Deduction of Oral Cancer Using Fuzzy Linear Regression

[1]S. Arulchinnappan, [2]K. Karunakaran and [3]G. Rajendran
[1]Department of Mathematics, SSM College of Engineering,
Komarapalayam, Namakkal (Dt), Tamilnadu,
[2]Department of Oral Maxillofacial Pathology, J.K.K. Nataraj,
Dental College, Komarapalayam, Namakkal(Dt), Tamilnadu,
[3]Department of Mathematics, Kongu Engineering College,
Perundurai, Tamilnadu, India

Abstract: Problem statement: To examine the risk factors of oral cancer at the earlier stage. Smoking, chewing, and drinking are the major risk factors which cause oral cancer considered as input variables. **Approach:** A case – control study was conducted at JKK Nataraj Dental College and Hospital, during the period from September 2007 to November 2009, in Namakkal District, Tamilnadu, India. Data collected were analyzed using Fuzzy Linear Regression. For this JAVA program was developed. **Results:** Using this fuzzy linear regression model Smoking, Drinking, and Chewing were identified as potent risk factors of oral cancer. **Conclusion:** Smoking, drinking, and chewing, are the most dangerous risk factors that will cause oral cancer. This study will help to improve the clinical practice, guidance for analyzing the risk factors of oral cancer.

Key words: Oral cancer, demographic data, fuzzy numbers, fuzzy membership function, fuzzy linear regression, Human Papilloma Virus (HPV), risk factors, clinical practice, cigarette smoking, alcohol drinking

INTRODUCTION

Oral cancer is one of the most common life threatening cancers all over the world, in particular Asian countries and tobacco is considered to be the most potent risk factor for oral cancer (Madani *et al*., 2010). Oral cancer is a significant cause of morbidity and mortality and consistently ranks as one of the top ten cancers worldwide, with broad differences in geographic distribution (Delavarian, 2010). Worldwide, oral cancer is reported to have the lowest survival rate. This has led to an increased concern over the role of cancer screening program (Mohd Dom Rosma *et al*., 2010). According to World Health Organization, of the diagnosed oral cancer worldwide around 40% occur in India, Pakistan, Bangladesh and Sri Lanka. India has one of the highest rates in the world; accounting for one-third of the total cancers and unfortunately this figure continues to rise. 80-90% of oral cancers are caused by tobacco use. Many epidemiological studies conducted over the last three decades in America, Europe and Asia have provided strong evidence of an association between alcohol and tobacco use and an increased risk of oral cancer. Low

socio-economic status is as well significantly associated with increased oral cancer risk in high and lower income-countries, across the world (Madani *et al*., 2010).

Tobacco is one of the most preventable causes of oral cancer. Although there is evidence that smoking (cigarette, cigar and pipe) is associated with oral cancer, the smokeless tobacco (often called chewing tobacco or spit tobacco) seems to be strongly associated with oral cancer. These findings are based primarily on the epidemiological association of tobacco use with increased incidence of oral cancer (Madani *et al*., 2010).

The co-factors are age, gender, ethnicity, smoking, lack of fruits and vegetables in diet, alcohol drinking, chewing, chronic irritation to the lining of the mouth, alcohol-containing mouth wash and Human Papilloma Virus (HPV) infection. Oral cancers are two to four times as common in men as in women. This may be because men are more likely to use tobacco and alcohol (Arulchinnappan *et al*., 2011).

The most common treatment is radiation therapy and chemotherapy. The new treatment methods in the field of cancer are introduced every day, but the decision making is the complex practice that should be done with extreme care and conscious. Sometimes

making decisions with initiative thinking may lead to wrong diagnosis and treatment. Methodological decision making is unfailing and will be the base for the decisions. Instead of identifying the stages and development of cancer, it is important to identify the risk possibility of oral cancer for prevention.

A risk factor is anything that may increase a personal chance of developing a disease. Some people with one or more risk factors never develop the disease, while others develop disease and have no known risk factors. But, knowing your risk factors to any disease can help to guide you into the appropriate actions, including changing behaviors and being clinically monitored for the disease.

