Predicting the predictor's variables of survival time for oral cancer with decision tree analysis

Wan Muhamad Amir W Ahmad [1,] *, Mohamad Shafiq Mohd Ibrahim [1], Rabiatul Adawiyah Rohim [1], Zalila Ali [2]

[1]School of Dental Sciences, Universiti Sains Malaysia, Kubang Kerian, Malaysia
[2]School of Mathematics Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia

ARTICLE INFO

ABSTRACT

Decision tree analysis is one of the famous analyses which assist the researcher to identify the associated factor that contributes to the certainty factor. In this study, we proposed a decision tree model with the high-risk factor and try to estimates the importance of every each predictor. The result reveals that the survival time for the patient is mostly depending on nerve invasion and the size of a tumor, followed by alcohol and ethnicity factor. This promising technique had led to a successful research and give the best results for the decision making especially for the decision maker. In conclusion, this analysis can provide a very useful for forecasting the survival time (in months) of oral cancer patients.

1. Introduction

Oral cancer is a disease resulting from abnormal cell growth in the mouth, lips, tongue or throat. Oral cancer is basically an abnormal growth found in the mouth region. This abnormal growth can be detected easily by a dentist at your regular check-up appointment. Oral cancer commonly occurs at the buccal mucosa (cheek), tongue, the floor of the mouth and lip. This cancer also can be detected early as it is presented in a part of your body that is so easily seen. Most of the oral cancers look like very similar to each other under the microscope and are called squamous cell carcinoma. However, there are another types of oral cancer which occur less commonly such as Kaposi's sarcoma. All over the world, the oral cancer is the 6th most common cancers and every year there were 350,000 new cases were reported (Parkin et al, 2005). It is also reported in National Cancer Registry Report (MOH, 2007) that the survival rate of oral cancer is lower than that of cervical cancer, skin melanoma, and breast cancer. In Malaysia, the incidence of oral cancer is prevalent among the Indian ethnic group compared to Malay and Chinese. Where this mouth and tongue cancers were among the 10 most common cancer among both male and female. The incidence of oral cancer is highest in Indian females where the ASR was 10.2/100,000 female populations. Of those cases reported with staging, only 35.4% of the cases were diagnosed at stage I and II. When detected early, oral cancer is almost always cured. But unfortunately, so many people still present with it at such a late stage. Oral cancer is more common in men than women because men tend to smoke more frequently.

In 2002, the American Cancer Society has estimated that the 28,900 of new cases of oral cancer was diagnosed and nearly 7,400 people will die from this disease. Over 90 percent of these tumors are squamous cell carcinomas, which arise from the oral mucosal lining. In spite of the ready accessibility of the oral cavity to direct examination, these malignancies still are often not detected until a late stage, and the survival rate for oral cancer has remained essentially unchanged over the past three decades (Neville and Day, 2002).

2. Data and methods

Data from the medical unit record, Hospital Universiti Sains Malaysia (HUSM) were reviewed and related information was extracted. The sampling frame was the list of patients which diagnosed with oral cancer admitted *to* in HUSM. The description of data as shown in Table 1.

A decision tree is a great and efficient method for classification, prediction and for facilitating decision making in sequential decision problems (Mesarić and Šebalj, 2016). This method had been used widely across many fields. In the medical field for example, often the decision maker will be faced with a

sequential decision problem involving decisions that lead to different outcomes depending on the chance. Decision process which involving a lot of sequential decisions will lead to the decision problem, this is because the decision becomes difficult to visualize and to implement.

Fig. 1 shows that decision trees are indispensable graphical tools in such settings, it's allowed for an intuitive understanding of the problem and aid for the optimal decision making. A decision tree is a graphical model describing decisions and their possible outcomes. A decision tree consists of three types of nodes (a) decision node (b) chance node (c) Endpoint node/Terminal node.

