THE FRAMEWORK FOR DIAGNOSIS ORAL CANCER PRIMARY TUMOR STAGE USING BAYESIAN REASONING

Zainab Abu Bakar[a], Fatihah Mohd[b], Noor Maizura Mohamad Noor[b], Zainul Ahmad Rajion[c,]

[a]Faculty of Computer and Mathematical Sciences, Universiti Teknology MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
zainab@tmsk.uitm.edu.my
[b]Computer Sciences Department, Faculty of Sciences and Technology, Universiti Malaysia Terengganu (UMT), 21030 Terengganu, Malaysia
mpfatihah@yahoo.com, maizura@umt.edu.my
[c]School of Dental Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia
zainul@kck.usm.edu.my

Abstract

Oral cancer, a common type of the head and neck cancer (H&N cancer), is categorized as one of the many oral and maxillofacial diseases. This cancer is universally increasing, becoming a serious problem worldwide. After the patients have reached remission, the number of oral cancer cases is still quite high. Therefore, early detection of such symptoms is very important for both the physicians and patients. The aim of this paper is to present the framework of Bayesian inference engine for the diagnosis of oral cancer. It applies the Bayesian model in expressing different probabilities of oral cancer primary tumor stage as a function of demographics profile, risk habits, clinical, radiology and histopathology data.

Keywords - Bayesian reasoning, diagnosis, oral cancer, primary tumor stage, probability:

1 INTRODUCTION

Cancer of the head and neck (H&N cancer) is the sixth most common cancer, accounting for about 643,000 new cases annually [1]. Oral cancer, a commonly diagnosed type of H&N cancer is increasing globally in incidence and growing seriously in many parts of the world. In 2009, oral cancer is mentioned as the sixth most common cancer in the world [2, 3]. An estimated 263,900 new cases and 128,000 deaths occurred in 2008 worldwide [4]. The highest rates are found in Melanesia, South-Central Asia and Central and Eastern Europe while the lowest rates are in Africa, Central America and Eastern Asia for both males and females [4]. Oral cancer in men is the most common cancer in some countries such as Sri Lanka, India, Pakistan and Bangladesh and may cause up to 25% of all new cases of cancer [2]. In the year of 2007, oral cancer in Malaysia has been reported as being the 21st most common cancer overall and 17th and 16th respectively in male and female occurrence. This ranked higher than 2006 in which it was 20th for female and 28th for male. The majority of the patients were Indian in which oral and tongue cancers were the most common in both male and female [5].

Most oral cancer cases are identified at a later stage where, treatment becomes less successful. It is very important to detect and diagnose such types of cancer at an earlier stage. Early detection helps doctors to provide necessary treatment which also benefit the patients' survival rate. Early detection is based upon awareness of early signs and symptoms [6]. The design of an effective diagnosis model is therefore an essential issue in cancer detection. Intelligent computing methods in medical domains is increasing rapidly due to the improvement effectiveness of these approaches as the problem solving and reasoning algorithm in decision support system especially in helping medical practitioners in their decision making. Inference engine (IE) as part of decision support system (DSS) which uses the knowledge on the system and the knowledge about the patient to generate conclusions regarding certain conditions which applies artificial intelligent algorithms as the problem solving and reasoning algorithm [7].

The rest of this paper is organized as follows - Section 2, overview of oral cancer; Section 3, Research Methodology presenting the framework proposed in this study; Section 4, reports the various informatics diagnosis applications in relation to oral cancer. Finally the paper concludes with the future scopes at Section 5.

2 ORAL CANCER

Oral cancer, a commonly diagnosed type of the head and neck cancer (H&N cancer), is classified as oral and maxillofacial diseases. Oral and maxillofacial diseases also include all oral, laryngeal and pharyngeal sites (see Fig. 1).

