DEVELOPMENT OF A PREDICTIVE FUZZY LOGIC MODEL FOR MONITORING THE RISK OF SEXUALLY TRANSMITTED DISEASES (STD) IN FEMALE HUMAN

Olajide Blessing Olajide1, Odenyi Olufemi Ayodeji2, Jooda Janet Olubunmi3, Balogun Monsurat Omolara4, Ajisekola Usman Adedayo5, Idowu Peter Adebayo6

1Lecturer, Department of Computer Science, Federal University, Wukari, Nigeria.
2Lecturer, Department of Computer Science, Osun State College of Technology, Esa Oke, Nigeria
3Lecturer, Department of Computer Engineering, Igboho Polytechnic, Igboho, Nigeria
4Lecturer, Department of Electrical and Computer Engineering, Kwara State University, Malete, Nigeria
5Student, Department of Computer and Information Science, Tai Solarin University of Education, Ijebu-ode, Nigeria
6Lecturer, Department of Computer Science and Engineering, Obafemi Awolowo University, Ile Ife, Nigeria

Abstract: The purpose of this study is to develop a classification model for monitoring the risk of sexually transmitted diseases (STDs) among females using information about non-invasive risk factors. The specific research objectives are to identify the risk factors that are associated with the risk of STDs; formulate the classification model; and simulate the model. Structured interview with expert physicians was done in order to identify the risk factors that are associated with the risk of STDs Nigeria following which relevant data was collected. Fuzzy Triangular Membership functions were used to map labels of the input risk factors and output STDs risk of the classification model identified to associated linguistic variables. The inference engine of the classification model was formulated using IF-THEN rules to associate the labels of the input risk factors to their respective risk of still birth. The model was simulated using the fuzzy logic toolbox available in the MATLAB® R2015a Simulation Software. The results showed that 9 non-invasive risk factors were associated with the risk of STDs among female patients in Nigeria. The risk factors identified were marital status, socio-economic status, toilet facility used, age at first sexual intercourse, practice sex protection, sexual activity (in last 2 weeks), lifetime partners, practice casual sex and history of STDs. 2, 3 and 4 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 no risk, low risk, moderate risk and high risk. The 2304 inferred rules were formulated using IF-THEN statements which adopted the values of the factors as antecedent and the STDs as consequent part of each rule. 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.

Key Words: STD, classification model, fuzzy logic, simulation, membership function, risk factor.

1.0 INTRODUCTION

Sexually transmitted diseases (STDs) are infections that can be transferred from one person to another during sexual activity. Common STDs are HIV human immunodeficiency virus, chlamydia, gonorrhoea, syphilis and trichomoniasis [1]. STDs could be either ulcerative or non-ulcerative. In developing countries, STDs exhibit a higher incidence and prevalence rates than developed countries [2]. The prevalence of STDs among Nigeria female youths is 17% [3]. In 2016, World Health Organisation (WHO) released its Global health sector strategy on STDs in year 2016 to 2021. The Strategy envisions that by year 2030, rates of congenital syphilis will be reduced to less than 50 cases per 100 000 live births in 80% of countries; and the incidence of infections with T. pallidum (syphilis) and N. gonorrhoeae (gonorrhoea) would have fallen by 90% globally between year 2018 and 2030. To achieve its goals, a critical component of the Global STD Strategy is strengthening STD surveillance and programme monitoring systems [1]. However, the global burden of STDs remains high [4]. Hence the research for more predictive expert systems for monitoring of STDs risk remains an open research.

