Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-8065

Reviewer A

Comments 1: How did the authors decide the hyperparameters for the machine learning (ML) models? Just using the default values provided by the packages seems to create a strawman argument since most ML experiments include hyperparameter search. Some of the seemingly poor performing models may actually perform better with proper hyperparameter tuning.

Reply 1: Thank you for the valuable comment. We have used the commonly functional methods, Grid Search algorithm, to optimize the hyperparameter automatically. After hyperparameter tuning, the machine learning (ML) models have performed better than before, as expected by the reviewer.

Change in the text 1: We have added more details in the Methods section (see page 6, line 127-138), and revised the relevant contents in the Abstract section (see page 2, line 44-46), the Results section (see page 7, line 165-175), Table 2 and Figure 3.

Comments 2: It seems that the distribution of parameters related to OCT of MH such as BASE tends to be small in external validation in internal validation and external validation.

Reply 2: IMH OCT parameters (BASE, MIN, H, MHI, THI, THI, and HFF) were not significantly different between the internal validation set and external validation set, as shown in the Result section and Table 1.

Change in the text 2: We have revised the Result section (page 7, line 162-164) and Table 1 accordingly.

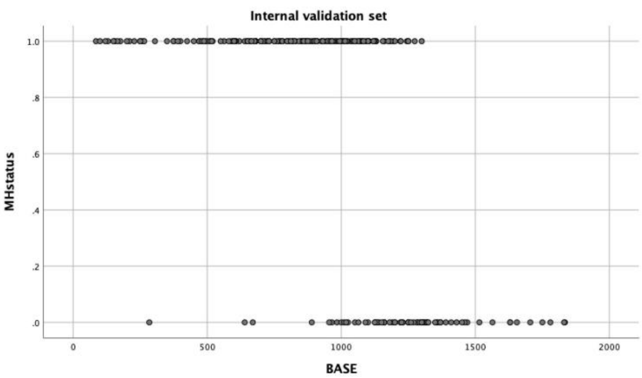
Comments 3: How is there a significant difference in each parameter? Please show p values.

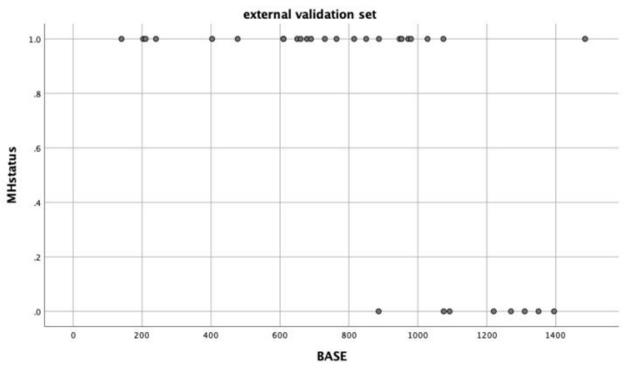
Reply 3: There is no significant difference in each parameter between internal validation set and external validation set. The p values are presented in Table 1.

Change in the text 3: We have added the p values in Table 1.

Comments 4: What is the scatter plot (for example, BASE) of IMH closed and non-closed in internal and external validation?

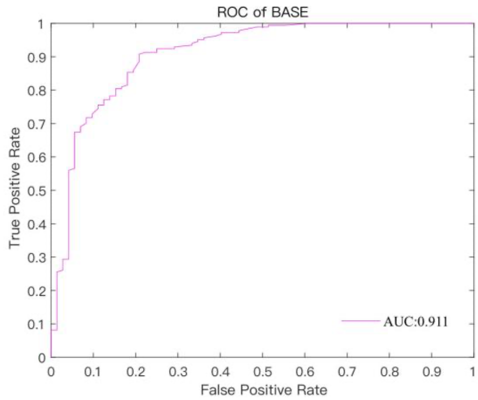
Reply 4: The scatter plots for BASE and MH status in internal and external validation are as follows:





Comments 5: What happens to the ROC curve when a simple prediction is made using only BASE?

