Oral Cancer Risk Prediction Model

WAN NURUL HANNAH NARRUDDIN[1], EKO SUPRIYANTO[1,2,3], MUHAMMAD NABIL MOHD WARID[1],
[1] School of Biomedical Engineering and Health Sciences, Faculty of Engineering, Universiti Teknologi Malaysia, MALAYSIA
[2]IJN-UTM Cardiovascular Engineering Centre, Universiti Teknologi Malaysia, MALAYSIA
[3]Advanced Diagnostics and Progressive Human Care Research Group, Universiti Teknologi Malaysia, MALAYSIA
eko@utm.my

Abstract: - Oral cancer is a serious disease that has become very common and affects a lot of people today. Early detection and risk assessment are essential to prevent this disease. This study aims to analyze the risks of developing the disease by using conventional equation method and machine learning. The risk factors are studied based on previous literature and the weightages for each factor are estimated based on the data. Machine learning method is utilized to further improve the efficiency of the risk prediction model. The results using the machine learning method have proven to be almost similar to the results using conventional equation method.

Key-Words: - Prediction model, Oral cancer, Risk factors

1 Introduction

Statistics have shown that oral cancer is the sixth most common human cancer, representing 3% of all types of cancer [1]. It has been reported that there are 300,373 new oral cancer cases and 145,353 cancer deaths globally in 2012 [2]. About 11% of global oral cancer incidence were from South East Asia, which alarmingly remained high for many years [3]. South East Asia is one of the regions with high mortality to incidence ratio in Asia, with 8508 mortality cases due to oral cancer reported in 2012 [4,5].

1.1 Background of the study

Oral cancer, caused by the growth of cancerous tissue in the oral cavity, is a multifactorial complex process that takes place when epithelial cells are affected by several genetic alterations [1]. Genetic alterations include variations of genetic control apparatus that happen as cells advance to a malignant state [6]. It is a result of several aspects such as mutagens. Mutagens contain carcinogens which are chemicals that can arise from a wide array of sources. There are three main categories of carcinogens, namely physical carcinogens (radiation), chemical carcinogens (including those from biological sources) and infectious carcinogens (viruses).

Mutagens can promote carcinogenesis by changing the structure of genetic material. This happens by generating chromosomal aberrations and abnormalities such as point mutations, deletions, insertions and rearrangements. Another cause of genetic abnormalities is due to the presence of oncogenic viruses. These viruses can be identified in a wide variety of malignant tumors. Oncogenes are also commonly known as tumor-causing genes that carries genes that causes normal cells to undergo malignant transformation. They can also be defined as genes derived from normal eukaryotic genes or proto-oncogenes that have undergone mutation. A retrospective study has described oncogene as any gene that encodes a protein that can transform cells and induce cancer in human or animal [2]. Activation of oncogenes are a result of gain-of-function mutation in the DNA.

Carcinogenesis can also be caused by epigenetic modification and methylation pattern changes [7]. These modifications can affect and change function without actually changing the structure of the sequence of genes. Previous studies have noted that both hypermethylation (by suppressor gene activation) and hypomethylation (by inappropriate oncogene activation) can also induce carcinogenesis [8]. Fig.1 illustrated the sequence of genetic and tumor alterations for further visualization on genetic alterations.

Significant studies have been conducted in the past to identify the factors that influences the emergence of oral cancer. This disease is highly associated with the consumption of alcohol and

smoking habits. It has been reported since the 1980s that the usage of alcohol and tobacco is unusually common among oral cancer patients [10]. Compared to non-smokers, heavy cigarette smokers have a 5- to 25-fold greater risk of contracting head and neck cancer [11]. Cancer Research UK have noted that for every 15 cigarettes smoked, the genetic material is altered, which eventually leads to cancer. Previous prospective and case-control studies have proven a 2- to 3-fold increased risk for oral cancer in people who consume 50g of alcohol a day when compared with non-smokers [12].

