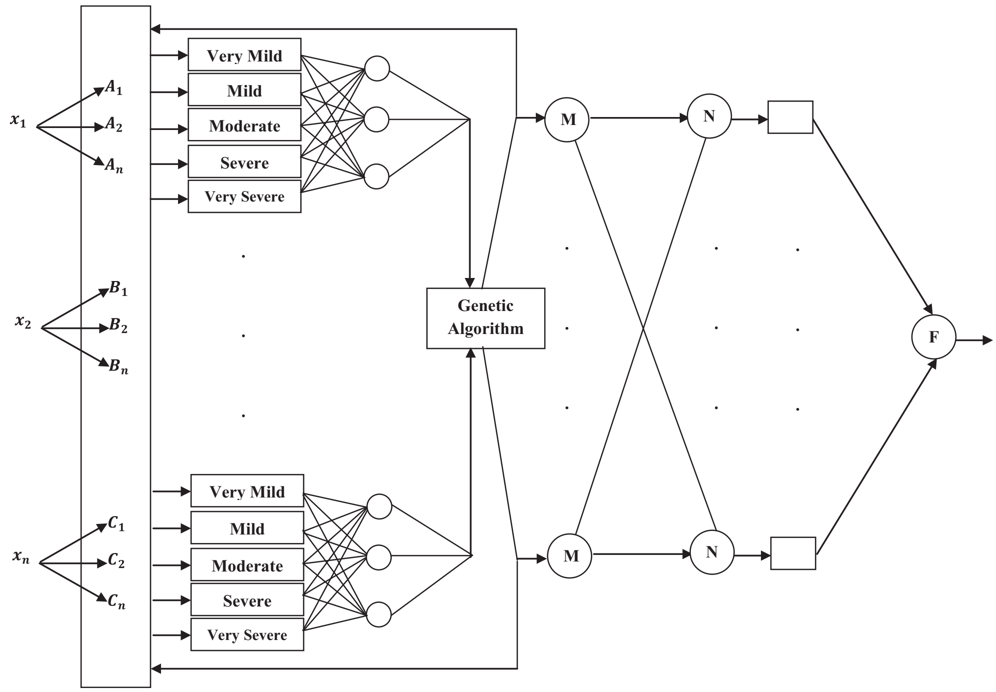

Figure 8 Combinatorial model for Genetic-Neuro-Fuzzy Inference System.

3.3. Decision support engine

The decision made by GENFIS is optimized by Decision Support Engine (DSE) which takes the output of GENFIS as input and tunes to fit any diagnostic case at hand [37]. The supporting components of DSE are cognitive filter which objectively influences the result of GENFIS on conformance basis with knowledge extracted from medical personnel, holding assumptions and beliefs with heuristics in medical field and emotional filter that refers to subjective feelings of a medical personnel based on physical and psychological elicited from patient. Emotional filter provides information that helps medical personnel to decide whether a diagnostic result from GENFIS is as a result of the patient's situation or environmental inhibiting factors.

4. Model simulation and evaluation

Simulation of the proposed model was done with a case study of 10 patients from Saint Francis Catholic Hospital Okpara-In-Land, Delta State, Nigeria. The procedure was observed in Matrix Laboratory (MATLAB) Version 7.9 environment, the result and evaluation of the simulations are reported in this section.

4.1. Simulation

In order to evaluate the performance of the proposed model, medical records of 100 patients' representing their state of health with respect to TB were formulated and stored as rules in the database. Each rule is made up of 24 input variables and an output variable. To determine the output value of each rule, the records were retrieved and assessed by domain experts in human respiratory diseases. Assessment was based on intensity of the input variables and the expertise of the (human) expert. Intensity of each variable represents its contribution to TB infection, rating was thereafter done based on linguistic terms shown in Table 2.

Table 2 Intensity of diagnosis variables.

Linguistic term	Very mild	Mild	Moderate	Severe	Very severe
Intensity	1	2	3	4	5

Result obtained from fuzzification of variables serves as input to the neural network. Each node in the network is a three-layered feed forward architecture which interacts with each other as shown in Fig. 8. Back-propagation algorithm with sigmoid function is used to train the NN for hidden and output layer neurons' transformation. The NN trained by the subsystem consists of 24 nodes at the input layer, each representing unique TB variables considered in this study. To determine an optimal number of variables needed for a diagnosis, GA component of the proposed GENFIS takes all variables as input and optimizes them into just N variables whose values have role to play in the diagnosis. Hence, the genotype is represented by a sequence of symptoms as described earlier.

