International Journal of Big Data and Analytics in Healthcare
Volume 5 • Issue 2 • July-December 2020

Fuzzy Logic-Based Predictive Model for the Risk of Sexually Transmitted Diseases (STD) in Nigeria

Jeremiah A. Balogun, Mountain Top University, Nigeria
Florence Alaba Oladeji, University of Lagos, Nigeria
Olajide Blessing Olajide, Federal University, Wukari, Nigeria
Adanze O. Asinobi, University of Ibadan, Nigeria
Olayinka Olufunmilayo Olusanya, Computer Science Department, Tai Solarin University of Education, Nigeria
Peter Adebayo Idowu, Department of Computer Science and Engineering, Obafemi Awolowo University, Nigeria
https://orcid.org/0000-0002-3883-3310

ABSTRACT

This study developed a classification model for monitoring the risk of sexually transmitted diseases (STDs) among females using information about non-invasive risk factors. Structured interview with physicians was done in order to identify the risk factors that are associated with the risk of STDs in Nigeria. The model was simulated using the fuzzy logic toolbox accessible in the MATLAB® R2015a Software. The results showed that nine non-invasive risk factors were associated with the risk of STDs among female patients in Nigeria. Two, three, and four triangular membership functions were appropriate for the formulation of the linguistic variables of the factors while the target risk was formulated using four triangular membership functions for the linguistic variables namely no risk, low risk, moderate risk, and high risk. The study concluded that the fuzzy logic model approach was adequate for predicting the risk of STDs based on the knowledge of the risk factors.

KEYWORDS

Classification Model, Fuzzy Logic, Membership Function, Risk Factor, Simulation, STD

1. INTRODUCTION

Sexually transmitted diseases (STDs) are infections that can be transferred from one person to another during sexual activity (Bryan, 2011). In developing countries like Nigeria, STDs exhibit a higher incidence and prevalence rates than developed countries (Adebowale, Titloye, Fagbamigbe, & Akinyemi, 2013). STDs remain a major public health challenge because of their health consequences and severe complications, especially among women who excessively bear their long-term consequences. Among, the most popular STDs are HIV/AIDS and Syphilis (Lakshmi and Isakki, 2017). HIV is a human immunodeficiency virus. It is the virus that can lead to acquired immunodeficiency syndrome or

DOI: 10.4018/IJBDAH.2020070103

AIDS if not treated (Ojunga, Peter, Otulo, Omollo, Edgar, 2014). Others include chlamydia, cancroid, pubic lice (crabs), genital herpes, hepatitis B, scabies, human papillomavirus (HPV), gonorrhea etc.

The Nigerian Health sector has set ambitious targets for providing essential health services to all citizens; improving the quality of decisions affecting treatment options for people at risk of STDs is very essential to reducing disease risk in Nigeria. Across the world, the number of cases of sexually transmitted diseases (STDs) are continuously on a steady increase due to the number of unprotected sex, lack of quality and timely sexual education and on most part ignorance. The need for an effective means of reducing the number of associated cases of STDs is necessary owing to the steady increase in the number of sex-related professions, such as sexual workers alongside those who engage in unprotected sex to mention a few.

Information and Communications Technology (ICT) has the capability of improving the quality, efficiency and safety of health care and allows health care providers to collect, store, retrieve, and transfer information by electronic means (Shekelle, Morton, & Keeler, 2006). Predictive research, which aims to predict future events or outcomes based on patterns within a set of variables, has become increasingly popular in medical research. Accurate predictive models can inform patients and physicians about the future course of an illness or the risk of developing illness and thereby help guide decisions on screening and/or treatment (Waijee, Higgins, & Singal, 2013).

Expert systems (also known as knowledge-based systems) are computer programs that aim to achieve the same level of accuracy as human experts when dealing with complex, poorly structured problems in a particular area, which allows non-specialists to use them for receiving answers to problems or for experts to gain decision support (Turban, Sharda, & Delen, 2010). The strength of these systems lies in their ability to use expert knowledge almost when an expert is unavailable. Expert systems make knowledge more accessible and help solve the problem of translating knowledge into practical useful results. As a result of this, fuzzy logic is used mainly to eliminate the uncertainties in human-oriented analysis as a way of processing complex, inaccurate, uncertain, and vague data.

Fuzzy logic is an extended set of traditional (Boolean) logic designed to access the concept of partial truth (Massad, Ortega, de Barros, & Struchiner, 2008). Fuzzy logic is associated with the morphology of logical inference, which can apply the approaches of human thinking to knowledge-based systems. Fuzzy logic is a computational approach based on *degrees of truth,* and not on the classic *truth or false* Boolean logic (1 or 0). This is because it is difficult to assign a natural language in absolute terms 0 and 1. Fuzzy logic uses 0 and 1 as extreme cases of truth, but also contains various states of truth represented by the meanings that lie between them. Fuzzy logic was introduced in order to solve the problem of imprecision and uncertainty, so as to improve tractability, robustness and low-cost solutions for real world problems (Sharareh and Xiao-Jun, 2009).

Fuzzy logic generally, uses simple rules to describe the system of interest and this makes it easy to implement. These are usually represented as a set of IF-THEN rules. In addition, Fuzzy logic has the ability to express the ambiguity of human thinking and translate expert knowledge into computable numerical data (Cheng, 2004). In developing nations in sub-Saharan Africa such as Nigeria, there are no systems in place that can help provide a means via which individuals who live a social life of casual sex can be able to access their risk of having STDs. Therefore, there is a need for the development of a system which requires the use of non-invasive risk factors for the assessment of STDs in an individual. Predictive models for the classification of the risk of STDs using fuzzy logic can allow physicians to make use of the risk factors assessed from individuals in order to determine the risk of STDs thus improving quality healthcare delivery.

2. RELATED WORKS

Idowu *et al.* (2019), worked on the classification of the CD4 count of HIV patients using machine learning algorithm. The study identified the factors that are associated with the estimation of CD4 count among HIV patients then collected relevant data. The classification model was formulated using decision trees (DT) then compared with the performance of support vector machines (SVM) and multi-layer perceptron (MLP). The results revealed that using the DT had a better performance compared with using SVM and MLP algorithms. The study was limited to the classification of CD4 count of HIV patients.

Khormehr and Maihami (2016), applied fuzzy logic to the development of a predictive model for the risk of STDs using invasive and non-invasive factors. The study identified a number of invasive and non-invasive parameters that were associated with the risk of STDs. The cuckoo search algorithm was applied to the reduction of the variables before the model was formulated using fuzzy membership functions. The results of the study showed that the predictive model showed an accuracy of 95% following validation. The study was limited to the use of both invasive and non-invasive risk factors for the identification of the risk of STDs.

Idowu *et al.* (2015), applied fuzzy logic theory to the development of a survival model for Sickle Aneamia Disease (SCD) among pediatrics in Nigeria. The study identified three (3) factors that were associated with severity of the survival of SCD among pediatrics. The study adopted the use of triangular membership functions for the formulation of the classification model for the severity of survival based on the identified factors. The results of the study showed that using the IF-THEN rules formulated and provided by experts, the outcome of the severity of SCD survival was easily assessed from pediatrics. The study was limited to the development of a classification model for the survival of SCD patients.

3. MATERIALS AND METHODS

In this study, fuzzy logic model was adopted for the development of a classification model which is required for the assessment of the risk of STD in an individual. Therefore, the various variables which are associated with the assessment of the risk of STD were identified from gynecologists selected from tertiary institutions in Nigeria. These information was elicited from the experts via the use of interview. The various variables which were associated with the risk of STD were referred to a risk factors and were further classified as socio-demographic and social lifestyle related information. The linguistic variables for each risk factor were identified and associated with a crisp value within a fixed interval which was used to measure the degree of association with the risk of STD. Therefore, the higher the crisp value then the higher the degree of association with the risk of STD. The crisp intervals were required for defining the association of the degree of membership of each linguistic variable with the risk of STD using membership functions. Table 1 shows information about the various risk factors that were identified to be associated with the risk of STD.

