Review of: "An Improved Hybrid Transfer Learning-Based Deep Learning Model for Alzheimer's Disease Detection Using CT and MRI Scans"

Maria Maddalena Autorino

Potential competing interests: No potential competing interests to declare.

GENERAL COMMENT:

The paper propose a deep learning/transfer learning method for the classification of medical images in the context of the Alzheimer diagnosis. In particular, it proposes to compare the performance of several neural network architectures and to show which one obtain the best result in term of accuracy, precision, recall and f1-score. The aim is clear but the paper can be improved adding more analyses and comments on the obtained results and the method limits. E.g.:
1) how the classification accuracy (and/or other metrics) varies by varying the test set disease class, since it could be interesting to understand if the NN is able to classify better one disease over the others. It helps also to catch any limits of the algorithm
2) to try some interpretability methods like the GradCAM one, that gives also an idea of the reliability of the NN and if and how the NN behaviour changes among different disease classes
3) to add some ensemble learning analysis to test the neural network robustness and reproducibility
Main other points should be also addressed or clarified in my opinion:

- the data set should be better described in an appropriate section or subsection. I found information a bit strewn in the paper. E.g., 1) I read in the title about CT and MRI images, but it is not clear which images form the dataset and in which proportion; 2) moreover, are the data of different resolutions? or acquired with different scanners?
- you write that the goal is to classify images in five classes (you write it in the Introduction section and even in the Conclusion and future scope section), but it seems that they are four (or three if you refers to the Figure 2)
- more detailed info on the baseline methods with which you compare your result, in order to understand clearly what is your improvement and to fairly compare the two methods; e.g., is the baseline method trained with the same dataset or the same data splitting? do they pre process the input data?
  A final table in the Result section could be informative, showing your best NN performance and the baseline performance, instead of reporting only the performance results of your trained NNs

DETAILED COMMENTS:

I would avoid abbreviations/acronyms to make the text more clear and fluent to any reader. E.g., in the *Abstract* section: 'MRI' in the 5th and the last third line or EMCI, MCI, LMCI in the third last line.
Or in the *Introduction* section: line 7, I would put the acronyms near to each abbreviated words to make clear the

associations, like: early mild cognitive impairment **(EMCI)**, mild cognitive impairment **(MCI)**, etc.). Indeed, in line 9 you refers to this acronyms.

Tables/figures:

- Tables: I suggest a more detailed caption/title for all the tables, describing their content
- Figure 1: I guess it is not necessary. It does not add information on your method or pipeline. A reference to a paper that describes how a convolutional neural network works could be enough
- Figure 2: do LMCI and AD belong to the same category in the orange classification box? In the introduction you state that you want the neural network to classify the images in 5 classes, but I see three or four classes in the orange classification box.
- Figure 3: I guess this picture can be better presented. It is a bit confusing. I would like to see the reference/original image and then the different transformation steps. Titles per each image are not necessary. A scheme in which each column and each row present the same transformation made on a specific image could be better. Even narrowing to one class of images (like the MCI one), since it seems that you do not differentiate the augmentation between different disease classes.
- figure 5 and figure 6 present the wrong caption

Keywords: I guess you can add *transfer learning* and *deep learning* keywords

Introduction section:

- I would merge *Introduction* and *literature review* sections in one fluent and concise section
- you are repeating the same period around the line 11

Proposed work and its experimental evaluation section:

- I would reorganize the section. You only have one sub section for *data augmentation* but I expected at least another subsection for *dataset description*. In particular, merging in this dataset subsection all the information about the data set that I found a bit strewn in the paper (data set origin, content, the disease classes, how you decide to split for the training, etc). you can also add more detailed information about the pre-processing pipeline of the data. Do you only normalize your data?
- I do not think images of the train/validation loss curves are necessary in the paper. Especially if you do not need particular comments on these curves
- The figures related to the neural network architectures (Fig. 5 and Fig. 7) can be also avoided since I have not read particular comments or changes that you made on them. Just the citation of the respective reference papers can be enough
- It seems that you mention in the paper some metrics like the ROC and also the confusion matrix that are not reported in the paper. Why? Confusion matrices reporting the several performances among the several disease classes could be interesting to show