Author’s response to reviews

Title: Semi-supervised Method for Image Texture Classification of Pituitary Tumors via CycleGAN and Optimized Feature Extraction

Authors:

Hong ZHU (zhuhong@xzhmu.edu.cn)

Qianhao FANG (301700911546@stu.xzhmu.edu.cn)

Yihe HUANG (huangyihe52647849@foxmail.com)

Kai XU (Xukai@xzhmu.edu.cn)

Version: 2 Date: 19 Feb 2020

Author’s response to reviews:

Thanks to the two reviewers for their valuable comments, here is a response to these comments.
Ps: All modifications in the manuscript are marked in red
Reviewer 1(Chenxi Huang, Ph.D.)
1 □ The format of the paper is confusing. The double columns sometime are split by the two-column-wide figures and hard to follow what the next line is. The authors can either switch to single column or stick to from-top-to-bottom-page order for the texts. (Answer: We have changed the double column of this paper to a single column as required.)
2 □ The Methods section now include unnecessarily detailed information on some of the well-established models, such as CycleGAN, DenseNet, Convolutional RNN, etc. There is no need to write and draw the architecture of these models; a citation should suffice unless significant adaptations are made. Motivations of why these models are used should be included though. If the authors really want to be self-contained, these should go to the supplementary. Otherwise the paper is difficult to follow and hard to know whether these are existing models or proposed by the authors. (Answer: We have modified this problem as required)
3 □ Please spell out the full name of the model abbreviations when they are first mentioned in the paper with reference. (Answer: We have spelt out the full name of the model abbreviations when they are first mentioned in the paper with reference.)
4 □ Please consolidate the figures. There are too many figures with overlapping information. The referencing of the figures are off in several places. Please check.(Answer: We have consolidated the figures as required)
5 □ The paragraphs on the data used, training/testing/tuning pipeline, performance metrics considered and baseline models to compare the new model with should be moved to methods section and not be combined with the results. (Answer: We have put the experiment platform and dataset part into methods section. Because we need to show the training process and final results when the cyclegan data is amplified, we put this part of the experiment into the results section. At the same time, we checked the relevant articles in your journal, and the structure is the same.)
When training the CycleGAN, you used 374 patients. How many were used for feature extraction and whether these patients were included in the CycleGAN model training? How many were used in the classification model development? Since the features were generated from the data, and if the same data/part of the data were also used to train the classifier, it is not surprising to get good results. My question is how generalizable of these features when used for classification of other unseen data. An illustration on how the data were split and used in each stage would be helpful to rule out any concern with data leakage and overfitting. (Answer: 152 patients used for feature extraction and these patients were included in the CycleGAN model training. The data used for the classification model is the feature data extracted by the feature extraction model: 152 patients with labels. Our experiment was mainly to expand the sequence of all 374 patients, including T1 and T2. However, only 152 of the 374 patients have labels. Therefore, we perform feature extraction on these 152 patients and classify the extracted feature sequences with CRNN. Feature extraction and classification are not the same model and there is no training with the same data set. At present we are only conducting experiments on pituitary tumors. There is no experiment with unseen data, but we estimate that the effect will not be bad. Every time we randomly divide the data set, we will check the data distribution of validation set and test set. Generally, we will take a test set with a soft and flexible number of pituitary tumors that is about 1:1, so the test results are more fair.)

Why did the authors choose to do 6-fold cross-validation (random-split)? And why did the authors not report the accuracy with the variation too? For comparison with the other baseline methods, where they evaluated on the same data as the new model? (Answer: When taking test set for testing, we need to make sure that the soft and tough number ratio of pituitary tumors is about 1:1. These six random split experiments were performed according to the test data ratio of 1:1. For comparison with other benchmark methods, they were evaluated on the same data as the new model.)

To test the performance of the generated features, the authors should consider the same classification methods as those used for the baseline models, including VGG, ResNet, DenseNet. (Answer: We have added the same classification methods as those used for the baseline models, including ResNet, DenseNet)

Please include the details of performance metrics used for each model in the Methods section before reporting the results. (Answer: We have added performance metrics in methods section)

Reviewer 2(Ming Huang)

The authors only use the accuracy and loss as the evaluation metrics, which is not sufficient and can be biased especially for unbalanced data. The authors should show precision, recall and F1-score of the classification results. (Answer: We have shown the relevant content in Table 2.)

In the Figures 17 and 18, the authors need to show the testing results to compare with the results of training and validation. (Answer: We have shown the relevant content in Table 1.)

In Table 2, why is the running time of CRNN longer than that of model combination of CRNN and DenseNet+ResNet? (Answer: The running time shown in the table is the time required to classify pituitary tumors. Since DenseNet + ResNet framework has extracted pituitary tumor features, the running time will be shorter than using the CRNN classification time directly.)