The socio-demographic information that are associated with the risk of STD, include: marital status of the individual, socio-economic status and toilet facility used. The marital status of an individual was defined using two (2) linguistic variables, namely: married and single. The risk of STD was identified as been more associated with been single than been married. Therefore, a crisp value of

Table 1. Description of the socio-demographic risk factors

Risk Factors	Linguistic Variables	Crisp Value
Marital Status	Married Single	0 1
Socio-Economic Status	Upper Class Middle Class Lower Class	0 1 2
Toilet Facility Used	Personal Shared	0 1

0 was used to define married while a crisp value of 1 was used to define single. The socio-economic status of an individual was defined using three (3) linguistic variables, namely: lower class, middle class and lower class. The risk of STD was identified as been more associated with individuals of the lower class, followed by middle class and least with individuals of the upper class. Therefore, a crisp value of 0 was used to define upper class individuals, 1 used to define middle class individuals while 2 was used to define lower class individuals. The type of toilet facility used by an individual was defined using two (2) linguistic variables, namely: personal and shared. The risk of STD was identified as been more associated with individuals who used shared toilet facility than those who used personal toilet facility. Therefore, a crisp value of 0 was used to define personal toilet facility while a crisp value of 1 was used to define shared toilet facility. Table 2 shows information about the various social-lifestyle related risk factors that were identified to be associated with the risk of STD.

Table 2. Description of the social-lifestyle related risk factors

Risk Factors	Linguistic Variables	Crisp Value
Age at First Sexual Intercourse	Never Above 18 years Between 14 and 18 years Below 14 years	0 1 2 3
Sex Protection Practice	Always Sometimes Never	0 1 2
Sexual Activity (in last 2 weeks)	Inactive Active	0 1
Lifetime Number of Sexual Partners	None or One More than One	0 1
Casual Sex Practice	No Yes	0 1
History of STD	No Yes	0 1

The social-lifestyle related information that are associated with the risk of STD, include: age at first sexual intercourse, sex protection practice, sexual activity (in last 2 weeks), lifetime number of partners and casual sex practice. The age at first sexual intercourse of an individual was defined using four (4) linguistic variables, namely: never, above 18 years, between 14 and 18 years and below 14 years. The risk of STD was identified as been more associated with individuals whom were below 14 years, followed by those whom were between 14 and 18 years, those whom were above 18 years and those whom had never had sex had the lowest association. Therefore, for the age of first sexual intercourse, a crisp value of 0 was used to define those that had never had sex, 1 to define those whom were above 18 years, 2 to define those whom were between 14 and 18 years and 3 to define those whom were below 14 years.

The practice of sex protection by an individual was defined using three (3) linguistic variables, always, sometimes and never. The risk of STD was identified as been more associated with individuals who never practices sex protection followed by those whom sometimes practices and the least among individuals whom always practices sex protection. Therefore, a crisp value of 0 was used to define individuals who always practices sex protection, 1 used to define individuals whom sometimes practiced sex protection while 2 was used to define individuals whom never practices sex protection. Sexual activity in the last 2 weeks by an individual was defined using two

(2) linguistic variables, namely: inactive and active. The risk of STD was identified as been more associated with individuals who had active sex than those who had an inactive sex life. Therefore, a crisp value of 0 was used to define those whom were inactive while a crisp value of 1 was used to define individuals whom were active.

The lifetime number of sex partners of an individual was defined using two (2) linguistic variables, namely: none or one and more than one. The risk of STD was identified as been more associated with individuals who had more than one lifetime sex partner than those who had none or one lifetime sex partner. Therefore, a crisp value of 0 was used to define individuals who had none or one lifetime sex partners while a crisp value of 1 was used to define individuals who had more than one lifetime sex partners. The practice of casual sex among individuals was defined using two (2) linguistic variables, namely: no and yes. The risk of STD was identified as been more associated with individuals who practiced casual sex than those who did not practice casual sex. Therefore, a crisp value of 0 was used to define individuals who did not practice casual sex life while a crisp value of 1 was used to define individuals who practices casual sex.

History of STD among individuals was defined using two (2) linguistic variables, namely: no and yes. The risk of STD was identified as been more associated with individuals who had a history of STD than those who did not have STD history. Therefore, a crisp value of 0 was used to define individuals who did not have history of STD while a crisp value of 1 was used to define individuals who had history of STD. Furthermore, the sum of the crisp values for all the risk factors was used to define intervals required for defining the risk of STD among individuals. Table 3 shows a description of the various intervals that were used for the classification of the risk of STD based on the crisp values of the risk factors selected.

Table 3. Classification of risk of STDs

Linguistic Variable	Crisp Interval	Crisp Value
No Risk	0	0
Low Risk	[1, 4]	1
Moderate Risk	[5, 9]	2
High Risk	[10, 13]	3

The total crisp values for the risk of STD was classified into four (4) classes using crisp intervals. A total crisp value of 0 was used to identify No Risk of STD, a total crisp interval of [1, 4] was used to identify Low Risk of STD, a total crisp interval of [5, 9] was used to identify Moderate Risk of STD while a total crisp interval of [10, 13] was used to identify High Risk of STD among individuals. Therefore, a crisp value of 0 was mapped to No Risk of STD, a crisp value of 1 was mapped to Low Risk of STD, a crisp value of 2 was mapped to Moderate Risk of STD while a crisp value of 3 was mapped to High Risk of STD.

Fuzzy Inference Systems (FIS) are widely used to model or control processes and can be developed based on expert knowledge or data. The fuzzy inference mechanism consists of three phases, namely: fuzzification, inference rules generation and defuzzification. Fuzzification was used for identifying linguistic variables of input variables as numerical inputs using mathematical membership functions with some degree of compatibility. The processing of inference rules generation involved the use of If-Then statements to communicate a combination of the values of the set of identified input variables with respect to values of the risk of STD. Defuzzification involves the conversion of the results of aggregated rules into a numerical value required for generating the linguistic variables of the risk of

STD. This procedure allows the use of fuzzy categories in the representation of words and abstract ideas of a person in the description of decision-making procedures.

The triangular membership function was used to formulate the linguistic values associated with each identified risk factor during the fuzzification process. This is because the triangular membership function was appropriate for specifying crisp values used for describing each linguistic value as the center apex point, *b* of each defined triangle according to an interval [a, b, c]. Equation (1) shows the relationship between the crisp interval [a, b, c] of each linguistic value suggested for this study and their respective fuzzified values:

$$linguistic\_value;\left[x : a,b,c\right] = \begin{cases} 0; x \leq a \\ \dfrac{x-a}{b-a}; a < x \leq b \\ \dfrac{c-x}{c-b}; b < x \leq c \\ 0; x > c \end{cases} \tag{1}$$

For this study, the crisp interval [a, b, c] that was formulated for the linguistic variables of the identified risk factors that are associated with the risk of STD are presented as follows. The crisp interval of triangular membership functions that was used to define linguistic variables with center 0 was [-0.5, 0, 0.5]. The crisp interval triangular membership functions that was used to define linguistic variables with center 1 was [0.5, 1, 1.5]. The crisp interval triangular membership functions that was used to define linguistic variables with center 2 was [1.5, 2, 2.5]. The crisp interval triangular membership functions that was used to define linguistic variables with center 3 was [2.5, 3, 3.5].

Therefore, risk factors with 2 linguistic variables used 2 triangular membership functions with centers 0 and 1; risk factors with 3 linguistic variables used 3 triangular membership functions with centers 0, 1 and 2; and risk factors with 4 linguistic variables used 4 triangular membership functions with centers 0, 1, 2 and 3. Also, the crisp interval [a, b, c] that was formulated for the linguistic values of the risk of STDs, namely: No Risk, Low risk, Moderate Risk and High Risk were [-0.5, 0, 0.5], [0, 5, 1, 1.5], [1.5, 2, 2.5] and [2.5 3 3.5] respectively with centers 0, 1, 2 and 3 respectively.

