Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-631

Reviewer A

I have read your manuscript with considerable interest and wish to provide a few comments that may strengthen the overall work.

Q1. My main concern is the number imbalance between your OSCC and control datasets. In my experience this can complicate many of the machine learning algorithms used. It would be advisable to explore some techniques to address this issue and do some further analysis - there may improvements in your models.

Response: ***Thanks for this insightful suggestion. We sincerely agree with the reviewer's idea regarding the imbalance between OSCC and non-OSCC control which might complicate the machine learning algorithms. Indeed, during our study, we have taken several measures to minimize the effects induced by such imbalance and enhance the reliability and reproducibility of our findings. Firstly, we enrolled the data from both age- and gender-matched non-OSCC patients by pre-set inclusion and exclusion criteria. Patient distribution in terms of age, gender, alcohol and smoking were similar between OSCC and non-OSCC cohorts. Secondly, we have set the weight of sample distribution in machine learning algorithm to address this sample imbalance. For example, when we set "class_weight=balanced" in the SVM model, the model might set the weight that was inversely proportional to the sample size of different categories. In addition, random forest, Xgboost and other machine learning algorithms also have similar settings (Bioinformatics. 2017;33(7):951-955; Bioinformatics. 2020;36(4):1074-1081). Indeed, this also reflected the advantages of machine learning approach in statistical analyses. Thirdly, all our classifiers had been trained by repeated 10-fold cross-validation using data from training cohort, and then their predictive performances were further evaluated in the validation***

cohort. Together, these abovementioned measures might in large part address the sample imbalance and empower the robustness of our findings.

Q2. Minor comments - the abbreviation PNI for prognostic nutritional index may be misleading as generally in association with OSCC PNI would be taken to mean perineural invasion. Could a different abbreviation be used?

Response: ***Prognostic nutritional index (PNI) is an emerging marker calculated by serum albumin concentration and total lymphocyte count in peripheral blood, which is initially designed to evaluate preoperative nutritional conditions and surgical complications in patients with gastrointestinal cancers (Nihon Geka Gakkai Zasshi. 1984; 85(9):1001-5). Furthermore, PNI has been reported to be a powerful and independent prognostic factor in several human cancers such as colorectal cancer, gastric cancer, lung cancer and esophageal cancer (Ann Surg Oncol. 2013;20(8):2647-54; PLoS One. 2015;10(9):e0136897; J Cancer Res Clin Oncol. 2017;143(7):1235-1242; Ann Surg. 2020;271(4):693-700). Thus, the abbreviation of prognostic nutritional index, PNI has been commonly used in the literature. We believe that it might be inappropriate to use another different abbreviation instead of PNI. Thanks.***

Q3. Some of the text on the figures is very hard/impossible to read even at full magnification. I think this will need to be addressed for any final version of the manuscript in accordance with the journal requirements. e.g., figure 2 A and B, figure 3, Supplementary figure 2.

Response: ***We are sorry for this. We have revised the figures to make the tests on figures sufficiently clear. Thanks.***

Reviewer B

The manuscript by Wu et al. uses machine-learning-driven approaches on common laboratory markers to develop a diagnostic and prognostic signature for oral squamous

cell carcinoma. In doing so, the authors used 13 hematologic and clinical chemistry markers including WBC, neutrophils, albumin and other values including ratios derived from these markers. The authors use 486 OSCC patients in a training and validation cohorts and apply 6 MBL claassifiers. The authors developed a novel prognostic nomogram and prognostic signature to predict survival. The bioinformatics approaches overall are solid however there are several concerns.

Q1. First, there are several issues with how this applies as a diagnostic signature. OSCC is often suspected on clinical exam of the oral cavity and diagnosed on in office biopsy so it is unclear how this aid the diagnosis of cancer. Furthermore, the role of such a signature as a screening tool is unclear as diagnosis was known apriori and values were obtained prior to surgery. In addition, several patients were excluded indication that this signature may not differentiate between patients with other types of cancers, patients with autoimmune or other inflammatory diseases or patients with certain medications. It is more likely than not that this signature detects an inflammatory/immune state that is not specific for OSCC.

Response: ***Thanks for these professional comments. We sincerely agree with the reviewer's idea that definitive diagnosis of OSCC depends on biopsy and histopathological examinations. Our diagnostic signature derived from pre-surgical blood parameters was aimed to validate the significant difference between OSCC and non-OSCC patients rather than to offer direct diagnostic evidence for OSCC. Moreover, this signature might provide additional information in term of nutrition, immunity and inflammation in OSCC, which may influence the choice and timing of treatment.***

During our patient enrollment, those with autoimmune or other inflammatory diseases or certain medications affecting blood parameters were excluded. Given the high complexity and heterogeneity of blood parameters among population, our classifiers based on routine blood parameters may not provide enough information to identify OSCC-specific inflammatory/immune status. However, large datasets from high-throughput multiomics analyses such as genome-wide RNA sequencing

and metabolomics in preoperative peripheral blood may facilitate the identification of OSCC-specific signature reflecting inflammatory/immune state. Based on the insightful ideas, we have talked about these points in the revised discussion.

