ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

The High Potential Prognostic Factors for Oral Squamous Cell Carcinoma (OSCC): A Neural Network (RBF) and Principle Components Analysis (PCA) Plot Approach

Received: 14 February 2023, **Revised:** 16 March 2023, **Accepted:** 20 April 2023

Wan Muhamad Amir W Ahmad1*, Muhammad Azeem Yaqoob2, Mohamad Nasarudin Adnan1, Farah Muna Mohamad Ghazali1, Nor Farid Mohd Noor3, Nor Azlida Aleng4, Nurfadhlina Abdul Halim5, Nuzlinda Abdul Rahman6, Noor Maizura Mohamad Noor7

1School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia.
Email: wmamir@usm.my, nasarudinadnan@student.usm.my, muna_ghazali@yahoo.com
2Oral Medicine Department, Shahida Islam Dental College, Lodhran, Pakistan.
Email: dr.axeem.sr@gmail.com
3Faculty of Medicine, Medical Campus, Universiti Sultan Zainal Abidin (UniSZA), Jalan Sultan Mahmud, 20400 Kuala Terengganu, Terengganu, Malaysia.
Email: faridmnoor@unisza.edu.my
4Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia.
Email: azlida_aleng@umt.edu.my
5 Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.
Email: nurfadhlina@usim.edu.my
6School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia.
Email: nuzlinda@usm.my
7Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia.
Email: maizura@umt.edu.my
*Corresponding author: wmamir@usm.my

Keywords
Radial basis function (RBF); principle components analysis (PCA); oral squamous cell carcinoma (OSCC)

Abstract
The aim of this research paper is to discuss the basic idea of the methodology building of the prognostic model for Oral Squamous Cell Carcinoma (OSCC) through the Radial Basis Function (RBF) model and Principle Component Analysis (PCA). This combined methodology is our main focus for the OSCC factor determination. Results from the RBF model will be verifying through PCA for validation purposes. Through the PCA plots, special characteristics or patterns of the tumor size will be assessed. The gained pattern will be studied for the characteristic association. This combined technique had led to successful research, giving the best results for decision making especially among the decision-maker and provide very important information for health science education. The utmost finding from this study, it provides a very useful information about the tumor size and their association and also gives an idea for further active treatment.

1. Introduction

Oral cavity squamous cell carcinoma (OCSCC) is the sixth most prevalent malignant tumour (Ajila, 2015) and is a virulent disease with up to 50% of mortality rate although modalities in treatment (Mehrotra, 2006). The incidence of OCSCC varies from region to region, habits, prevention and awareness. Worldwide GLOBOCAN in 2012 estimates that incidence of new cases of oral cavity cancer were 300373. Furthermore, estimation of oral cavity cancer is highest in South Asia World Health Organization (WHO). The mortality rate of oral cancer was 2.7 per 100000 worldwide (Ferlay, 2015). Oral cancer dynamic risk factors have been identified which include tobacco, alcohol, betel nut, human papilloma virus (HPV)

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

environmental factors, infection agent, life style, and epigenetic changes (Gillison, 2000). An essential difficulty with better survival rate of OCSCC patients is that, majority of cases are diagnosed in the late stage. In this view, it is essential that estimation be developed for the easy determination of risk factor (Yaaqob, 2019). The outcome of patient's prognostic factors is one of the pivotal challenges in the management of oral squamous cell carcinoma (OSCC) (Almangush, 2015). TNM staging classification and site of tumor are commonly used the criteria for prognosis and management of OSCC (Dissanayaka, 2012). The prognosis of OSCC is not consistent although tumor site, stage and treatment is similar. In view of these challenges, it has been suggested that assessing the high potential prognostic factors could help the clinicians in decision making during management that leads towards improvement of prognosis of OSCC (Brandwein, 2005).

2. Material and Methods

The archive of medical unit record of Hospital Universiti Sains Malaysia (USM) was reviewed and related information was extracted. A total of 16 eligible cases were selected from the list of a patient diagnosed with OSCC. The selected variables are shown in Table 1 as follows:

Table 1: Data Description of the variable in study.