Finally, the inference rules for the inference engine of the fuzzy model was formulated based on the combination of the possible linguistic values of the identified 9 risk factors that are associated with the risk of STDs. This was done by determining the permutation of all linguistic variables of the 9 risk factors according to Equation (2):

$$Number\ \ of\ \ Rules = \prod_{i=1}^{9} n\left(linguistic\ \ variables\right)_{risk-factor,\ i} \tag{2}$$

where $n\left(linguistic\ \ variables\right)_{risk-factor,\ i}$ is number of linguistic variables of risk factor *i*.

According to the equation shown above, the number of rules inferred for the risk of STDs is equal to the product of the number of linguistic variables of all identified risk factors. From the identified risk factors, it was observed that 6 risk factors have 2 linguistic variables, 2 risk factors have 3 linguistic variables and 1 risk factor has 4 linguistic variables. According to Equation (2), the total number of rules inferred for the fuzzy logic model required for risk of STDs would require 2304 ($= 2^6 \times 3^2 \times 4^1$) rules in total. Therefore, 2304 rules were inferred for determining the risk of STDs based on the 9 identified associated risk factors. The rules were inferred by using *If-Then statements* which were used to combine the various possible linguistic variables of the identified risk factors as the antecedent (the *If part*) while their respective output linguistic value of the risk of STDs (the *Then part*) were adopted as the consequent part of the rules. Table 4 shows the description of some of the rules that were inferred from the collection of 2304 rules required for the assessment of the risk of STD based on the combination of the linguistic variables of the 9 associated risk factors.

Table 4. Description of some of the rules inferred for STD risk

V1	V2	V3	V4	V5	V6	V7	V8	V9	STD Risk
Married	Upper	Personal	Never	Always	Inactive	0 or 1	No	No	No Risk
Married	Upper	Personal	Never	Always	Inactive	0 or 1	No	Yes	Low Risk
Married	Upper	Personal	Never	Always	Inactive	0 or 1	Yes	No	Low Risk
Married	Upper	Personal	Never	Always	Inactive	1 or more	No	No	Low Risk
Single	Lower	Shared	Below 14	Never	Active	1 or more	Yes	Yes	High Risk

4. RESULTS

This section presents the results of this study which includes the fuzzification of the linguistic variables of the risk factors associated with the risk of STDs using the triangular membership function. As stated earlier, the number of triangular membership function adopted for the fuzzification of the risk factors were based on the number of linguistic variables of each risk factor. The triangular membership functions were formulated as function of the values of a, b and c as identified in their respective interval [a, b, c]. For each linguistic value, the value of *a* was used to define the left base point, *c* to identify the right base point and *b* to identify the central apex point which corresponds to the crisp value of the linguistic variable.

The results showed that all triangular membership functions adopted for the fuzzification of the risk factors associated with the risk of STDs were formulated based on the central values of the crisp intervals defined for each linguistic variables, namely: 0, 1, 2 and 3. Also, the risk of STDs was also formulated using 4 triangular membership functions with centers 0, 1, 2 and 3 for No Risk, Low Risk, Moderate Risk and High Risk respectively.

The triangular membership function with center 0 was formulated based on the interval [-0.5, 0, 0.5] according to Equation (3) such that the extreme points had crisp values of -0.5 and 0.5. The triangular membership function with center 1 was formulated based on the interval [0.5, 1, 1.5] according to Equation (4) such that the extreme points had crisp values of 0.5 and 1.5. The triangular membership function with center 2 was formulated based on the interval [1.5, 2, 2.5] according to Equation (5) such that the extreme points had crisp values of 1.5 and 2.5. The triangular membership function with center 3 was formulated based on the interval [2.5, 3, 3.5] according to Equation (6) such that the extreme points had crisp values of 2.5 and 3.5:

$$crisp-value\_0;\left[x:-0.5,0.0,0.5\right]=\begin{cases} 0; x \le -0.5 \\ \dfrac{x+0.5}{0.5}; -0.5 < x \le 0 \\ \dfrac{0.5-x}{0.5}; 0 < x \le 0.5 \\ 0; x > 0.5 \end{cases} \quad (3)$$

$$crisp-value\_1;\left[x:0.5,1.0,1.5\right]=\begin{cases} 0; x \le 0.5 \\ \dfrac{x-0.5}{0.5}; 0.5 < x \le 1 \\ \dfrac{1.5-x}{0.5}; 1 < x \le 1.5 \\ 0; x > 1.5 \end{cases} \quad (4)$$

$$crisp-value\_2;\left[x:1.5,2.0,2.5\right]=\begin{cases} 0; x \le 1.5 \\ \dfrac{x-1.5}{0.5}; 1.5 < x \le 2 \\ \dfrac{2.5-x}{0.5}; 2 < x \le 2.5 \\ 0; x > 2.5 \end{cases} \quad (5)$$

$$crisp-value\_3;\left[x:2.5,3.0,3.5\right]=\begin{cases} 0; x \le 2.5 \\ \dfrac{x-2.5}{0.5}; 2.5 < x \le 3 \\ \dfrac{3.5-x}{0.5}; 3 < x \le 3.5 \\ 0; x > 3.5 \end{cases} \quad (6)$$

The fuzzification of the fuzzy inference system of the fuzzy logic model required for the risk of STD was simulated using the MATLAB Fuzzy Logic Toolbox available in the MATLAB R2015a. The source code for the simulated fuzzy logic model excluding the 2304 rules is presented in the Appendix. The source code consists of the specification of the fuzzy model which were defined under the *system* component of the code which contained the model type as *Mamdani*, number of inputs as *9*, number of output as *1*, number of rules as *2304*, and method as *minimum* fuzzy value, or method as *maximum* fuzzy value, implication method as *minimum*, aggregation method as *maximum* while the defuzzification method was *centroid*. Following this, the description of the fuzzification of the 9 input alongside the risk of STDs were specified based on their name, range of crisp values, number of membership functions within range alongside the interval of each triangular membership function used to formulate the linguistic variables. Figure 1 shows the screenshot of the source code of the fuzzy logic model for risk of STDs using the MATLAB Fuzzy Logic Toolbox as viewed via notepad++ software.

Figure 1. Screenshot of source code of Fuzzy logic model for risk of STDs

```
[System]
Name='Fuzzy Logic Predictive Model for the Risk of STDs'
Type='mamdani'
Version=2.0
NumInputs=9
NumOutputs=1
NumRules=2304
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='Marital-Status'
Range=[-0.5 1.5]
NumMFs=2
MF1='Married':'trimf',[-0.5 0 0.5]
MF2='Single':'trimf',[0.5 1 1.5]

[Input2]
Name='Socio-economic-status'
Range=[-0.5 2.5]
NumMFs=3
MF1='High-class':'trimf',[-0.5 0 0.5]
MF2='Middle-class':'trimf',[0.5 1 1.5]
MF3='Low-class':'trimf',[1.5 2 2.5]

[Input3]
Name='Toilet-facility-use'
Range=[-0.5 1.5]
NumMFs=2
MF1='Personal':'trimf',[-0.5 0 0.5]
MF2='Shared':'trimf',[0.5 1 1.5]

[Input4]
Name='Age-at-first-sexual-intercourse'
Range=[-0.5 3.5]
NumMFs=4
MF1='Never':'trimf',[-0.5 0 0.5]
```

The results of the simulation of the model for the marital status is expressed in Figure 2 such that the interval [-0.5, 0.5] with center 0 was used to model married while [0.5, 1.5] with center 1 was used to model single. The results of the simulation of the model for socio-economic status is shown in Figure 3 such that the interval [-0.5, 0.5] with center 0 was used to model high class, [0.5, 1.5] with center 1 was used to model middle class and [1.5 2.5] with center 2 was used to model low class. The results of the simulation of the model for toilet facility use is shown in Figure 4 such that

