ORAL CANCER SUSCEPTIBILITY ASSESSMENT USING PYTHON

Md Musharraf RK, Mohamed Affan M[2], Mohammed Saqui T[3], Dr.Deba Roti Das[4]

Department of Mathematics, School of Advance Sciences

Vellore Institute of Technology, Vellore –632014, Tamil Nadu, India.

E-mail: mdmusharraf.rk2022@vitstudent.ac.in , mohamedaffan.m2022@vitstudent.ac.in, mohammedsaqui.T2022@vitstudent.ac.in, debaroti.das@vit.ac.in

Abstract - A type of cancer that affects the lining of the lips, mouth, or upper throat is called oral cancer. It is estimated that people who use alcohol and tobacco both have a 15 times higher risk of developing oral cancer than those who do not use them. Using statistical application techniques, susceptibility assessment has been done for a selected population based on oral health and habit questionnaire. Methodically collected data has been initially divided into age based small subgroups. A processed data based on If-Then rules with cer- tain conditions has been used to investigate patients' susceptibility to oral pre- cancers and squamous cell carcinoma. To achieve a more accurate prediction of the disease outcome and the associated condition, similarity measures using the Jaccard similarity coefficient has been incorporated. Based on age, gender, demographics, and oral health hygiene, the data set have been clustered in to various groups having different degree of susceptibility to oral cancer. As a result, strategies for policy making via real-life questionnaire data handling will be developed using this analytical approach.

Introduction:

One of the most prevalent cancers is oral cancer, especially in Asia, where it accounts for around 66% of all instances of oral cancer (OC), with an estimated 248,360 new cases and 131,610 fatalities each year. The prevalence of OC is highest among Indian men, and it looks to be spreading around the world [1]. The overall 5-year survival rate of OC patients is between 50 and 60%, despite improvements in surgical techniques, adjuvant radiation, and chemotherapy. Interdisciplinary therapy and early detection can increase the success rates of OC treatment [2]. The most well-known risk factors for OC are HPV-16/18, betel nut use, cigarette smoking, drinking alcohol, and chewing tobacco [3,4,5]. With over 100 million users in India and Pakistan alone, Southeast Asia is thought to account for over 90% of the world's smokeless tobacco usage burden [6]. Annually, nearly 2–3 percent of oral lesions show diseased conversion. Long term effect studies clinically imply that OLK accompanying harsh dysplasia is more liable until revolution into OSCC [7]. Additionally, it has been demonstrated that a person's risk of developing oral cancer is increased when they smoke, consume alcohol, and practise poor oral hygiene because these behaviours affect the resident microbiota that is necessary for maintaining the environment's homeostasis in the mouth . Chronic inflammation and infection are the main factors in cancer pathogenesis [8,9] .

Literature review:

Intake of tobacco is the essential reason of demise and disability worldwide. it's miles acquired from Nicotiana tabacum. When extensively classified, to- bacco is both smoked or smoke less tobacco "a big type of commercially or non-commercially available merchandise and combinations that include tobacco because the principal constituent and are used both orally or nasally with out combustion" [10]. The amount, frequency, and length of smoke less tobacoo exposure are inversely correlated with the occur- rence and severity of tobacco-related oral lesions. Symptomatic exposure may result in OLK [11]. A dosages relationship has been observed between Smoke less tobacco use and an elevated chance of oral cancer across the USA [12]. The roles of Betel quid and Smoke less tobacco in the Indian people are widely understood [13]. In a meta - analytic review, bidi smoking showed a higher odds ratio (OR) than smoking cigarettes when tobacco use is taken into account [14]. A person was thought to have the highest risk of developing cancer if they were subjected to each smoking agents on a regular basis due to their synergistic effects. A study reported that betel nut use could prevent oropharynx and esophageal cancer without consuming cigarettes. We assessed chewing as a possible carcinogen when OSCC was taken into account. An oral pre-cancer (OSF) may develop even without tobacco usage, according to a research, if irritation from frequent chewing causes it [15]. The intensity and severity of chewing betel nut without tobacco was found to have a dosages relation that raised the risk of OSF [16]. Chewing betel nut, whether with or without cigarettes, is regarded as a standalone cancer-causing risk [17]. There are numerous activities you can engage in to campaign of oral health in India [18].

Repeatedly, it was discovered that the Indian population's good oral healthcare awareness, attitude, and sanitary measures practise were directly related to their economic factors. Although poor oral hygiene was thought to be a separate OSCC-causing disease [19,20] . Fuzzy logic is frequently used to interpret ambiguous knowledge that is contained in a system and incorporates hazy human judgement in issues that are not taken into account by any traditional computing techniques. The precision in the meaning of an output and gaining an understanding of a complicated system with tolerance for imprecision are what make fuzzy logic's practical applications so beautiful [21]. By using fuzzy numbers, complex systems can be modelled utilising a higher level of abstraction based on known information and experience [22]. When considering real-life situations, it is frequently observed that the number of addictive products consumed daily is not always fixed, and that people who are illiterate or who lack proper knowledge frequently are unable to provide precise information on the year in which an addictive habit first developed [23].