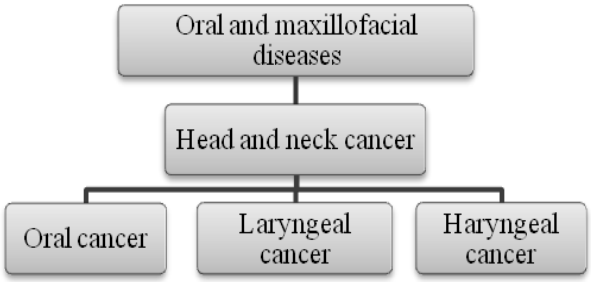

Fig. 1. Oral and maxillofacial diseases

This cancer is largely related to lifestyle, with major risk factors being tobacco and alcohol misuse. In addition to smoking and betel quid chewing, the use of smokeless tobacco has been strongly linked to oral cancer [1, 3, 4, 8]. Worldwide, smoking accounts for 42% of deaths from cancers of the oral cavity while heavy alcohol consumption accounted for 16% of the deaths. Furthermore, poor oral hygiene, periodontal disease, chronic candidiasis, human papilloma virus (HPV) and herpesvirus also contributed to risk factors of oral cancer [9]. Oral or mouth cancer most commonly involves the tongue. It may also occur on the tongue, buccal mucosa, palate, floor of mouth, maxilla, lip, cheek, mandible, tonsil, parotid gland, oropharynx and other unspecified parts.

3 RELATED WORKS

In many years, various extensive efforts have been applied to aid medical expert and people who suffered from cancer to survive in their life. In oral cancer, an informatics application has been developed to help users make better decision as shown in Table 1.

Table 1. Oral Cancer Informatics Applications

Author	**Application**	**Data description**	**Methodology**
Bianchi, et al. [10]	expert system for the diagnosis of pseudotumorous lesions and tumors of the jaws	considered 92 patients with benign space-occupying and fully documented lesions of the jaw bones from different radiologic exams	adapt-m used a formula based on bayes' theorem to calculate the "a posteriori" probability of a diagnosis in the presence of a symptom.
Firriolo and Levy [11]	expert system for the histopathologic diagnosis of salivary gland neoplasms	diagnosing 15 types of primary salivary gland neoplasms using histopathology report	incorporated multiple subprogram methods including: data-driven and goal-directed rule-based reasoning, linear pattern recognition, and

			bayesian classification
Guvenir and Emeksiz [12]	expert system for the differential diagnosis of erythemato-squamous diseases	using clinical and histopathological parameters obtained from a patient to diagnose and differentiate six types of erythemato-squamous diseases	used three classification algorithms: nearest neighbour classifier, naive bayesian classifier and voting feature intervals-5
Borra et al. [13]	decision-support system for diagnosis in oral pathology	simulate the diagnosis of 43 known cases of oral bone disease	Bayes' theorem connected to a relational database was developed using the C++ programming language
Dom et al. [14]	oral cancer prediction model for Malaysian sample	demographic profiles, risk habits and genetic markers from 84 oral cancer patients and 87 controls	a statistical prediction models. Used logistic regression and fuzzy regression
Rosma et al. [15]	predict the likelihood of an individual in developing oral cancer	based on knowledge of risk habits and demographic profiles	used fuzzy neural network and fuzzy regression
Exarchos et al. [16]	a multilevel and multiscale approach for the prediction of oral cancer reoccurrence	using heterogeneous data from 41 patients diagnosed with OSCC that have reached remission.	analysis tool using data mining technique. The classifiers examined within this study were bayesian networks artificial neural networks (ANN), support vector machines (SVM), decision trees (DT) and random forests
Exarchos et al. [17]	modelling the progression of oral cancer.		
Sharma [18]	data mining techniques to evaluate their performance in predicting the malignancy in the patients	the database included 1025 patients with 36 attributes	applied multilayer perception neural network model, and tree boost model,
Sharma and Om [19]	early detection and prevention of oral cancer using data mining		
Zainab et, al, current study	decision support system to diagnose oral cancer primary stage	demographic profiles risk habits, clinical, and histopathology report	applied bayesian reasoning method and open-cases based.

