International Journal of Engineering & Technology, 7 (3.28) (2018) 76-79

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Using Decision Tree and Response Surface Methodology to Review Chronic Periodontitis Profile in Pakistan: a Preliminary Case Study

Wan Muhamad Amir W Ahmad[1*], Nasar Um Min Allah[1], Mohamad Shafiq Mohd Ibrahim[1], Mustafa Mamat[3], Nor Azlida Aleng[2], Adam Husein[1], Mohamad Arif Awang Nawi[1], Muhammad Azeem Yaqoob[1]

[1]School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia
[2]School of Informatics and Applied Mathematics (PPIMG), 21030 Kuala Terengganu, Terengganu, Malaysia
[3]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UNISZA), 21030 Kuala Terengganu, Terengganu, Malaysia
**Corresponding author E-mail: wmamir@usm.my*

Abstract

Chronic periodontitis is among the most prevalent oral diseases in the world. Apart from its destructive outcomes in the oral cavity, there is a strong evidence of the potential association between chronic periodontitis and systemic diseases such as cardiovascular disease, diabetes, and preterm low birth weight. In this paper, the potential contributing factors which lead to chronic periodontitis will be determined among the patients attending Islamabad Dental Hospital by response surface methodology (RSM) and decision tree (DT) analysis. The results of this study show that the prevalence of severe chronic periodontitis is high in patients who have not brushed their teeth (23.5%) and for the case without the severity of chronic periodontitis (61.1%). Whereas, from the RSM analysis, it is shown that there is an association between chronic periodontitis, diabetic status, and frequency of smoking. However, using DT analysis, age and frequency of brushing were the best predictors of severe chronic periodontitis. This present study concludes that by using RSM and DT analysis can predict a better forecasting result in future for the decision making among the decision maker.

Keywords: *Chronic Periodontitis; Response Surface Methodology; Decision Tree*.

1. Introduction

Chronic periodontitis is an infectious inflammatory disease caused by the bacteria of the dental plaque that affects the protective and supportive tissues (the gingiva, the periodontal ligament, cementum and the alveolar bone) around the tooth [1]. The progressive destruction of the supporting tissues of the tooth may result in an increased tooth mobility, oral discomfort and loss of function, halitosis and eventually tooth loss. This disease is an infection of the global prevalence and mainly affect the adult population. Epidemiological data suggests that the prevalence of chronic periodontitis increases with age which is correlated with the increase in extent and severity of the disease. Approximately, 35.7% of adults aged 30 to 39 years experienced greater than 3 mm of attachment loss compared with 89.2% for those individuals aged 80 to 90 years [2]. Since the accumulation of dental plaque at the tooth gingival margins are the main cause of the periodontal disease, the growth and maturation of which are considered to have a primary role in the initiation, persistence and progression of the chronic periodontitis. Beside this, there are certain risk factors which are involved in the susceptibility of the progressive periodontitis. These include cigarette/tobacco smoking, genetic factors, some systemic diseases (e.g. diabetes mellitus), the presence of specific bacteria within the microbial plaque and psychosocial factors (e.g. stress, depression) [3, 4].

In statistical analysis, data mining is an effective and practical approach for identifying key associations in data obtained from numerous perspectives. This approach is normally used to extract a significant information from large data sets and to present it in easy-to-interpret visualizations. The decision tree (DT) models, first introduced in 1960's is the powerful and nonparametric statistical method commonly used in data mining to analyse complex data and induce the DT which is used to make classifications or predictions [5, 6]. Decision trees are a class of predictive data mining tools, which predict either a categorical or continuous response variable. More recently, the decision tree methodology has become popular in medical and dental research [7-9]. They get their name from the structure of the models built [10]. A series of decisions are made to segment the data into homogeneous subgroups. This is also called recursive partitioning. When drawn out graphically, the model can resemble a tree with branches [10].

2. Methodology

Data from medical unit of record of Islamabad Dental Hospital (a teaching Hospital of Islamabad Medical and Dental College) were review and related information was extracted. A total of 221 eligible patients were selected from the list of a patient diagnosed with oral squamous cell carcinomas admitted to Islamabad dental hospital. The selected variables are shown in Table 1.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1: Description of data

No.	Variables	Explanation of User Variables
1.	Freq_Smoking	Frequency of smoking 1 = Non-Smoker 2 = Ex-Smoker 3 = Occasional 4 = LightSmoker 5 = HeavySmoker
2.	Diabetic_Status	Diabetes Status: 1=Yes, 0= No
3.	Chperio_Status	Severity of Chronic Periodontic: 1 = Yes, 0 = No
4	Age	Age of Patient in Year
5.	Freq_Brushing	Frequency of Brushing

2.1. Contour plot and surface plot

Contour plot and the surface plot is being used to examine the relationship between one or more response variables and a set of quantitative experimental variables or factors. In this study, a first-order model takes the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ where $(\beta_0, \beta_1, \beta_2)$ are regression coefficients, (x_1, x_2) are independent or predictor variables which are consist of the frequency of smoking and diabetic status, ε is random error and y is a dependent or response variables (chronic pro status). The analysis can be performed by MINITAB. At the first stage of analysis by using MINITAB, we have to define the custom response surface design for our datasets and follow by analysis response surface design. After running the analysis, MINITAB will produce an important information that related to study model.