During simulation, binary-matrix vectors with length $\boldsymbol{n}$ were created. Each element of the vector corresponds to specific diagnostic parameters in NN. The binary bit of a diagnostic parameter is determined using:

$$b_i = \begin{cases} 1 & \text{Selected} \\ 0 & \text{Ignored} \end{cases} \quad (17)$$

Table 3 Parameters used in finding optimal solution.

S/No	Parameters	Value
1	Number of generations	20
2	Number of individuals	40
3	Crossover probability	0.55
4	Mutation probability	0.35

Hence, if the value of a parameter is equal to 1 then the variable will be selected otherwise not. In this simulation, the GA used selected values given in Table 3 to find optimal solution and once a complete set of optimal value for the parameters is reached, the GA processes stops. High mutation probability used in the parameter settings is a brain behind bringing new individuals at each stage hence, to avoid a scenario whereby best combination will only be from initial individuals who passed first selections.

4.2. Result and evaluation

Medical records of 10 patients from the Hospital of our case study were taken as testing data for the system. Result obtained from simulation indicates that GA component of the proposed model can extract a maximum of 13 out of 24 parameters as the best combination. As shown in Table 4, selected parameters show that the model is sensitive to radiographic variables. The variables were selected for all cases with diagnosis result above 50%. Diagnosis of records R04, R09, and R10 by the model shows that patients possessing such attributes have tuberculosis infection.

Result of this simulation procedure was validated by human experts and their response using metrics proposed in some previous studies. Human responses presented in Table 5 shows an expert either accepts the model's result by assigning "√" or rejects it with "x". In [36], sensitivity is a quality of Neuro-Genetic model used to check the effects of selected parameters on a trained NN with evaluation function.

Given a True Positive value (TP) that represents the number of patients with tuberculosis as agreed by both model and human expert, a True Negative value (TN) indicating the number of patients where agreement could not be reached by both model and human expert, and TNR as the total number of records; the sensitivity and accuracy of GENFIS are:

$$\text{Sensitivity} = \frac{TP}{TNR} * 100\% \tag{18}$$

$$\text{Accuracy} = \frac{TP + TN}{TNR} * 100\% \tag{19}$$

From Table 5, the sensitivity and accuracy of the proposed GENFIS are 60% and 70%.

5. Conclusion

The use of soft computing techniques in medical diagnosis cannot be overemphasized as they have greatly imparted the processes in medical diagnosis and aided an increase in diagnosis accuracy. This novel study demonstrated how an aggregation of such technique can assist in the diagnosis of TB. In the approach, cognitive and emotional filters were adapted to take care of contextual factors that often affect medical expert during diagnosis of diseases in the traditional and conventional ways.

This study shows that a combination of the soft computing methods can offer a more effective system of medical diagnosis with improved system accuracy. For instance, [32] validated a fuzzy-based expert system for tuberculosis diagnosis with 61% accuracy. Also, in [24], a neuro-fuzzy decision support model for therapy of heart failure was conducted and a sensitivity analysis conducted shows that diagnosis done by the model has a high concordance of 60.72% with physician's diagnosis at an accuracy of 57.14%. Unlike the proposed model, existing systems does not have a thorough scope in terms of data set or diagnosis depth and on a general sense, the model exhibits a relatively higher performance when compared with existing systems hence, depicting a more reliable results.

Table 4 Results from selected parameters in the proposed model.

Id	A_1	A_2	A_3	A_4	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	C_1	C_2	C_3	C_4	C_5	C_6	D_1	D_2	D_3	D_4	Result
R01	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0.350
R02	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0.296
R03	0	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0.381
R04	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	0	1	0	1	1	1	1	1	0.814
R05	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.174
R06	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0.313
R07	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0.280
R08	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0.412
R09	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	1	0.660
R10	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	1	0.717

Table 5 Validation of GENFIS simulation results.