The usage of Information Communication Technology (ICT) in health has the potential to improve the quality of health care as stated by many researchers who supported the belief. 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 such as fuzzy logic 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 [5]. Fuzzy logic is a basic concept for embedding structured human knowledge into workable algorithms that constitutes fuzzy models [6]. Fuzzy logic is an alternative to probability theory in which outcome represents the
degree to which it leads to true or false. Fuzzy logic will be beneficial to work on incomplete, vague and uncertain problems. It totally depends upon membership functions based upon optimum results for given application. The conventional approach based on crisp sets to evaluate the results but with fuzzy logic we step forward to implement with fuzzy sets for approximate reasoning. So, it behaves as mimic nature as of human thinking to deal with specific problem. Fuzzy logic totally based on fuzzy if-then rules to produce an output for an application. Fuzzy rule based inference approach combines all three basic representation as matching, inference and combination to meet the final response. There are two ways to demonstrate fuzzy rules based on fuzzy mapping and fuzzy implication. The role of fuzzy mapping lies when we don’t know all the values of the parameters in specific application whereas fuzzy implication focused on approximate reasoning with different statements. The purpose of fuzzification is to convert crisp sets to fuzzy sets. Therefore fuzzy logic can generate an intelligent machine which resembles the working as of human brain. To tackle the emergent need to society in medical field, the contribution of fuzzy logic is growing rapidly in every sector of medicine [7]. Following this, the study intends to apply fuzzy logic modelling to the development of a predictive model which can be used to assess the risk of STDs among young females.

2.0 LITERATURE REVIEW

[8] developed an expert system to diagnose asthma by assigning parameters with fuzzy logic. The whole representation was performed in Iran and concluded with 100% specificity and 94% sensitivity. In the same vein, [9] developed an expert system for adult asthma diagnosis using Neuro-Fuzzy fitting tool with SOM, LVQ and BPNN algorithms. The back-propagation considered to be the best among all at epoch 9 by providing 535 samples. Another classification was done by [10] for asthma and chronic disease with MATLAB tool profiler with different classification algorithms as NN and LM which provides 99.41% correctly classified results with specificity 100% and sensitivity 99.28%. [11] worked on diagnosing of liver disease on the basis of expert system collaborated with fuzzy logic. According to them, the representation was given with Mamdani approach by using 3 inputs and 1 output variables to identify the risk factors.

[12] proposed a model to diagnose diabetes disease based on PCA and ANFIS methodology with 8 input features implemented in MATLAB. The expert system showed 89.47% accurate with 85% sensitivity, 92% specificity and 0.262 root mean square error. [13] proposed a model of combining three different diseases covered in one fuzzy expert approach. He worked on dental, heart and anemia problem with 11 input parameters for diagnosing multiple diseases. The system proved 94.627% accuracy. [14] applied deep learning for the processing of images of cassava plant for the detection of plant diseases. The study collected a dataset of cassava diseases images taken at a field in Tanzania. The images were analyzed using a transfer leaning to train a deep convolutional neural network to identify 3 diseases and 2 types of pest damages. The best model achieved an overall accuracy of 93% for data not used in the training process. The results also showed that the transfer learning approach for image recognition of field images offers a fast, affordable, and easily deployable strategy for digital plant disease detection.

Also, Dynamic neural network (DNN) was used to detect time-varying occurrences of tremor and dyskinesia from time series data acquired from EMG sensors and triaxial accelerometers worn by Parkinson patient [15]. Other predictive fuzzy logic models combine different Artificial Intelligence AI and Machine Learning ML approaches for reproducing intelligent human reasoning process [16]. By using information fusion, hybrid models combine heterogeneous ML approaches and improve quality of reasoning for complex regression and classification problems [17]. Neurofuzzy systems combine neural network and fuzzy logic paradigms to avoid the limitations of neural network explanations to reach decision and limitations of fuzzy logic to automatically acquire the rules used for making decisions [18].

3.0 METHODOLOGY

To actualise this work, the research was divided into three: identification of ten (10) different risk factors associated with the risk of STDs among young female human, membership functions was used for the formulation of the fuzzy logic model for each risk factor identified alongside the output target risk of STDs and the resulting fuzzy logic model is simulated using MATLAB R2015a software.