Figure 2. Fuzzification of marital status

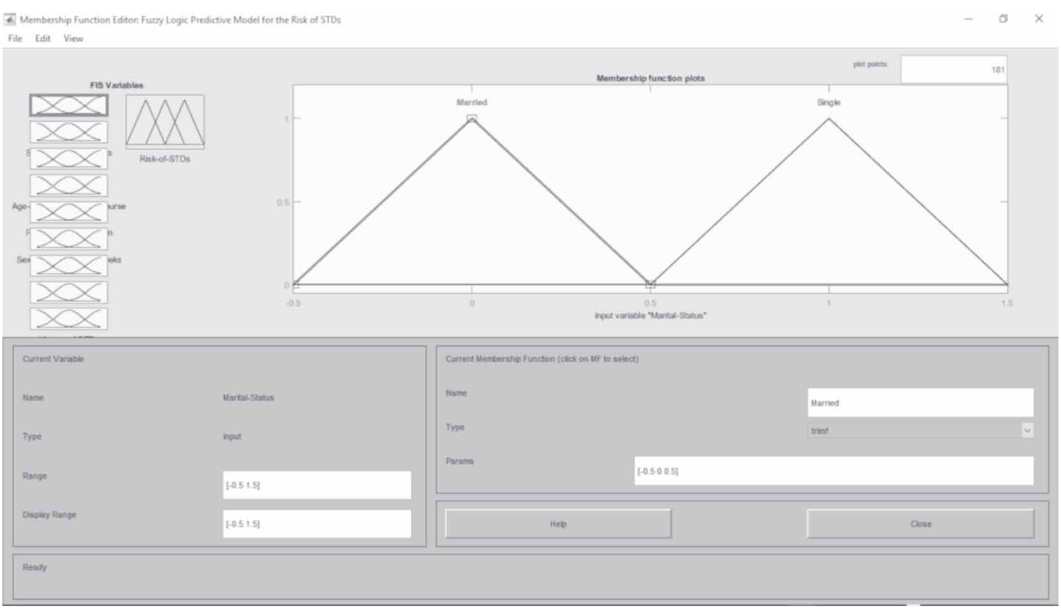


Figure 3. Fuzzification of socio-economic status

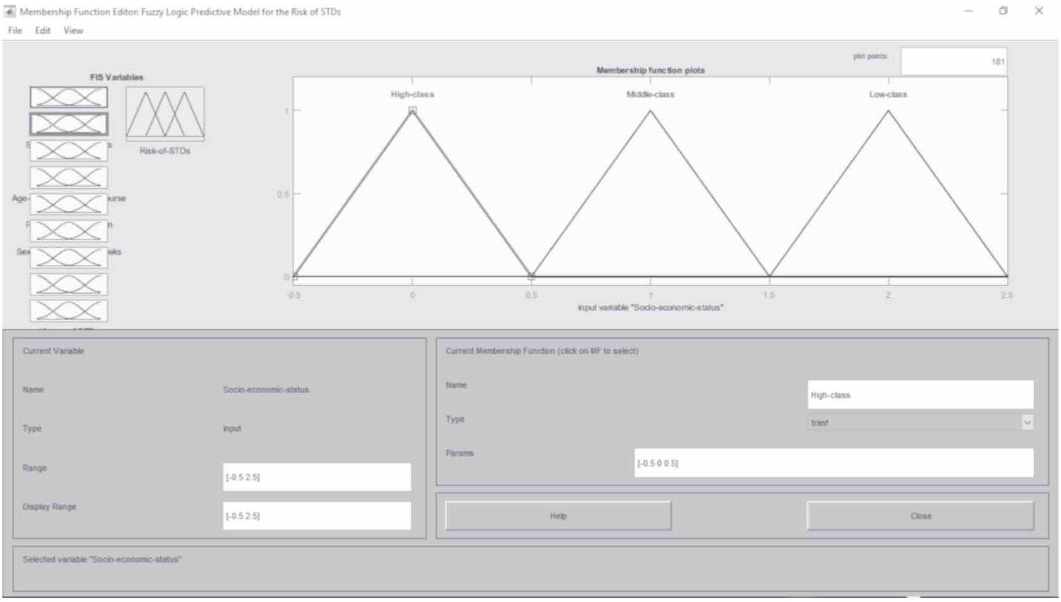
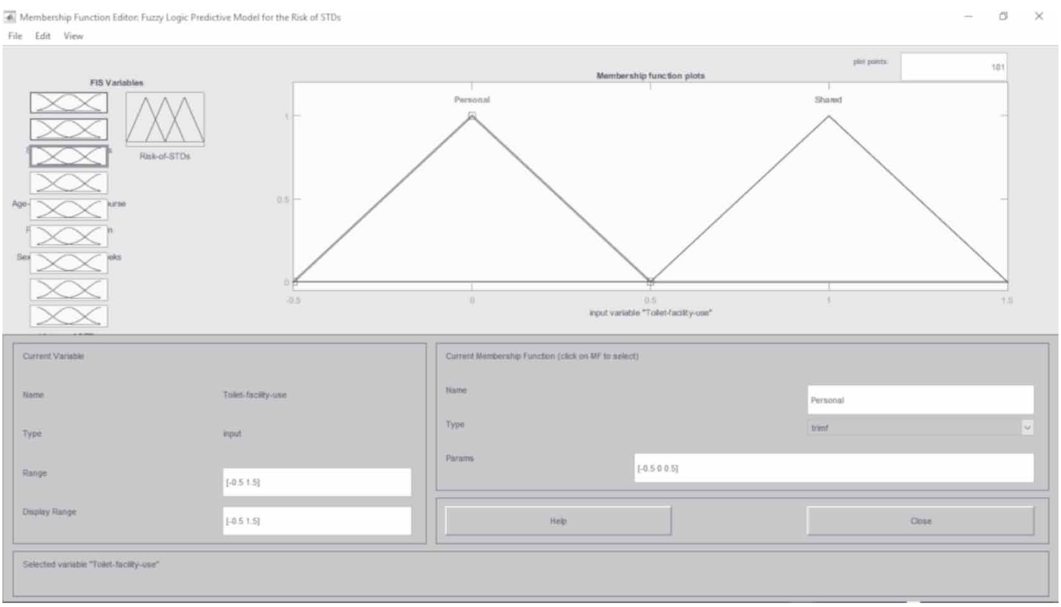


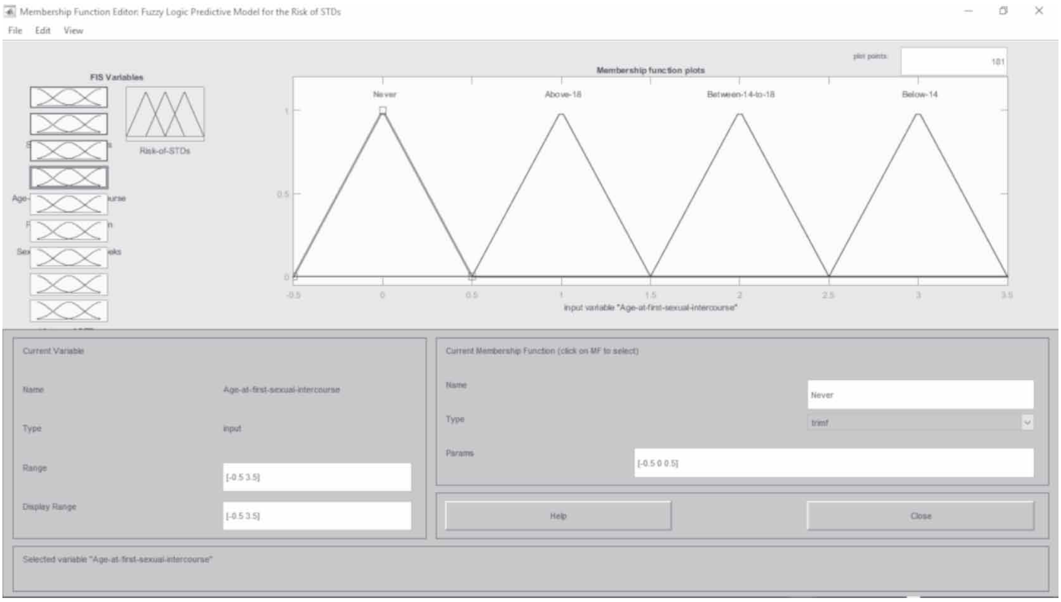
Figure 4. Fuzzification of toilet facility used



the interval [-0.5, 0.5] with center 0 was used to model personal while [0.5, 1.5] with center 1 was used to model shared.

