Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-22-734

Reviewer A

Comment 1: Acronyms should be introduced at first occurrence (SDB, EEG, etc.).

Reply 1: Thanks for your comments. I added the introduction of these abbreviations on page 4, line 8&13.

Comment 2: Why did these patients get a CT scan? This could be relevant for the study if they differ between the OSA and non-OSA groups.

Reply 2: Thanks for noticing. CT images were mainly retrospectively collected from patients who visited our hospital complaining of sleep apnea. They underwent PSG to confirm the diagnosis of OSA and wanted to undergo surgical treatment, and preoperative upper airway CT was performed.

CT images of individuals in the control group were mainly derived from patients who came to our hospital with a diagnosis of leukoplakia of the vocal cords and were going to be treated surgically. These patients underwent preoperative CT of the neck to define the extent of the lesion. PSG was performed to clarify whether those patient had OSA to assess the risk from subsequent anesthesia. The control group was mainly selected retrospectively from this group of patients with a PSG diagnosis of AHI <10. If AHI was ≥10, the patient was included in the experimental group. We added this explanation to the methods section, on page 6, line1-11.

Comment 3: Why is AHI not reported for controls? You state that all participants underwent PSG examination, so this should be available.

Reply 3: Many thanks for your advice. We calculated the AHI for the control group, added in Supplementary Table 2.

Comment 4: If the AHI is available in all participants, the authors could optimize the deep neural networks as a regression problem instead, which could offer some diagnostic detail. However, this might not be ideal if there are few with moderate OSA as you claim. Why did you choose classification over regression?

Reply 4: Thank you for your comments. The vast majority of patients with OSA are combined with structural problems of the maxillofacial or upper airway, but anatomical factors can not fully explained AHI and may be associated with functional factors. The aim of this study was to perform primary screening for OSA in patients with CT of the neck, and further refinement of PSG monitoring for confirmatory diagnosis and phenotypic assessment is required in high-risk patients. However, the use of regression tasks to predict AHI does provide more information about the physician's diagnosis, but this may require the inclusion of factors other than anatomical factors (e.g., physiological factors), which we will further investigate in the future. We also discuss this section in detail in the limitation, located on page 17, lines 15-19.

Comment 5: All variables in equations should be explained explicitly. For equation 1, it seems that the model output prediction is denoted as both logitsi and logits si, which should be fixed.

Reply 5: Thank you for pointing out the problem, which was indeed an oversight on our part and has been corrected in the original article.

Comment 6: How was the training data split into training and validation during cross-validation?

Reply 6: Thank you for your notice. We divide the dataset into five equal parts and then select one of these five equal parts as the test set, and the rest as the training and validation sets. The above steps are repeated five times, and finally, the average result of the five times is used to report the performance. I have added the detailed steps in the Methods section, on page 9, lines15-20.

Comment 7: Why was each reconstruction method only used separately? You could potentially increase performance by 1) concatenating all 18 views in one model, or 2) ensemble the models using an average prediction.

Reply 7: Thank you for your comments. We have added it (all 18 views) to the results and found some improvement in performance.

Comment 8: I would recommend including a recent more extensive study (n = 1,366) on predicting OSA from facial photography in your discussion of other methods (https://doi.org/10.1109/JBHI.2021.3078127). This method also uses software that solves the issue of calibration and standardization.

Reply 8: Thanks for your advice. This is an excellent piece of literature and I have added it to the discussion section, located on page 15, line1-2, of the article, added to reference38.

Comment 9: Why is patient privacy protection mentioned for facial photography and not for CT reconstructions? These both include clear identifiers (skin) and rich health data, which really requires more protection.

Reply 9: Thanks for your reminder. It is true that either facial photography or CT reconstruction includes clear identifiers (skin) and rich health data, which do require more information protection. I have added this part of the discussion on page 17, line 11-13.

Comment 10: Finally, the authors propose that this method could be used to screen for OSA in patients that are scheduled for a CT scan for various reasons. Population-based screening methods with low precision have a problem of generating many false positives that are highly costly to further investigate. You should discuss if this method has sufficient precision for this purpose or if other follow-up studies are necessary.