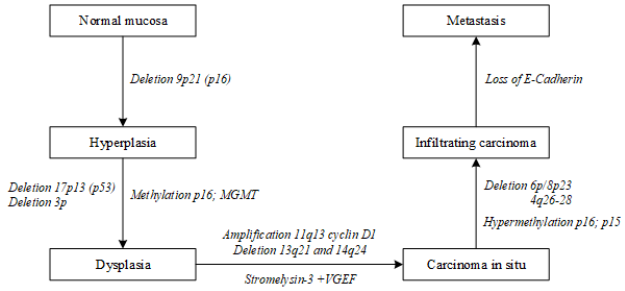

Fig. 1. Sequence of genetic and tumor/stroma alterations implicated in origin of oral cancer [9].

Another risk factor that plays a crucial role in the progression of oral cancer is gender and family history. It has been significantly stated that men have doubled the risk of getting oral cancer because men have lower immune system and have been practicing the use of tobacco since the early ages. Generally, people with family history would be more exposed to a certain disease as the genetic alterations mentioned previously that could promote carcinogenesis are heritable.

Age is also a risk factor as DNA repair mechanism slows down as we age. Other risk factors that could directly affect the immune system and lead to oral cancer includes unprotected sex, presence of Human Papillomavirus (HPV), food, chronic stress and norepinephrine (NE) levels. Unprotected sex exposes one to HPV, which is a virus infection that can be obtained through skin-to-skin contact is a major risk factor linked to cancer. There is sufficient evidence for the causal role of HPV16 E6 in squamous cell carcinoma in the oral cavity [2]. Food could also increase the risk of getting oral cancer because some food might contain carcinogens. Chronic stress has also been identified as one of the causes of tumor growth as it can increase NE levels and affect immune function [13]. Prior studies have shown oral cancer cells behavior can be affected by stress as well [14].

1.2 Problem statement

An existing oral cancer risk assessment test that can be found online has its drawbacks. It is not user friendly and can only be used by healthcare professionals. Additionally, a certain fee must be paid before using the assessment test. Hence, this study aims to develop a new method to calculate the risk of getting oral cancer that is more user-friendly and more convenient in terms of setting up appointments with physicians directly for further diagnosis and treatments once the user is assessed using the risk calculator. The outcome of this study is to produce a better model than the existing oral cancer risk assessment test that could overcome all the limitations that the current model conveys.

2 Method and Material

2.1 Basic Risk Calculation Model

Initially, studies on risk factors of oral cancer are being reviewed and the weightages for each factor are estimated based on previous literature, as shown in Fig. 2. The risks are divided into three energy levels, low, medium and high energy level.

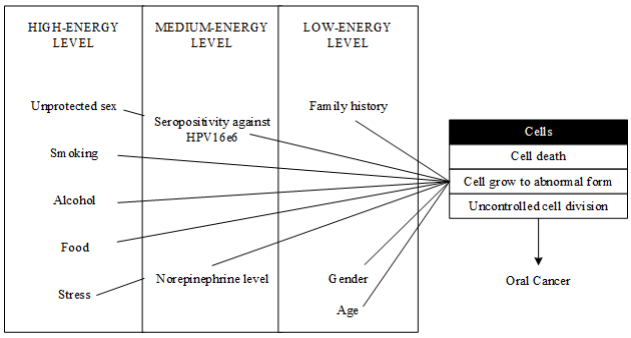

Fig. 2. Risk factors

As illustrated in Fig. 3, it can be summarized that two major factors of oral cancer are immune system and genetic. The risk factors are further discussed in the following subsections.