The results of the simulation of the model for age at first sexual intercourse is expressed in Figure 5 such that the interval [-0.5, 0.5] with center 0 was used to model never, [0.5, 1.5] with center 1 was used to model above 18 years, [1.5 2.5] with center 2 was used to model between 14 and 18 years and [2.5 3.5] with center 3 was used to model below 14 years. The results of the simulation of the

Figure 5. Fuzzification of age at first sexual intercourse



model for practice sex protection is shown in Figure 6 such that the interval [-0.5, 0.5] with center 0 was used to model always, [0.5, 1.5] with center 1 was used to model sometimes and [1.5 2.5] with center 2 was used to model never.

The results of the simulation of the model for sexual activity (in the last 2 years) is expressed in Figure 7 such that the interval [-0.5, 0.5] with center 0 was used to model inactive and [0.5, 1.5] with center 1 was used to model active. The results of the simulation of the model for lifetime partners is

Figure 6. Fuzzification of practice sex protection

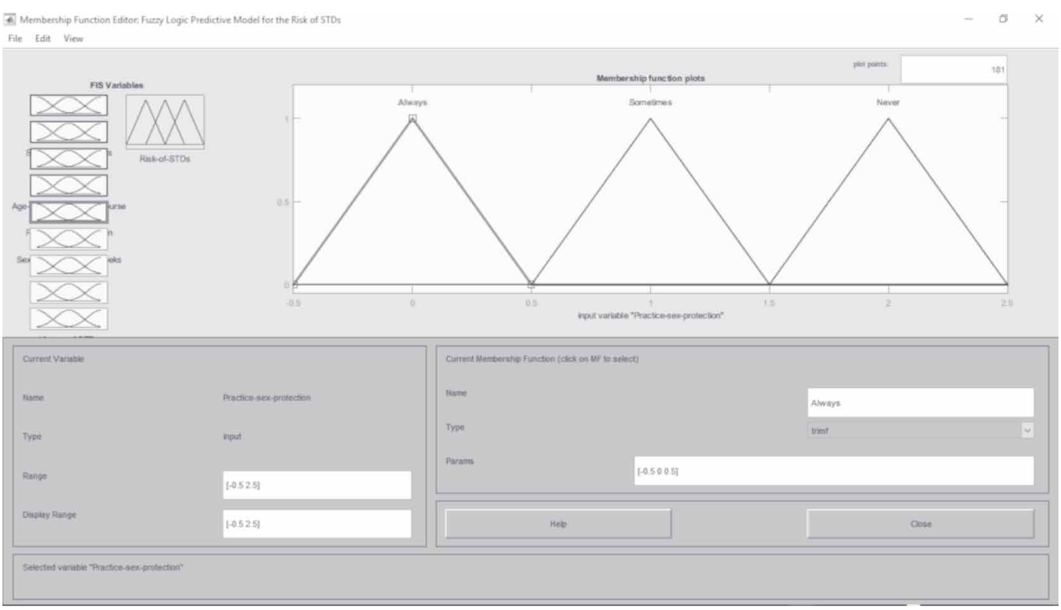
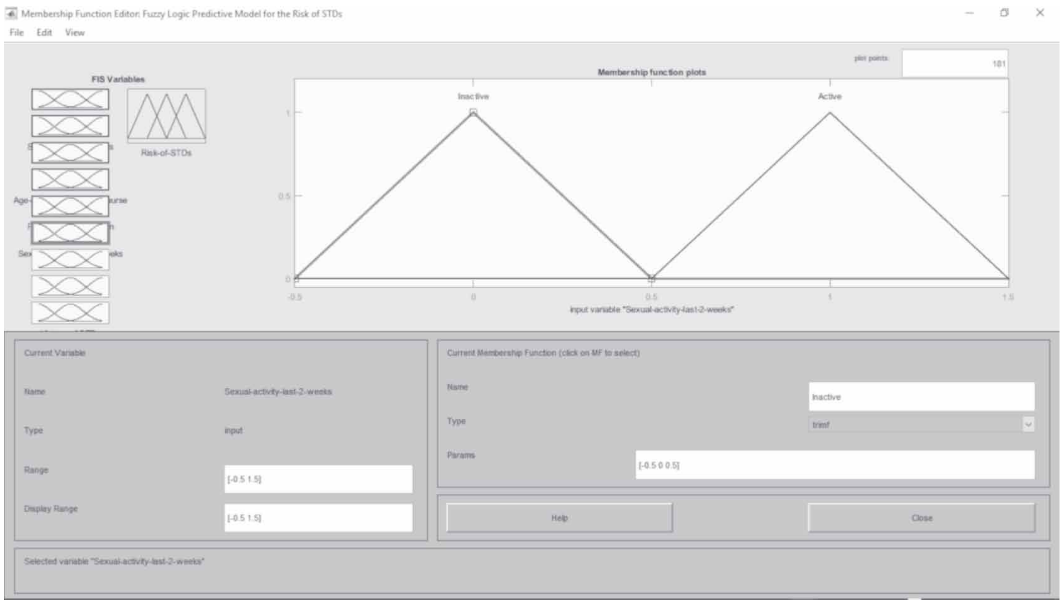


Figure 7. Fuzzification of sexual activity (in last 2 weeks)



shown in Figure 8 such that the interval [-0.5, 0.5] with center 0 was used to model none or one and [0.5, 1.5] with center 1 was used to model more than one. The results of the simulation of the model for practice casual sex is shown in Figure 9 such that the interval [-0.5, 0.5] with center 0 was used to model no and [0.5, 1.5] with center 1 was used to model yes.

The results of the simulation of the model for history of STDs is expressed in Figure 10 such that the interval [-0.5, 0.5] with center 0 was used to model no and [0.5, 1.5] with center 1 was used