Reply 10: Thanks for your comments. The population-based screening methods are not very accurate and do result in false positives thus making further investigation costly. In the future, a larger sample size is needed, as well as the inclusion of populations attending other centers for model training to progress one for screening. Patients with false positives have anatomic risk factors and we recommend close follow-up and may still have OSA as they gain weight and age. this section is added in limitation, page 17, lines 15-22.

Reviewer B

Major concern:

My major concern with the methodology is the input which involve, reconstruction of the CT images and training of the algorithm based on six directional views for skin, skeleton and airway, which is simple and can be conducted easily with pretrained networks. Instead 3D segmentation of the airway should have been applied which is the mainstay for assessing morphological parameters of the upper airway. The study had made more sense if automated segmentation of the anatomical structures was performed and combined with sleep apnea detection parameters. Better deep learning methodologies consisting of volumetric (segmentation) data and already exist for diagnosing and predicting OSAS.

Reply: Thank you for your comments. Our initial thought was to perform an initial screening of patients who may have OSA and then perform a PSG to confirm whether they have OSA, so we wanted to simplify our work as much as possible. Your comments are very meaningful, and we will do related research in our future work in the hope of providing more valuable help to physicians in their diagnosis. This part of the work that can continue to be expanded is added in limitation, page 18, lines 5-7.

Other concerns:

1. The complete manuscript needs to be edited by a professional English editor.

Reply: Thanks for your advice. We have had the manuscript edited by a professional English editor.

2. Page 4 line 16-17- use recent references

Reply: Thanks for noticing. We have replaced them with recent references.

3. Introduction should be structured by an introductory paragraph followed by what is known and unknown about the topic at hand. Thereafter, aim to answer the unknown.

Reply: Thanks for your reminder. We reorganized the structure of the introduction.

4. Methodology: Sample size calculation should be added. To me the sample looks really small for DL.

Reply：Thanks for your comment. we selected different sample sizes for training to obtain F1 values and thus determine whether the sample size was sufficient. We randomly selected 20% of the data as the test set and 10% to 100% of the remaining data as the training set, respectively. The training mode was all 18 views plus the fusion method. The results of the sample size calculation are shown in Figure 6. The performance of the model did not improve significantly when the training set exceeded 154 (70%) of the data, which proves that the size of our dataset was sufficient. We have added the method on page 10-11, lines 21-22&1-2; the result on page 12, lines 1-3.

5. What was the reason for PSG of control group.

Reply: Thanks for your comment. CT images of individuals in the control group were mainly derived from patients who came to our hospital with a diagnosis of leukoplakia of the vocal cords and were going to be treated surgically. These patients underwent preoperative CT of the neck to define the extent of the lesion. PSG was performed to clarify whether those patient had OSA to assess the risk from subsequent anesthesia. The control group was mainly selected retrospectively from this group of patients with a PSG diagnosis of AHI <10. If AHI was ≥10, the patient was included in the experimental group. We added this explanation to the methods section, on page 7, line5-11.

6. Page 6- Elaborate the methodology conducted in Mimics for viewing and saving the directional views. Plus Mimics 19 is too old, I would suggest using the newer version.

Reply: Thanks for your reminder. We recaptured the image using mimics version 21.0.

7. Provide references in the text under “data processing” for networks, adam optimizer etc. wherever required.

Reply: Thank you for your reminder. This was an oversight on our part and the relevant information has been added to the references 41&42.

8. The “results” section is too limited. Needs to be expanded with further testing such as significance testing for fusion methods, timing of prediction etc.

Reply: Thank you for your comments. We have included the inference time for different methods in our results. (Table 4). In addition, we believe that the AUC curves of different methods can reflect the validity of the method to some extent, so we only compare the AUC curve values of different methods.

9. The results also show that the performance parameters as described in table 3 are below par and do not represent very good scoring for prediction. Which again could have been avoided if volumetric information instead of photographic information was recorded and a sample size calculation was performed.

Reply: Thank you for your comments. We have added the results using 18 views in Table 2 and found that the performance of the model has improved to some extent. For the volume calculation, which is currently not possible due to our current technical limitations, we plan to do related research in future work in the hope of providing more valuable help to physicians in their diagnosis. We have also written this suggestion in the outlook, located on page 18, lines 5-7.