2.2 Databases from literature

2.2.1 Age

Previous studies have shown that oral cancer incidence rate increases with age. Fig. 4 portrays the incidence rate of oral cancer by age. The greatest increase of oral cancer incidence affects the 30-54 and 55-69 age groups [15]. Most patients were diagnosed at the age of 62, and two-thirds of patients were mostly over the age of 55. This was likely

because the DNA repair mechanism activity decreases significantly as age increases. As one ages, spontaneous mutations and rearrangements will accumulate with time. Consequently, this will lead to the deregulation of transcription, impaired stress response and diminished function of DNA repair genes. These events will result in even more decline of DNA repair efficiency and fidelity and further aggravates the functional decline of the mechanism [16].

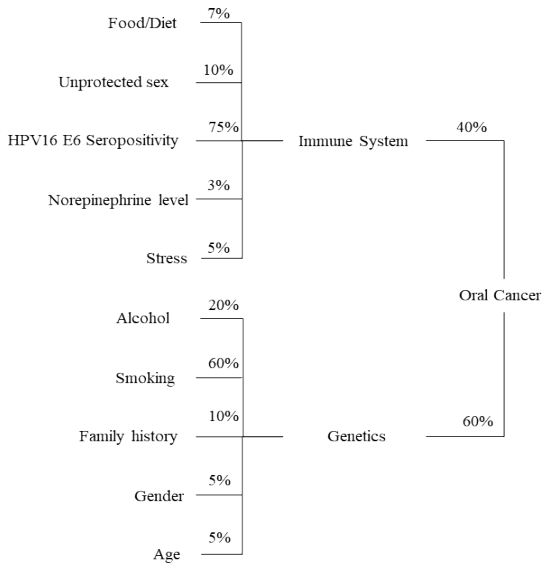

Fig. 3. Decision tree

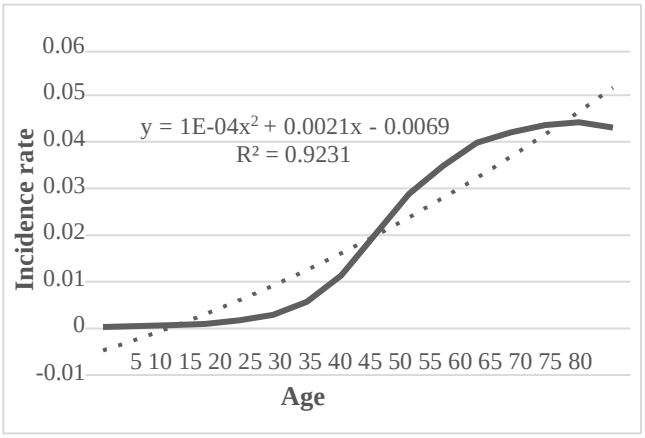

Fig. 4. Oral cancer incidence rate by age

2.2.2 Gender

It has been long noted that cancer affects men evidently more than women. Some studies indicated that this was a result of men having weaker immune system than women and men have been practicing the use of cigarettes ever since decades ago. This can be associated to the socio-economic factors such as income levels, education, availability of proper healthcare which plays a role in the development of this disease. Fig. 5 illustrates the comparison of oral cancer incidence rate between male and female. Male shown higher incidence rate compared to female which only show insignificant increment from 1975 to 2014.

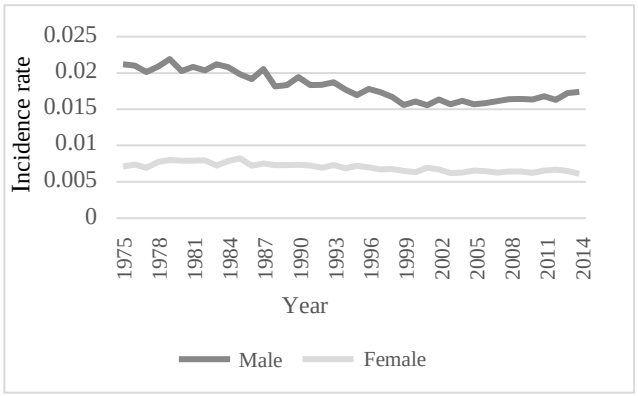

Fig. 5 Incidence rate of oral cancer by gender

2.2.3 Smoking habits

Chemicals in tobacco smoke cause genetic changes in mouth cavity cells which can lead to the development of oral cancer. Tobacco use increases the risk of oral cancer by exposing the mouth to these carcinogenic chemicals, either during inhalation while smoking or through direct contact while chewing tobacco products [11,17-19]. Fig. 6 shows the incidence rate of oral cancer for smokers for both gender male and female. Both genders show an increment in incidence rate as the age increases.