No.	Variables	Explanation of user variables
1	Smoking	Smoking Status 1 = Never, 2 = Stop, 3 = Current
2	Tumour_Size	Tumour Site 1 = Gum, 2 = Tongue, 3 = Cheek, 4 = Lip
3	Betel_Quid	Betel Quid Status 1 = Never, 2 = Stop, 3 = Current
4	Tumour_Size	Tumour Size 1 = <2cm, 2 = 2cm - 4cm, 3 = >4cm

2.1 Radial Basis Function (RBF)

Radial Basis Function (RBF) consists of an input layer, one or several hidden layers and an output layer. The neurons in (RBF) are generally grouped into layers. Signals flow in one direction from the input layer to the next, but not within the same layer (Mehrotra 2006). In this research, the output node is fixed at one since there is only one independent variable. Thus, for the Radial Basis Function (RBF) with N input nodes, H hidden nodes, and one output node, the values $\hat{y}$ are given by $\hat{y} = g\left(\sum_{j-1}^{H} w_j h_j + w_0\right)$. Where w_j an output weight from hidden node j to the output node is w_0 the bias for the output node, g is an activation function. The values of the hidden node h_j, $j = 1 \ldots H$ are given by $h_j = k\left(\sum_{i-1}^{N} v_{ji} x_j + v_{j0}\right) j = 1,\ldots,H$. Here, v_{ji} the output *weight* from input node i to hidden node j, v_{j0} is the bias for hidden node j, x_i is the independent variables where $i = 1 \ldots N \ldots$ and k is

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

an activation function. The architecture of the (RBF) model is illustrated in Figure 1.

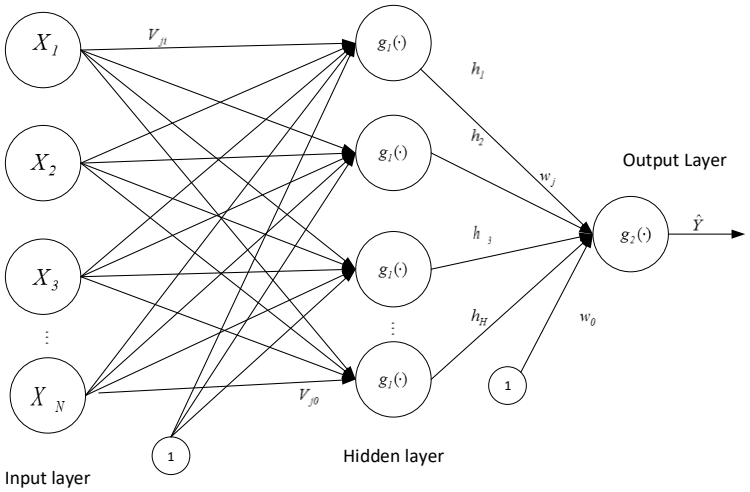

Figure 1 Radial Basis Function (RBF) model with N input nodes, H hidden nodes and one output node.

2.2 Principle Component Analysis (PCA)

PCA is a well-established multivariate analysis technique which concerned with data sets that have more than one dependent variables (Khaled, 2006). The PCA model can be represented as $u_{m\times1} = W_{m\times d} X_{d\times1}$ with u, *m*-dimensional vector, is a projection of *x*-the original *d*-dimensional data vector $(m < d)$. It can be presented that the m projection vectors that maximize the variance u, called the principal axes, are given by the eigenvectors $e_1, e_2, .., e_m$ of the data set's covariance of matrix S, where $s = \frac{1}{n-1}\sum_{i-1}^{n}(\chi-\mu)(\chi-\mu)^T$, corresponding to the m largest non-zero eigenvalues $\lambda_1, \lambda_2, ..., \lambda_m$. The data set's covariance matrix *S* is given as $s = \frac{1}{n-1}\sum_{i-1}^{n}(\chi-\mu)(\chi-\mu)^T$,where μ is the mean vector of χ. The eigenvectors e_i can be found by solving the set of equations: $(S-\lambda_i I)e_i = 0 \; i = 1,2,...,d$ where λ_i are the eigenvalues of *S*. After calculating the eigenvectors, they are sorted by the magnitude of the corresponding eigenvalues. Then, the *m* vectors with the largest eigenvalues are chosen. The PCA projection matrix is then calculated as $W = E^T$, where E has the *m* eigenvectors. Here *W* is a $m\times d$ matrix. The dimensionality reduction is achieved by calculating the first few principal components representing the highest variance in the components of the input feature vector, without the need to perform any transformations on the input space (Khaled, 2006).

3. Results

The model Radial Basis Function (RBF) was constructed based on the recommendation proposed by IBM SPSS Modeler 18.0 to ensure that this model fits the data. From the result, the accuracy analysis of the RBF is 75%

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

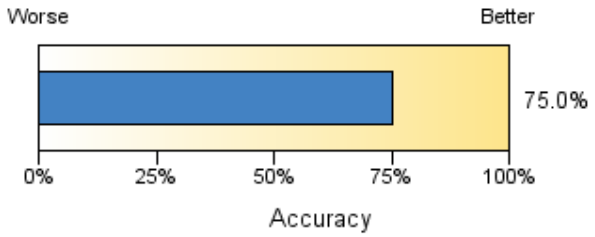

Figure 2 The Accuracy of Radial Basis Function (RBF).