Proposed method:

The proposed method is demonstrated schematically (Figure. 1) and represented by the following steps:

Subject selection

A study of the rural population in Tamil Nadu, India, was chosen for this study. In spite of its cultural diversity, this population has not received much attention as far as their oral health is concerned. Although this group is well renowned for the variety of cultures represented in the area, limited data recognized about their oral health.

Data collection strategy

A number of variables were collected, including age, gender, education level, oral lesions, alcohol consumption, tobacco smoking, tobacco chewing, chewing of betal nuts and leaves, and brushing habits. Prior written informed consent from all patients and from the concerned authorities was taken. An oral health habit-related questionnaire was used to interview 359 patients (201 males and 158 females) who attended the hospital's outpatient department for oral health treatments.

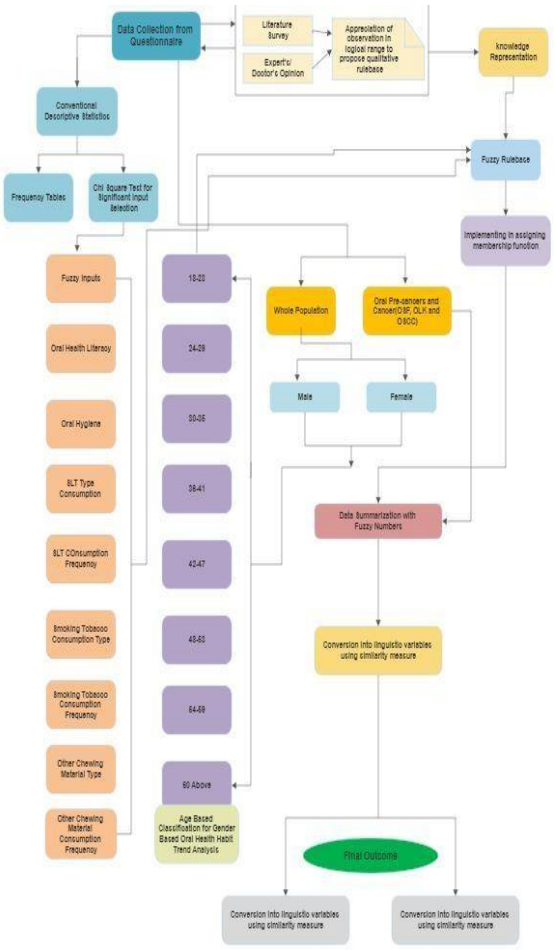

Figure 1 Proposed methodology on application of fuzzy consensus in assessing oral cancer susceptibility

Data Analysis Strategy

Conventional statistical analysis

In order to estimate risk and to select primary features for use in the risk assessment, clinical-epidemiological data from the study subjects were analyzed statistically using Python version 3.0 and scipy module. Pearson's chi-square test was used to compare each input variable between patients with and without oral lesions. and the cutoff significance was established at pi 0.01.The degree of freedom is 1.

Defining fuzzy numbers and rulebase generation

To map linguistic variables from numeric data for fuzzy reasoning, a fuzzy inference rule and membership function must be defined[36]. Eight parameters were considered during the fuzzy rulebase generation. health literacy (X1), Brushing (X2), Smokeless tobacco type (X3), Smokeless tobacco frequency (X4), smoking type (X5), smoking frequency (X6), chewing habit type (X7) and chewing habit frequency (X8). Using physician intuition and literature review information, the rulebase was compiled[11-25]. Each gender was then again separated into eight age groups in years i.e., 18–23, 24–29, 30–35, 36–41,42–47, 48–53, 54–59 and 60 above for age associated trend analysis. A fuzzy scale of 0–1 was used to describe input variables quantitatively through intuition in Table 1.

Summarization of data

Based on age and gender, the data were summarized according to frequencies. In order to summarize the data, the frequency of patients for each age group and gender was sorted out in accordance with each rule, then the relative frequency of each type was multiplied by the corresponding fuzzy scale. To get the final fuzzy number which summarizes the condition of each group quantitatively, each fuzzy number was then added for each group. The data from 36 patients diagnosed with OSF, OSLK and OSCC were used to

Table	**Fuzzy scale(in 0-1)**	
Variable	Linguistic scale	Fuzzy scale
Oral health literacy (X1)	poor	(0, 0, 0.50)
[based on schooling level in	Satisfactory	(0, 0.50, 1)
Indian education system]	Good	(0.50, 1, 1)
Oral hygiene (X2) [assessed on	poor	(0, 0, 0.50)
brushing habit with toothpaste, powder,	Satisfactory	(0, 0.50, 1)
kala-manjan, lal-manjan,daantan, etc,]	Good	(0.50, 1, 1)
SLT type (X3) [based on	Low	(0, 0, 0.50)
materials used like tambaku,Gudaku and related products)	Moderate	(0, 0.50, 1)
	High	(0.50, 1, 1)
SLT frequency (X4)	High	(0, 0, 0.50)
[frequency of consumption Of products divided into type	Moderate	(0, 0.50, 1)
i,e, occasional, less than 2/day,less than 2-5/day, more than 10/day	Low	(0.50, 1, 1)
Smoking type (X5) [based	Low	(0, 0, 0.50)
on materials used like Beedi and cigarette]	Medium	(0, 0.50, 1)
	High	(0.50, 1, 1)
Smoking frequency (X6)	Low	(0, 0, 0.50)
[frequency of consumption of products divided into	Moderate	(0, 0.50, 1)
type, i.e. occasional,less than 2/day, 2-5/day, 5-10/day and more than 10 day]	High	(0.50, 1, 1)
Chewing habit type (X7)	Low	(0, 0, 0.50)
[based on the product used	Medium	(0, 0.50,1)
like betel quid and others consumption)	High	(0.50, 1, 1)
Chewing habit frequency(X8)	Low	(0, 0, 0.50)
[frequency of	Moderate	(0, 0.50, 1)
consumption of products	High	(0.50, 1, 1)
divided into few types. i.e.occasional, less than 2/day,less than 5/day,and more than 10/days		