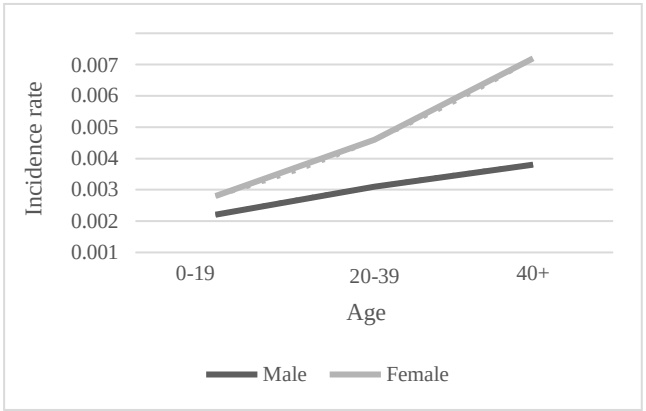

Fig. 6. Incidence rate of oral cancer for smokers

2.2.4 HPV16E6 Serology Status

Approximately 15 types of Human Papillomavirus (HPV) are associated with cancer. Research has proven that HPV indirectly lead to cancer, by

altering the genetic material, which may cause the cells to become cancerous, once infected. The primary route of obtaining HPV is through oral sex, because usually they are transmitted sexually. Serology status as shown in Fig. 7 proved that HPV may influence the rate of oral cancer.

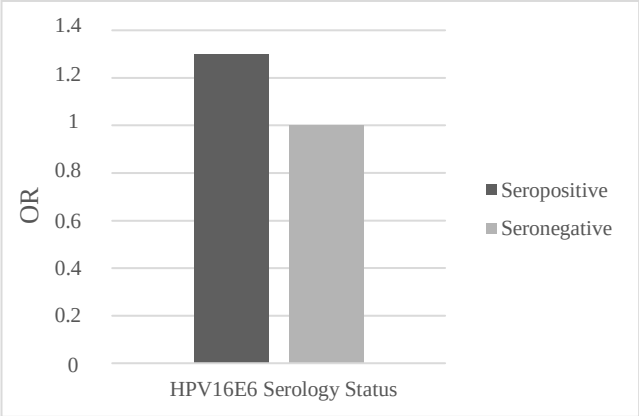

Fig. 7 Serology status of HPV16E6

2.2.5 Norepinephrine levels

Stress hormones, including norepinephrine and epinephrine, are produced as response by the body towards emotional, mental and physical pressure (see Fig. 8). This will cause the heart rate, blood sugars and blood pressure level to increase. Previous literatures have shown that stress plays a significant role in inducing cancer. The association of cancer with psychological stress can develop in several ways. Certain bad habits that increase the risk of cancer may developed when under stress, for example, smoking, drink alcohol, or overeating.

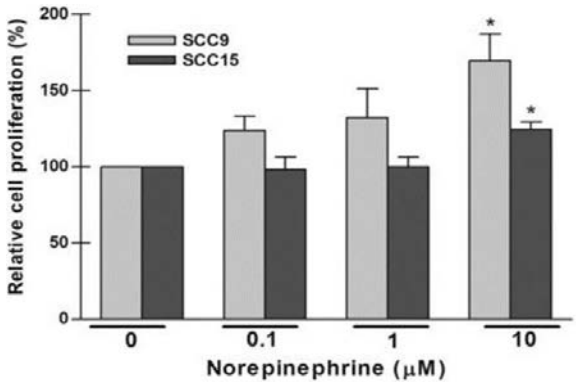

Fig. 8. Norepinephrine levels in cancerous cells

2.2.6 Unprotected Sex

By the age of 25, approximately 90% sexually active people are at risk of genital HPV, from which 2-3% will develop visible genital warts. The prevalence of oral cancer due to unprotected sex is shown in Fig. 9.