Record Id	R01	R02	R03	R04	R05	R06	R07	R08	R09	R10
Result	35.0%	29.6%	38.0%	81.4%	17.4%	31.3%	28.0%	41.2%	66.0%	71.7%
GENFIS	√	√	√	√	x	√	x	x	√	√
Human	x	√	√	√	√	√	x	√	√	√

So well so good, this study demonstrates how triangular membership functions of fuzzy logic system can be employed to define linguistic labels and neural networks were introduced for self-tuning and adaptation in case of new situations. Also, for effective selection of optimal input parameters, Genetic Algorithm has been incorporated. Yet, the effectiveness of the model has only been validated using casual TB records because the rule base of the knowledge base was formulated. To further validate the results of this model, sufficient real-life records of TB patients can be obtained from medical clinics in order to generate rules and training data sets.

Acknowledgments

We thank the management of St. Francis Catholic Hospital Okpara-In-Land, Delta State, Nigeria, for their utmost cooperation in the area of data collection. We also appreciate the efforts Dr. Ahmed of Obafemi Awolowo University Teaching Hospital Complex, Wesley Guild Unit, Ilesa, Osun State, Nigeria, for his assistance on rule-base formulation and proposed model performance evaluation.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.aci.2015.06.001.

References

[1] C.A. Robert, J.E. Buikstra, The Bio-Archaeology of Tuberculosis: A Global Perspective on a Re-Emerging Disease, University Press of Florida, Gainesville, Fl, 2009.
[2] David M. Morens, Gregory K. Folkers, Anthony S. Fauci, The challenge of emerging and re-emerging infectious diseases, Nature 430 (2004) 242–249.
[3] World Health Organization, WHO Country Cooperation Strategy: Federal Republic, 2002.
[4] World Health Organization, Global Tuberculosis Control, WHO Report 2008, Geneva, 2008.
[5] P. Sudre, G.T. Dam, A. Kochi, Tuberculosis: a global overview of the situation today, Bull. World Health Organ. 70 (1992) 149–159.
[6] American Thoracic Society, Diagnostic standards and classification of tuberculosis in adults and children, Am. J. Respir. Crit. Care. Med. 161 (2000) 1376–1395.
[7] Syed Tasleem, Gulraiz Ali, Farzana Mahdi SikandarHasan, Molecular diagnosis of tuberculosis: a new primer design, Iran. J. Clin. Infect. Dis. 4 (2) (2009) 105–108.
[8] Young S. Kim, Donald E. Shelton, Gregg Barak, Studying juror expectations for scientific evidence: a new model for looking at the CSI myth, J. Am. Judges Assoc. 47 (1–2) (2011) 8–18.
[9] S.A. Oke, A literature review on artificial intelligence, Int. J. Inf. Manage. Sci. 19 (4) (2008) 535–570.
[10] V.P. Masnikosa, The fundamental problem of an artificial intelligence realisation, Cybernetes 27 (1) (1998).
[11] W.E. Halal, Artificial intelligence is almost here, On the Horizon, The Strategic Planning Resource for Education Professionals, vol. 11(2), 2003.
[12] Y. Peng, X. Zhang, Integrative data mining in systems biology: from text to network mining, Artif. Intell. Med. 41 (2) (2007) 83–86.
[13] X. Zhou, B. Liu, Z. Wu, Y. Feng, Integrative mining of traditional chinese medicine literature and medline for functional gene networks, Artif. Intell. Med. 41 (2) (2007) 87–104.
[14] J. Sutojo, E. Mulyanto, V. Suhartono, KecerdasanBuatan, Penerbit ANDI Publisher, Yogyakarta, 2011.
[15] A. Ajith, Nature and scope of AI techniques, in: H.S. Peter, T. Richard (Eds.), Handbook of Measuring System Design, John Wiley & Sons, Ltd., USA, 2005.
[16] V. Cecilia, K. Fabio, Intelligent and expert systems in medicine – a review, in: XVIII CongresoArgentino de Bioingeniería SABI 2011 – VII Jornadas de IngenieríaClínica Mar del PlataSara, 28–30 of September 2011, 2011, pp. 326–331.
[17] N. Mehdi, Y. Mehdi, Designing a fuzzy expert system for diagnosing the hepatitis B intensity rate and comparing it with adaptive neural network fuzzy expert system, in: Proceedings of the World Congress on Engineering and Computer Science (WCECS), vol. 11, San Francisco, USA, 2009.