Table 1: Fuzzy scale for the input variables (in 0–1).

construct a fuzzy disease expert system for patients with oral pre-cancers and cancers. The generated fuzzy numbers were then used for further decision making. In some cases, linguistic interpretation of the mid value of the generated fuzzy set was ambiguous, therefore similarity measures were introduced. Tables 4, 5 for males, females respectively, show the linguistic outputs that were discovered to have changed after taking similarity measures into consideration. These outputs are where If-Then rules to predict disease susceptibility in specific conditions were ultimately obtained.

Result and discussion

Descriptive statistics of the whole data:

Table 2 contains specific observations made when evaluating the addictions that respondents claimed to have a connection to oral carcinogenesis. Table 3 shows the variables that were further taken into account when determining overall clinical outcomes and have confidence intervals. In the whole population, there is substantial evidence of a relationship between daily smoking tobacco (Pearson chi square 2.86), use of SLT (Pearson chi square 0.00143), betel nut chewing (Pearson chi square 1.35) and the presence of oral lesions. Low literacy rates (52.37) people were with education up to 8th standard of Indian education system) in conjunction with debilitating addictive habits were also found to be associated with oral precancers and OSCC occurrence in this area. Daily

brushing habit and higher education showed a protective effect on oral lesion occurrence. However, daily alcohol intake did not show any statistical correspondence with disease prevalence in this population.

Interpretation of the summarized data

The majority of recent research focuses on institutional studies that examine the frequency and distribution of various oral diseases and perform risk estimation for each particular possible causes [38,39]. This investigation hence sought to identify a significant relationship between a number of clinico-epidemiological characteristics related to oral literacy

Variables	Frequency	Percentage
Gender		
Male	201	55.99
Female	158	44.01
Education		
Illiterate	77	21.45
Primary	27	7.52
Medium	84	23.40
Secondary	61	16.99
Higher Secondary	52	14.48
Graduate	50	13.93
Higher than Graduate	8	2.23
Brushing		
Yes	347	96.66
No	12	3.34
Smoking Tobacco		
Yes	65	18.11
No	294	81.89
Smokeless Tobacco		
Yes	122	33.98
No	237	66.02
Betel Nut		
Yes	103	28.69
No	256	71.31
Alcohol		
Yes	111	30.92
No	248	69.08

Table 2: Table of variables, frequency, and percentage

levels, oral habits, and hygiene, as well as oral genetic susceptibility in various ages.This method merely employed a traditional risk estimation technique to extract the basic features. For the purpose of understanding such correlations via trend analysis of the dataset, the systematic fuzzy mapping of characteristics and linguistic transformation of allocated membership functions shown in tables (Tables 4, 5), were employed. Thus, disease risk was evaluated under certain settings utilising built-in If-Then rules. In addition to predicting oral pre-cancers and malignancies (OTH),the guidelines were also inferred to predict oral problems. Tables 4, 5were created by the linguistic transformation of fuzzy numbers for male, female respectively, in the framework of appraising characteristic summary towards the disease conditions. The eight characteristics (X1–X8) all were combined. The regulations are written in a way that one of them can be described as follows : Considering oral hygiene literacy and very good oral hygiene, males between 18 and 23 were more likely to experience oral complications other than pre-cancers and cancers (OTH), as well as very low Smokeless tobacco consumption and low smokeless tobacco consumption frequency, and low smoking tobacco type and low smoking tobacco consumption frequency as well as low chewing material type and low chewing material frequency.The perspective was found to be significantly different when the trend between the male and female population of various age groups was evaluated (Tables 4 and 5). Additionally, it more clearly illustrated age-based feature segmentation. Only women between the ages of 18 and 23, 30 and 41, and 54 and 59 showed greater oral health literacy than men, whereas in the Indian community, men are thought to have higher education than

women [40]. SLT type and consuming frequency were significantly lower in females in this cohort than in males, which there was an increase from age 42 years (Table 4). Again, it is well recognised that OLK and cigarette use are related [10]. The current research has also shown new data to support the If-Then fuzzy rule for the development of such pre-cancer from 40 years, when male smokers' smoking habit and type started to rise from 36 years (Tables 4). Increased prevalence of harmful oral behaviours, such as Smokeless tobacco use, was also observed