The purpose of this study is to present the use of fuzzy regression models in the prediction of oral cancer susceptibility as a function of demographic function (age, gender and ethnicity), the risk habits (cigarette smoking, alcohol drinking, tobacco chewing, sunlight, mouthwash, chronic irritations and diet).

Fuzzy regression models have been the main predictive modeling for a long time. Predictive regression models characterize the relationship between inputs and outputs using linear equations for a linear function. In order to extend the capability of regression models, non-linear transformations are often applied to the model inputs or output.

Fuzzy regression prediction model: A fuzzy set is mathematical theory for describing the interested variables from uncertain factors or variables like seasonal inflows. The relationship between input and output variables is defined from fuzzy rule, according to human processes in thinking and decision. In addition, fuzzy rules are relatively easy to explain and understand. Recently, the fuzzy model was accepted to describe the relationship of the uncertain variables (Hassan *et al*., 2010). Often, the calibration processes of the fuzzy model were performed by manual adjusting (trial and error) the membership functions and rule bases (Teerawat *et al*., 2011).

A Fuzzy regression model is a non parametric model that can be used to explain the variation of a dependent variable Y in terms of the variations of the independent variable X as $Y = f(x)$ where $f(x)$ is a linear function. Fuzzy regression provides a means for handling regression problem lacking the significant amount of data and with vague relationships between the explanatory and response variables. It was first introduced by Tanaka in 1982. A fuzzy Linear Regression model expresses the fuzzy regression coefficients as fuzzy number in the interval form. The estimated dependent variable Y is also a fuzzy number since the regression coefficients are fuzzy numbers.

There are two approaches to fuzzy regression. The first approach also known as the possibilistic regression is based on minimizing fuzziness as an optimal criterion. The second which is based on the least squares of errors as a fitting criterion. In fuzzy regression, deviation between observed values and estimated values are assumed to be due to system fuzziness or fuzziness of regression coefficients (Mohd Dom Rosma *et al*., 2010).

In this study the fuzzy regression model used is based on Tanaka's possibility regression describes the response variable Y as:

$$\tilde{Y} = A_0 + A_1 x_1 + A_2 x_2 + ... + A_j x_j + ... A_k x_k ...$$

where Y is the fuzzy output, $X = [X_1, X_2, ..., X_K]^T$ is the real valued input vector of independent variables and each regression co efficient A_J, j=0,1……..k was assumed to be asymmetric triangular fuzzy number with centre α_j and half width $c_j, c_j > 0$.

The Fuzzy Regression equation is considered as:

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X$$

Where:

$$\tilde{Y} = (\underline{y}, \overline{y}), \tilde{A}_0 = (\underline{a}, \overline{a}), \tilde{A}_1 = (\underline{b}, \overline{b})$$

The normalized equations can be written as:

$$\sum_{i=1}^{n} Y_i = n\tilde{A}_0 + \tilde{A}_1 \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = \tilde{A}_0 \sum_{i=1}^{n} X_i + \tilde{A}_1 \sum_{i=1}^{n} X_i$$

The above equations can be reduced to two crisp system of simultaneous equations having upper and lower bound of Fuzzy Co-efficient values and as follows:

$$\sum_{i=1}^{n} \overline{y}_i = n\overline{a} + \overline{b} \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} \underline{y}_i = n\underline{a} + \underline{b} \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i \overline{y}_i = \overline{a} \sum_{i=1}^{n} X_i + \overline{b} \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} X_i \underline{y}_i = \underline{a} \sum_{i=1}^{n} X_i + \underline{b} \sum_{i=1}^{n} X_i^2$$

Fuzzifing the dataset: First, in order to represent a continuous fuzzy set, we need to express it as a function and then map the elements of the set to their degree of membership. Triangular membership functions are used to represent fuzzy sets because of its simplicity, easy comprehension and computational efficiency. Membership functions are usually predefined by experienced experts. They also can be derived through automatic adjustments. We use three linguistic terms (low, medium and high). The three fuzzy membership values are produced.