4 RESEARCH METHODOLOGY

The proposed framework portrays the methodology for developing a model for the diagnosis of oral cancer primary tumor stage using bayesian model. The framework is referred as DOCS framework as shown in Fig. 2. It comprises of three phases. Firstly, data collection is done by reviewing patients and

doctors records. Secondly, data pre-processing encompasses multi sub-phases: data interpretation, data integration, remove noisy data, complete missing value, and data inconsistency. Lastly, knowledge based. It comprises of a database, Bayesian inference engine and result. The database consists of the demographic details of the patients, the observed clinical symptoms and oral diseases data. The inference engine consists of Bayesian reasoning algorithms. The inference engine of the system is design to diagnose oral cancer from differential diagnosis result. It consists of a set of parameters needed for diagnosis as presented in section 4.1.

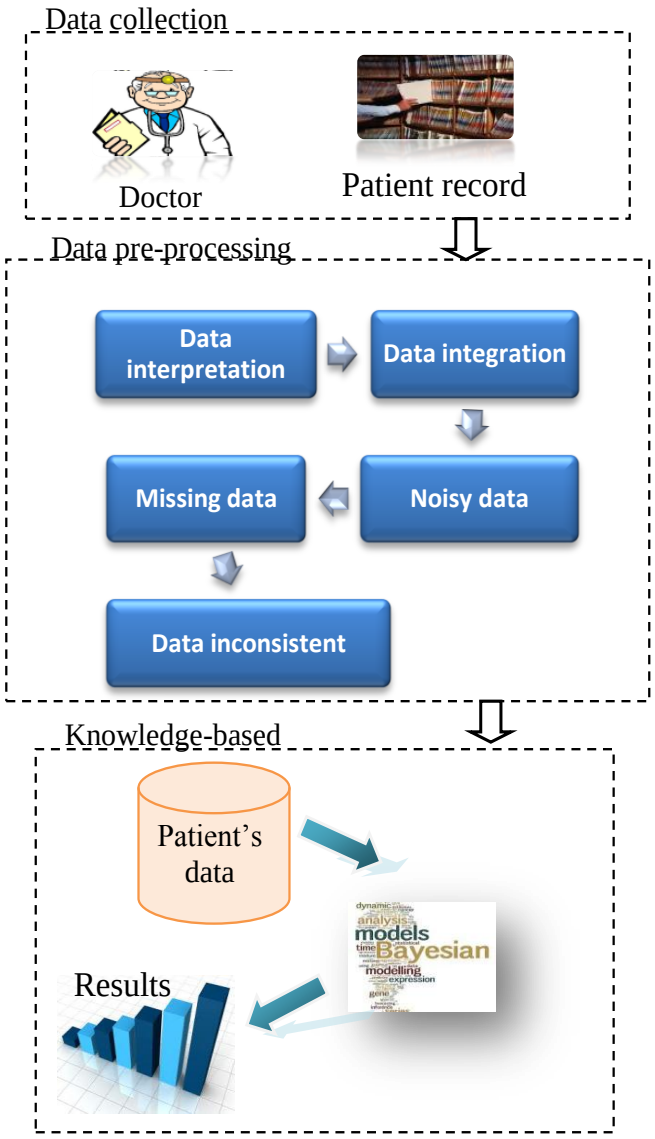

Fig. 2. Framework of Bayesian inference engine for diagnosis of oral cancer

4.1 Variables Database

The database includes all the fields that are required for the research, pertaining to oral cancer from the records of the cancer registries, OPD data sheet, radiology report and histopathology lab report of the hospitals. Table 2 shows 34 attributes which are identified to generate as oral cancer dataset with categorical variables and description.