Variable	Type of variable	Meaning of variable	Chi-square value	degree of freedom	p- value	
Age	bifurcated (0/1)	age more/less than 40	23.31	1	0.00	Null Hypothesis is rejected.
Education	bifurcated (0/1)	Education Morethan/less than standard 8th	14.45	1	0.00	Null Hypothesis is rejected.
smoking Tobacco	bifurcated (0/1)	yes/no	2.86	1	0.09	Failed to reject the null hypothesis
smokless tobacco	bifurcated (0/1)	yes/no	0.00143	1	0.97	Failed to reject the null hypothesis
Betal Nut chewing	bifurcated (0/1)	yes/no	1.35	1	0.24	Failed to reject the null hypothesis
Brushing	bifurcated (0/1)	yes/no	0.68	1	0.41	Failed to reject the null hypothesis

Table 3: Significant variables obtained from the chi-square test

Age group	Oral health literacy	Oral hy-giene	SLT types con-sump-tion	SLT con-sump-tion fre-quency	Smoking tobacco type	Smoking tobacco con-sump-tion fre-quency	chewing material type	chew ing mate rial fre-quen cy	Disease Chances
18–23	Satis-factory	Good	Low	Low	Low	Low	Low	Low	OTH
24–29	Satis-factory	Good	Low	Low	Low	Low	Low	Low	OTH
30–35	Poor	Good	Low	Low	Low	Low	Low	Low	OTH
36–41	Poor	Good	Low	Low	Medium	Moderate	Low	Low	OTH
42–47	Poor	Good	Low	Moderate	Low	Low	Medium	Low	OTH
48–53	Poor	Good	Moderate	Moderate	Medium	Moderate	Low	Low	OTH
54–59	Poor	Good	Moderate	Moderate	Low	Low	Medium	Low	OTH
60 Above	Poor	Good	Moderate	Moderate	Medium	Moderate	Low	Low	OSF

Table 4: Oral health and habit trend obtained through linguistic conversion of fuzzy numbers in male

	Oral health literacy	Oral hygiene	SLT types con-sump-tion	SLT con-sump-tion fre-quency	Smoking to-bacco type	Smoking tobacco con-sump-tion fre-quency	chewing material type	chewing material frequency	Disease Chances
18–23	Good	Good	Low	Low	Low	Low	Low	Low	OTH
24–29	Poor	Good	Low	Low	Low	Low	Low	Low	OTH
30–35	Poor	Good	Low	Low	Low	Low	Low	Low	OTH
36–41	Poor	Good	Low	Low	Low	Low	Low	Low	OTH
42–47	Poor	Good	Low	Low	Low	Low	Medium	Low	OTH
48–53	Poor	Good	Low	Low	Low	Low	Low	Low	OTH
54–59	Poor	Good	Low	Low	Low	Low	Medium	Low	OTH
60 above	Poor	Poor	Low	Low	Low	Low	Medium	Moderate	OTH

Table 5: Oral health and habit trend obtained through linguistic conversion of fuzzy numbers in female.

in males in the 36-year-old age group (Table 5).

The findings of the other studies can be confirmed by the results of the present investigation, which showed a relationship between high smoking tobacco use and Smokeless tobacco intake, poor oral hygiene, low oral health literacy, and high age in the case of OSCC prevalence.

Conclusion

In order to add value to the results from traditional statistical methods, fuzzy rule-base methods have been used to define specific associations between important clinic epidemiological characteristics and their conceivable effects on disease output in a given dataset. Oral pre-cancers and OSCC were found to be common in the investigated community for a variety of causes, including low literacy rates and crippling addictions. Additionally, a fuzzy If-Then rule trend analysis of oral health and habits revealed disparities in awareness attitude across age groups based on gender.The suggested methodology can also predict the likelihood of illness susceptibility in specific conditions. In contrast to conventional statistical methods that predict disease chances in rigid quantitative values, the novelty of the proposed approach relies on consideration of the uncertainty of circumstances involved in disease incidence and incorporation of physician's intuition in real-life situations. The population of a particular demographic was involved in the new dimension of questionnaire data management, and the same methodology can be used for other demographic circumstances as well. Additionally, it is the first of its kind and can assist physicians and decision-makers in

implementing therapies and habit-preventing techniques.

Appendix A. If-Then Rules for oral health literacy

If Age group 18–23 and Education level is Il or Med Then Oral health literacy is Poor
If Age group 18–23 and Education level is Pr or S or HS Then Oral health literacy is Satisfactory
If Age group 18–23 and Education level is G or HG Then Oral health literacy is Good
If Age group 24–29 and Education level is Il or Med or Pr Then Oral health literacy is Poor
If Age group 24–29 and Education level is S or HS Then Oral health literacy is Satisfactory
If Age group 24–29 and Education level is G or HG Then Oral health literacy is Good
If Age group 30–35 and Education level is Il or Pr or M or S Then Oral health literacy is Poor
If Age group 30–35 and Education level is HS or G Then Oral health literacy is Satisfactory
If Age group 30–35 and Education level is HG Then Oral health literacy is Good
If Age group 36–41 and Education level is Il or Pr or M or S or HS Then Oral health literacy is Poor
If Age group 36–41 and Education level is G Then Oral health literacy is Satisfactory
If Age group 36–41 and Education level is HG Then Oral health literacy is Good
If Age group 42–47 and Education level is Il or Pr or M or S or HS Then Oral health literacy is Poor
If Age group 42–47 and Education level is G Then Oral health literacy is Satisfactory
If Age group 42–47 and Education level is HG Then Oral health literacy is Good
If Age group 48–53 and Education level is Il or Pr or M or S Then Oral health literacy is Poor
If Age group 48–53 and Education level is HS Then Oral health literacy is Satisfactory
If Age group 48–53 and Education level is G or HG Then Oral health literacy is Good
If Age group 53–59 and Education level is Il or Pr or M or S Then Oral health literacy is Poor
If Age group 53–59 and Education level is HS Then Oral health literacy is Satisfactory
If Age group 53–59 and Education level is G or HG Then Oral health literacy is Good
If Age group >60 and Education level is Il or Pr or M or S Then Oral health literacy is Poor
If Age group >60 and Education level is HS Then Oral health literacy is Satisfactory
If Age group >60 and Education level is G or HG Then Oral health literacy is Good