Q2. Second, no tests for co-dependency of variables were performed. In fact, many of the variables, by definition are co-dependent. Consequently, this classifier may be biased based on the data imputed.

Response: ***Thanks for your insightful suggestion. Multicollinearity is generally caused by co-dependency of variables which may cause overfitting and invalidation of predictive ability of the model. Previous reports have found that several supervised machine learning algorithms such as RF, SVM and Xgboost were resistant to multicollinearity interference and had the excellent predictive performance (Comput Struct Biotechnol J. 2014; 13:8-17.). In addition, LASSO as another machine learning algorithm also has the ability to reduce the multicollinearity of the variables in model. In our study, the prognostic model was ultimately composed of 6 blood parameters (lymphocyte count, platelet, A/G, LMR, PNI and SII) by LASSO analysis. We performed multicollinearity analysis on these six parameters by 'car' package in R, and the variance inflation factor (VIF) was used to detect multicollinearity, the result was as follows.***

PNI	lymphocyte	platelet	A_G_ratio	LMR	SII
2.349151	2.170416	1.617303	1.207404	1.610358	1.993593

According to the criteria of statistics (Qual Quant. 2007; 41:673-690), VIF<10 or 5 indicated that there were no obvious multicollinearity between variables. Furthermore, the predictive model has been verified internally and externally to show good and stable prediction performance.

Q3. Third, it remains difficult to assess this as a prognostic signature without additional disease and treatment details. This classifier may simply represent the inflammatory state present with larger disease burdens. Furthermore, there are not details of extent of surgery, chemotherapy and/or radiotherapy to account for differences in risk score.

Furthermore, it remains unclear why clinical stage is only included and not pathological stage, which may be more prognostic. Finally, what staging system was used for classify extent of disease. Finally, it is unclear if this this prognostic indicator also reflects the comorbidity state and if other comorbidity indexes (e.g., Charlson) performs similarly.

Response: ***Thanks for these professional suggestions. It has been increasingly recognized that cancer-associated inflammation and nutritional status of patients are key determinants of disease progression and patient survival in several human cancers (Cell. 2011; 144(5):646-74; Eur Urol. 2011; 59(6):923-8.). Our prognostic signature consisted of six blood parameters such as lymphocyte, platelet, PNI, LMR, SII, A/G, which reflected pre-treatment nutritional and inflammatory state of OSCC patients. We sincerely agree with this reviewer with regard to these parameters like extent of surgery, chemotherapy and/or radiotherapy which might significantly affect patient prognosis. However, the research purpose of our study was to develop the signature for prognostication before treatment, which was highly beneficial for individualized treatment planning and selection. This was also supported by previous extensive studies regarding these parameters prior to treatment in cancer prognostication (J Cancer Res Clin Oncol. 2009; 135(12):1783-90; Cancer. 2015; 121(4):545-55; Br J Cancer. 2018; 118(2):248-257; Oral Dis. 2020; 26(5):903-911 ;). Moreover, the treatment heterogeneity among our patients further complicated our efforts to determine the effects of various treatments on the risk score. Of course, we can't rule out that diverse treatments have influence on the prognostic prediction of our signature. Here, we used machine learning algorithms to integrate the six preoperative blood parameters and developed a novel and potent prognostic signature for OSCC. We have discussed this point in revised manuscript.***

As suggested by this reviewer, we revised the nomogram by adding pathological grade as a prognostic factor. In addition, pathological grade has been added in the decision curve analyses. Please see revised Figure 8.

As described in Materials part, we have indicated that AJCC 7th edition staging system was utilized in our study to classify the disease extent.

As suggested by this reviewer, comorbidity indexes like Charlson reflected comorbidity state of patients and had prognostic significance (J Chronic Dis. 1987; 40(5):373-83.). However, in our study, our prognostic signature mainly reflected the nutritional, inflammation/immune status of OSCC patients without attention to comorbidity. Indeed, these comorbidities might influence the blood parameter before surgery and patients with these comorbidities were excluded. Further studies may be needed to determine the prognostic values of comorbidity indexes in OSCC.

Reviewer C

The study used machine learning algorithms and is based on a good number of cases. My comments:

* The title is somehow confusing, please re-write it more simply.

Response: ***Thanks for your suggestion. We have revised the title into "Identification of diagnostic and prognostic signatures derived from preoperative blood parameters for oral squamous cell carcinoma" and made it more simply and easily understandable.***

* in the abstract, "…and 200 age/gender-matched non-OSCC patients.". Please make this clearer.