MATERIALS AND METHODS

Setting of study: This research work was conducted based on case-control study at JKK Dental college hospital during the year sep. 2007-Nov 2009 in Tamilnadu, India.

Study of population: The majority of the subjects were textile mill and saego factory workers employed in and around Namakkal. In addition workers from allied textile industries, transport workers and electrical power plant employees were examined.

The original study population was comprised of 221 industrial workers.

All participants were 35 years of age or older.

All examinations were conducted with the help of specially trained dentists.

Identifying information, such as age, gender (demographic profile) of patients and information on cigarette smoking, alcohol drinking, tobacco chewing (oral cancer risk habits) was recorded and used as input variables.

The output refers to the health condition of the patients as either healthy or unhealthy.

The demographic and disease variables of patients that were reported to be associated risk factors to oral cancer were used as the predictor variables in developing the fuzzy linear regression models.

The data set was either 'cancer' (1) or 'healthy' (0). Variable descriptions and the membership function of lower, upper bound values are shown in Table 1.

After calculating the Lower and Upper bound of fuzzy co-efficient, the result for the crisp system of linear equation is:

$$\sum X = 66, \sum \underline{y} = 6.56, \sum \overline{y} = 8.41, \sum X\overline{y} = 50.79,$$
$$\sum X\underline{y} = 39.23, \sum X^2 = 4356$$

Fuzzy linear regression algorithm:

Stage 1: Set up stage:

Step 1: To find membership function.
Step 2: Data converted to fuzzification.

Stage 2: Problem solving stage:
Step 3: Date preprocessing.

Find fuzzy regression problem:

Step 4: Find lower and upper bound values.
Step 5: Find the result

Program using Java:

```
import java.io.*;
import java.lang.*;
class Fuzzy
{
public static void main(String args[])
{
try
{
DataInputStream
dis=new DataInputStream(System.in);
String input;
double
x5,x6,x7,x8,y5,y6,y7,y8,z5,z6,z7,z8,a1,a2,b1,b2;
System.out.print("\n Enter x5: ");
input=dis.readLine();
x5=Integer.parseInt(input);
System.out.print("\n Enter x6: ");
input=dis.readLine();
x6=Integer.parseInt(input);
System.out.print("\n Enter x7: ");
input=dis.readLine();
x7=Integer.parseInt(input);
System.out.print("\n Enter x8: ");
input=dis.readLine();
x8=Integer.parseInt(input);
System.out.print("\n Enter y5: ");
input=dis.readLine();
y5=Integer.parseInt(input);
System.out.print("\n Enter y6: ");
input=dis.readLine();
y6=Integer.parseInt(input);
System.out.print("\n Enter y7: ");
input=dis.readLine();
y7=Integer.parseInt(input);
System.out.print("\n Enter y8: ");
input=dis.readLine();
y8=Integer.parseInt(input);
System.out.print("\n Enter z5: ");
```

```
input=dis.readLine();
z5=Integer.parseInt(input);
System.out.print("\n Enter z6: ");
input=dis.readLine();
z6=Integer.parseInt(input);
System.out.print("\n Enter z7: ");
input=dis.readLine();
z7=Integer.parseInt(input);
System.out.print("\n Enter z8: ");
input=dis.readLine();
z8=Integer.parseInt(input);
b1=((z5*x8)-(z8*x5))/((y5*x8)-(y8*x5));
a1=(z5-(y5*b1))/x5;
b2=((z6*x7)-(z7*x6))/((y6*x7)-(y7*x6));
a2=(z6-(y6*b2))/x6;
System.out.print("\n\tLower Bound Values: \n");
System.out.print("\t\t"+(float)a1+"\t\t"+(float)b1);
System.out.print("\n\tUpper Bound Values: \n");
System.out.print("\t\t"+(float)a2+"\t\t"+(float)b2);
}
catch(Exception e)
{
System.out.print(e);
}
}
}
```

Output:

Lower bound values:

> 0.6409091   0.0075

Upper bound values:

> 0.7227273   0.0075

RESULTS

In this fuzzy linear regression model the lower bound values are $\underline{a} = 0.641$ and $\underline{b} = 0.007$. The Upper Bound Values are $\overline{a} = 0.722$ and $\overline{b} = 0.008$. The slope of the lines are increasing and decreaing.