Appendix B. If-Then Rules for oral hygiene

If Age group 18–23 and Brushing with ToothPaste or Powder Then Oral Hygiene Good
If Age group 18–23 and Brushing with KM or LM or Dantan Then Oral Hygiene Satisfactory
If Age group 18–23 and Brushing with Others Then Oral Hygiene Poor
If Age group 24–29 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 24–29 and Brushing with Powder or KM or LM Then Oral Hygiene Satisfactory
If Age group 24–29 and Brushing with Others or Dantan Then Oral Hygiene Poor
If Age group 30–35 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 30–35 and Brushing with Powder Then Oral Hygiene Satisfactory
If Age group 30–35 and Brushing with KM or LM or Dantan or Others Then Oral Hygiene Poor
If Age group 36–41 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 36–41 and Brushing with Powder or KM or LM or Dantan Then Oral Hygiene Satisfactory
If Age group 36–41 and Brushing with Others Then Oral Hygiene Poor

If Age group 42–47 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 42–47 and Brushing with Powder or KM or LM or Dantan Then Oral Hygiene Satisfactory
If Age group 42–47 and Brushing with Others Then Oral Hygiene Poor
If Age group 48–53 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 48–53 and Brushing with Powder or KM or LM or Dantan or Others Then Oral Hygiene Poor
If Age group 54–59 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group 54–59 and Brushing with Powder or KM or LM or Dantan or Others Then Oral Hygiene Poor
If Age group >60 and Brushing with ToothPaste Then Oral Hygiene Good
If Age group >60 and Brushing with Powder or KM or LM or Dantan or Others Then Oral Hygiene Poor

Appendix C. If-Then Rules for Smokeless tobacco type consumption

If Age group 18–23 and SLT type 4 and above Then susceptibility high
If Age group 18–23 and SLT type any 3 Then susceptibility moderate
If Age group 18–23 and SLT type any 2 or 1 or occasional Then susceptibility Low
If Age group 24–29 and SLT type of 4 types or more Then susceptibility high
If Age group 24–29 and SLT type any 3 or 2 Then susceptibility moderate
If Age group 24–29 and SLT type 1 or occasional Then susceptibility low
If Age group 30–35 and SLT type all 4 Then susceptibility high
If Age group 30–35 and SLT type any 3 or 2 or 1 Then susceptibility moderate
If Age group 30–35 and SLT type occasional Then susceptibility low
If Age group 36–41 and SLT type all 4 or 3 or 2 Then susceptibility high
If Age group 36–41 and SLT type 1 Then susceptibility moderate
If Age group 36–41 and SLT type occasional Then susceptibility low
If Age group 42–47 and SLT type 4 or 3 or 2 or 1 Then susceptibility high
If Age group 42–47 and SLT type occasional Then susceptibility moderate
If Age group 48–53 and SLT type 4 or 3 or 2 or 1 Then susceptibility high
If Age group 48–53 and SLT type occasional Then susceptibility moderate
If Age group 54–59 and SLT type 4 or 3 or 2 or 1 Then susceptibility high
If Age group 54–59 and SLT type occasional Then susceptibility moderate
If Age group >=60 and SLT type 4, 3, 2 Then susceptibility very high
If Age group >=60 and SLT TYPE 1 Then susceptibility high
If Age group >=60 and SLT TYPE occasional Then susceptibility moderate

Appendix D. If-Then Rules for smoking tobacco type consumption

If Age group 18–23 and smoking type cigarette Then low
If Age group 18–23 and smoking type Beedi Then Medium
If Age group 18–23 and smoking type Both Then High
If Age group 24–29 and smoking type cigarette Then low
If Age group 24–29 and smoking type Beedi Then Medium
If Age group 24–29 and smoking type Both Then High
If Age group 30–35 and smoking type Beedi or Cigarette Then Medium
If Age group 30–35 and smoking type Both Then High
If Age group 36–41 and smoking type Beedi or Cigarette Then Medium
If Age group 36–41and smoking type Both Then High
If Age group 42–47 and smoking type Bedi or Cigarette Then Medium

If Age group 42–47 and smoking type Both Then High
If Age group 48–53 and smoking type Beedi or Cigarette Then Medium
If Age group 48–5 and smoking type Both Then High
If Age group 54–59 and smoking type Cigarette Then Medium
If Age group 54–59 and smoking type Beedi or Both Then High
If Age group >=60 and smoking type Cigarette Then Medium
If Age group >=60 and smoking type Beedi or Both Then High