[18] O.W. Samuel, M.O. Omisore, B.A. Ojokoh, A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever, Exp. Syst. Appl. 40 (2013) (2013) 4164–4171.
[19] A. Imianvan, F. Anosike, C. Obi, An expert system for the intelligent diagnosis of HIV using fuzzy cluster means algorithm, Glob. J. Comp. Sci. Technol. 11 (12) (2011) 73–80.
[20] B.K. Bose, Modern Power Electronics and AC Drives, Prentice-Hall, Upper Saddle River, NJ, 2002, pp 44.
[21] J. Zhao, B.K. Bose, Evaluations of membership functions for fuzzy logic controlled induction motor drive, in: Proceedings of the IEEE, 2002.
[22] B.A. Ojokoh, M.O. Omisore, O.W. Samuel, T.O. Ogunniyi, A fuzzy logic based personalized recommender system, Int. J. Comp. Sci., Inf. Technol. Sec. 2 (5) (2012) 1008–1015.
[23] T.O. Oladele, J.S. Sadiku, R.O. Oladele, Coactive neuro-fuzzy expert system – a framework for diagnosis of malaria, Afric. J. Comput. ICT 7 (2) (2014) 174–188.
[24] O.U. Obot, F.M. Uzoka, O.C. Akinyokun, J.J. Andy, A neuro-fuzzy decision support model for therapy of heart failure, Int. J. Med. Eng. Inform. 6 (4) (2014) 319–334.
[25] A. Kadhim, A. Alam, H. Kaur, Design and implementation of fuzzy expert system for back pain diagnosis, Int. J. Innov. Technol. Creat. Eng. 1 (9) (2011) 16–22.
[26] A. Barwad, P. Dey, S. Susheilia, Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology, Cytomet. B Clyn Cytom. 82 (2012) 107–111.
[27] P. Dey, A. Lamba, S. Kumari, N. Marwaha, Application of an artificial neural network in the prognosis of chronic myeloid leukemia, Anal. Quant. Cytol. Histol. 33 (6) (2012) 335–339.
[28] K. Vahid, A.M. Gholam, Artificial Intelligence in Medicines, Information Technology Department, School of Engineering, Terbiat Moderas University Tehran, Iran, 2009.
[29] A.J. Ramirez, B.D. Knoester, B.H.C. Cheng, P.K. McKinley, Applying genetic algorithms to decision making in autonomic computing systems, in: Proceedings of the 6th International Conference on Autonomic Computing and Communications (ICAC'09), Barcelona, Spain, June 15–19, 2009, 2009.
[30] E. López-González, M.A. Rodríguez-Fernández, C. Mendaña-Cuervo, The logistic decision making in management accounting with genetic algorithms and fuzzy sets, Mathw. Soft Comput. 7 (2000) 229–241.
[31] K. Rajeswari, M. Shindalkar, N. Thorawade, P. Bhandari, DSS using apriori algorithm, genetic algorithm and fuzzy logic, Int. J. Eng. Res. Appl. 3 (4) (2013) 132–136.
[32] X.Y. Djam, Y.H. Kimbi, A medical diagnostic support system for the management of hypertension, J. Sci. Multidisc. Res. 3 (2011) 16–30.
[33] V. Ekong, E. Onibere, A. Imianvan, Fuzzy cluster means system for the diagnosis of liver diseases, Int. J. Comp. Sci. Technol. 2 (3) (2011).
[34] J. Ermine, Expert System: Theory and Practice, Prentice Hall of India, New Delhi, 1995.

[35] H.L. Fang, Genetic Algorithms in Timetabling Scheduling, Doctoral Thesis, University of Edinburgh, UK, 1994.
[36] K. Priya, T. Manju, R. Chitra, Predictive model of stroke disease using hybrid neuro-genetic approach, Int. J. Eng. Comp. Sci. 2 (3) (2013) 781–788.
[37] L. Bechoo, R. Chandrahauns, An empirical analysis on performance evaluation of decision support system (DSS) in business process reengineering, Int. J. Adv. Res. Comp. Sci. Softw. Eng. 3 (7) (2003) 1093–1097.
[38] H.G. Sol, Expert systems and artificial intelligence in decision support systems, in: Proceedings of the Second Mini-Euro Conference, 17–20 November 1985, Springer, Lunteren, The Netherlands, 1987, pp. 1–2.
[39] A. Agunwamba, D. Bloom, A. Friedman, M. Ozolins, L. Rosenberg, D. Steven, M. Weston, Nigeria: The Next Generation: Literature Review, Havard School of Public Health, British Council of Nigeria, 2009, pp. 57.