

Machine Learning Based Risk Prediction Models for Oral Squamous Cell Carcinoma Using Salivary Biomarkers

Yi-Cheng WANG^a, Pei-Chun HSUEH^b, Chih-Ching WU^b, and Yi-Ju TSENG^{a,c,1}

^a*Department of Information Management, Chang Gung University, Taoyuan, Taiwan*

^b*Graduate Institute of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan*

^c*Department of Information Management, National Central University, Taoyuan, Taiwan*

Abstract. Tumor-associated autoantibodies can be used as biomarkers for detecting different types of cancers. Our objective was to use machine learning techniques to predict high-risk cases of oral squamous cell carcinoma (OSCC) with salivary autoantibodies. The optimal model was using eXtreme Gradient Boosting (XGBoost) with the area under the receiver operating characteristic curve (AUC) of 0.765 ($p < 0.01$). Thus, applying machine learning model to early detect high-risk cases of OSCC could assist the clinic treatment and prognosis.

Keywords. autoantibody, machine learning, oral squamous cell carcinoma

1. Introduction

In Taiwan 2018, there were about 8,100 newly diagnosed oral cancer cases and 3,000 deaths. Over 90% of oral cancer patients were diagnosed with oral squamous cell carcinoma (OSCC). Autoantibodies are an immunological response that is induced by tumor cells, and these immune biomarkers can help to detect an early stage of high-risk populations [1]. This study tried to develop a machine learning based risk prediction model of OSCC using salivary autoantibodies.

2. Methods

All participants were recruited in Chi-Mei Medical Center (Liouying, Taiwan) from September 2008 to December 2012. The individuals signed the informed consent form before the screening and permitting the use of saliva samples. The individuals were divided into high-risk and low-risk groups by clinical or pathological diagnosis. We used ELISA to detect and measure the levels of 10 salivary autoantibodies such as ANXA2, CA2, HSPA5, ISG15, KNG1, MMP1, MMP3, p53, PRDX2, and SPARC. The demographic (age and gender) and behavior (smoking, alcohol consumption, and betel nut chewing) data were also included in the model development.

¹ Corresponding Author, Yi-Ju Tseng; E-mail: yjtseng.info@gmail.com.

Logistic regression, random forest, support vector machine (SVM) with a radial basis function kernel, and eXtreme Gradient Boosting (XGBoost) were used to develop the prediction models. The data was divided into 75% training and 25% test sets. We randomly split the training-test dataset 100 times and used 5-fold repeated 5 times cross-validation approach to optimize the model parameter of each round of evaluation. We evaluated the performance of the models by using the area under the receiver operating characteristic curve (AUC) and measured the variable importance. The AUCs of the models were compared with repeated measure analysis of variance (rANOVA) and Holm-Bonferroni method post hoc tests between algorithms. All analyses were performed with R software. A two-sided $p < 0.05$ was considered statistically significant.

3. Results

A total of 190 high-risk and 147 low-risk cases were enrolled in this study. High-risk group comprised older age, more alcohol consumption, more betel nut chewing, and higher levels of salivary autoantibodies. The optimal model for predicting high-risk for OSCC was constructed using XGBoost with AUC of 0.765 ($p < 0.01$). The AUCs of random forest, logistic regression, and SVM with a radial basis function kernel were 0.758, 0.743, and 0.725, respectively. The top three ranking variables in XGBoost by variable importance match are betel nut chewing, MMP3, and age.

4. Discussion

We developed a prediction model based on salivary autoantibodies and demographic characteristics, which effectively predicts high-risk cases for developing OSCC by detecting salivary autoantibodies. However, our model has several limitations such as the small number of the study population and from a single institution.

5. Conclusion

We successfully established a risk prediction model for high-risk cases of OSCC. Using salivary autoantibodies as biomarkers to early detect high-risk cases of OSCC can assist the clinical practice that would help to reduce morbidity and mortality from OSCC.

Acknowledgments

This research was supported by Chang Gung Memorial Hospital (CMRPD3K0011) and the Ministry of Science and Technology, Taiwan (MOST 109-2636-E-182-001).

References

- [1] Macdonald IK, Parsy-Kowalska CB, Chapman CJ. Autoantibodies: Opportunities for Early Cancer Detection. *Trends in Cancer*. 2017 Mar;3(3):198-213.