Appendix E. If-Then Rules for intake frequency of Smokeless tobacco, smoking tobacco and chewing products

If Age group 18–23 and frequency Occasional or less than 2 The intake is low
If Age group 18–23 and frequency less than 5 or less than 10 The intake is moderate
If Age group 18–23 and frequency more than 10 The intake is high
If Age group 24–29 and frequency Occasional The intake is low
If Age group 24–29 and frequency less than 2 or 5 or 10 The intake is moderate
If Age group 24–29 and frequency more than 10 The intake is high
If Age group 30–35 and frequency Occasional or less than 2 The intake is low
If Age group 30–35 and frequency less than 5 The intake is Moderate
If Age group 30–35 and frequency less than 10 or more than 10 The intake is high
If Age group 36–41 and frequency Occasional or less than 2 The intake is low
If Age group 36–41 and frequency less than 5 The intake is moderate
If Age group 36–41 and frequency less than 10 or more than 10 The intake is high
If Age group 42–47 and frequency Occasional or less than 2 or 5 or 10 The intake is moderate
If Age group 42–47 and frequency more than 10 The intake is high
If Age group 48–53 and frequency Occasional less than 2 or 5 or 10 The intake is moderate
If Age group 48–53 and frequency more than 10 The intake is high
If Age group 54–59 and frequency Occasional or less than 2 or 5 or 10 The intake is moderate
If Age group 54–59 and frequency more than 10 The intake is high
If Age group >=60 and frequency Occasional or less than 2 or 5 or 10 or more than 10 The intake is high

Age group	Oral health literacy	SLT types cousurnption	SLT consumption frequency	Smoking tobacco type	Smoking tobacco consumption frequency	chewing material type	chewing material frequency	Oral hygiene	Y (disease chance)
18-23 (M)	0.3, 0.518, 0.73	0.006, 0.05, 0.097	0.037, 0.081, 0.147	0.013, 0.081, 0.147	0.031,0.065, 0.116	0.088, 0.206, 0.238	0.022, 0.072, 0.131	0.738 0.988, 1	0.004, 0.016, 0.349
24-29 (M)	0.304, 0.533, 0.764	0.056, 0.110, 0.164	0.061,0.117, 0.168	0.009,0.028, 0.131	0.056, 0.112, 0.205	0.103, 0.224, 0.252	0.040, 0.100, 0.159	0.493, 0.743, 0.995	0.006, 0.009, 0.34

30-35 (M)	0.081, 0.190, 0.440	0.089, 0.177, 0.262	0.097, 0.185, 0.259	0.016, 0.113, 0.194	0.060, 0.073, 0.153	0.113, 0.234, 0.242	0.044, 0.105, 0.161	0.484, 0.734, 0984	0.011, 0.027, 0.354
36-41 (M)	0.055, 0.161, 0.411	0.136, 0.25, 0.356	0.225, 0.331, 0.394	0.060, 0.305, 0.492	0.301, 0.441, 0.475	0.093, 0.237, 0.288	0.102, 0.174, 0.288	0.483, 0.725, 0.975	0,0.005, 0.339
42-47 (M)	0.008, 0.084, 0.360	0.242, 0.358, 0.454	0.258, 0.373, 0.462	0.023, 0.2, 0.354	0.192, 0.281, 0.354	0.170, 0.362, 0.385	0.204, 0.3,0.384	0.462, 0.731, 0.962	0.15, 0.026, 0.348
48-53 (M)	0.034, 0.097, 0.347	0.278, 0.420, 0.557	0.313, 0.455, 0.568	0.057, 0.318, 0.523	0.290, 0.420, 0.523	0.136, 0.318, 0.364	0.193,0.284, 0.364	0.42, 0.631, 0.881	0.038, 0.068, 0.386
54-59 (M)	0.043, 0.091, 0.341	0.256, 0.378, 0.469	0.262, 0.384, 0.488	0.085, 0.293, 0.415	0.232,0.335, 0.415	0.195, 0.402, 0.414	0.207,0.311, 0.415	0.463, 0.695, 0.945	0.041, 0.073, 0.394
60 above (M)	0.019, 0.055, 0.305	0.310, 0.454, 0.565	0.319, 0.463, 0.574	0.231, 0.518, 0.574	0.287, 0.430, 0.574	0.148, 0.305, 0.315	0.162,0.240, 0.315	0.481, 0.722, 0.972	0.123, 0.191, 0.468
18-23 (f)	0.383, 0.657, 0.901	0, 0.003,0.009	0,0.003, 0.009	0, 0, 0	0, 0, 0	0.043, 0.136 0.191	0.018,0.049, 0.099	0.741, 0.987, 0.991	0,0,0.333
24-29 (f)	0.223, 0.443, 0.696	0.024, 0.048, 0.072	0.036, 0.066, 0.09	0, 0, 0	0, 0, 0	0.060, 0.157, 0.205	0.033,0.087, 0.139	0.463, 0.696, 0.946	0,0,0.333
30-35 (f)	0.013, 0.16 0.303	0.025, 0.053, 0.081	0.01, 0.06, 0.200	0, 0, 0	0, 0, 0	0.075, 0.188, 0.256	0.038, 0.109, 0.178	0.444, 0.675, 0.925	0.004, 0.008, 0.341
36-41 (f)	0.019, 0.051, 0.301	0.051, 0.106, 0.162	0.102, 0.157, 0.204	0, 0, 0	0, 0, 0	0.093, 0.231, 0.278	0.111, 0.181, 0.241	0.50,0.75,1	0.006, 0.012,0.345
42-47 (f)	0.026, 0.057, 0.302	0.188, 0.281, 0.365	0.193, 0.286, 0.375	0, 0.021, 0.042	0.021, 0.031, 0.042	0.156, 0.365, 0.417	0.219,0.323, 0.417	0.427, 0.666, 0.917	0.014, 0.028,0.361
48-53 (F)	0.006, 0.020, 0.270	0.090, 0.167, 0.243	0.153, 0.230, 0.306	0, 0.070, 0.138	0.069, 0.104, 0.138	0.083, 0.181, 0.194	0.097,0.146, 0.194	0.458, 0.688, 0.938	0.009, 0.019, 0.361