2.2.7 Food

World Health Organization (WHO) has eloquently stated that dietary deficiencies or imbalance intake are the cause of approximately 35-55 % of human cancers and 15 % of oral cancers [20,21]. It has been proven that the presence of carcinogens in good and food additives could directly induce carcinogenesis. Altered dietary habits could also induce in vivo synthesis of carcinogens [22].

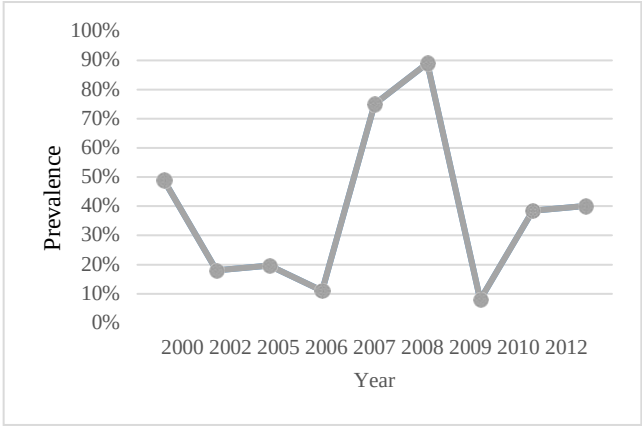

Fig. 9. Prevalence of oral cancer due to unprotected sex

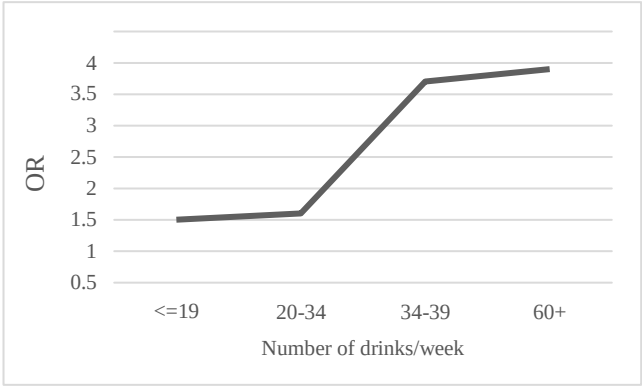

Fig. 10. Odds ratios of oral cancer by number of drinks taken per week

2.2.8 Alcohol consumption

It has been found that alcohol has a dehydrating effect that affects the cell walls, which enhance the ability of carcinogens to permeate mouth tissues. In addition, the body's natural ability to use antioxidants for cancer prevention is reduced due to nutritional deficiencies associated with heavy drinking. Fig. 10 displays the Odds Ratios (OR) for

drinking associated with oral cancer. The adjusted OR shows that the odds of oral cancer increased with increasing level of alcohol consumption (drinks/week).

2.2.9 Family history

First-degree family history is directly associated with the risk of oral and pharyngeal cancer. The risk is greater when there is more than one relative affected, and independent from alcohol and tobacco consumption. Family history of lung cancers and skin melanoma are also associated with the risk of oral and pharyngeal cancer.

The Odds Ratios (OR) for family history associated with oral cancer as shown in Fig. 11 proved that family with oral and pharyngeal cancer may increase the risk for oral cancer.

