Peer Review File

Article Information: https://dx.doi.org/10.21037/tcr-22-2750

Reviewer A

Comment 1: Is there any preprocessing done on images prior to feature extraction?

Reply 1:

Thank you for your comments. In fact, we had preprocessed the images before the feature exaction. The preprocessing on images is an essential step before the radiomics analysis. The add of this part would complete this article. Specific operations are as follows:

Firstly, we matched different target images of the same patient spatially by Elastix software package (v. 4.10, http://elastix.isi.uu.nl/index.php). We registered DWI and ADC images successively based on T_2WI images, which could ensure that the three sequences had the same resolution, field of vision and orientation. Then, we standardized the images of each patient to improve the texture recognition rate before feature extraction. The T_2WI, DWI, and ADC images of each patient were resampled to $1\times1\times1$ cm^3 voxel sizes to normalize the voxel spacing. The voxel intensity discretization was accomplished by fixing the bin width at 25 HU to reduce imaging noise and standardize the intensity. Z-score normalization was carried out for different sequences of each patient to reduce the impact of inconsistent image parameters on the changes of radiomics features, and the voxel intensity was converted into a distribution with a mean of 0 and a standard deviation of 1. Finally, we used an open-source radiomics software (FAE, v.0.4.0) which was programmed with Python and used NumPy, pandas, and scikit-learning modules. to extract radiomics features from each sequence after image preprocessing was completed.

Changes in the text:

We have modified our text as advised (see Page 4, line 145-148 and line 160-167).

Comment 2: Please explain what Z-score is and give a reference.

Reply 2:

Thank you for your question. Z-score normalization is a process in which the difference between measured values and the average is divided by standard deviation. It can use different magnitude-of-magnitude data as a Z-score of the same measurement for comparison. By converting the voxel intensity into a distribution with a mean value of 0 and a standard deviation of 1 through Z-score, the influence of image parameter inconsistency on the change of radiomics features can be reduced. (*doi: 10.1158/0008-5472.CAN-17-0339*)

Changes in the text:

We have modified our text as advised (see Page 4, line 164-167 and reference NO.20).

Comment 3: Please give a reference for LASSO.

Reply 3:

Thank you for your question. Least absolute shrinkage and selection operator (LASSO) is a classifier that realizes feature selection by adding the L1 constraint of feature weight to logistic regression, which is indeed a L1 regularization method. It is essentially a logistic regression model, that can use selected features for logistic regression modeling and output forecast probabilities for each sample. By adjusting a constraint weight, it can simultaneously perform feature selection and model building. Above all, LASSO is a linear regression algorithm that can not only solve the problem of overfitting in radiomics, but also extract useful features by parameter reduction. (*PMID: 20808728*)

Changes in the text:

We have modified our text as advised (see Page 8, line 319-322 and reference NO.36).

Comment 4: In M&M, Construction of Clincially Relevant Model: Please include what the clinically relevant variables are.

Reply 4:

Thank you for your comments. We regret for the lacking of the explicitly state about the used clinical variables in clinical model. We have added the clinical variables to be analyzed in the corresponding section of the article according to your suggestion. The clinical variables included age, PSA at admission, lesion location distribution, biopsy GG, and PI-RASD v2.1 score for each patient.

Changes in the text:

We have modified our text as advised (see Page 5, line 195-196).

Comment 5: Results Section: Could you please include what the 20 radiomics features selected to build the final radiomics model.

Reply 5:

Thank you for your leerily suggestion. By listing the selected features in the radiomics model, readers can have a clearer understanding of the construction process of radiomics, which can provide certain references for subsequent studies and increase the transparency of our model. We added a table 5 in the part of results to describe the 20 radiomics features selected in the final radiomics model and explained the implications of radiomics features a little in the discussion.

Changes in the text:

We have modified our text as advised and added a table 5 (see Page 6, line 250-253; Page 8, line 328-332 and table 5).

Reviewer B

Comment 1: "The main contribution(s) should be clarified more explicitly in the introduction, highlighting the added values compared to previous related work. A related work section is missing (maybe add one). I'm not sure if there is page limit concerns, but to me the listed related work in paragraph 1 of page 2 needs to be extended and more related work should be mentioned and the added value of this study should be better highlighted."