Table 1: Description of data

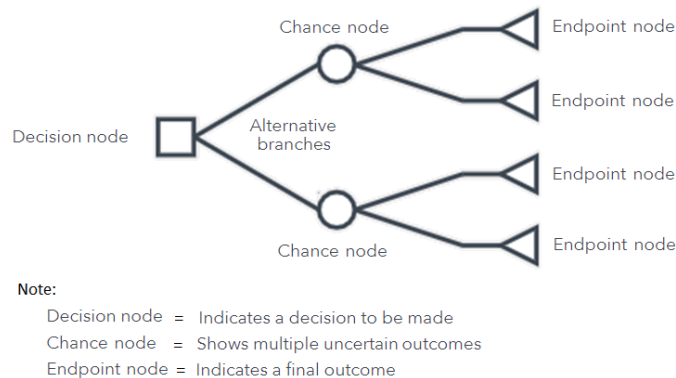| Number	Variables	Explanation of user variables
1.	Time	Time in months
2.	Smoking	Smoking Status [0= Never, 1= Yes]
3.	Alcohol	Alcohol used [1= Never, 2= Stop, 3=Current]
4	Size	Tumor Size
		[1= Less than 2, 2= Greater than 2 less than 4, 3=Greater than 4]
5.	Nerve	Nerve invasion [0= No, 1= Yes]

Fig. 1: Decision trees are graphical models for describing sequential decision problems

3. Results and discussion

Firstly, a decision tree model allows us to develop a classification system that predicts observation based on a set of decision rules. Secondly, the process will automatically include in its rue only that attributes that really matter in making a decision. Attributes that do not contribute to the accuracy of the tree are ignored. This can yield very useful information about the data and be used to reduce the data relevant fields before training another learning technique, such as a neural net. In this section, we perform an analysis weight by considering betel quid factor. So, the results gained in this section is controlling the betel quid perspective.

According to the Fig. 2, the top five predictors are ranking to their contribution are nerve invasion, tumor size, alcohol consumption, ethnicity, and smoking. Using the CHAID method, nerve invasion and tumor size are the best predictors of survival time. In total there were four predictors that the model deemed important.

According to the decision tree analysis in Fig. 3, the survival time for the oral cancer is about 31 months. We can see that the first split is by nerve invasion. Records where the nerve invasion with "Yes" status is assigned to Node 2 with predicted time survive up to 12 months. While nerve invasion with "No" is assigned to Node 1 with predicted time survive up to 42 months. Notice how the model has divided these patients into two sub-categories (Nodes 3 and 4), based on the reading of tumor size.

On average the time survival for patients is quite higher for the patients which do not have nerve invasion, with the tumor size is greater than 4 cm (it was estimated 81 months). The fourth split shows the nodes (Node 5, Node 6, Node 7 and Node 8). Node 5 (Never/stop smoking) and Node 6 (current smoking) is split under the smoking factor, it was predicted that 25% of patient with current smoking, on average has 46 months of survival time.

Node 7 and Node 8 are split by the factor of alcohol. The estimated time for patients which having tumor size greater than 4 cm and never taken alcohol, the mean time for surviving is around 77 months while, for patients which having tumor size greater than 4 cm and they are already stopped from taking alcohol, the mean time for them is around 87 months.