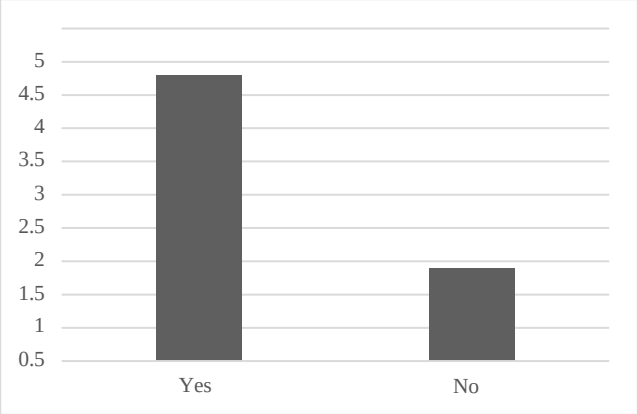

Fig. 11. Odds ratios of oral cancer by family history

2.3 Rule/Algorithm

2.3.1 Conventional equation

Table 1 shows calculation of the risk percentage using conventional equation. Taking a female patient age of 27 years old as an example, with a positive family history, unprotected sex practice, smoker, drinks alcohol, unhealthy diet as well as stress. Additionally, she has negative HPV16E6 and normal Norepinephrine level. Final oral cancer risk calculation of 50.07% is obtained.

Table 1. Calculations using conventional method
*Patient inputs may vary

Risk	Relative percentage (%)	Weightage (%)	Patient Input	Final Risk (%)
Age	5	60	27	0.27
Gender	5	60	Female	3.00
Family history	10	60	Yes	6.00
HPV16E6 seropositivity	75	40	Seronegative	0.00
Norepinephrine level	3	40	Normal	0.00
Unprotected Sex	10	40	Yes	4.00
Smoking	60	40	Yes	24.00
Alcohol	20	40	Yes	8.00
Food/Diet	7	40	Unhealthy	2.80
Stress	5	40	Yes	2.00
TOTAL				50.07

2.3.2 Machine learning

Machine learning is a field of computer science that enable the computers to learn without being explicitly programmed. In this study, machine learning is applied by utilizing the Neural Network (NN) tool in MATLAB. The NN tool requires the user to select the number of layers and neurons to be applied in the neural network. Two layers and ten neurons are selected for this risk prediction model and the neural network undergo multiple trainings until the results obtained had little to no difference with the results calculated with the conventional equation. Fig. 12 illustrates the NN tool with its layer and neuron respectively.

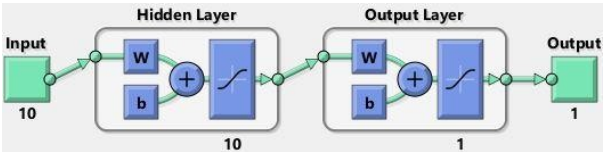

Fig. 12. Neural network consisting of 2 layers and 10 neurons

3 Results and Discussions

3.1 Comparison between using conventional equation and machine learning

Table 2. Comparison of risk calculated using conventional and machine learning

No	Sample	Risk calculated	Risk using machine learning	Error
1	A	80.22	80.21	0.01
2	B	71.27	71.28	- 0.01
3	C	50.33	50.31	0.02
4	D	72.41	83.96	- 11.55

5	E	50.47	50.43	0.04
6	F	51.5	51.33	0.17
7	G	48.53	47.71	0.82
8	H	30.80	30.69	0.11
9	I	44.46	44.15	0.31
10	J	12.68	13.78	- 1.10
11	K	77.59	79.68	- 2.09
12	L	37.05	36.05	1.00
13	M	18.88	17.86	1.02
14	N	48.93	49.09	- 0.16
15	O	76.97	76.96	0.01
16	P	46.99	48.67	- 1.68
17	Q	40.84	39.87	0.97
18	R	84.17	83.74	0.43
19	S	50.07	50.02	0.05
20	T	21.77	21.41	0.36

Table 2 displays the comparison between risk calculated and risk using machine learning for 20 samples. 14 out of 20 samples show small error which is less than 1. The other 6 samples show error more than 1. From the results obtained can be seen that measuring the risk using machine learning can help in predicting the risk for oral cancer instead of using conventional risk calculation. Machine learning gives real-time predictions and processes data more accurately than using conventional equation method. It also provides a continuous quality with large and complex process environments and the process of automation of tasks is easily possible.