Figure 8. Fuzzification of lifetime partners

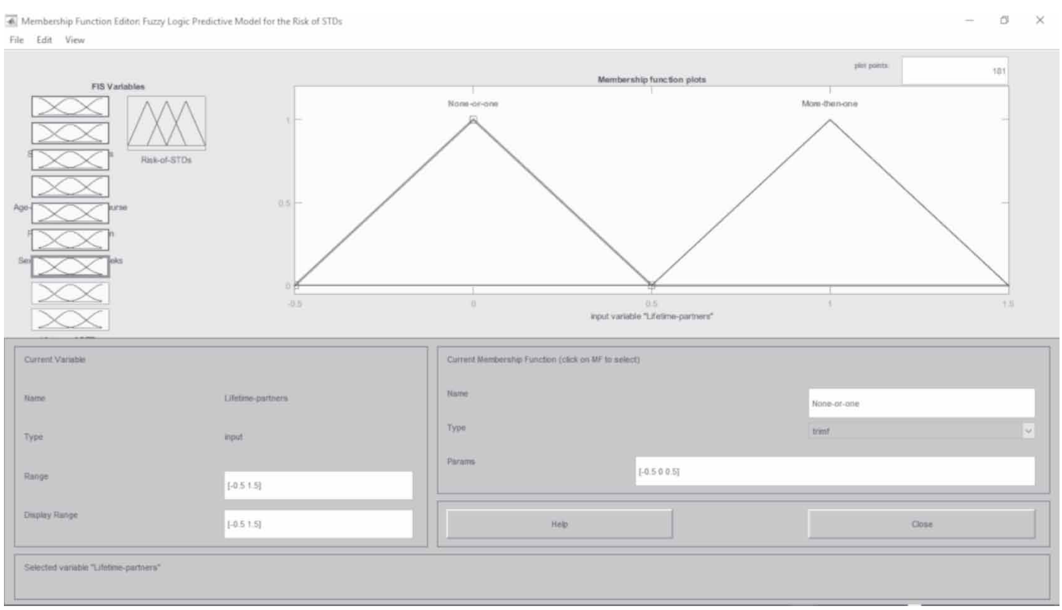


Figure 9. Fuzzification of practice casual sex

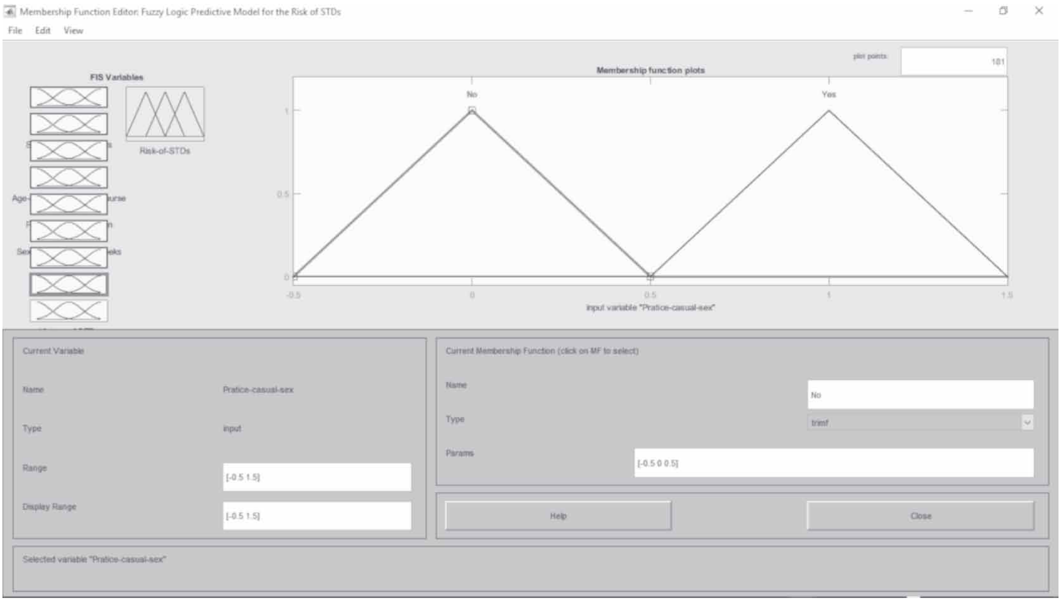
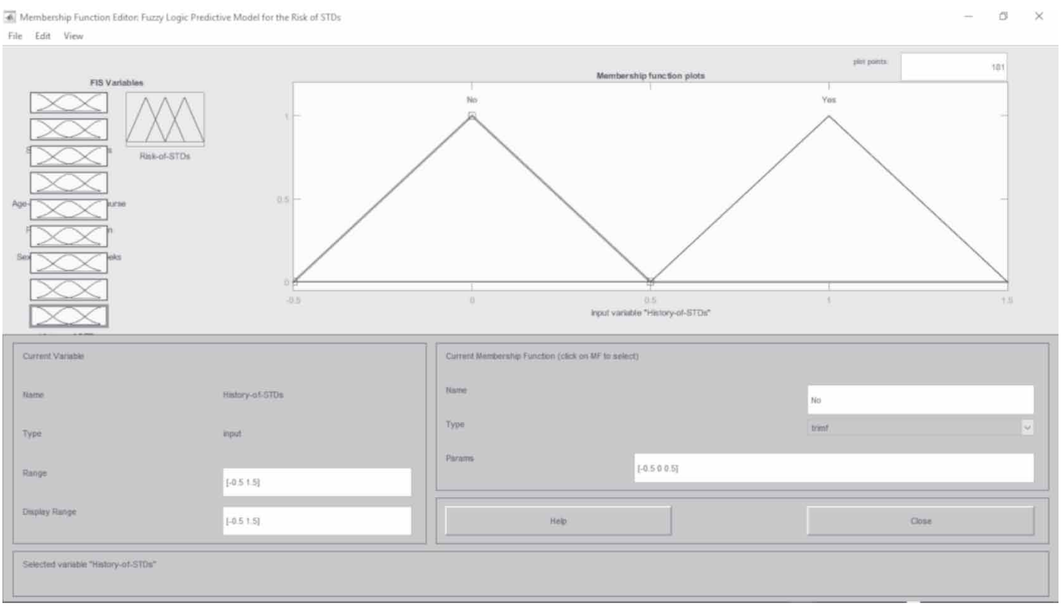


Figure 10. Fuzzification of history of STDs



to model yes. The results of the simulation of the model for risk of STDs is shown in Figure 11 such that the interval [-0.5, 0.5] with center 0 was used to model no risk, [0.5, 1.5] with center 1 was used to model above low risk, [1.5 2.5] with center 2 was used to model moderate n risk and [2.5 3.5] with center 3 was used to model below high risk.

Figure 12 shows the screenshot of the results of the simulation of the 2304 rules that were inferred for the assessment of the risk of STDs based on information about the 9 risk factors. The

Figure 11. Fuzzification of risk of STDs

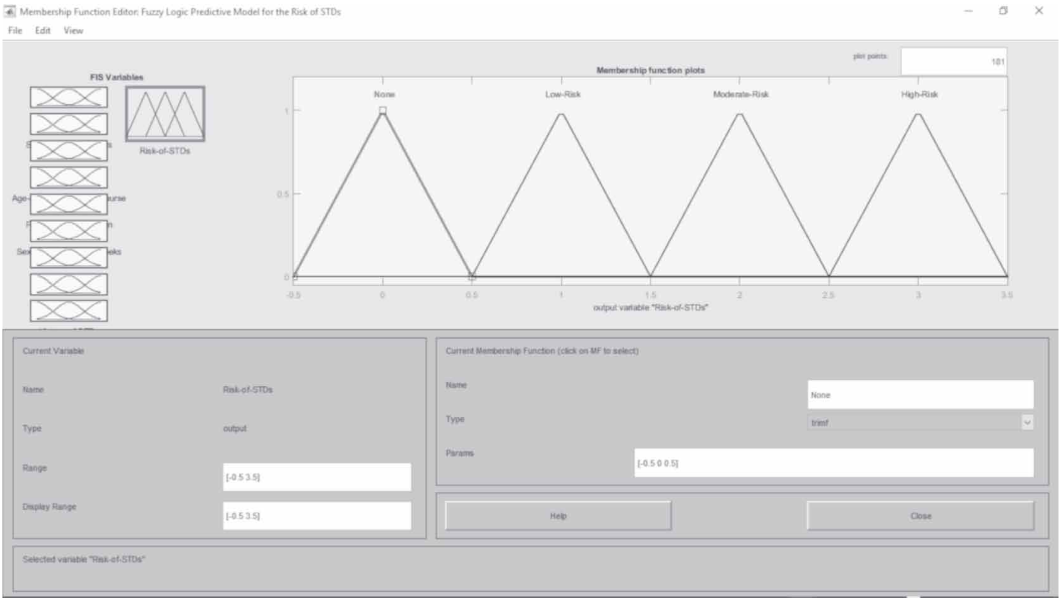
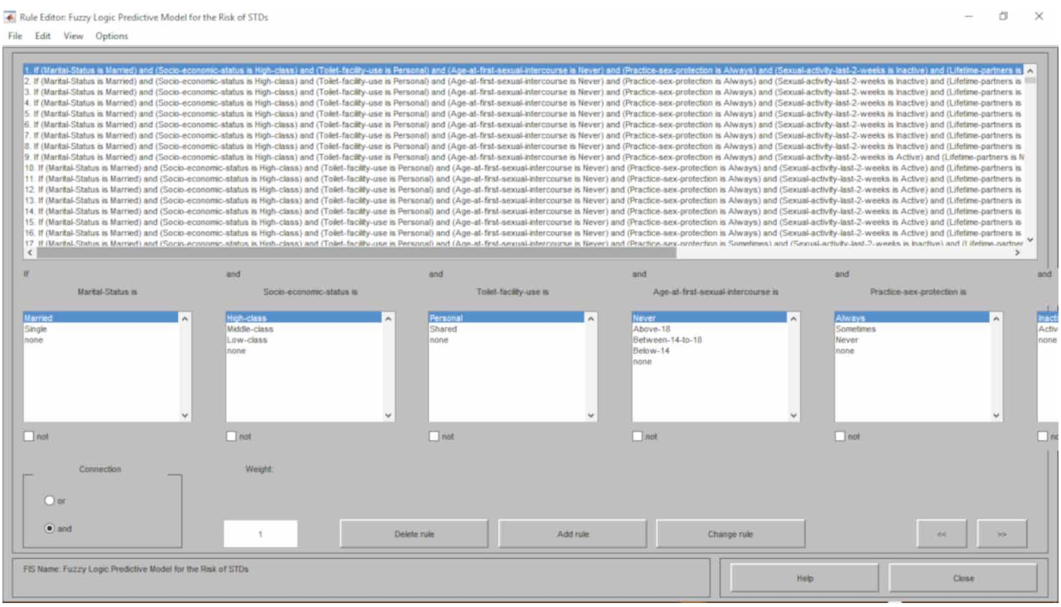