2.2. Decision tree of chronic periodontitis

In this study, the decision process begins with the status of chronic periodontitis. For the age factor, the split had divided into three groups. The first group was focused on the age, which is less than 30 years old. The second group focused on 30 to 43 years old and the third group was focused on those patients who are at the age of more than 43 years old. The second split was the frequency of brushing. Decision trees are indispensable graphical tools in such settings, which allow for an intuitive understanding of the problem and aid for the optimal decision making. The conceptual framework of our proposed decision tree is outlined as follows.

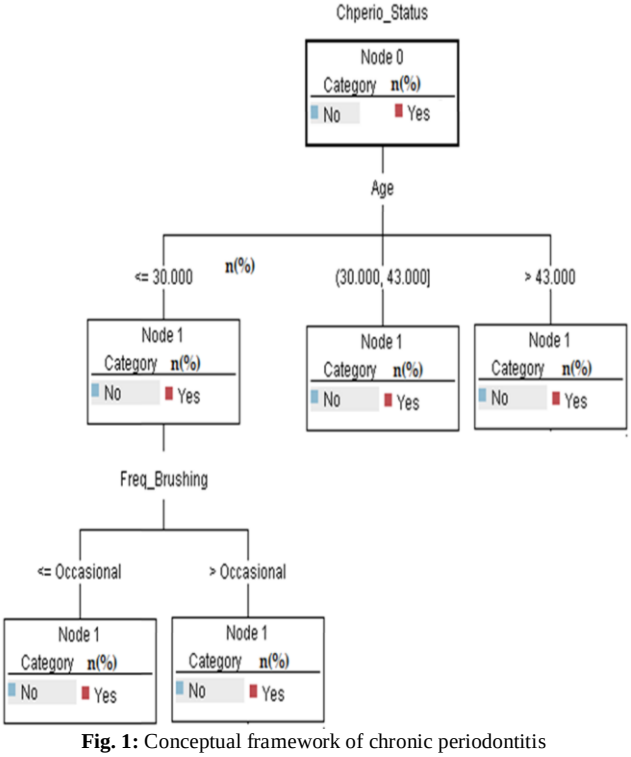


Fig. 1: Conceptual framework of chronic periodontitis

3. Results and discussion

Table 2: Descriptive statistics for severity of chronic periodontics

	The Severity of Chronic Periodontics	
	No (%)	Yes (%)
Frequency of Smoking		
Non-Smoker	143(64.7%)	Non-Smoker
Ex-Smoker	1 (0.5%)	Ex-Smoker
Occasional	6(2.7%)	Occasional
Light Smoker	4(1.8%)	Light Smoker
Heavy Smoker	15(6.8%)	Heavy Smoker
Frequency of Brushing		
None	135(61.1%)	None
Occasional	18(8.1%)	Occasional
Once a day	16(7.2%)	Once a day
Diabetic Status		
No	167(75.6%)	No
Yes	2(0.9%)	Yes

Data were analyzed using SPSS for the window. The distribution and frequencies were examined and expressed as frequencies and percentage. A total of 221 eligible patients were recruited for the study. The patients with the characteristics of severity status can be evaluated easily through the Table 2. The severity of chronic periodontics most occurs among the patients which do not brush their teeth, it has been estimated 14.5% of the total of the frequency of brushing.

3.1. Contour plot and surface plot

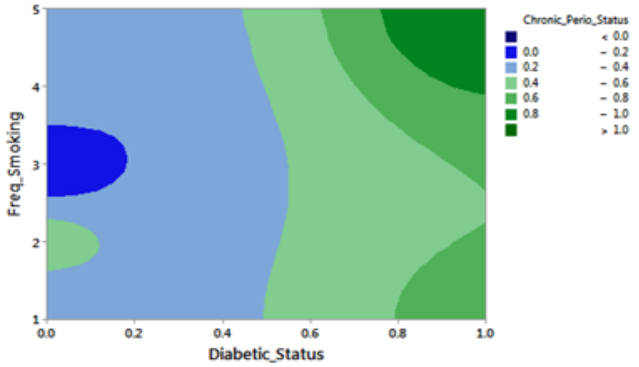


Fig 2: Contour plot for chronic periodontitis versus diabetic and smoking status

To explore the underlying association between chronic periodontitis, diabetic status and frequency of smoking, let us define the model as follows:

Y' = -1.379 + 0.044 Freq_Smoking + 2.531 Diabetic_Status (1)

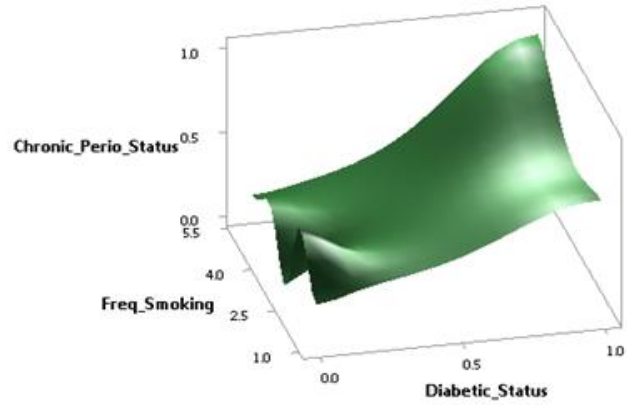


Fig. 3: Contour plot for chronic periodontitis versus diabetic and smoking status

The counter and surface plots indicate that the chronic periodontitis is mostly found among patients that have diabetic with light or heavy smoker. This area appears at the upper right corner of the

plot. This scenario explains that chronic periodontitis has an association with diabetic patients and smoking status. If both factors occur together, the interaction effect will be highly significant and lead to the chronic periodontitis status. From both potential variables, we can see clearly how the pattern of chronic periodontitis get increasing significantly, moving together with the increasing of the frequency of smoking and from non-diabetic patients to diabetic patients.

3.2. Decision tree of chronic periodontitis

The model was constructed based on the recommendation proposed by IBM SPSS Modeler 18.0 to ensure that this model fits the data. To determine the suitability of the model, the adj. p-value from the Chi-Square were evaluated for each split (a smaller p-value indicates a better quality of the split). A decision tree model has been designed for the chronic periodontitis status patient which consist of two predictors variable. The process will automatically include in its rule, only that attributes that really matter in making a decision. The attributes that do not contribute to the accuracy of the tree are ignored. This allows us to develop a classification system that predicts observation based on a set of decision rules. Figure 4 shows the top two predictors in a ranking according to their prediction important to chronic periodontitis, the highest contribution is age factor and followed by frequency of brushing.