The Radial Basis Function (RBF) architecture is composed of the number of inputs, hidden and output nodes. There are four selected variables, which were smoking, tumour site, betel quid and tumour size. The output node in this study is one node since we have one dependent variable which is refer to tumour size. The studied data was partitioned into three parts which are training partition size (60%), testing partition size (30%) and validation (10%). Below is the path which using SPSS Modeler to obtain the neural network analysis.

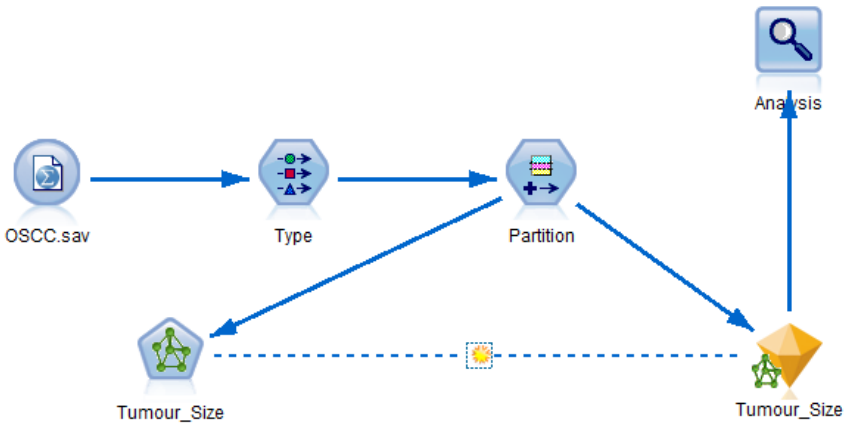

Figure 3 Radial Basis Function (RBF) path in SPSS Modeller.

Since we consider all three independent variables (Figure 4) as inputs for the RBF, then input nodes are three nodes and as tumour size is considered as the output (output node). We then apply the RBF to find the best number of hidden nodes (automatically computer number of units).

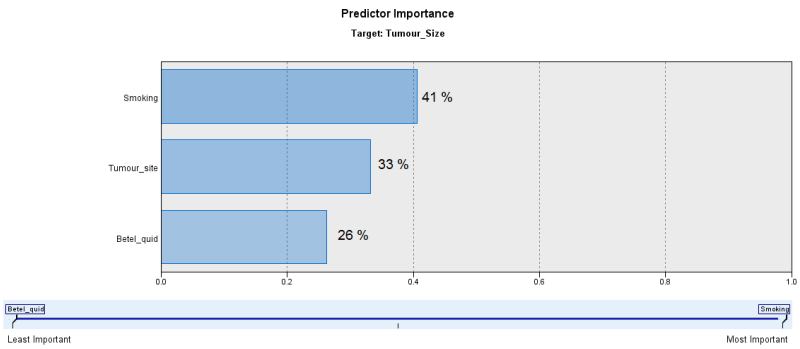

Figure 4 Predictor Importance for RBF in SPSS Modeler.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Predictor important through the RBF procedure shows that smoking factor 41% contribute to the size of tumour. Tumour site contribute about 33% to tumour size, while betel quid consumption contributes 26% towards tumour size.