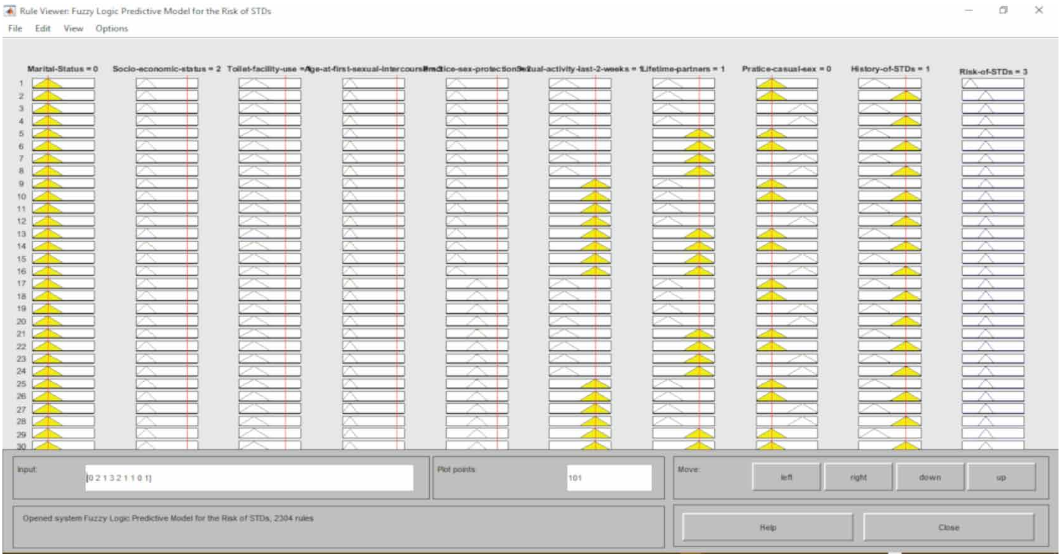
Figure 12. Inferred rules of the Fuzzy inference engine for risk of STDs



results showed that each rule inferred is unique and does not contain linguistic variables occurring in the same pattern of any of the other rules defined. Therefore, for any given set of rules *r* and *s* within the 2304 rules there is no rule *r* that has the same linguistic variables for a set of risk factors as indicated by another rule *s*.

Figure 13 shows the results of the graphical region occupied by the assessment of the fuzzy logic model for the risk of STDs with respect to the values of the linguistic variables selected for the

Figure 13. Validation of rules inferred for the risk of STDs



risk factors. It presents the results of the defuzzification of the aggregation of the crisp values of the linguistic variables tested over the 2304 rules.

The input provided included the values: 0, 2, 1, 3, 2, 1, 1, 0 and 1 which consistent with the linguistic variables, namely: married for marital status, low for socio-economic status, shared for toilet facility used, below 14 years for age of first sexual intercourse, never for practice sex protection, active for sexual activity (in last 2 weeks), more than one for lifetime partners, no for practice casual sex and yes for history of STDs respectively. These values were consistent with the values of the linguistic variables for the 9 risk factors as indicated by rule number 783. The results showed that since the defuzzified crisp value revealed a value of 5 which falls within the triangular membership function with interval [2.5, 3, 3.5] then the risk of STD is high.

5. DISCUSSIONS

Based on the results of this study, the assessment of the risk of STDs for an individual can be assessed based on knowledge about associated risk factors. In this study, a fuzzy logic model was developed based on the development of a multiple input and single output (MISO) Mamdani fuzzy logic-based model for the risk of STDs based on information collected about the values of 9 risk factors. The risk factors that were identified in this study to be associated with the risk of STDs were all non-invasive risk factors which consisted of socio-demographic and social-lifestyle related information. The study showed that the values of the linguistic variables of the risk factors of associated with the risk of STDs have varying levels of association which was assigned based on crisp values. Thus, the degree of impact of each linguistic variable on the risk of STDs were identified as a function of the crisp value assigned which laid between a numbers of intervals. Therefore, the risk of STDs was determined by an interval which corresponded to the total sum of the crisp values of the linguistic variables of the risk factors that are associated with the risk of STDs.

6. CONCLUSION

This study developed a fuzzy logic-based model for the prediction of the risk of STDs among young female adults in Nigeria. This study concluded that using information about the risk factors that are associated with the risk of STDs, fuzzy logic modeling was adopted for predicting the risk of STD based on knowledge about the risk factors. Triangular membership functions were used to formulate the linguistic variables of 9 risk factors alongside the target risk of STDs. 2304 rules were extracted about the knowledge of the values of the risk factors required for predicting the risk of STDs. The study recommends that additional efforts be put place into the identification of additional non-invasive risk factors required for the early detection of STDs among individuals. Also, data about risk factors associated with STDs should be collected so that data mining and machine learning algorithms can be adopted for the development of objective predictive classification models which do not depend on rules elicited from experts which may be limited by bias.

REFERENCES

Adebowale, A. S., Titiloye, M., Fagbamigbe, A. F., & Akinyemi, O. J. (2013). Statistical Modeling of Social Risk Factors for Sexually Transmitted Diseases among Female Youths in Nigeria. *Journal of Infectious Diseases in Developing Countries*, *7*(1), 17–27. doi:10.3855/jidc.2272 PMID:23324816

Bryan, C. (2011). Infectious Disease Chapter Eight Sexually Transmitted Disease. In Microbiology and Immunology On-line. University of South Carolina School of Medicine.

Idowu, P.A., Aladekomo, T.A., Williams, K.O., & Balogun, J.A. (n.d.). Predictive Model for Likelihood of Survival of Sickle Cell Anaemia (SCA) among Peadiatric Patients using Fuzzy Logic. Transactions in Networks and Communications, 3(1), 31–44.

Idowu, P. A., Balogun, J. A., Akinyemi, I. O., & Egejuru, N. C. (2018). A Classification Model for the CD4 Count of HIV Patients using Decision Trees Algorithm. In Computer Science Advances: Research and Applications. NOVA Science Publishers Inc.

Khormehr, A., & Maihami, V. (2016). A Novel Fuzzy Expert System Design for Predicting Still-birth. *International Journal of Computers and Applications*, *138*(4), 33–38. doi:10.5120/ijca2016908789

Lakshmi, G. S., & Isakki, P. (2017). A Comparative Study of Data Mining Classification Technique using HIV/AIDS and STDs Data. *International Journal of Innovative Research in Computer and Communication Engineering*, *5*(1), 134–139.

