An Improved Asynchronous Tuberculosis Diagnosis System using Fuzzy Logic Mining Techniques

Morgan O. Obi [1], Eke B. O. [2], Asagba P. O. [3]
Department of Computer Sciences, University of Port Harcourt, Choba, Nigeria.

Abstract - *Tuberculosis is an air borne sickness that could easily be transmitted through numerous mediums like sneezing, coughing, making a song, speaking and so forth. It results from a bacterium named Mycobacterium tuberculosis. Improper diagnosis of this disease can lead to increased fatality and further spread of the disease. This work tends to proffer a diagnostic system that will aid in fast and accurate diagnosis of this disease which will aid in early treatment and isolation of carrier to curtail further spread of the disease which according to World Health Organization (WHO), kills over 4,000 people each day. The proponents made use of Fuzzy Logic Mining Techniques to model uncertainty inherent in diagnosis and implement the system making use of asynchronous techniques which improves the performance of the system and produce results of diagnosis without delay.*

Keywords — *Defuzzification, Fuzzification, Inference System, Linguistic Variables, Matlab, Membership Function.*

I. INTRODUCTION

Tuberculosis is a bacterial disease which in humans is usually caused by an organism called Mycobacterium *tuberculosis (M. tuberculosis).* The common symptoms of active tuberculosis include chronic cough with blood stained sputum, chest pain, fatigue, night sweats, weight loss, fever, nail clubbing (a state in which the ends of the fingers and toes become wide and thick) [1]. The term "consumption" is most times used to portray tuberculosis because of the weight loss experienced by people infected with the disease. Infection of other organs can cause a wide range of symptoms [2].

In step with the Institute of Medicine, Diagnostic mistakes account for about 15 percent of errors that upshot in harm to patients.

Making the right decision in therapeutic conclusion is pivotal as it determines the prognosis to be made for a patient. In such cases, the procedure of reaching a decision should be based on well refined facts and not instincts alone. The unwholesomeness and death rates of tuberculosis patients are sometimes because of the combination of inadequate expertise and the complexity involve in medical practice. Most current diagnostic systems are not readily available for patients to use; this hinders early identification and immediate initiation of treatment for patients with tuberculosis.

Diagnosis based totally on cultured specimens is the reference well-known; however, outcome takes weeks to achieve. Gradual and insensitive diagnostic strategies hampered the worldwide control of tuberculosis, and scientists are looking for early detection strategies, which remain the basis for curtailing tuberculosis [3]. Consequently, the need arises to develop an artificial intelligent system that would facilitate medical professionals diagnose the disease accurately.

Artificial intelligent system has been used efficiently for diagnosing numerous sicknesses. But, little attempt has been undertaken to enhance its diagnosis accuracy.

In an effort to enhance the diagnosis precision, this study brings to light an improved Fuzzy Logic system making use of asynchronous technology which cuts down delay.

II. LITERATURE REVIEW

A fuzzy expert system for the management of malaria designed by [4] attempted to incorporate fuzzy techniques and develop a fuzzy expert system for the management of malaria. Here, the study revealed that the use of fuzzy logic for medical diagnosis provides an efficient way to assist inexperienced physicians to arrive at the final diagnosis of malaria more quickly and efficiently. The developed system provided decision-support platform to assist malaria and Tuberculosis researchers, physicians and other health practitioners in Tuberculosis and malaria endemic regions [4].

Still in the same line of thought, [5] designed a decision-support model for diagnosing tropical diseases using fuzzy logic. The aim of their research was to detect the disease of a patient based on the patient’s complaints and also the level of severity of the patient’s complaints. In another related work, “cognitive analysis of multiple sclerosis utilizing fuzzy cluster means” designed by [6], neuro solutions and crystal reports were used for neural network analysis and graphical representation to aid in the diagnosis of multiple sclerosis, which eliminates the challenges posed by the shortage of medical experts. Several publications have successfully explained the benefits and challenges in using expert system for medical diagnosis. [7] Developed a Fuzzy Expert System for malaria diagnosis. This aids in rapid and accurate diagnosis of malaria. The system was developed to provide supports for medical practitioners and assists malaria researchers to deal with the vagueness, imprecision and time-consuming found in traditional laboratory diagnosis of malaria, and provide accurate output based on the input data.