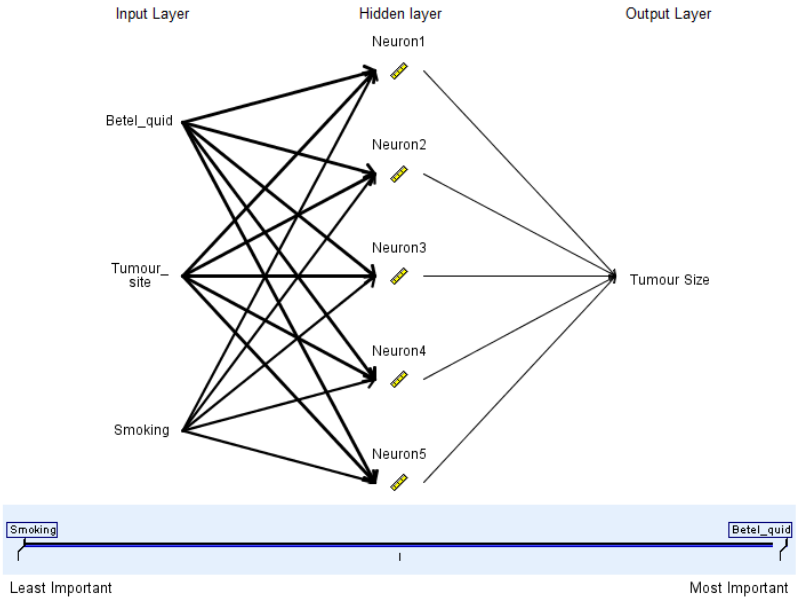

Figure 5 The architecture of the best (RBF) model with one hidden layer, three input variables, seven hidden nodes and one output node.

Table 2: The results of Mean Error and Mean Absolute Error for Training and Testing.

Input Variables	Mean Correct Training	Mean Correct Testing	Mean Correct for Validation
Betel quid, Tumour Site, Smoking	82.9%	83.1%	60.8%

This paper extended the idea of RBF neural network to examine the significant variables that influencing the tumour size. The output node in this study is one node since we have one dependent variable which is the tumour size. To find the appropriate number of hidden nodes and best combination of input variables, the model selection strategies is set by recommendations proposed by SPSS modeler. This is to ensure that the suggested model fits the principle component analysis (PCA). Below is the result obtained from the principle component analysis.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

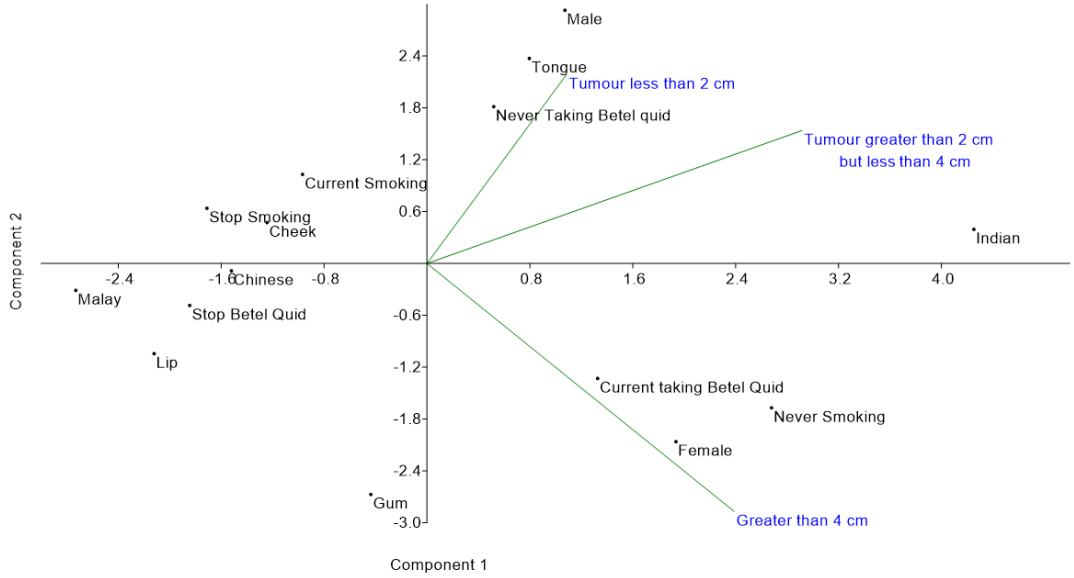

Figure 6 Principle component analysis plot according to the tumour size.

Figure 6 shows the principle components analysis plot on both dimensions (classification tumour size and classification of tumour site with a classification of taking betel quid and classification of smoking status between the male and female). According to result that suggested by the principle components analysis plot above, the tumour size less than 2 cm is near toward tumour site at tongue, never taking betel quid and male. Then, the tumour size greater than 2 cm but less than 4 cm is near toward the Indian. Besides that, the tumour size greater than 4 cm is associated with current taking betel quid, never smoking and female. At the overall of an assessment, there was no having tumour size for Malay and Chinese, no stop betel quid, no tumour site at the gum, cheek, and lip and smoking status was stop smoking and current smoking were found.