4. Conclusion

The main objective of this research paper is to determine the associated factors for oral cancer data. On top of that, we try to find the most influenced factor in rank using decision tree analysis. Result reveals that the survival time for the patient is mostly depending on nerve invasion and the size of a tumor, followed by alcohol and ethnicity factor. This promising technique had led to a successful research and give the best results for the decision making especially for the decision maker.
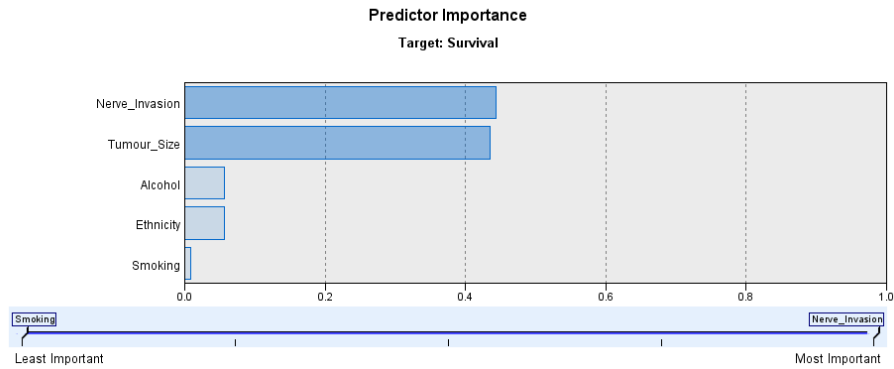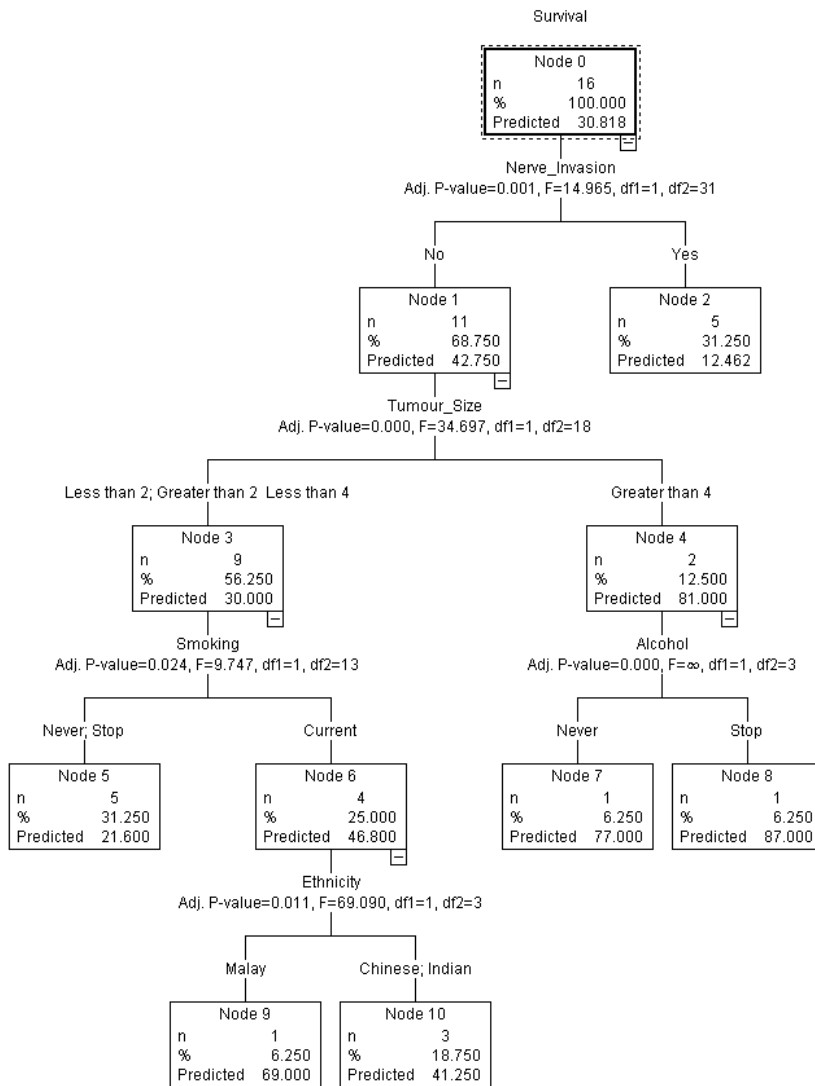
Fig. 2: Predictor important

Fig. 3: Decision tree for oral cancer classification

Acknowledgment

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, Kampus Kesihatan).

References

Mesarić J and Šebalj D (2016). Decision trees for predicting the academic success of students. Croatian Operational Research Review, 7(2): 367-388.

MOH (2007). National cancer registry report Malaysia cancer statistics. Ministry of Health Malaysia, Putrajaya, Malaysia.

Neville BW and Day TA (2002). Oral cancer and precancerous lesions. CA: A Cancer Journal for Clinicians, 52(4): 195-215.

Parkin DM, Bray F, Ferlay J, and Pisani P (2005). Global cancer statistics, 2002. CA: A Cancer Journal for Clinicians, 55(2): 74-108.