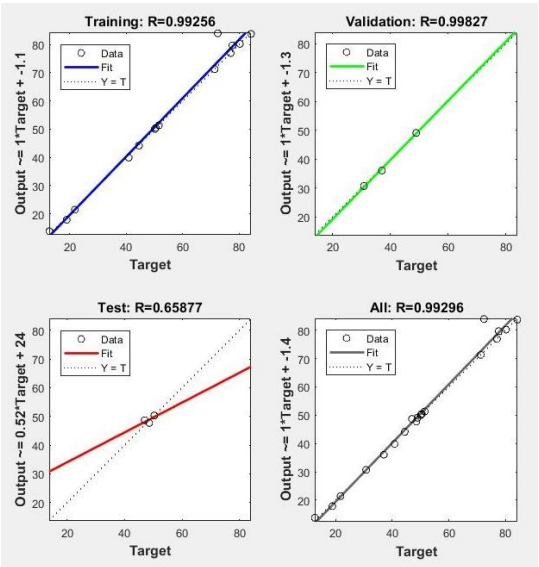

Fig. 13 Regression graphs of neural network after multiple trainings

Fig. 13 shows the regression graphs of neural network after multiple trainings being applied. Three trainings show regression value near to one which highlighted the NN performance. 20 random samples were selected, and their data was inserted into the risk prediction model by conventional equation and machine learning on MATLAB. The data obtained using the machine learning method had significant error values but after several trainings using the NN tool, the numbers were minimized as shown in Fig. 14. This proves the efficacy of utilizing machine learning method in predicting risks of getting oral cancer.

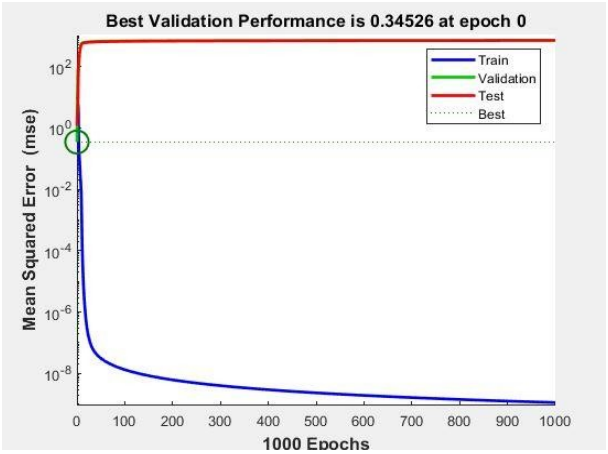

Fig. 14. Performance graph of Neural Network after multiple trainings

4 Conclusion

Oral cancer is a very common disease and hence it is pivotal to get the risk of developing the disease calculated. A normal person does not have to be smoking or consuming alcohol to be infected with oral cancer. The risk prediction model comes in handy in solving this problem as a method of prevention. The model developed in this study, however, might still have its limitations and needs improvements in future studies.

5 Acknowledgment

The work in preparation for this paper has been funded by PT Karya Indonesia Cerdas, Indonesia, through a research grant (R.J130000.7651.4C223).

References:

[1] J. Campo-Trapero, J. Cano-Sánchez, B. Palacios-Sánchez, J. Sánchez-Gutierrez, M. A. González-Moles, and A. Bascones-Martínez, Update on molecular pathology in oral cancer and precancer, *Anticancer Res.*, Vol. 28, No. 2 B, 2008, pp. 1197–1205.

[2] Ferlay J, Soerjomataram I, Dikshit R, Eser S,

Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015 Mar 1;136(5):E359-86.

[3] Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral oncology. 2009 Apr 1;45(4-5):309-16.