4. Summary and Discussion

The main focus of this paper is to illustrate the association of tumor size, with gender, site of tumor and habit of risk factors including smoking, alcohol and betel quid users. For this purpose, we performed the PCA plot in this study. From the PCA plots, tumor less than 2 cm was associated with male, never used betel quid and present on tongue. Moreover, tumor greater than 4cm were associated with female and never smoking habits. As a conclusion, tumor size was not associated with habit of smoking and betel quid user among Malaysian population. This indicated that less than 2cm tumor size is associated with male as compare to female which are associated with increase tumor size. Although the habit of smoking and betel quid associated with OCSCC, within this study we have not found any relation with size of tumor to smoking habit and betel quid users. These finding were consistent with study in which traditional risk factors were not associated with OSCC (Vargas, 2012). Furthermore, these results suggested that, it is important to screening the patients without the potential risk factors, since this observation may help to determine other features associated with the oral cancer, enabling an adequate management strategies, monitoring and decision making. Further studies recommended for assessing the tumor size association factors which might play a pivotal role in better prognosis of OCSCC.

5. Acknowledgments

The authors would like to thank the Ministry of Higher Education Malaysia for Fundamental Research Grant Scheme with Project Code: FRGS/1/2022/STG06/USM/02/10 and Universiti Sains Malaysia (USM).

References

[1] Ajila V, Shetty H, Babu S, Shetty V, Hegde S. Human Papilloma Virus Associated Squamous

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

Cell Carcinoma of the Head and Neck. J Sex Transm Dis, 2015.

[2] Almangush A, Coletta RD, Bello IO. A Simple Novel Prognostic Model for Early Stage Oral Tongue Cancer. Int J Oral Maxillofac Surg, 2015.

[3] Brandwein-Gensler M, Teixeira MS, Lewis CM. Oral Squamous Cell Carcinoma: Histologic Risk Assessment, but Not Margin Status, is Strongly Predictive of Local Disease-Free and Overall Survival. The American Journal of Surgical Pathology, 2005.

[4] Dissanayaka WL, Pitiyage G, Kumarasiri PVR, Liyanage RLPR, Dias KD, Tilakaratne WM. Clinical and Histopathologic Parameters in Survival of Oral Squamous Cell Carcinoma. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 2012.

[5] Ferlay J, Soerjomataram I, Dikshit R. Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. Int J Cancer, 2015

[6] Gillison ML, Koch WM, Capone RB. Evidence for A Causal Association Between Human Papillomavirus and A Subset of Head and Neck Cancers. Journal of The National Cancer Institute, 2000.

[7] Mehrotra R, Yadav S. Oral Squamous Cell Carcinoma: Etiology, Pathogenesis and Prognostic Value of Genomic Alterations. Indian J Cancer, 2006.

[8] W.M.A.W. Ahmad., M.A. Yaqoob, Nurfadhlina Abd Halim, N.A. Aleng, Z. Ali and R.Hassan. Computer Programming Based: An Optimized Prognostic Model of The Fuzzy Survival Oral Cancer Using Sas. Journal of Science and Arts, 2018.

[9] W.M.A.W. Ahmad, M.S. Ibrahim, R. A. Rohim, Z. Ali. Predicting The Predictor's Variables of Survival Time for Oral Cancer With Decision Tree Analysis. International Journal of Advanced and Applied Sciences, 2018.

[10] W.M.A.W. Ahmad, N.A. Aleng, Z. Ali, M. S. M. Ibrahim. Statistical Modeling via Bootstrapping and Weighted Techniques Based on Variances. Engineering, Technology & Applied Science Research, 2018.

[11] W.M.A.W. Ahmad, N. A. Aleng, Y. Dasril, N.A Halim, R. Hasan, Z. Ali, A.Baharum & S.Zakaria. International Journal of Multidisciplinary Research and Modern Education (IJMRME), 2017.

[12] W. M. A. W. Ahmad, R. A. A. Rohim, Y. Norhayati, N. A. Aleng Z. Ali. Developing A New Dimension of an Applied Exponential Model: Application in Biological Sciences. Engineering, Technology & Applied Science Research, 2018.

[13] Yaaqob MA, Hairon SM, Kannan TP, Rahman NA. Factors Associated with Mortality of Oral Squamous Cell Carcinoma (OSCC) Patients Managed at Hospital USM. Archives of Orofacial Science, 2019.

[14] L. Khaled and V.Roa Vemuri. An Application of Principal Component Analysis to The Detection and Visualization of Computer Network Attacks. Annales Des Telecommunications, 2006.

[15] Vargas-Ferreira F, Nedel F, Etges A, Gomes APN, Furuse C, Tarquinio SBC. Etiologic Factors Associated with Oral Squamous Cell Carcinoma in Non-Smokers and Non-Alcoholic Drinkers: A Brief Approach. Brazilian Dental Journal, 2012.