III. METHODOLOGY

The proposed model uses Fuzzy logic data mining techniques. It seeks to take advantage of the unique capability of fuzzy logic systems in modeling uncertainties to improve the diagnosis of tuberculosis.

An obvious characteristic of medical data is that it is majorly collected for patient health care with secondary consideration for research [8]. As a result, medical data contain many features that create problems for data mining techniques and they most often exist in a format not suitable for direct application of these techniques.

For this reasons, data mining processes are applied to extract useful knowledge that could help for decision-making in tuberculosis diagnosis.

The proposed system will assist medical practitioners in the diagnosis of tuberculosis severity and prescribe appropriate medications based on the patient complaints, the signs and symptoms are input into the system, and diagnosis is done using the knowledgebase gotten from the fuzzy rules.

The Fuzzy Inference System has eleven attributes selected for the research based on their relevance to the problem of discuss. These attribute were used as input variables and an output variable was produced. The inputs made use of the following linguistic variables (mild, moderate, severe and very severe, yes, no) where applicable to describe the membership function. Whereas the output variable is the nature of the disease.

The input parameter of the system under consideration is shown in table 1 and the graphical representation of the input and output variables is shown in Fig. 1.

Table 1: Attributes and their description for MATLAB simulation

SN	ATTRIBUTES	DESCRIPTION	DATA TYPE
1	Cough	Cough for more than 3 weeks	Nominal
2	Chest pain	Pain around the chest	Nominal
3	Hemoptysis	Sputum mixed with blood tint	Nominal
4	Fever	Unusual increase in temperature	Nominal
5	Weight loss	Whether the patient reduce in weight	Nominal
6	Headache	Whether the patient have headache	Nominal
7	Night sweats	Whether the patient sweats unusually at night	Nominal
8	Loss of appetite	Whether the patient is losing weight	Nominal
9	Fatigue	Whether the patient is getting unusually tired	Nominal
10	Age	Patient's age in years	Numeric
11	HIV	Patient's HIV status	Nominal

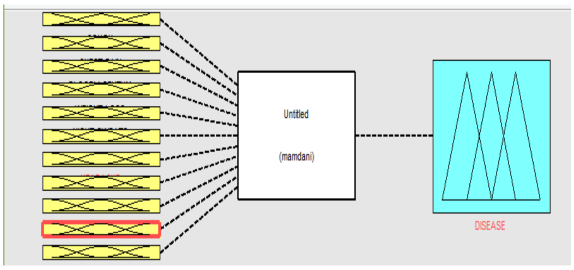

Fig. 1 Graphical representation of the input and output

A. Algorithm for Fuzzy Diagnosis of Tuberculosis

The developed pseudocode and flowchart for the fuzzy diagnostic process of tuberculosis is given below:

Step 1: Input patient's observed signs and symptoms (x) into the system. Where x represents the number of signs and symptoms obtained from the patient.

Step 2: For the symptoms and signs supplied by the patient, search through the knowledge base for matching disease (d) having same signs with patients own.

Step 3: Assign linguistic variables, mild, moderate, severe very severe to the patients presenting symptoms. Where Mild = 1, Moderate = 2, Severe = 3, Very Severe = 4.

Step 4: Apply fuzzy rules.

Step 5: Record fuzzy inputs into their related weighing factors to decide their degree of membership.

Step 6: Establish the rule base evaluating (non-zero minimum values).

Step 7: Establish the firing strength of the rules R.

Step 8: Calculate the degree of truth R, of each rules by evaluating the non-zero minimum value.

Step 9: Calculate the intensity of the disease.

Step 10: Output degree of severity.