3.1 Identification of Risk Factors for STDs

Crisp values were given to each risk factor assessed as a way of quantifying the response to each risk factor by a user. Therefore, higher values were given to intervals that increased the risk while lower values were given to intervals that reduced the risk of STDs. Hence, each crisp interval were assigned a linguistic value that was used as a nominal tag for which each fuzzy membership function was required to be formulated based on the values of the crisp interval. The ten risk factors identified are: marital status, socioeconomic status, toilet facility use, age at first sexual intercourse, practice sex protection, sexual activities (in last two weeks), life time number of partners, practice casual sex, history of STDs and risk of STDs. All these risk factors are allocated crisp values and linguistic variables as shown in Table 1.
Table 1: STD Risk Factors Linguistic Variables and Crisp Values.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Linguistic Variable</th>
<th>Crisp Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Status</td>
<td>Married</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Single</td>
<td>1</td>
</tr>
<tr>
<td>Socio-Economic Status</td>
<td>High Class</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Low Class</td>
<td>2</td>
</tr>
<tr>
<td>Toilet Facility Use</td>
<td>Personal</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Shared</td>
<td>1</td>
</tr>
<tr>
<td>Age at First Sexual Intercourse</td>
<td>Never</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Above 18 years</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Between 14 and 18 years</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Below 14 years</td>
<td>3</td>
</tr>
<tr>
<td>Practice Sex Protection</td>
<td>Always</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Sometimes</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Never</td>
<td>2</td>
</tr>
<tr>
<td>Sexual Activity (in Last 2 weeks)</td>
<td>Inactive</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>1</td>
</tr>
<tr>
<td>Lifetime Number of Partners</td>
<td>None or One</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>More than One</td>
<td>1</td>
</tr>
<tr>
<td>Practice Casual Sex</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>History of STDs</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Risk of STDs</td>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Low Risk</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>High Risk</td>
<td>3</td>
</tr>
</tbody>
</table>

3.2 Fuzzy Logic Model Formulation for Risk Factors of STDs

A triangular membership function of the value \((a, b, c)\) was formulated to depict interval of \(a \leq b \leq c\) such that the parameters are numeric valued. The interval of this parameter was used to define the crisp interval within which each crisp value required for calling the linguistic variable was assigned. Therefore, 2 or 3 triangular membership functions were formulated for each risk factor that was identified in this study based on the mathematical expression in equation (1).

Using 2 or 3 triangular membership functions, the

\[
\text{Variable}_{\text{label}(x,a,b,c)} = \begin{cases} 
0; & x \leq a \\
\frac{x-a}{b-a}; & a < x \leq b \\
\frac{c-x}{c-b}; & b < x \leq c \\
0; & x > c 
\end{cases}
\]  

(1)

Where \(x\) is a numeric value to be fitted into a crisp interval of \((a, b, c)\) to get the variable label. Labels of the identified risk factors were formulated using the crisp intervals of \((-0.5, 0.5), (0.5, 1.5)\) and \((1.5, 2.5)\) to model the linguistic variables for 0, 1 and 2 respectively such that the values 0, 1 and 2 became the center \(b\) of each interval as shown in Table 2.

3.3 Fuzzification of the risk of STDs

The triangular membership function was used to formulate the fuzzy logic model for the target variable by assigning crisp values of 0, 1, 2 and 3 to the target class labels, namely: No risk, low risk, Moderate risk and High risk using the intervals \((-0.5, 0.5), (0.5, 1.5), \) \( (1.5, 2.5)\) and \((2.5, 3.5)\) respectively. Therefore, four (4) triangular membership functions were used to formulate the fuzzy logic model required to describe the 4 labels of the target class that was used to describe the risk of STDs using the identified crisp as shown in Table 3. In order to construct the knowledge base of the classification model using fuzzy logic, a number of IF-THEN rules were used by combining the risk factors as the precedence while the risk of STDs was used as the consequent variable. The following rule was inferred for this work:

\[
\text{IF (Present Age = "Above 30 years") AND (Marital Status = "Married") AND (Socio-economic Status = "High Class") AND (Toilet facility Used = "Personal") AND (Age at first sexual intercourse = "Never") AND (Practice sex protection = "Yes") AND (Sexual Activity in last 2 weeks = "Inactive") AND (Lifetime number of partners = "None or One") THEN (Risk of STDs = "No Risk").}
\]
Table 3: Formulation of the Risk of STDs

<table>
<thead>
<tr>
<th>Target Class</th>
<th>Interval</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Risk</td>
<td>(-0.5, 0.5)</td>
<td>-0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Low Risk</td>
<td>(0.5, 1.5)</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Moderate Risk</td>
<td>(1.5, 2.5)</td>
<td>1.5</td>
<td>2</td>
<td>2.5</td>
</tr>
<tr>
<td>High Risk</td>
<td>(2.5, 3.5)</td>
<td>2.5</td>
<td>3</td>
<td>3.5</td>
</tr>
</tbody>
</table>

The developed fuzzy logic model was then simulated in MATLAB software to investigate the design’s appropriateness.