54-59 (f)	0, 0.026, 0.276	0.145, 0.211, 0.263	0.132, 0.197, 0.263	0, 0, 0	0, 0, 0	0.184, 0.395, 0.421	0.210, 0.316, 0.421	0.421, 0.632, 0.880	0.018 0.053,0.881,0.368
60 above(f)	0.01, 0.03, 0.28	0.23, 0.34,0.43	0.23, 0.33,0.40	0.03, 0.04, 0.04	0.02, 0.03, 0.04	0.2, 0.42, 0.44	0.26, 0.38,0.48	0.19, 0.38, 0.63	0.053, 0.08, 0.386

Appendix F. Assigned membership functions for the fuzzy set

Age group	Education	smokeless type	smokeless frequency	Smoking type	Smoking frequency	Chewing type	Chewing frequency	Oral hygiene	Y (disease chance)
18–23 (M)	0.945	0.972	0.976	0.905	0.973	0.888	0.976	0.988	0.974
24–29 (M)	0.986	0.976	0.965	0.904	0.972	0.881	0.972	0.984	0.974
30–35 (M)	0.979	0.945	0.945	0.918	0.975	0.869	0.97	0.985	0.974
36–41 (M)	0.973	0.971	0.945	0.712	0.947	0.886	0.951	0.985	0.974
42–47 (M)	0.966	0.94	0.932	0.927	0.956	0.821	0.956	0.983	0.975
48–53 (M)	0.966	0.969	0.976	0.874	0.96	0.841	0.957	0.965	0.968
54–59 (M)	0.968	0.936	0.943	0.874	0.947	0.836	0.958	0.982	0.966
60 above (M)	0.98	0.974	0.976	0.896	0.973	0.837	0.955	0.984	0.936
18–23 (F)	0.966	0.942	0.942	0.83	0.937	0.907	0.972	0.988	0.973
24–29 (F)	0.981	0.962	0.964	0.832	0.937	0.903	0.973	0.982	0.973
30–35 (F)	0.981	0.965	0.988	0.832	0.937	0.906	0.972	0.977	0.974
36–41 (F)	0.981	0.969	0.944	0.832	0.937	0.887	0.935	0.985	0.974
42–47 (F)	0.98	0.959	0.959	0.857	0.953	0.838	0.953	0.974	0.973
48–53 (F)	0.986	0.944	0.951	0.903	0.962	0.884	0.949	0.98	0.971
54–59 (F)	0.952	0.94	0.945	0.832	0.937	0.84	0.957	0.949	0.974
60 above (F)	0.985	0.947	0.928	0.851	0.952	0.849	0.94	0.943	0.966

Appendix G. Jaccard similarity measure for linguistic output selection

References:

[1] Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249.

[2] Yang, C.Y.; Yeh, Y.M.; Yu, H.Y.; Chin, C.Y.; Hsu, C.W.; Liu, H.; Huang, P.J.; Hu, S.N.; Liao, C.T.; Chang, K.P.; et al. Oral Microbiota Community