Table 2: Categorical variables and its description

No.	Variable	Description

1.	Age	Numeric
2.	Gender	Nominal
3.	Ethnicity	Polynomial
4.	History of non head and neck cancer	Nominal
5.	Smoking	Nominal
6.	Category smoking	Polynomial
7.	Duration	Numeric
8.	Quid chewing habit	Nominal
9.	Category quid	Polynomial
10.	Duration	Numeric
11.	Alcohol	Nominal
12.	Category drinking	Polynomial
13.	Duration	Numeric
14.	Neck lump	Nominal
15.	Difficulty in chewing / swallowing	Nominal
16.	Loss of appetite	Nominal
17.	Dramatic weight loss	Nominal
18.	Hoarseness of voice	Nominal
19.	Bleeding	Nominal
20.	Painless ulceration > 14 days	Nominal
21.	Burning sensation in the mouth	Nominal
22.	Painful	Nominal
23.	Swelling	Nominal
24.	Numbness	Nominal
25.	Date of diagnosis (year)	Numeric
26.	Type of cancer	Polynomial
27.	Site	Polynomial
28.	Size	Polynomial
29.	Clinical spread	Polynomial
30.	Lymph node involvement	Nominal
31.	Histological type	Polynomial
32.	Differentiation (SCC)	Polynomial
33.	TNM stage	Polynomial
34.	Overall stage	Polynomial

4.2 Bayesian inference engine

The study is applying Bayesian model to generate probability analysis for the diagnosis of the disease. Bayesian method is a model for reasoning about uncertainty. It is based on causal model applying the Bayes's theorem which was developed from the reasoning adopted by an eighteenth century clergyman, the Rev. Thomas Bayes. In this theorem, all the information which is used to make inferences about the attributes of interest in the population is obtained from the sample [20]. So it is present as (1),

$$p(A \mid B) = \frac{p(B \mid A) \times p(A)}{p(B \mid A) \times p(A) + p(B \mid \sim A) \times p(\sim A)} \quad (1)$$

Eq. (1) can be generalized to take into account both multiple diseases A1, A2, ..., Am and multiple findings B1, B2, …, Bn as in (2),

$$p(Ai \mid B1, B2, \ldots Bn) = \frac{p(B1 \mid Ai) \times p(E2 \mid Ai) \times \ldots \times p(Bn \mid Ai) \times p(Ai)}{\sum_{k=1}^{m} p(B1 \mid A_k) \times p(B2 \mid A_k) \times \ldots \times p(Bn \mid A_k) \times p(A_k)} \quad (2)$$

Eq. (2) describes A as the fact primary tumor stage and B is a particular combination of attributes. The probability of the primary tumor stage, given the multiple attributes B, is equal to the probability of B supposing the primary tumor exists, multiplied by the overall probability of existence of the symptoms B, in which the primary tumor stage exists.

5 CONCLUSION

Decision support system as a component of decision making tools enables the use of knowledge from the system and the knowledge about the patient to generate conclusions regarding certain conditions. This capability applies intelligent computing method (ICM) as the problem solving and reasoning algorithm. Bayesian theorem which is within ICM is utilized as the theoretic basis and to generate differential diagnosis outcome via expressing different probabilities of oral cancer primary tumor stage as a function of multi parameters attributes (demographics profile, risk habits, clinical, radiology and histopathology data). It also employs the method that integrates prior probability and conditional probability to determine the posterior probability. The integration of ICM could be applied to achieve more accuracy in decision making tasks, to reduce the weaknesses of single ICM and to improve the quality of decision making.

6 ACKNOWLEDGEMENTS

This study has been supported in part the Exploratory Research Grant Scheme (ERGS) 600_RMI/ERGS 5/3 (3/2011) under the Malaysia Ministry of Higher Education (MOHE) and Universiti Teknologi MARA (UiTM) Malaysia. The authors would like to acknowledge all contributors, technical members at Hospital Universiti Sains Malaysia (HUSM) who have helped and greatly assisted in the completion of the study.

7 REFERENCES

[1] S. Fedele, "Diagnostic aids in the screening of oral cancer," Head & Neck Oncology, vol. 1, pp. 5, January 2009.

[2] S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," Oral Oncology, vol. 45, pp. 309-316, April-May 2009.