The fuzzy linear regression equations are:

$$Y = a + bX$$
$$(\underline{y}, \overline{y}) = (\underline{a}, \overline{a}) + (\underline{b}, \overline{b})X$$
$$(6.56, 8.41) = (0.641, 0.722) + (0.0075, 0.008)X$$

From this fuzzy linear regression model it was identified that smoking, drinking and chewing are the major risk factors which cause oral cancer.

DISCUSSION

Oral cancer is the sixth most common cancer for both sexes in the general population and the third most common cancer in developing nations. Oral cancer is well known to occur in the age group of 35 and above. Oral cancer constitutes the most life threatening of all dental condition. Unfortunately, most malignant oral tumors are not detected until they are in advanced stages. Oral cancer is a major health problem in tobacco users all over the world. The early stage development of oral cancer is a matter of great concern in this study. The study attempts to evaluate how the fuzzy method can improve the early detection of risk factors which all can cause oral cancer.

CONCLUSION

The oral cancer risk factors are found to be more prevalent. The purpose of this study is to evaluate the ability of a fuzzy regression model to predict the likelihood of an individual in developing oral cancer based on knowledge of their risk habits and demographic profiles. The data connected with oral cancer is fuzzified and the result obtained using fuzzy linear regression and JAVA program. The result shows that smoke, drink and chewing, are the most dangerous risk factors that will cause oral cancer. Through this innovative Fuzzy Linear Regression Algorithm, it is possible to identify the risk factors of oral cancer. The findings of this study suggest that fuzzy regression models provide good alternative to human expert prediction in predicting oral cancer susceptibility.

REFERENCES

Arulchinnappan, S., K. Karunakaran and G. Rajendran, 2011. The use of fuzzy correlation to identify people at risk of oral cancer. Eur. J. Sci. Res., 52: 332-338. http: //www.eurojournals.com/EJSR_52_3_07.pdf

Delavarian, Z., A. Pakfetrat, F. Falaki, M. Pazouki 2010. The role of viruses in oral squamous cell carcinoma in young patients in khorasan (Northeast of Iran). J. Applied Sci., 10: 981-985, http: //www.doaj.org/doaj?func=abstract&id=662952

Hassan, L.H., M. Moghavvemi and H.A.F. Mohamed, 2010. Takagi-sugeno fuzzy gains scheduled pi controller for enhancement of power system stability. Am. J. Applied Sci., 7: 145-152. DOI: DOI: 10.3844/AJASSP.2010.145.152

Madani, A.H., A.S. Jahromi, M. Dikshit and D. Bhaduri, 2010. Risk Assessment of Tobacco Types and Oral Cancer. Am. J. Pharmacol. Toxicol., 5: 9-13. DOI: 10.3844/AJPTSP.2010.9.13

Madani, A.H., M. Dikshit, D. Bhaduri, A.S. Jahromi and T. Aghamolaei, 2010. Relationship between selected socio-demographic factors and cancer of oral cavity. Case Con. Stud. Cancer Inform., 9: 163-169. 10.4137/CIN.S4774

Rosma, M.D., A.K. Sameem, A. Basir, I.S. Mazlipah and M.D. Basir, 2010. The use of artificial intelligence to identify people at risk of oral cancer: empirical evidence in Malaysian University. Int. J. Sci. Res. Educ., 3: 10-20. http://www.ijsre.com/Vol.,%203_1_-Rosma%20et%20al.pdf

Thongwan, T., A. Kangrang and S. Homwuttiwong, 2011. An estimation of rainfall using fuzzy set-genetic algorithms model. Am. J. Eng. Applied Sci., 4: 77-81. DOI: 10.3844/AJEASSP.2011.77.81