B. Components of Fuzzy Inference System (FIS)

Our developed fuzzy logic system has an architecture presented in Fig. 2 below. The development involves fuzzification, inference engine and defuzzification. Our model uses fuzzy logic rather that Boolean logic. It was developed based on the following FIS key components:

i. Knowledge base
ii. Fuzzification.
iii. Fuzzy Inference.
iv. Defuzzification.

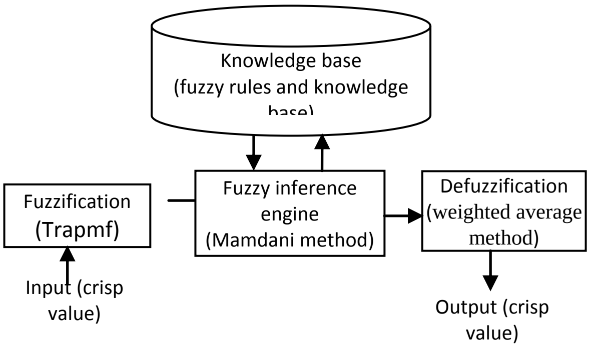

Fig. 2 Architecture of Proposed Fuzzy Logic

IV. RESULTS AND DISCUSSION

The proposed system was simulated using MATLAB (R2010a) using the Fuzzy Logic Design Toolbox version 7.10.0. using the Root Sum Square inference engine and the output was defuzzified using the Weighted Average method and was implemented using Hypertext Preprocessor and MYSQL as the database and WAMP server as the server technology.

Fig. 8 to Fig. 10 shows the graphical construction of the surface view for some inputs against output and the Inference Diagram viewer of Fuzzy Sets generated in the MATLAB Fuzzy Logic Toolbox is shown in Fig. 11.

A. Knowledge Base

This is the component of the system that contains the system's knowledge. It contains succinct description of domain expert's knowledge in tuberculosis symptoms and diagnosis.

The knowledge-based for the proposed model were collected from journals, books, medical records, the Internet and through administered questionnaires/interviews with medical experts. Other sources of the data used in this work include: Records from UCI Repository, Mayo clinic (www.mayoclinic.org), Isabel symptom checker (https://symptomchecker.isabelhealthcare.com), NetDoctor (www.netdoctor.co.uk) etc. We compared the records from all these sources to ensure that the dataset being used is up-to-date.

B. Fuzzification

Fuzzification is the technique of changing a crisp value into a fuzzy value. This is achieved with different types of membership function. There are several types of membership function that can be used for fuzzification process. Some of which are: triangular curve, trapezoidal curve, singleton functions, generalized bell shaped, polynomial curves and Gaussian curves. The trapezoidal membership was selected in the MATLAB software to simulate our work.

The trapezoidal curves depend on four (4) parameters which is given by

$$f(x;a,b,c,d) = \begin{cases} 0 & for\ x < a \\ \dfrac{x-a}{b-a} & for\ a \le x < b \\ 1 & for\ b \le x < c \\ \dfrac{d-x}{d-c} & for\ c \le x < d \\ 0 & for\ d \le x \end{cases}$$

Or more compact as

$$f\ (x;a,b,c,d) = \max\left(\min\left(\frac{x-a}{b-a},1,\frac{d-x}{d-c}\right),0\right)$$

Fuzzification of data is prepared by plotting input parameters into the horizontal axis and projecting vertically to the upper boundary of membership function to determine the degree of membership.

The first step in the development of fuzzy logic based system is to construct fuzzy sets for the parameters. On the basis of domain data collected, both input and output parameters selected for this research were described with the following linguistic variables (mild, moderate, severe and very severe, yes, no). The range of value for each linguistic variable is shown in table 2 below:

Table 2: Range of Fuzzy values

Linguistic variables	**Actual values**	**Linguistic variable Abbreviation**
Mild	0 - 3	MI (1)
Moderate	2 - 5	MO (2)
Severe	4 - 7	SV (3)
Very severe	6 - 10	VS (4)

The fuzzy rules for the model were developed using the record gotten from our research about tuberculosis and interaction with medical experts. The knowledge-base of the designed system has 36 fuzzy rules. These rules together with other rules that specialist in the medical profession will supply to the system as new cases surfaces where used in the system design.