4.0 RESULTS

In this study, 2, 3 and 4 triangular membership functions were used to formulate the fuzzy logic model for the labels of each risk factor with centers 0 and 1; centers 0, 1 and 2; and centers 0, 1, 2 and 3 respectively for the risk factor labels. Mathematical representation of the fuzzy logic model formulation using the triangular membership function for each of the labels is presented in equation (2) to (5):

\[
\text{Crisp} - \text{center}_0(x; -0.5, 0, 0.5) = \begin{cases} 
0; x \leq -0.5 \\
\frac{x + 0.5}{0.5}; -0.5 < x \leq 0 \\
\frac{0.5 - x}{0.5}; 0 < x \leq 0.5 \\
0; x > 0.5 
\end{cases} (2)
\]

\[
\text{Crisp} - \text{center}_1(x; 0.5, 1, 1.5) = \begin{cases} 
0; x \leq 0.5 \\
\frac{x - 0.5}{0.5}; 0.5 < x \leq 1 \\
\frac{1.5 - x}{0.5}; 1 < x \leq 1.5 \\
0; x > 1.5 
\end{cases} (3)
\]

\[
\text{Crisp} - \text{center}_2(x; 1.5, 2, 2.5) = \begin{cases} 
0; x \leq 1.5 \\
\frac{x - 1.5}{0.5}; 1.5 < x \leq 2 \\
\frac{2.5 - x}{0.5}; 2 < x \leq 2.5 \\
0; x > 2.5 
\end{cases} (4)
\]

\[
\text{Crisp} - \text{center}_3(x; 2.5, 3, 3.5) = \begin{cases} 
0; x \leq 2.5 \\
\frac{x - 2.5}{0.5}; 2.5 < x \leq 3 \\
\frac{3.5 - x}{0.5}; 3 < x \leq 3.5 \\
0; x > 3.5 
\end{cases} (5)
\]

Also, the classification of the risk of STDs was classified into 4 linguistic variables, namely: No risk, Low risk, Moderate risk and High risk using crisp values with centers of 0, 1, 2 and 3 respectively. Using the 4 triangular membership functions stated in equations (6) to (9), the linguistic variables of the risk of STDs was formulated:

\[
\text{Crisp} - \text{No_risk}(x; -0.5, 0, 0.5) = \begin{cases} 
0; x \leq -0.5 \\
\frac{x + 0.5}{0.5}; -0.5 < x \leq 0 \\
\frac{0.5 - x}{0.5}; 0 < x \leq 0.5 \\
0; x > 0.5 
\end{cases} (6)
\]

\[
\text{Crisp} - \text{low_risk}(x; 0.5, 1, 1.5) = \begin{cases} 
0.5\left(1 - \frac{x}{0.5}\right); 0.5 < x \leq 1 \\
0.5\left(1 - \frac{1.5 - x}{0.5}\right); 1 < x \leq 1.5 \\
0; x > 1.5 
\end{cases} (7)
\]

\[
\text{Crisp} - \text{moderate_risk}(x; 1.5, 2, 2.5) = \begin{cases} 
0; x \leq 1.5 \\
0.5\left(1 - \frac{x}{0.5}\right); 1.5 < x \leq 2 \\
0.5\left(1 - \frac{2.5 - x}{0.5}\right); 2 < x \leq 2.5 \\
0; x > 2.5 
\end{cases} (8)
\]

\[
\text{Crisp} - \text{high_risk}(x; 2.5, 3, 3.5) = \begin{cases} 
0; x \leq 2.5 \\
0.5\left(1 - \frac{x}{0.5}\right); 2.5 < x \leq 3 \\
0.5\left(1 - \frac{3.5 - x}{0.5}\right); 3 < x \leq 3.5 \\
0; x > 3.5 
\end{cases} (9)
\]