Dynamics Associated with Oral Squamous Cell Carcinoma Staging. Front. Microbiol. 2018, 9, 862.
[3] Tenore, G.; Nuvoli, A.; Mohsen, A.; Cassoni, A.; Battisti, A.; Terenzi, V.; Della Monaca, M.; Raponi, I.; Brauner, E.; De Felice, F.; et al. Tobacco, Alcohol and Family History of Cancer as Risk Factors of Oral Squamous Cell Carcinoma: Case-Control Retrospective Study. Appl. Sci. 2020, 10, 3896.
[4] Rao, S.V.K.; Mejia, G.; Roberts-Thomson, K.; Logan, R. Epidemiology of Oral Cancer in Asia in the Past Decade—An Update (2000–2012). Asian Pac. J. Cancer Prev. 2013, 14, 5567–5577.
[5] Chaturvedi, P.; Singh, A.; Chien, C.Y.; Warnakulasuriya, S. Tobacco Related Oral Cancer. BMJ 2019, 365, l2142.
[6] Imam, S.Z.; Nawaz, H.; Sepah, Y.J.; Pabaney, A.H.; Ilyas, M.; Ghaffar, S. Use of Smokeless Tobacco among Groups of Pakistani Medical Students—A Cross Sectional Study. BMC Public Health 2007, 7, 231.
[7] Kurokawa H, Matsumoto S, Murata T, et al. Immunohistochemical study of syndecan-1 down-regulation and the expression of p53 protein or Ki-67 antigen in oral leukoplakia with or without epithelial dysplasia. J Oral Pathol Med 2003;32(9):513–21.
[8] Ganly, I.; Yang, L.; Giese, R.A.; Hao, Y.; Nossa, C.W.; Morris, L.G.T.; Rosenthal, M.; Migliacci, J.; Kelly, D.; Tseng, W.; et al. Periodontal Pathogens Are a Risk Factor of Oral Cavity Squamous Cell Carcinoma, Independent of Tobacco and Alcohol and Human Papillomavirus. Int. J. Cancer 2019, 145, 775–784.
[9] O'Grady, I.; Anderson, A.; O'Sullivan, J. The Interplay of the Oral Microbiome and Alcohol Consumption in Oral Squamous Cell Carcinomas. Oral Oncol. 2020, 110, 105011.
[10] Bokor-Bratic´ M, Vucˇkovic´ N. Cigarette smoking as a risk factor associated with oral leukoplakia. Arch Oncol 2002;10(2):67–70.
[11] Boffetta P, Hecht S, Gray N, Gupta P, Straif K. Smokeless tobacco and cancer. Lancet Oncol 2008;9(7):667–75. [12] Carr AB, Ebbert JO. Interventions for tobacco cessation in the dental setting. A systematic review. Community Dent Health 2007;24(2):70–4.
[12] Rodu B, Cole P. Smokeless tobacco use and cancer of the upper respiratory tract. Oral Surg Oral Med Oral Pathol Oral Radiol Endodontol 2002;93(5):511–5.
[13] Jayalekshmi PA, Gangadharan P, Akiba S, Koriyama C, Nair RR. Oral cavity cancer risk in relation to tobacco chewing and bidi smoking among men in Karunagappally, Kerala, India: Karunagappally cohort study. Cancer Sci 2011;102(2):460–7.
[14] Rahman M, Sakamoto J, Fukui T. Bidi smoking and oral cancer: a meta-analysis. Int J Cancer 2003;106:600–4.
[15] Yang YH, Ho PS, Lu HM, Huang IY, Chen CH. Comparing dose–response measurements of oral habits on oral leukoplakia and oral submucous fibrosis from a community screening program. J Oral Pathol Med 2010;39(4):306–12.
[16] Jacob BJ, Straif K, Thomas G, et al. Betel quid without tobacco as a risk factor for oral precancers. Oral Oncol 2004;40(7): 697–704.
[17] Muwonge R, Ramadas K, Sankila R, et al. Role of tobacco smoking, chewing and alcohol drinking in the risk of oral cancer in Trivandrum, India: a nested case-control design using incident cancer cases. Oral Oncol 2008;44(5):446–54.
[18] Paul B, Basu M, Dutta S, Chattopadhyay S, Sinha D, Misra R. Awareness and practices of oral hygiene and its relation to sociodemographic factors among patients attending the general outpatient department in a tertiary care Hospital of Kolkata, India. J

Family Med Prim Care 2014;3(2):107–11.
[19] Chandra Shekar B, Reddy CVK, Manjunath B, Suma S. Dental health awareness, attitude, oral health-related habits, and behaviors in relation to socio-economic factors among the municipal employees of Mysore city. Ann Trop Med Public Health 2011;4(2):99–106.
[20] Guha N, Boffetta P, Wunsch Filho V, et al. Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies. Am J Epidemiol 2007;166(10):1159–73.
[21] Singh H, Gupta MM, Meitzler T, et al. Real-life applications of fuzzy logic. Adv Fuzzy Syst 2013.
[22] Dubois D, Prade HM. Fuzzy sets and systems: theory and applications. Academic Press; 1980
[23] Elderd BD, Dukic VM, Dwyer G. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. Proc Natl Acad Sci 2006;103(42):15693–7.
[24] Patil PB, Bathi R, Chaudhari S. Prevalence of oral mucosal lesions in dental patients with tobacco smoking, chewing, and mixed habits: a cross-sectional study in South India. J Family Community Med 2013;20(2):130. [39] Sujatha D, Hebbar PB, Pai A. Prevalence and correlation of oral lesions among tobacco smokers, tobacco chewers, areca nut and alcohol users. Asian Pac J Cancer Prev 2012;13(4):1633–7.
[25] Kingdon GG. Does the labour market explain lower female schooling in India? J Develop Stud 1998;35(1):39–65.
[26] Rani M, Bonu S, Jha P, Nguyen S, Jamjoum L. Tobacco use in India: prevalence and predictors of smoking and chewing in a national cross sectional household survey. Tobacco Control 2003;12(4), e4 (1–8).
[27] Bokor-Bratic´ M, Vuc˘kovic´ N. Cigarette smoking as a risk factor associated with oral leukoplakia. Arch Oncol 2002;10(2):67–70.