Table 4 below shows sample dataset for tuberculosis. The dataset used for training the system has 11 attributes with a summarized total of 20 instances. The attributes are cough, chest pain, bloody sputum, weight loss, drenching night sweats, fever, headache, loss of appetite, fatigue, age and HIV status. The input parameters were used to generate the rules for the tuberculosis rule base.

Table 3: Attributes and abbreviation

Attributes	Abbreviation
Cough > 3 weeks	CO
Chest pain	CP
Bloody sputum	BS
Weight loss	WL
Drenching night sweat	NS
Fever	FV
Headache	HA
Loss of appetite	LA
Fatigue	FA
Age	AG
HIV status	HIV

Table 4: Dataset of tuberculosis in Indexed format

SN	IF											THEN
	CO	CP	BS	WL	NS	FV	HA	LA	FA	AG	HIV	RESULT
1	3	0	3	3	1	3	0	3	0	0	0	VS
2	2	0	1	1	1	1	0	1	0	0	0	MI
3	0	0	0	0	0	0	3	3	0	0	2	MI
4	1	1	3	1	3	3	1	0	0	0	1	SV
5	2	0	2	2	2	2	0	2	0	0	0	MO
6	4	0	3	3	3	3	0	0	0	0	0	VS
7	4	0	1	3	2	4	0	2	0	0	0	VS
8	1	0	1	1	1	1	0	1	0	0	0	MI
9	3	0	4	1	3	3	0	3	0	0	0	SV
10	2	0	3	2	4	2	0	1	0	0	0	SV
11	4	0	2	2	2	2	0	1	2	0	0	VS
12	3	0	1	3	2	3	0	2	2	0	0	SV
13	2	0	2	1	1	2	0	2	1	0	0	MO
14	1	0	2	1	1	1	0	2	2	0	0	MI
15	1	0	2	1	2	2	0	2	1	0	0	MO
16	3	0	3	3	2	3	0	2	2	0	0	SV
17	2	0	3	3	0	2	2	0	0	0	2	MO

C. Determining Membership Functions

As stated above, we made used of 11 input parameters to assess the possibility of a patient having tuberculosis. The graphical representation of four input variables and the output variables is shown in Fig. 3 - 7.

Cough

Measurement unit in this input is the number of days the patient suffers from constant coughs which has not gotten better.

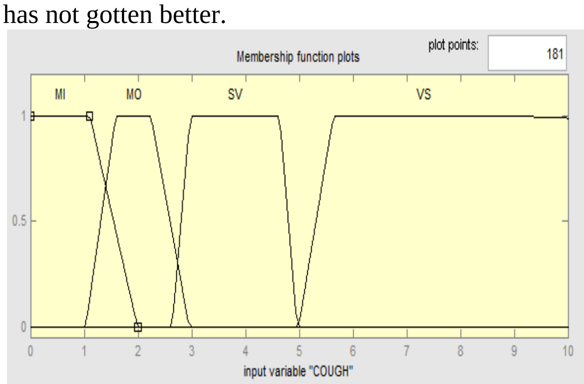

Fig. 3 Input variable cough

Chest Pain

Measurement unit in this input, is the number of days the patient suffers from chest pain.

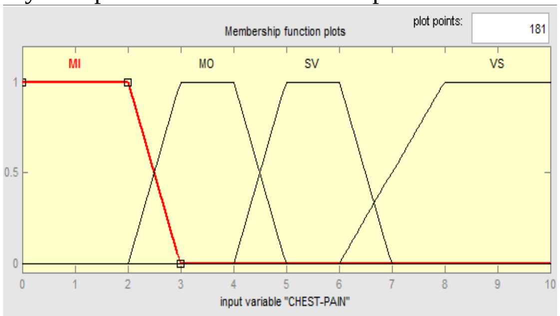

Fig. 4 Input variable chest pain

Bloody Sputum

Measurement unit in this input, is the number of days the patients suffers from persistent bloody sputum which has not gotten better.