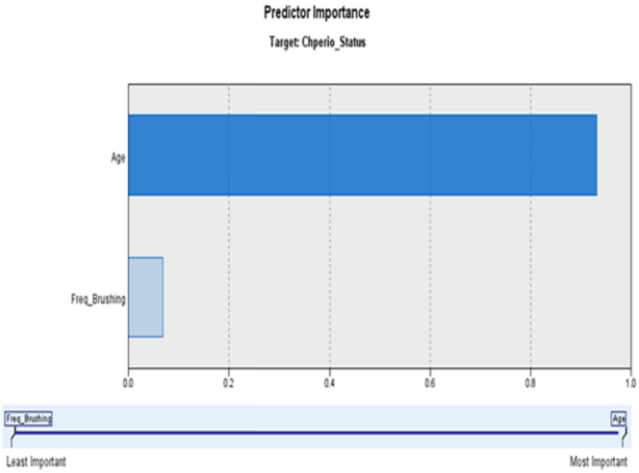


Fig. 4: Decision tree for oral cancer classification

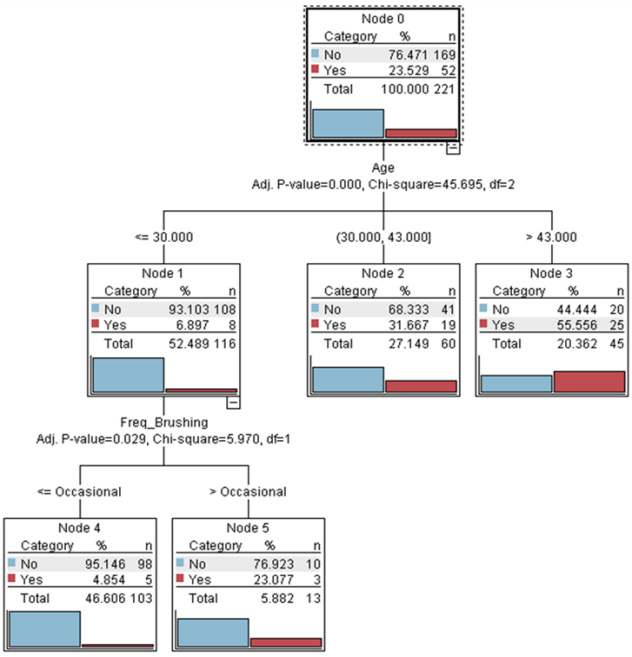


Fig. 5: Decision tree for oral cancer classification

The decision tree model was constructed based on the recommendation proposed by IBM SPSS Modeler 18.0. To ensure that this model fits the data, we based on the smallest p-value which gained from each split. According to the decision tree analysis, the first split of chronic periodontitis is based on age factor which is given by p-value = 0.000, Chi-square = 45.695, d.f. = 2. The second split focuses on the frequency of brushing. The adjusted p-value = 0.029, Chi-square = 5.970, d.f. = 1. A significant p-value from the first and second layer shows the better quality of a split. From Figure 5, a chronic periodontitis status is assigned to Node 0. The first layer is divided into three categories of age group. The first group is patients with age less than 30, the second group is the patients with the age between 30 to 43and the third group is the group of patients with the age of more than 43 years old. From the total sample studied, 52 (23.53%) of the patients having chronic periodontitis. Most of the patients came from age group which more than 43 years old. The trend of getting chronic periodontitis increased from Node 1 (age less than 30) till Node 3(age more than 43). At the age of more than 43, there was 25 (55.6%) cases record. This indicated that the awareness of chronic periodontitis is lack among patients at the age of more than 43. Interestingly, the decision tree show there is 5(4.8%) patients at the age of less than 30 which come from Node 4 with the status of brushing occasionally. Patients with the status of more brushing than occasionally showing the decreasing trend of chronic periodontitis.

4. Conclusion

This paper examines the factor influence that leads to chronic periodontitis among periodontitis patients. Basically, this paper provides only a preliminary overview of the factor that has an association with periodontitis and chronic periodontitis. The first part of this paper is on descriptive statistics, the second part is emphasizing on contour and response surface methodology and the third is focused on decision tree analysis descriptive statistics reveal some important information about the patient's background of the study. From Table 2, we can see the distribution of periodontics and chronic periodontics occurs in patients with non-smoking status, the frequency of brushing and diabetic status. In this study, the prevalence of severe chronic periodontitis is high in patients who have not brushed their teeth, it is about 23.5% and for the case without the severity of chronic periodontitis is 61.1%. The second part of the analysis is to perform the response surface methodology and contour plot. We have to be very cautious and keep in mind that these plots are based on a model equation (1), the suggested model must be adequate before interpreting the plots. It is surprising that in through this model, it appears that there is an association between chronic periodontitis, diabetic status and frequency of smoking. The results from this section that the chronic periodontitis is mostly found among patients that have diabetic with light or heavy smoker. Those patients with periodontitis disease are at high risk of chronic periodontitis. This two factors will be highly significant and lead to the chronic periodontitis status. The pattern from the contour plot give a preliminary idea to make an inference, the same interpretation goes to response surface methodology method. The third part of the analysis is a decision tree analysis. The result from the decision tree reveals the findings more explicitly due to use of diagram classification for periodontitis and chronic periodontitis patients. Using the CHAID method, age and frequency of brushing were the best predictors of severe chronic periodontitis. From the results obtained, we can predict a better forecasting result in future for the decision making among the decision maker. From the analysis, a recommendation on the following study can be done by as adding some more potential variable which can be used together in order to optimize the gained output. This could include a higher level of a statistical analysis combination of different methodology, which leads to the higher precise and accuracy of the results. With more detailed data, the advantage of the covariate estimation will be more precise.

Acknowledgement

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing the research funding (RUI Grant no.1001/PPSG/8012278, School of Dental Sciences, Health Campus).

References

[1] Shaddox LM, & Walker CB (2010), Treating chronic periodontitis: Current status, challenges, and future directions. Clinical, Cosmetic and Investigational Dentistry, 2, 79-91.
[2] Burt B (2005), Position paper: Epidemiology of periodontal diseases. J. Periodontol, 76(8), 1406-1419.
[3] Newman MG, et al. (2011), Carranza's clinical periodontology. Elsevier Health Sciences.
[4] Pihlstrom BL (2001), Periodontal risk assessment, diagnosis and treatment planning. Periodontol, 2000 25, 37-58.
[5] Quinlan JR (2014), C4. 5: Programs for machine learning. Elsevier.
[6] John Lu, Z (2010), The elements of statistical learning: Data mining, inference, and prediction. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(3), 693-694.
[7] Dima S et al., (2018), Decision Tree Approach to the Impact of Parents' Oral Health on Dental Caries Experience in Children: A Cross-Sectional Study. Int J. Environ Res. Public Health, 5(4).
[8] López-Vallverdú JA, Riaño D, & Bohada JA, (2012), Improving medical decision trees by combining relevant health-care criteria. Expert Systems with Applications, 39(14), 11782-11791.
[9] Rohlin M, & Mileman PA (2000), Decision analysis in dentistry the last 30 years. J. Dent, 28(7), 453-468.
[10] Textbook ES (2013), StatSoft, Inc. Tulsa, StatSoft.