Response: ***We have added more detailed information about non-OSCC patients in revised abstract as required.***

* in the abstract, please explain the parameters that were included in the analyses.

Response: ***We have added the details about those included parameters in revised abstract as required.***

* in the Introduction, please note that lip cancer is not any more part of OSCC according to the TNM AJCC 8.

Response: ***Thanks for your professional suggestion. We have revised the sentence by***

deleting lip cancer in the revised introduction. Thanks.

* in the Methods, please add some details about "LASSO regression".

Response: ***We have added more details about "LASSO regression" in the revised Methods part***

* The Discussion is good. The recent studies on machine learning in oral cancer need to be discussed briefly.

Response: ***Thanks for this valuable suggestion. We have added more information about studies on machine learning in oral cancer in the revised Discussion part.***

*Please improve the language of the manuscript (from beginning to Results. Discussion is ok).

Response: ***We have checked throughout the whole manuscript and tried our best to correct the errors as much as possible. The revised manuscript has been sent to one of our colleagues who is a native English speaker for manuscript editing. We hope that our revised manuscript might be suitable for publication.***

Reviewer D:

This paper used six machine learning classifiers and 13 parameters to develop diagnostic and prognostic signatures for OSCC. The results are impressive. The authors also addressed the application, limitation, and biological explaining of the predictive signatures. Nice work.

I have some questions for this paper:

Q1. Page 2, line 5, what is the PNL score?

Response: ***PNL score (platelet–neutrophil–lymphocyte score) is a new prognostic risk score designed based on combined preoperative platelets, neutrophils, and lymphocytes from our team (Oral Dis. 2019; 25(4):1057-1066.). This paper has been cited as appropriate.***

Q2. Could authors, please provide the information of the non-OSCC patients to show their diversity or similarity? Are the criteria for patient inclusion also applicable to non-OSCC controls?

Response: ***Thanks for your professional suggestion. The detailed information of non-OSCC patients has been listed in Table 1. The criteria for OSCC patient inclusion in terms of other diseases and drug affecting blood parameters were also applicable to non-OSCC controls.***

Q3. In Fig 2C, the view angle of the PCA plot cannot distinguish non-OSCC from OSCC samples very well. Did the authors try to use a 2D plot or change the view angle of the 3D plot?

Response: ***Thanks for your professional idea. We have revised the Figure 2C into a new 3D plot.***

Q4. Page 7, line 1, the dataset was randomly divided into training and testing sets. Did authors repeat the entire procedure to check if the predictive signatures are consistent in different runs?

Response: ***Thanks for your insightful comments. The random sampling method adopted by 'caret' package stratified random sampling based on mixed data, which made sample distribution between training dataset and testing dataset consistent with the overall equilibrium. This randomly sampling approach was beneficial to ensure the final effect of model training. Hence, we don't repeat the procedure to check whether the predictive signatures are consistent in different runs.***

Q5. Did the authors try to use the bootstrap sampling approach in machine learning models?

Response: ***Thanks for this professional suggestion. Bootstrap sampling approach is the sampling method of reversion to estimate the variance of statistics and interval estimation in nonparametric statistics. In this study, we employed the 10-fold cross-validation method instead of bootstrap sampling in our machine learning models***

including Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boosting (Xgboost), Naive Bayes (NBs), Neural Network (NN).

Q6. In constructing the prognostic model, the authors used LASSO-Cox regression analysis to identify 6 parameters highly associated with favorable and unfavorable survival. A risk score was calculated based on the six parameters. Next, the authors divide patients into a high and low-risk group based on the risk score and the Kaplan-Meier method to show that the high-risk group has shorter DFS and OS than the low-risk group. The risk score itself already implies favorable/unfavorable survival information. Therefore, the high-risk group will get shorter DFS/OS inevitably.

Response: ***Thanks for this insightful comment. We identified 6 parameters significantly associated with favorable and unfavorable survival by LASSO-Cox regression. The risk scores for each patient were calculated based on these six parameters including lymphocyte count, platelet, PNI, LMR, SII and A/G. Among them, lymphocyte count, PNI, LMR and A/G were associated with favorable survival, while platelet and SII were associated with unfavorable survival. Moreover, each parameter has a different weight in the prognostic model. Thus, the high-risk group will not necessarily result in a shorter DFS/OS.***

Q7. The version of the packages should be listed in addition to the R version.

Response: ***The version of the R packages has been added as suggested.***

Q8. I also encourage the authors to release their data and code to the public to repeat and extend the study.

Response: ***Thanks for this idea. The data and code used throughout the whole study is available from the corresponding author upon reasonable request.***