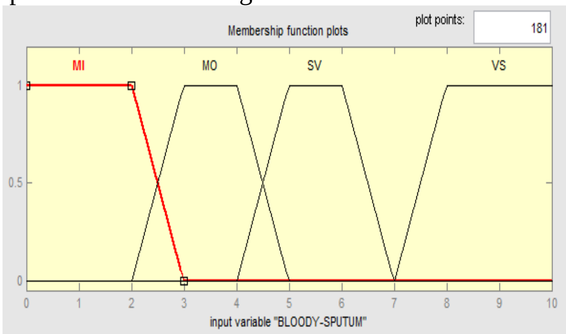

Fig. 5 Input variable bloody sputum

Weight Loss

Measures the percent of weight loss in a patient.

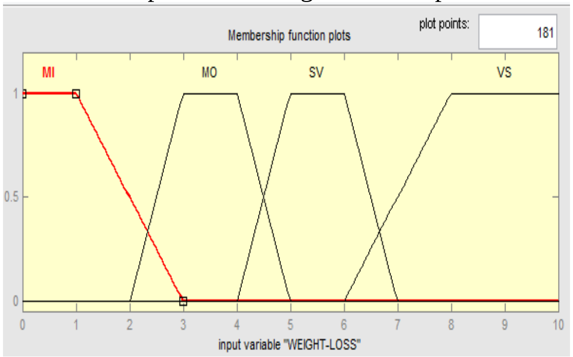

Fig. 6 Input variable weight loss

Output

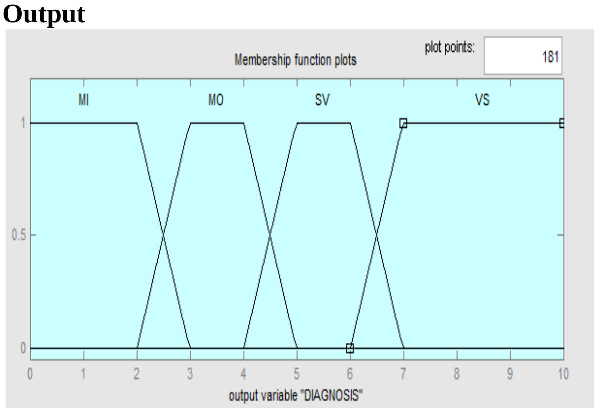

Fig. 7 Output Variable of Disease

D. FUZZY INFERENCE

The inference engine controls how the rules are applied on the facts. This is the part of the rule-based system that makes inferences. It chooses which rules are fired by facts and controls general execution. Also, it matches the facts against the rules to see what rules are valid. The proposed system makes use of forward chaining reasoning to determine the rules that fires. It would make use of the facts given by the patient to diagnose the result of the patient.

Fuzzy inference rules were constructed based on the defined input and output variables. Some of the rules that make up the rule base for the inputs and output model are shown below.

if (cough is SV) and (bloody-sputum is SV) and (weight-loss is SV) and (night-sweat is MI) and (fever is SV) and (loss-appetite is SV) then (diagnosis is VS)

if (cough is MO) and (bloody-sputum is MI) and (weight-loss is MI) and (night-sweat is MI) and (fever is MI) and (loss-appetite is MI) then (diagnosis is MI)

if (headache is SV) and (bloody-sputum is SV) and (HIV-status is Yes) then (diagnosis is MI)

if (cough is MI) and (chest-pain is MI) and (bloody-sputum is SV) and (weight-loss is MI) and (night-sweat is SV) and (fever is SV) and (headache is MI) and (HIV-status is No) then (diagnosis is SV)

if (cough is MO) and (bloody-sputum is MO) and (weight-loss is MO) and (night-sweat is MO) and (fever is MO) and (loss-appetite is MO) then (diagnosis is MO)

if (cough is VS) and (bloody-sputum is SV) and (weight-loss is SV) and (night-sweat is SV) and (fever is SV) and (loss-appetite is SV) then (diagnosis is VS)