Massad, E., Ortega, N., de Barros, L., & Struchiner, C. (2008). *Fuzzy Logic in Action: Applications in Epidemiology and Beyond* (Vol. 232). Berlin, Germany: Springer. .10.1007/978-3-540-69094-8

Ojunga, N., Peter, M., Otulo, W., Omollo, O., & Edgar, O. (2014). The Application of Logistic Regression in Modeling of Survival of Chances of HIV-Positive Patients under Highly Active Antiretroviral Therapy (HAART): A case of Nyakach District, Kenya. *Journal of Medicine and Clinical Sciences*, *3*(3), 14–20.

Plerou, A., Vlamou, E., & Papadopoulos, B. (2016, May 8). Fuzzy Logic Models in Epidemic Control. *Journal of Precision Medicine*, *1*(1), 1–8.

Shekelle, P. G., Morton, S. C., & Keeler, E. B. (2006). Costs and benefits of Health Information Technology. *Evidence Report/technology Assessment*, *132*, 1–71. doi:10.23970/AHRQEPCERTA132 PMID:17627328

Turban, E., Sharda, R., & Delen, D. (2010). *Decision Support and Business Intelligence Systems*. Prentice Hall.

Waijee, A. K., Higgings, P. D. R., & Singal, A. G. (2013). A Primer on Predictive Models. *Clinical and Translational Gastroenterology*, *4*(44), 1–4.

APPENDIX

Source Code of Fuzzy Logic Model for Risk of STDs

```
[System]
Name='Fuzzy Logic Predictive Model for the Risk of STDs'
Type='mamdani'
Version=2.0
NumInputs=9
NumOutputs=1
NumRules=2304
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'
[Input1]
Name='Marital-Status'
Range=[-0.5 1.5]
NumMFs=2
MF1='Married':'trimf',[-0.5 0 0.5]
MF2='Single':'trimf',[0.5 1 1.5]
[Input2]
Name='Socio-economic-status'
Range=[-0.5 2.5]
NumMFs=3
MF1='High-class':'trimf',[-0.5 0 0.5]
MF2='Middle-class':'trimf',[0.5 1 1.5]
MF3='Low-class':'trimf',[1.5 2 2.5]
[Input3]
Name='Toilet-facility-use'
Range=[-0.5 1.5]
NumMFs=2
MF1='Personal':'trimf',[-0.5 0 0.5]
MF2='Shared':'trimf',[0.5 1 1.5]
[Input4]
Name='Age-at-first-sexual-intercourse'
Range=[-0.5 3.5]
NumMFs=4
MF1='Never':'trimf',[-0.5 0 0.5]
MF2='Above-18':'trimf',[0.5 1 1.5]
MF3='Between-14-to-18':'trimf',[1.5 2 2.5]
MF4='Below-14':'trimf',[2.5 3 3.5]
[Input5]
Name='Practice-sex-protection'
Range=[-0.5 2.5]
NumMFs=3
MF1='Always':'trimf',[-0.5 0 0.5]
MF2='Sometimes':'trimf',[0.5 1 1.5]
MF3='Never':'trimf',[1.5 2 2.5]
[Input6]
```

```
Name='Sexual-activity-last-2-weeks'
Range=[-0.5 1.5]
NumMFs=2
MF1='Inactive':'trimf',[-0.5 0 0.5]
MF2='Active':'trimf',[0.5 1 1.5]
[Input7]
Name='Lifetime-partners'
Range=[-0.5 1.5]
NumMFs=2
MF1='None-or-one':'trimf',[-0.5 0 0.5]
MF2='More-then-one':'trimf',[0.5 1 1.5]
[Input8]
Name='Pratice-casual-sex'
Range=[-0.5 1.5]
NumMFs=2
MF1='No':'trimf',[-0.5 0 0.5]
MF2='Yes':'trimf',[0.5 1 1.5]
[Input9]
Name='History-of-STDs'
Range=[-0.5 1.5]
NumMFs=2
MF1='No':'trimf',[-0.5 0 0.5]
MF2='Yes':'trimf',[0.5 1 1.5]
[Output1]
Name='Risk-of-STDs'
Range=[-0.5 3.5]
NumMFs=4
MF1='None':'trimf',[-0.5 0 0.5]
MF2='Low-Risk':'trimf',[0.5 1 1.5]
MF3='Moderate-Risk':'trimf',[1.5 2 2.5]
MF4='High-Risk':'trimf',[2.5 3 3.5]
```

Jeremiah A. Balogun is an Assistant Lecturer in the Department of Computer Science and Mathematics at Mountain Top University (MTU), MFM Prayer City, Ibafo, Ogun State. He has B.Sc. (Computer Science with Mathematics) and M.Sc. (Computer Science) both from Obafemi Awolowo University (OAU), Ile-Ife, Osun State in Nigeria. He is presently running his Ph.D. (Computer Science) at the same institution (OAU). He has published over twenty (20) scientific articles, contributed to book chapters. His interests include: Preventive Medicine, Classification Modeling, Database Design and Modeling, Data Mining and Machine Learning.

Florence Oladeji is a PhD. Holder in Computer Science. Her area of research has been in Data Communication, Information Security and Health Information Mining. Her ambition in life is to impact others positively and lead women into computing. My hobbies are programming, reading and counseling.

Olajide Blessing Olajide is a COREN certified computer engineer that got his B.Tech and M.Tech from Ladoke Akintola University of Technology, Ogbomoso, Nigerian. He is currently undergoing his Ph.D study in the same school. He is presently a lecturer at Federal University, Wukari, Nigeria.

Adanze Onyenonachi Asinobi is a Senior Lecturer in Paediatric Nephrology in the Department of Paediatrics, College of Medicine, University of Ibadan, Ibadan Nigeria. She is also a Consultant Paediatric Nephrologist to the University College Hospital, Ibadan Nigeria. She holds an MB,BS degree, MSc, the Fellowship of the National Postgraduate Medical College of Nigeria, FMC(Paed) and is a Fellow of the International Society of Nephrology (ISN Fellow).She is currently, the President-Elect of the Nigerian Association of Nephrology, a member of Paediatric Association of Nigeria and of many Societies affiliated with Child Health. Her main research interest is in Paediatric Nephrology, Childhood Hypertension and General Paediatrics. Her other area of interest is infectious diseases many of which cause high morbidity and mortality in sub-Saharan Africa and cause damage to the kidneys.

Olayinka Olufumilayo Olusanya is a lecturer in the Department of Computer Science, Tai Solarin University of Education, Ijebu-Ode, Ogun State, Nigeria. She holds undergraduate, Masters and Ph.D Degrees in Computer Science from the Olabisi Onabanjo University, Ago Iwoye, Nigeria, University of Wolverhampton, United Kingdom and University of Ilorin, Ilorin, Nigeria. Her research interest is in Software Quality Assurance, Health Informatics, Software Testing, Security, Artificial Intelligence and Project Management. She is a member of Nigeria Computer Society.

Peter Adebayo Idowu is a Senior Researcher in the Department of Computer Science and Engineering, Obafemi Awolowo University, Nigeria. His research focus is on applied computing that is application of computing to address and solve health related problems in Sub Saharan Africa. He has over 60 publications to his credit in local journals, international journals and referred conference proceedings. He is currently researching into HIV/AIDS, Disease modelling and cloud computing in health care delivery. He is also a Member of British Computer Society, Internet Society Nigeria Chapter, Nigerian Computer Society, Computer Professional Registration Council of Nigeria, Nigerian Young Academy, International Geospatial Society, International Association of Engineers and International Federation of Information Processing WG 9.4. His research interest includes Health Informatics, Data Modelling, Software Engineering, GIS, and Informatics. Within the last five years, Dr Idowu has successively trained over 15 graduate students (both Master and Doctoral students). He is blessed with three Research Associates; Praise, Vicky and Peter. He enjoys reading and driving.