Reply 1:

We all appreciate your clear and detailed feedback and comments. In the introduction, we have explored the different approaches to the assessment of ECE in recent years, from MRI imaging to structured report that used to describe ECE, and the applications of radiomics. The advantages and disadvantages of different methods to assess ECE werre introduced respectively. In addition, it focused on the application of radiomics in PCa in some studies and further highlighted the potential application value of the research of radiomics in ECE.

Changes in the text:

We have modified our text as advised (see Page 2, line 51-61 and line 67-78).

Comment 2: "The only classifier used in this study is the LASSO method which is indeed a L1 regularization method. I strongly recommend using other classifiers such as SVMs and ensemble trees (e. g., random forest) which are shown to be prominent for multivariate classification tasks. I believe adding this comparison will make the study more noticeable by the audience."

Reply 2:

Thank you for your leerily suggestion. According to your suggestion, we have added different classifiers for modeling, including decision tree (DT), logistics regression (LR), random fotest (RF), support vector machine (SVM), and LASSO. By the comparison among different classifiers, it can enhance the depth of the article and increase the interest of readers.

LASSO is a classifier that realizes feature selection by adding the L1 constraint of feature weight to logistic regression. It is essentially a logistic regression model, that can use selected features for logistic regression modeling and output forecast probabilities for each sample.By comparing the results of these five classifiers, the superiority of LASSO classifier was highlighted. According to the research, LASSO algorithm can ensure that all important radiomics features can not only be recognized effectively, but also avoid the overfitting problem of classification task. So that, it can get better results. In addition, RF, DT and other classifiers have good performance for nonlinear problems. But in the execution of multicollinearity tasks, they may have some problems such as difficulties in processing missing data, over-fitting and ignoring the correlation among each feature attribute, leading to unsatisfactory output results. For example, in this study, the AUCs of RF and DT classifers in the training set were very high, reaching 1.000, but in the validation set, the AUCs were low, only 0.619-0.804, which may be due to the phenomenon of overfitting.

(*doi: 10.1111/spc3.12579*).

Changes in the text:

We have modified our text as advised and added a table 4 (see Page 5, line 186-188; Page 6, line 244-250; Page 8, line 317-327 and table 4).

Comment 3: "For most of the classification problems, accuracy and AUC metrics are incapable of representing true negative rates. I strongly recommend providing precision-recall (PR) curves and AUPR (area under PR curve) as a further metric to better compare the methods."

Reply 3:

Thanks for your clear and detailed suggestion. ROC curve and PR curve are both statistical methods used to evaluate classification and diagnostic performance. ROC curve is a diagonal line from (0,0) to (1,1) with sensitivity as the ordinate and 1-specificity as the abscissa. PR curve is a trapezoidal curve with sensitivity as the abscissa and positive predictive value as the ordinate. Previous literature has proved that there is a one-to-one correspondence between the ROC and PR curve points of the same diagnostic method in the same data set. However, ROC curve and PR curve have completely different responses to the change of prevalence. According to some studies, when the positive rate of the study samples is close to 50%, the results of ROC curve and PR curve are similar, and the evaluation of ROC curve is more specific, so it is recommended to use ROC curve for diagnostic evaluation. On the other hand, the positive rate in samples is relatively low, less than 20%, especially when it is less than 5%, it is recommended to use PR curve for sample evaluation. (*doi:10.1111/j.1365-2753.2005.00598.x; 10.1016/j.jclinepi.2015.02.010; 10.3760/cma.j.cn112150-20220104-00007.*)

In this study, the ECE positive rate was about 45.9%, which is closer to 50%. Therefore, we prefer to use ROC curve to evaluate the samples. However, we do not know whether the ROC curve results will produce errors with the PR curve results. Such exploration can be verified in future studies.

According to the above explanation, we have not modified according to the reviewer's opinion.

Comment 4: "Table 4 only provides comprehensive comparisons of all the methods only in training cohort. Authors should provide all the comparisons also on the validation set!"

Reply 4:

We are deeply sorry for the confusion caused to the reviewer's understanding because of our mistake. We have completed the comprehensive comparisons of all the methods in both training and validation set in table 6.

Changes in the text:

We have modified our text as advised (see table 6).

Comment 5: "Generally, the image quality of the figures is a bit off. I strongly recommend to redraw the diagrams using more elegant tools (e. g., Python seaborn library) and use dot/dash styles to make it readable in black and white screens/hard copies as well."

Reply 5:

We are deeply sorry for the poor image quality. We have copied and rearranged the image to ensure the quality of the image.

Changes in the text:

We have modified our text as advised (see Figure 2).