if (cough is VS) and (bloody-sputum is MI) and (weight-loss is SV) and (night-sweat is MO) and (fever is VS) and (loss-appetite is MO) then (diagnosis is VS)

if (cough is MI) and (bloody-sputum is MI) and (weight-loss is MI) and (night-sweat is MI) and (fever is MI) and (loss-appetite is MI) then (diagnosis is MI)

if (cough is SV) and (bloody-sputum is VS) and (weight-loss is MI) and (night-sweat is SV) and (fever is SV) and (loss-appetite is SV) then (diagnosis is SV)

if (cough is MO) and (bloody-sputum is SV) and (weight-loss is MO) and (night-sweat is VS) and (fever is MO) and (loss-appetite is MI) then (diagnosis is VS)

if (cough is VS) and (bloody-sputum is MO) and (weight-loss is MO) and (night-sweat is MO) and (fever is MO) and (loss-appetite is MI) and (fatigue is MO) then (diagnosis is VS)

if (cough is SV) and (bloody-sputum is MI) and (weight-loss is SV) and (night-sweat is MO) and (fever is SV) and (loss-appetite is MO) and (fatigue is MO) then (diagnosis is SV)

As can be seen, the rules consists of two parts: the IF part, called the predecessor or preface or condition and the THEN part called the resulting or conclusion or activity. A rule could have multiple precursors joined with the aid of the keywords AND (conjunction), OR (disjunction) or a mix of each.

The inference engine uses the Root Sum Square technique to evaluate the rules that fired in the rule base. The Root Sum Square is given by as:

$$X_{RSS} = \sqrt{\sum_{n=1}^{N} |X_n|^2} \quad (1)$$

X_n are estimations of various guidelines which have a similar conclusion in the fuzzy rule-base, that is, X = value of fired rule.

The square root of the sum of the square of each conclusion that fires is taken which gives a fuzzy value for each of the conclusion [7].

$$X_{mild} = \sqrt{x_{mi\,1}^2 + x_{mi\,2}^2 + x_{mi\,3}^2 + \dots + x_{min}^n} \quad (2)$$

$$X_{moderate} = \sqrt{x_{mo\,1}^2 + x_{mo\,2}^2 + x_{mo\,3}^2 + \dots + x_{mon}^n} \quad (3)$$

$$X_{severe} = \sqrt{x_{sv1}^2 + x_{sv2}^2 + x_{sv3}^2 + \dots + x_{svn}^n} \quad (4)$$

$$X_{very\ severe} = \sqrt{x_{vs1}^2 + x_{vs2}^2 + x_{vs3}^2 + \dots + x_{vsn}^n} \quad (5)$$

The value gotten for each severity is passed on to the deffuzification stage where a crisp value is gotten using the weighted average deffuzification method.

E. DEFUZZIFICATION

This involves changing fuzzy output back into numerical values for system action. The result from the inference engine is expressed as a crisp value which is more precise than the fuzzy result. The input to the defuzzification process is a fuzzy set while the output of the defuzzification process is a crisp value. Four commonly used defuzzication methods are: Maximum Membership Method or Height Method, Centroid Method, Weighted Average Method or Symmetric Membership Method and Min Max Method or Middle of Maximum (Chakraverty). In this work, we made use of the Weighted Average method for deffuzification. The Weighted Average method is expressed as:

$$Z^* = \frac{\sum \mu_{\bar{c}}(z).Z}{\sum \mu_{\bar{c}}(z)} \quad (6)$$

where Z^* is the crisp output and $\mu_{\tilde{c}}(z)$ is the aggregated membership function and Z is the center of membership function.

$$crisp\ value = \frac{(X_{mi} * Z_{mi}) + (X_{mo} * Z_{mo}) + (X_{sv} * Z_{sv}) + (X_{vs} * Z_{vs})}{(X_{mi} + X_{mo} + X_{sv} + X_{vs})} \quad (7)$$