[3] I. Van Der Waal and R. De Bree, "Second primary tumours in oral cancer," Oral Oncology, vol. 46, pp. 426-428, June 2010.

[4] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," CA: A Cancer Journal for Clinicians, vol. 61, pp. 69-90, March-April 2011.

[5] Z.A. Omar and N.S.I. Tamin, NCR Report 2007. 2011.

[6] WHO, "Global status report on noncommunicable diseases 2010," vol., 2011.

[7] E.A. Mendonca, "Clinical decision support systems: perspectives in dentistry," J Dent Educ, vol. 68, pp. 589 - 97, 2004.

[8] W. Kruaysawat, W. Aekplakorn, and R.S. Chapman, "Survival time and prognostic factors of oral cancer in Ubon Ratchathani Cancer Center," J Med Assoc Thai, vol. 93, pp. 278-84, 2010.

[9] A.A. Razak, N. Saddki, N.N. Naing, and N. Abdullah, "Oral cancer presentation among Malay patients in Hospital Universiti Sains Malaysia, Kelantan," Asian Pacific J Cancer Prev, vol. 10, pp. 1131-1136, 2009.

[10] S.D. Bianchi, G. Girelli, and G. Ramieri, "The prototype of an expert system for the diagnosis of pseudotumorous lesions and tumors of the jaws: ADAPT-M. Archiviazione e Diagnosi Automatica di Pseudotumori e Tumori delle ossa Mascellari," Radiol Med. , vol. 91, pp. 219-225, 1996.

[11] F.J. Firriolo and B.A. Levy, "Computer expert system for the histopathologic diagnosis of salivary gland neoplasms," Oral Surg Oral Med Oral Pathol Oral Radiol Endod, vol. 82, pp. 179-86, 1996.

[12] H.A. Güvenir and N. Emeksiz, "An expert system for the differential diagnosis of erythemato-squamous diseases," Expert systems with applications, vol. 18, pp. 43-49, January 2000.

[13] R.C. Borra, P.M. Andrade, L. Correˆa, and M.D. Novelli, "Development of an open case-based decision-support system for diagnosis in oral pathology," Eur J Dent Educ, vol. 11, pp. 87-92, 2007.

[14] R.M. Dom, S.A. Kareem, B. Abidin, R.L.R. Jallaludin, S.C. Cheong, and R.B. Zain., "Oral cancer prediction model for Malaysian sample," Austral - Asian Journal of Cancer, vol. 7, pp. 209-214, October 2008.

[15] M.D. Rosma, A.K. Sameem, A. Basir, I.S. Mazlipah, and M.D. Norzaidi, "The use of artificial intelligence to identify people at risk of oral cancer: empirical evidence in Malaysian university," International Journal of Scientific Research in Education (IJSRE), vol. 3, pp. 10-20, 2010.

[16] K.P. Exarchos, G. Rigas, Y. Goletsis, and D.I. Fotiadis, A multilevel and multiscale approach for the prediction of oral cancer reoccurrence, in XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010, P. D. Bamidis and N. Pallikarakis, Editors. 2010, Springer Berlin Heidelberg. p. 588-591.

[17] K. Exarchos, Y. Goletsis, and D. Fotiadis, "Multiparametric decision support system for the prediction of oral cancer reoccurrence," Information Technology in Biomedicine, IEEE Transactions on, vol. PP, pp. 1-1, 2011.

[18] N. Sharma, "Comparing the performance of data mining techniques for oral cancer prediction," in Proc. ICCCS'11, Rourkela, Odisha, India., 2011, pp. 433-438.

[19] N. Sharma and H. Om, "Framework for early detection and prevention of oral cancer using data mining," International Journal of Advances in Engineering & Technology vol. 4, pp. 302-310, 2012.

[20] A. Petrie, J.S. Bulman, and J.F. Osborn, "Further statistics in dentistry Part 9: Bayesian statistics," Br Dent J, vol. 194, pp. 129-134, February 2003.