Reply 5: Binary logistic regression analysis was derived for prediction of IMH status using BASE in internal validation set. The ROC curve of BASE was showed as follow, with an AUC of 0.911.



Change in the text 5: We have added the AUC value and ROC curve of BASE in the Result section (see page 8, line 173-175) and Figure 5.

Comments 6: I think that the definition of MH close or open differs depending on the manuscripts. Cite the appropriate articles for the definition of MH close.

Reply 6: Thanks for your helpful comments, we have cited the reference for the definition of closed MH (reference 16).

Change in the text 6: We have cited the reference as required in the Method section (see page 5, line 109, reference 16).

Comments 7: Does it include MH due to high myopia? How about a lamellar hole? Please clarify.

Reply 7: Full-thickness idiopathic MH (IMH) eyes with at least one-month follow-up were retrospectively recruited, while those with macular holes secondary to trauma, high myopia, macular edema, epiretinal membrane, retinal detachment, or retinoschisis were excluded.

Change in the text 7: We have revised the details of inclusion in the Method section to make it more clear (see page 4, line 82).

Comments 8: The abstract should state which technique worked best, whether you used a test and a validation cohort.

Reply 8: The best-working technique was random forest in the internal validation set, and it was tested in the external validation set. We have described the relevant content in the Abstract section as follows:

In the internal validation, the mean area under the receiver operating characteristic curves (AUCs) of the five ML algorithms were 0.882-0.951. The AUC, accuracy, sensitivity, and specificity of the best-performing algorithm (i.e., random forest, RF) were 0.951, 0.892, 0.973, and 0.904, respectively. In the external validation, the AUC of RF was 0.940, with an accuracy of 0.875, a specificity of 0.875, and a sensitivity of 0.958. (see page 2, line 43-46)

Comments 9: The results section should literally state for which machine learnings the prediction worked best.

Reply 9: The best-working machine learning algorithm was random forest. We have described the relevant content in the Results section as follows:

In the internal validation, the mean AUCs of the five ML algorithms were 0.882-0.951 (95%CI: 0.789-0.993), with mean accuracies of 0.857-0.892 (95%CI: 0.812-0.940), mean sensitivities of 0.865-0.934 (95%CI: 0.805-0.970), and mean specificities of 0.804-0.973 (95%CI: 0.659-1.000). Among these ML algorithms, RF achieved the best performance. The mean AUC of RF was 0.951 (95%CI: 0.908-0.993), with a mean accuracy of 0.892 (95%CI: 0.844-0.940), a mean sensitivity of 0.904 (95%CI: 0.856-0.952), and a mean specificity of 0.973 (95%CI: 0.938-1.000). (see page 8, line 165-169)

Reviewer B

Comments 1: It would be useful to mention that it would be important to validate 3d images, more useful in surgery;

It would be important to increase the external validation group with different eye characteristics (different ages and degrees of myopia for example) ;

Mention the importance of external data validation with different oct devices

Reply 1: Thank you for these valuable comments, we have revised the relevant literature in the Discussion section as follows (marked in red):

There are several limitations of this study. Firstly, the presented model relies on the manual measurement of preoperative macular OCT parameters. Nevertheless, the acceptable repeatability and reproducibility of these manual measurements on OCT have been demonstrated in previous studies.(31,32) Secondly, since 3D-OCT images may be more useful than 2D-OCT images in IMH surgery, future investigation should be performed to validate the IMH outcome prediction using 3D-OCT images. Another limitation is that the number of patients included in the dataset was limited, especially in the external validation set, because only eight patients without primary closure after surgery were available; however, the cases in the training set and the external validation set were enrolled from different ophthalmic centers, and the promising performance in external validation also suggests the excellent adaptability of our ML model. Furthermore, to improve the generalizability of our ML model, it is important to recruit IMH with different characteristics (such as age and myopia severity) and collect images obtained from different OCT devices for development and validation of the ML model in future work.

Change in the text 1: We have revised the relevant literature as suggested in the Discussion section (see page 11, line 240-242, line 245-248).