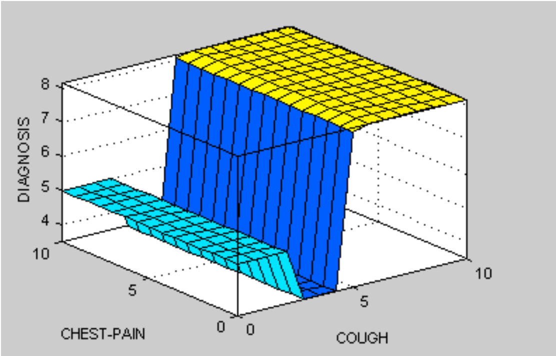

Fig. 8 Surface view for the inputs Chest pain and cough against the diagnosis.

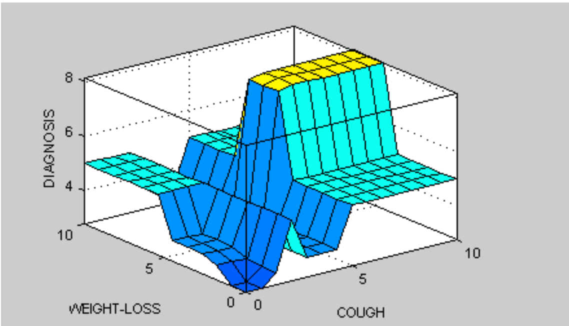

Fig. 9 Surface view for the inputs weight loss and cough against the diagnosis.

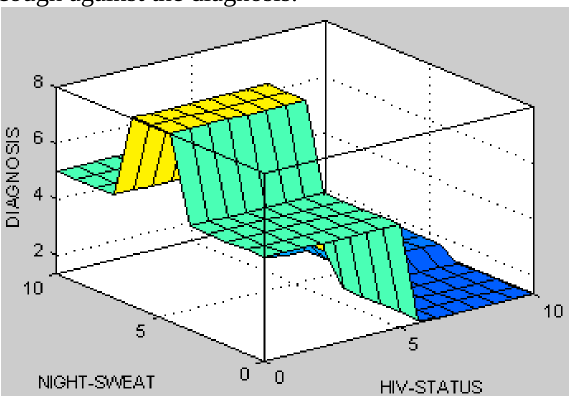

Fig. 10 Surface view for inputs night sweat and HIV status against diagnosis

REFERENCES

[1] WHO. (2016). Chest Radiography in Tuberculosis Detection. Switzerland: Knut Lönnroth.

[2] Gerald, L. M., John, E. B., & Raphael. (2010). Mandell, Douglas, and Bennett's principles and practice of infectious diseases. Philadelphia:Churchill Livingstone/Elsevier.

[3] Mahmoud, R. S., Shahaboddin, S., Shahram, G., Teh, Y. W., Aghabozorgi, S., Laiha, M. K., et al. (2014). RAIRS2 a new expert system for diagnosing tuberculosis with real-world tournament selection mechanism inside artificial immune recognition system. International Federation for Medical and Biological Engineering.

[4] Djam, X.Y. and Y.H. Kimbi. (2011), A Decision Support System for Tuberculosis Diagnosis. Pacific Journal of Science and Technology. 12(2):410-425.

[5] Olabiyisi,S. O, Omidiora, E. O, Olaniyan, M. O & Derikoma, O. A Decision Support System Model for Diagnosing Tropical Diseases Using Fuzzy Logic. Afr J. of Comp & ICT. Vol 4, No. 2. Issue 2: 1-6 (2011).

[6] Imianvan, A. A and Obi J, C (2012). Cognitive analysis of multiple sclerosis utilizing fuzzy cluster means. International Journal of Artificial Intelligence & Applications (IJAIA), 3(1).

[7] Onuwa, B. O. (2014). Fuzzy Expert System for Malaria Diagnosis. Oriental Journal of Computer Science and Technology , 7(2), 273 - 284.

[8] Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2012). Medical Data Classification with Naïve Bayes Approach. International Technology Journal , 1166- 1174.