[4] Alwan A, MacLean DR, Riley LM, d'Espaignet ET, Mathers CD, Stevens GA, Bettcher D. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. The Lancet. 2010 Nov 27;376(9755):1861-8.

[5] Ng CJ, Teo CH, Abdullah N, Tan WP, Tan HM. Relationships between cancer pattern, country income and geographical region in Asia. BMC cancer. 2015 Dec;15(1):613.

[6] A. Krishna, S. Singh, V. Kumar, and U. Pal, Molecular concept in human oral cancer, *Natl. J. Maxillofac. Surg.*, Vol. 6, No. 1, 2015, pp. 9.

[7] R. Pillai, P. Balaram, and K. S. Reddiar, Pathogenesis of oral submucous fibrosis. Relationship to risk factors associated with oral cancer, *Cancer*, Vol. 69, No. 8, 1992, pp. 2011–20.

[8] S. L. B. Rosas et al., Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients, *Cancer Res.*, Vol. 61, No. 3, 2001, pp. 939–942.

[9] E. I. Auerkari, Methylation of tumor suppressor genes p16(INK4a), p27(Kip1) and E-cadherin in carcinogenesis, *Oral Oncology*, Vol. 42, No. 1, 2006, pp. 5–13.

[10] J. Califano et al., Genetic progression model for head and neck cancer: implications for field cancerization., *Cancer Res.*, Vol. 56, No. 11,1996, pp. 2488–92.

[11] A. R.,Cancer of the mouth, *New York Med. J.*, No. 102, 1915, pp. 1–2.

[12] D. Goldenberg et al., Habitual risk factors for head and neck cancer, *Otolaryngology - Head and Neck Surgery*, Vol. 131, No. 6, 2004, pp. 986– 993.

[13] H. K. Seitz and F. Stickel, Molecular mechanisms of alcohol- mediated carcinogenesis, *Nat. Rev. Cancer*, Vol. 7, No. 8, 2007, pp. 599–612.

[14] P. H. Thaker et al., Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma, *Nat. Med.*, Vol. 12, No. 8, 2006, pp. 939–944.

[15] D. G. Bernabé, A. C. Tamae, É. R. Biasoli, and S. H. P. Oliveira, Stress hormones increase cell proliferation and regulates interleukin-6 secretion in human oral squamous cell carcinoma cells, *Brain. Behav. Immun.*, Vol. 25, No. 3,2011, pp. 574–583.

[16] N. C. Institute, National Cancer Insitute, *Mortality*, 1992, pp. 96– 99.

[17] V. Gorbunova, A. Seluanov, Z. Mao, and C. Hine, Changes in DNA repair during aging, *Nucleic Acids Res.*, Vol. 35, No. 22, 2007, pp. 7466– 7474.

[18] W. J. Blot et al., Smoking and Drinking in Relation to Oral and Pharyngeal Cancer, 1988, pp. 3282–3287.

[19] J. E. Muscat, J. P. Richie, S. Thompson, and E. L. Wynder, Gender differences in smoking and risk for oral cancer, *Cancer Res.*, Vol. 56, No. 22, 1996, pp. 5192–5197.

[20] I. L. A. Ribeiro, J. J. de Medeiros, L. V. Rodrigues, A. M. G. Valença, and E. de A. Lima Neto, Factors associated with lip and oral cavity cancer, *Rev. Bras. Epidemiol.*, Vol. 18, No. 3, 2015, pp. 618–629.

[21] I. Romieu, Diet and breast cancer, *Salud Publica Mex.*, Vol. 53, No. 5, 2011, pp. 430–439.

[22] International Agency for Research on Cancer, "World cancer report 2014," 2014.

[23] N. Taghavi and I. Yazdi, Type of food and risk of oral cancer, *Arch Iran Med*, Vol. 10, No. 2, 2007, pp. 227–232.

[24] N. Johnson and D. Ph, Tobacco Use and Oral Cancer : A Global Perspective, Vol. 65, No. 4, 1991, pp. 328–339.