4.1 Simulation Results of the Predictive Fuzzy logic Model for STDs

The results of the simulation of the model for the marital status is shown in Figure 1 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 2 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 3 such that 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 shown in Figure 4 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 model for practice sex protection is shown in Figure 5 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 shown in Figure 6 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 shown in Figure 7 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 8 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 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 risk of STDs is shown in Figure 10 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 risk and [2.5, 3.5] with center 3 was used to model below high risk.

![Figure 1: Fuzzification of Marital Status](image)

![Figure 2: Fuzzification of Socio-Economic Status](image)
Figure 3: Fuzzification of Toilet Facility Used

Figure 4: Fuzzification of Age at First Sexual Intercourse
Figure 5: Fuzzification of Practice Sex Protection

Figure 6: Fuzzification of Sexual Activity (in last 2 weeks)
Figure 7: Fuzzification of Lifetime Partners

Figure 8: Fuzzification of Practice Casual Sex
4.2 Discussion of Results

This study developed a classification model which can be used for monitoring the risk of sexually transmitted diseases (STDs) based on information provided about associated risk factors in female human. The study identified 9 non-invasive risk factors required for the classification of the risk of STDs among female patients. Each risk factor was defined using a number of linguistic variables for which central crisp values were assigned based on the association with the risk of STDs. The higher the association of the linguistic variable then the higher the central crisp values assigned.

The crisp values for each of the identified risk factor was done by allocating the values 0, 1, 2 and 3 to some risk factors with 4 values; values of 0, 1 and to some risk factors with 3 values; or the values 0 and 1 to binary risk factors in increasing order of association with the risk of STDs. Each risk factor was discretized into 2, 3
or 4 parts such that the values of 0 and 1 or 0, 1 and 2 or 0, 1, 2 and 3 were allocated to each linguistic variable defined. Therefore, crisp intervals with centers of 0, 1, 2 and 3 were used to define the labels of the identified risk factor using triangular membership functions to identify labels in intervals [-0.5 0.5], [0.5 1.5], [1.5 2.5] and [2.5 3 3.5] respectively.

For the purpose of establishing a relationship between the identified non-invasive risk factors identified, 2304 rules were inferred from the experts in order to determine the relationship between the risk factors identified and the risk of STDs. In order to construct the knowledge base of the classification model using fuzzy logic, a number of IF-THEN rules were used by combining the risk factors as the precedence while the risk of STDs was used as the consequent variable. Using the risk factors that were identified for assessing the risk of STDs, the process of inference rule generation was achieved.

5.0 CONCLUSION

This study developed a fuzzy logic-based model for the classification of the risk of STDs in order to mitigate the effect of STDs on the productivity of female in Nigeria using non-invasive techniques. 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. The study also concluded that the 9 related factors identified were marital status, socio-economic status, toilet facility used, age at first sexual intercourse, practice sex protection, sexual activity (in last 2 weeks), lifetime partners, practice casual sex and history of STDs. 2, 3 and 4 sexual activity (in last 2 weeks), lifetime partners, practice sex protection, marital status, socio-economic status, toilet facility used, age at first sexual intercourse, practice sex protection, sexual activity (in last 2 weeks), lifetime partners, practice casual sex and history of STDs. 2, 3 and 4 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 no risk, low risk, moderate risk and high risk. The 2304 inferred rules were formulated using IF-THEN statements which adopted the values of the factors as antecedent and the risk of STDs as consequent part of each rule. The developed fuzzy logic model was simulated to demonstrate its capabilities in predicting the risk of STDs in order to mitigate the effect of STDs on the productivity of female. The study also, recommends that additional efforts be put place into the identification of additional non-invasive risk factors required for the early detection of STDs among females. 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 models which do not depend on expert rules elicited from experts which may be limited by bias.

REFERENCES


